SCUOLADIDOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Information Systems and Communications

Dottorato di Ricerca in / PhD program Computer Science Ciclo/ Cycle XXIX

Curriculum in (se presente / if it is)

Borrower Risk Assessment in P2P Microfinance Platforms

Cognome / Surname Jamal Uddin Nome / Name Mohammed

Matricola / Registration number 787876

Tutore / Tutor: Prof. Dr. Giuseppe Vizzari

Cotutore /Co-tutor:
(se presente/if there is one)

Supervisor: Prof. Dr. Stefania Bandini
(se presente / if there is one)

Coordinatore / Coordinator: Prof. Dr. Stefania Bandini

ANNO ACCADEMICO / ACADEMIC YEAR 2015/2016

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Stefania Bandini for the continuous support of my PhD study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my PhD study.

Besides my supervisor, I would like to thank my tutor Professor Guiseppe Vizzari for his insightful comments and encouragement, but also for the hard question which helped me to widen my research from various perspectives.

Moreover, I would like to express my sincere gratitude to Dr. Mahmood Osman Imam, Professor of Finance, Dhaka University, who has been always played a key role in encouraging and coordinating me in this whole project. I am very grateful to him for his invaluable support, time, suggestions and guidance throughout this period of research.

Besides, I am indebted to all of my friends and colleagues of the LINTER lab at the Department of Computer Science, Systems and Communication (DISCO) for their support throughout my doctoral study. I thank my fellow lab mates Luca, Andrea and Mizar for the stimulating discussions, for the time we were working together before deadlines, and for all the fun we have had in the last three years.

I would like to thank Andrea Sedini for his support during this program. Moreover, I would like to thank Dr. Kazi Ahmed Nabi, Professor of Finance and Banking, University of Chittagong, Professor Dr. M. A. Baqui Khalily, Ex-Executive Director of Institute of Microfinance (InM), University of Dhaka, and Professor Dr. M. Jahangir Alam Chowdhury, Professor of Finance & Executive Director of Center for Microfinance and Development, University of Dhaka for their contributions in defining the risk rating rules, Mr. Md. Fazlul Kader, Deputy Managing Director of Palli Karma-Sahayak Foundation (PKSF) and Mr. Md. Shahid Ullah, Executive Director of Development Initiative for Social Advancement (DISA) in understanding the current practices of borrower's risk assessment.

My sincere thanks also go to Firoj bhai, Lucky, Salim and many more. Without their precious support it would not be possible to conduct this research.

Last but not the least, I would like to thank my family for supporting me spiritually throughout writing this thesis and my life in general.

This thesis is dedicated to the memory of my beloved father Abdur Razzak, whom I lost during this PhD period and who will always remain in my thoughts and prayers.

Table of Contents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Despite its recent fast growth in fame and money raised, Peer-to-Peer (P2P) lending remains understudied and young field in academia. Peer-to-Peer[1] (P2P) lending is an Internet-based[2] platform of financial transactions where borrowers place requests for loans online and private lenders fund them directly or indirectly[3](Bachmann et al., 2011; Everett, 2015; Herrero-Lopez, 2009; H. Wang & Greiner, 2011). Most of the direct P2P lending models like Prosper (USA), Zopa (UK), Smartika (Italy) operate nationally without the support from any intermediary and capture the retail market for consumer loans and credit card loans globally, more particularly in the US (Weib, Pelger, & Horsch, 2010).However, most of the indirect models like Kiva, Zidisha, MyC4 operate globally with the support of local agents, known as field partners, who manage borrowers locally and get them connected to the platforms. The main focus of this study is on the models who aim to connect people (here users of the platforms) through lending to alleviate poverty. Hence, they are the indirect P2P models who facilitate providing microcredit services to less privileged people, especially poor and marginal groups who do not have access to formal financial services because of lacking collateral, steady employment and a verifiable credit history (Bauchet, Marshall, Starita, Thomas, & Yalouris, 2011).

1.1 Motivation

These emerging web-based platforms help microfinance overcome the challenge of sustainability with operational efficiency and cost effectiveness(Kauffman & Riggins, 2012; Uddin, Vizzari, & Bandini, 2015a). Which motivated us in these particular P2P models is how lenders choose borrowers or loan applicants with the given information on the platforms. Although most of the models among indirect P2P platforms operate as prosocial lending models who give emphasis on social values through the services and lenders do not receive any interest on their lending except taking back the principal amount they lent, they should concern about the risk of lending at least for the loan principal amount. Lenders always face challenges in choosing a borrower from many candidates on such platforms, particularly for individual lenders who are not expert in lending. Moreover, they are provided with little information which lacks the details of the financial aspects, particularly risk assessment of the loan applicants. Such lacking makes lending decision really a tough job for lenders. In this context, Jenq, Pan, & Theseira (2012) argued that

> "…As experienced lenders may be less prone to rely on their implicit attitudes, we view this as indirect evidence that implicit discrimination may explain part of our findings, suggesting that future research on this line will be valuable…"

[1]Also referred to as Person-to-Person lending, People-to-People lending, social lending, or P2P lending. We will use peer-to-peer lending and P2P lending interchangeably.

[2]Also referred to as web-based, and online. The term 'Online' will be used in this paper.

[3] Direct P2P Model allows borrowers and lenders to connect directly, eliminating conventional intermediaries (bank or other financial intermediary), to provide for greater access to credit at a lower cost; Indirect P2P Model typically provides capital to developing markets by connecting borrowers and lenders through local intermediaries or field partners (Hassett et al., 2011). The details about P2P platforms are described in section 2.

In Jenq et al.(2012), the authors argued that despite the limitation of the paper in providing direct evidence on the extent to which observed bias is attributable to explicit or implicit discrimination, they are able to show that lenders with more experience on Kiva (P2P lending platform) are less likely to fund loans in a pattern consistent with lender bias on physical characteristics. This argument lead them to interpret this as indirect evidence that *greater lender familiarity with the choice problem reduces the lender's tendency to rely on implicit mental processes – although it could also be evidence that more committed lenders simply have a different type of preference.*
While measuring the experience on Kiva by the number of months a lender has been a Kiva member and by the total number of loans made, and investigating if experience affects lender biases, it is found that all else equal, loans with more attractive (overweight) borrowers are, on average, funded by lenders with relatively less (more) experience on Kiva. Moreover, need and trustworthiness are the factors that are considered by the less experienced lenders to fund borrowers. Hence, greater experience appears to be related to a lower degree of bias towards physical and subjective attributes of borrowers.

In this paper, the authors clearly identify the need of understanding the default risk of borrowers as follows:

> "… Kiva lenders face two potential considerations. First, they are likely to care about the social impact of their loan and, all else equal, we may expect them to prefer borrowers who would maximize social impact, such as borrowers who appear more needy than others. Second, while Kiva lenders are really donors, recovery of the loan principal is important since a recovered principal allows the 're-gift' of the principal to a new borrower, promoting an additional charitable goal. Kiva lenders should therefore pay attention to borrower profitability and default risk. As virtually all the loans on Kiva eventually receive full funding, we analyze the speed at which loans are funded as a proxy for the relative attractiveness of a given loan…"

According to Galak, Small, & Stephen(2011), this context constitutes a new hybrid decision form which is called prosocial lending. This is hybrid since it consists of both financial and prosocial characteristics. On one hand, from financial perspective, it shares many characteristics with conventional financial decision making. On the other hand, from prosocial perspective, its stated purpose is to help others. The decision to lend is financial in nature: the principal of the loan is returned to the lenders (assuming the loan does not go into default) and many investment-like metrics are provided to the decision maker (e.g., field partner rating, historic default rate, loan duration, etc.). All of these features could compel the lenders to treat the decision in a more calculative manner, which could make psychological/emotional drivers ineffective(see more in Small, Loewenstein, & Slovic, 2007).As lender's decision is both financial in nature as well as prosocial, risk assessment of borrower might help lender to assess the borrower more efficiently from financial perspective.

1.2 Statement of Problem

Different risk management tools are practiced in the sector but most of them are for group borrowers and risk rating of borrowers is not provided to the lenders on indirect P2P platforms[4]. This lack of missing information on borrower risk assessment is surprising since

[4]Risk rating with credit score is available in most of the direct P2P platforms who operate nationally like Prosper, Zopa which are out of the scope of this study (Ceyhan, Shi, & Leskovec, 2011; Slavin, 2007; H. Wang & Greiner, 2011). However, this study focuses only on the indirect P2P lending models who operate globally like Kiva, Zidisha(Hassett et al., 2011) and they have no such risk rating.

credit scoring could help the online indirect P2P model's lenders to evaluate the loan applicants more efficiently and thereby could match their lending risk perception with the degree of risk associated with a particular loan applicant.

Holding many promises like disintermediation of expensive traditional financial intermediary, easy access of *unbankable* borrowers to the financial services, new economies of scale, lower financing cost, this innovative online P2P lending also carries some challenges- default rates, regulatory requirements, and leveraging social capital. An inherent risk, the focus of this study, exists in a pseudonymous online environment of P2P lending where most of the individual lenders are not professional investors which causes serious information asymmetry problems (Assadi & Ashta, 2010; Bruett, 2007; Hawkins, Mansell, & Steinmueller, 1999; Heng, Meyer, & Stobbe, 2007; Jeong, Lee, & Lee, 2012; Klafft, 2008; Magee, 2011; Slavin, 2007; Tan & Thoen, 2000; H. Wang, Greiner, & Aronson, 2009). Therefore, loan default and loan fraud would be the most fundamental concern, among others, with lending money unsecured to complete strangers over the Internet (H. Wang et al., 2009). For example, Prosper.com failed to predict the delinquency rates which were higher than expected (H. Wang & Greiner, 2011). The inference on this problem made by several research studies is its vulnerability to adverse selection (Berger & Gleisner, 2009; Freedman & Jin, 2008). In this case, borrower information and its accuracy are critical for lenders to assess a borrower's credit risk. However, obtaining and verifying borrower information would increase the operation cost considerably. It is more acute in online indirect P2P lending platforms that are serving globally in general, developing countries in particular. In addition, being a new innovative business model, online P2P lending platform is under the most influential challenges to overcome the regulatory issues as well as to replicate the social network learned from off-line solidarity lending (H. Wang & Greiner, 2011). Among the challenges, the problem with the borrower's or loan applicant's information is critical to the web-based lenders to remain active in such platforms and to sustain them in the long term in the promotion of novel goal, reducing global poverty. Moreover, it is serious, in Figure 1, because no individual credit risk rating is provided directly or indirectly by the field partners or by such lending platforms resulting bearing the default risk lies absolutely with the lenders who ultimately refinance the field partners.

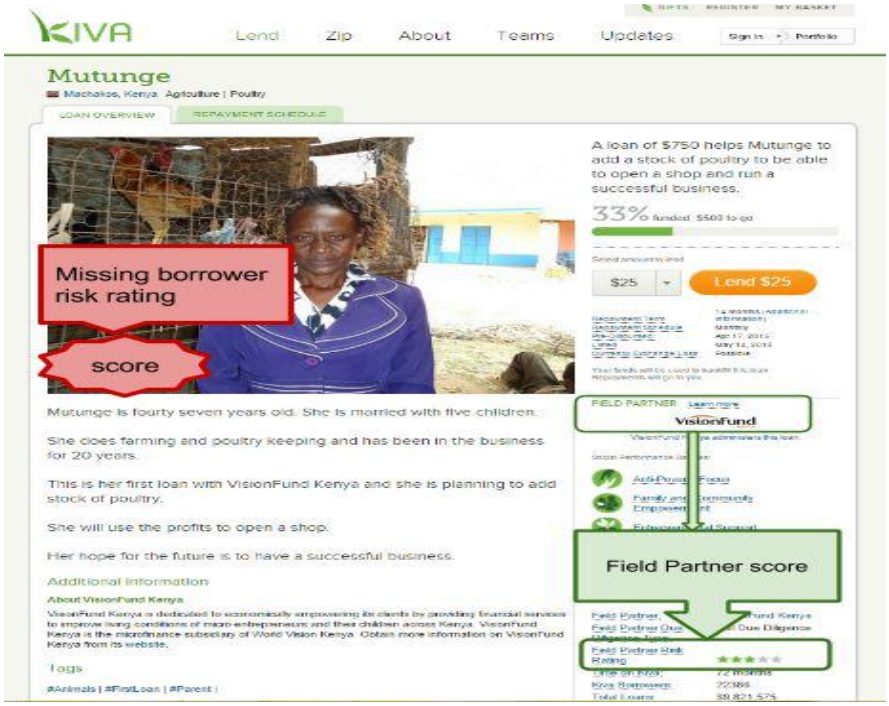


Figure 1.1 *An example of Kiva loan request description.*

Selecting borrower is the challenging task to online microcredit lenders as individual borrower's profile does not provide any risk rating on the platform except the microfinance intermediaries' aggregate risk indicators (depicted on the right bottom corner in Figure 1) and the information that these intermediaries (filed partners in Kiva model) screen/assess the borrowers before being posted and made available to the lenders (to kiva platform). Moreover, the platforms merely keep typical advices for lenders/end users to diversify their portfolios through lending to more than one borrower via different field partners as well as in different countries and/or sectors. However, the borrower's risk, which is missing on the models (indicated in Figure 1.1), remains critical to the aggregate lenders or individual lenders who ultimately refinance the field partners in the platforms. To address this problem, Kiva model has been chosen as the most leading one to represent the borrower or loan applicant's profile in a scientific manner which is not only solve the problem of borrower's information in Kiva but also in other models that have the same problem in this sector.

1.3 Research Objectives

The prime emphasis is on the specific problem- borrower risk assessment in P2P microfinance platforms- that supports the tendency of experienced lenders for not relying on their implicit attitudes like unintentional or subconscious choices which has been viewed as an indirect evidence by Jenq et al.(2012) that lenders behave rationally based on the merits of the loan request or loan proposal. Therefore, the prime objective is (a) to build a CBR system for borrower risk assessment in online indirect P2P microfinance platforms and to suggest how risk assessment, especially credit scoring can be useful to online P2P micro-lenders. In order to achieve this objective it needs cases in which solution part is missing. Therefore, solving the problem of missing solution in case structure another sub-objective is (b) to develop an expert-based risk rating model using spreadsheet coding.

1.4 Research Contributions

With the research work, there are two value additions: expert-based risk rating and CBR-based risk rating. As a dominant risk rating approach in microfinance till now, expert-based risk rating can assess borrower risk in microfinance very well(Bunn & Wright, 1991). However, despite its good practice in traditional brick-and-mortar models, it does not fit well with online P2P models due to its inherited limits like no learning, need maintenance, computationally expensive, high user requirements, not completely automated, and not applicable to large scale operations. As a result, the expert-based rating (risk scoring) has been used, in Figure 2, for providing the solution to a proper set of representative relevant cases to use in CBR-based risk rating which is completely automated with incremental learning that will lead to act as bootstrapping for improving the system with more predictive power.

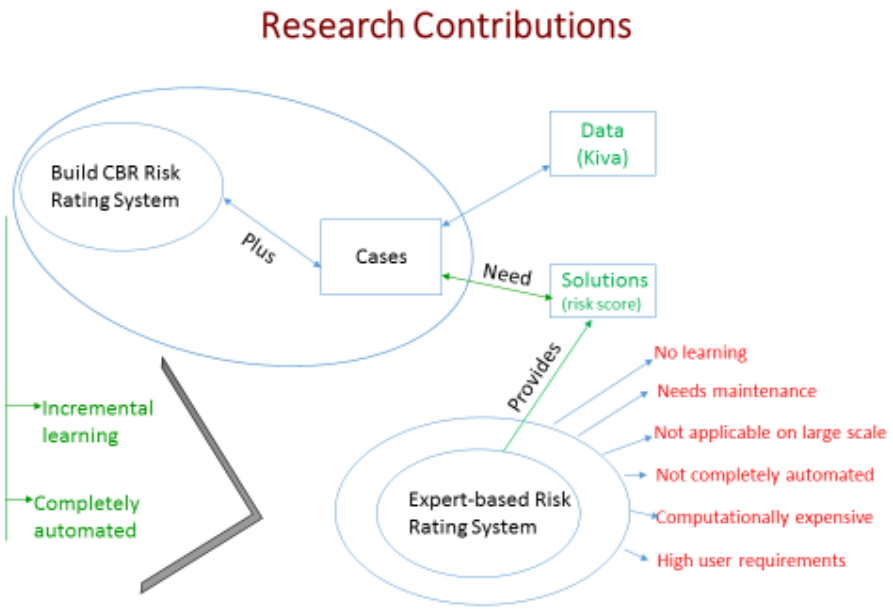


Figure 1.2*Research contributions*

Therefore, from this research work, there are two contributions: (a) an expert based risk rating model (but not fully automated and not applicable on Kiva scale), and (b) a CBR based risk rating model with a proper set of initial relevant cases and a similarity function tuned according to (a).

1.5 Methodological Approach

The scope of research is the online indirect P2P models that operate globally. The Kiva model has been chosen as the largest and leading one (Hassett, Bergeron, Kreger, & Looft, 2011) as a case study which allows to access its open source data for study. The Kiva XML data have been recovered using XQuery to organize a database for past loans of individual borrowers (unit of analysis) with representative numbers and then examination method has

been used for identifying relevant and readily extractable features for the sample of past individual borrower loans.

The CBR approach has been chosen(see details in Chapter 4) as a prime methodology to assess borrower risk in online indirect P2P models. Because no other statistical models fit well with the unique nature of borrower profiles in microcredit where the nature of borrower characteristics demands for special knowledge and little relevant data exist in online indirect P2P lending platforms for global borrowers. Hence, the CBR system works as a statistical model to improve the results (risk rating/prediction) of judgemental or expert rules through the bootstrapping process(Bunn & Wright, 1991, p.505). Secondary data from Kiva open source database (build kiva) have been used for the study because the under taken research objective (borrower risk rating using statistical model) demands for historical data on which new case base can be developed. The database of Kiva is large enough to qualify the requirements of large size database for CBR application. From the database of Kiva, only African and Asian zones have been chosen with 45 countries which count more than 50% coverage of this database in terms of country covers (83 countries). The reason for choosing these two zones are the homogeneity in terms of loan size and nature of borrower's activities. Only individual borrower's loan data have been chosen as a unit of analysis skipping group borrower data for selected variables. All the information necessary to define a case description are available and also the final outcome is known (the information about the actual repayments), but no actual risk rating is present and therefore all cases would be missing the solution part. To solve this cold boot problem, a strategy is adopted to select a reasonable number of past loans that are sufficiently representative of all the selected countries, economic sectors for the funded activities, kind of borrowers, and actually rate them (filling thus the solution part of the case) employing expert rules for rating the risk associated with loan requests in developing countries, coded into a spreadsheet.

The expert rules (see details in chapter 6)have been chosen for using knowledge-based rating (Baklouti; Ibtissem & Bouri, 2013) for providing the missing part "solution (risk scoring)" to make the loan cases complete to use in CBR approach. Because knowledge/judgement-based rating works well where the opacity problem and little data exist. The context of microcredit lending in online P2P platforms especially in developing countries conforms both opacity issue and little data availability for which expert rules are justified. Therefore, a constrained expert model or integrated model has finally been chosen combining expert-based manual model (expert-judgment approach or knowledge-based approach) with automated statistical model (CBR approach). Other models like group lending approach, dynamic incentives, or collateral substitutes that work well for borrowers in group lending but do not fit with the risk assessment of individual lending in microcredit system (see Chapter 3 for comprehensive methodological approach).

Using this expert-based models *credit scoring* has been carried out for a set of representative cases of 107 loans, and then they have been used in the CBR system as complete loan cases to run the system for assessing new loan applicants. The CBR rating has been tested with a set of test loan cases (75 cases from holdout sample from 2014) for evaluating its predictive power. The CBR system considered as low risk borrow requests 52 of them where 77%of which were correctly repaid and has correctly predicted 9 (60%) of 15 default loan cases. The system turned out to be quite conservative, since requests considered risky often turned out to be correctly repaid, but in general results are encouraging.

1.6 Thesis Organization

The discussion of the Thesis work is organized as follows:

While this first chapter presents an introduction of the study by identifying the research problem and specific objectives along with a general overview for its methodological approach and contributions, the second chapter proposes a thorough discussion of the state of the art on borrower risk rating in online P2P microcredit lending model and then it ends up with the conclusion for credit scoring for online indirect P2P microcredit borrowers.

The third chapter states the methodological approach that sets the research work. It describes why a CBR approach has been chosen for assessing borrower risk in P2P microfinance platforms and how the CBR system fits with the use of supplementary expert-based risk scoring. Also it discusses about data and their collection techniques and finally mentions about a database and user interface linked to the system.

The fourth chapter introduces CBR as an approach and describes the process of CBR and its application in finance for credit scoring. Also, this chapter gives a clear idea how CBR approach can be applied to borrower risk assessment in online P2P microfinance platforms.

The fifth chapter states the design and the development of CBR System and its proprietary database along with technical details. Then, this chapter ends up with the implications of the CBR system and effectiveness of the credit score.

The sixth chapter presents the scenario of credit (risk) scoring in finance and then discusses different credit scoring models in microfinance. Finally, it illustrates how spreadsheet based credit scoring has been developed using expert rules for online indirectP2P microfinance platforms.

The seventh chapter analyses the results of the initial training set of relevant cases and evaluates the results of the test set for testing the predictive power of the model (CBR-base risk scoring model).

The eighth and final chapter summarizes the thesis discussion, specifies the contributions made in this research and provides future directions in this line of research.

Chapter 2

State of the Art: Borrower Risk Rating in P2P Microcredit Lending Model

Peer-to-Peer[5] (P2P) lending is an Internet-based[6] platform of financial transactions where borrowers place requests for loans online and private lenders fund them directly or indirectly[7](Bachmann et al., 2011; Everett, 2015; Herrero-Lopez, 2009; H. Wang & Greiner, 2011).This new digital intermediary taking the advantage of web 2.0 was emerged on the basis of microcredit principles (Herrero-Lopez, 2009; Magee, 2011).This has eventually grown in recent years, especially after Zopa[8] and Prosper[9] as an alternative platform of traditional saving and investment and later spread in Europe and Asia (Jeong et al., 2012; Magee, 2011; Slavin, 2007; Yum, Lee, & Chae, 2012). The growing platform has captured consumer loans globally, more particularly in the US (Weib et al., 2010) and has quickly drawn significant attention from the mainstream media and academia in several disciplines (Bachmann et al., 2011; Light, 2012; H. Wang & Greiner, 2011).Despite its recent fast growth in fame as well as money raised, P2P lending remains a field underscore and understudied in research area(H. Wang et al., 2009).

New digital intermediation and the re-intermediation of traditional financial intermediaries offer new benefits as well as new challenges (Berger & Gleisner, 2009; Hawkins et al., 1999).The most popular selling value of this digital innovation is that disintermediation of expensive traditional financial firms for which borrowers can avail cheaper loans without collateral while lenders still can earn better return from their investments (Jeong et al., 2012; Klafft, 2008; Magee, 2011; Slavin, 2007; H. Wang et al., 2009). Another remarkable advantage of this platform is the access of *unbankable* borrowers or ones with low credit scores to the financial services through microfinance approaches that rely upon social collateral (Bruett, 2007).Besides, outreach coverage by Internet has created new economies of scale, and the lower financing costs have contributed to cost reductions for the microlending platforms (Ashta & Assadi, 2010; Magee, 2011; Slavin, 2007).Holding many promises, this innovative online P2P lending also carries some challenges- default rates, regulatory requirements, and leveraging social capital. An inherent risk, the focus of this study, exists in a pseudonymous online environment of P2P lending where most of the individual lenders are not professional investors which causes serious information asymmetry problems (Heng et al., 2007; Klafft, 2008; Tan & Thoen, 2000; Steelmann, 2006 in Berger & Gleisner, 2009). Therefore, loan default and loan fraud would be the most fundamental concern, among others, with lending money unsecured to complete strangers over the Internet(H. Wang et al., 2009). For example, Prosper.com failed to predict the delinquency rates which were higher than expected (H. Wang & Greiner, 2011). The inference on this

[5]Also referred to as Person-to-Person lending, People-to-People lending, social lending, or P2P lending. We will use peer-to-peer lending and P2P lending interchangeably.

[6]Also referred to as web-based, and online. The term 'Online' will be used in this paper.

[7] Direct P2P Model allows borrowers and lenders to connect directly, eliminating conventional intermediaries (bank or other financial intermediary), to provide for greater access to credit at a lower cost; Indirect P2P Model typically provides capital to developing markets by connecting borrowers and lenders through local intermediaries or field partners (Hassett et al., 2011). The details about P2P platforms are described in section 2.

[8]http://www.zopa.com, the first P2P lending site in 2005 in UK.

[9]http://prosper.com, the largest P2P lending platform in 2006 in the US.

problem made by several research studies is its vulnerability to adverse selection (Berger & Gleisner, 2009; Freedman & Jin, 2008). In this case, borrower information and its accuracy are critical for lenders to assess a borrower's credit risk. However, obtaining and verifying borrower information would increase the operation cost considerably. It is more acute in online indirect P2P lending platforms that are serving globally in general, developing countries in particular. In addition, being a new innovative business model, online P2P lending platform is under the most influential challenges to overcome the regulatory issues as well as to replicate the social network learned from off-line solidarity lending(H. Wang & Greiner, 2011).

As our prime emphasis is on borrower's risk, the tendency of experienced lenders not relying on their implicit attitudes like unintentional or subconscious choices is viewed as an indirect evidence by Jenq et al.(2012) that lenders behave rationally based on the merits of the loan request or loan proposal suggesting that future research on this line will be valuable. Because while measuring the experience on Kiva by the number of months a lender has been a Kiva member and by the total number of loans made, and investigating if experience affects lender biases, it is found that all else equal, loans with more attractive (overweight) borrowers are, on average, funded by lenders with relatively less (more) experience on Kiva. Moreover, need and trustworthiness are the factors that are considered by the less experienced lenders to fund borrowers. Hence, greater experience appears to be related to a lower degree of bias towards physical and subjective attributes of borrowers.

Following this underscore and under-researched field, we are motivated for a comprehensive literature review that gives an overview of online P2P lending in general and emphasises the-state-of-the-art on borrower's risk in online P2P microcredit lending models in particular. Our purpose is to synthesize the research contributions previously done and to explore a precise idea of the borrower selection in online P2Plending platforms by the lenders, especially the individual lenders who are not expert in lending, and to recommend how risk assessment, especially credit scoring can be useful to online P2P micro-lenders. Our main target groups are readers from information society and financial management communities in particular, but social business in a broader context.

The chapter is organized as follows, we first describe the paper selection approach and then we introduce the online P2P microcredit lending market. The main part of this review covers borrower selection in online global-based indirect P2P platforms and lender's decision making, and borrower indebtedness, delinquency rate and information system. We then discuss credit risk management practices and recommend a risk assessment system (credit scoring) for the lenders to assess individual borrower's risk. Finally, a short conclusion summarizes the main points of the contribution and indicates the next steps of our research.

2.1 Paper Selection Approach

During the period of 6 months from September 2015 to February 2016we conducted a keyword-search in Google Scholar with the terms "P2P Lending","P2P Microcredit Lending Platforms", "Peer-to-Peer Microcredit Lending", "ICT-based Microcredit Lending Models", "Microcredit Borrower Risk", "Borrower Risk in Online P2P Lending", "Credit Scoring in Microfinance" and included further articles through backward-and-forward search. Because the research topic is contemporary and most journals and conferences open their databases for search engines like Google Scholar, we assumed that any search bias would be limited and therefore abstained from journal search(Bachmann et al., 2011). To supplement this search bias, if any, we also searched SSRN, Science Direct and Springer for the same.

In order to reach our objective, following the methodology proposed by(Peters, Howard, & Sharp, 2012), we developed a relevance tree in order to build an initial structure for the

intended literature search that guides our search process. This relevance tree helped us to identify those areas that we needed to search immediately (underlined) and those that we particularly needed to focus on (starred) (Figure 2.1). We followed the steps suggested by (Saunders, Saunders, Lewis, & Thornhill, 2011) for creating the relevance tree. At the very beginning, we set our research question as: "How is borrower's risk measured in and embedded to online P2P microcredit lending model?" which is then categorized into two major subcategories: Direct P2P models and Indirect P2P models. Figure 1 shows the relevance tree that helps us to further proceed with the literature searching.

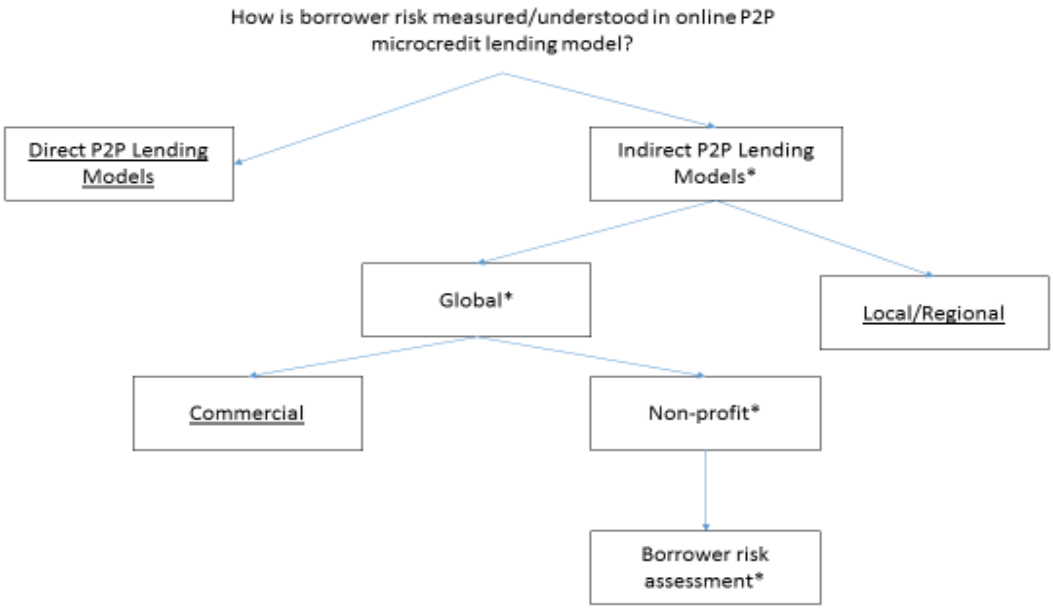


Figure 2.1*Relevance Tree*

We found 90 papers, Figure 2.2, from the years 1970 to the beginning of 2016[10]. Most of the studies found started from 2006 onwards (P2P period & refinement period of the P2P model) when merger between microfinance and web 2.0 took place to open a new lending platform with its transparency, connectedness and affordability option for anyone on the Internet to lend directly to the active poor(Coleman, 2007). The years earlier than year 2006 (pre-development of P2P) covered here to capture the issues like group-lending, credit scoring, credit information system as credit risk management strategies in microcredit. Following the relevance tree we reviewed the papers for a particular research direction in online P2P lending(J. W. R. T. Watson, 2002). While reviewing the papers on online P2P lending models we considered only non-profit or pro-social online indirect P2P lending platforms operating globally, like Kiva.org, where lenders provide loans without any interest, and the platforms create revenues from donations, optional lender fees and other sources (www.kiva.org). Having viewed the papers in different contexts, we evaluated for-profit online direct P2Pplatforms, like Prosper.com, as well as other regional or local non-commercial models only for understanding purpose as we found very few studies in indirect P2P lending with non-commercial background.

[10]The details of the publications have been given in table 1 as appendix 1 at the end of this chapter.

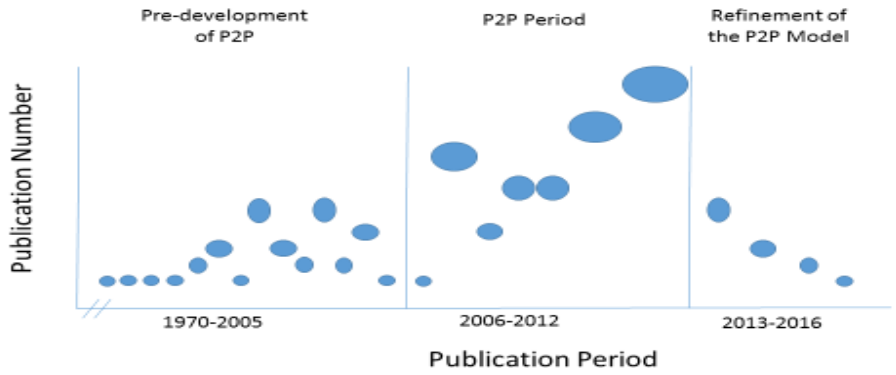


Figure 2.2*Publication Over the search period*

2.2 P2P Microcredit Lending Market

The practice of personal lending exists long before the formal institutional lending within family members or known persons or communities. The Internet has leveraged the concept beyond the family and known communities over the largest network globally where individual members become a lender, a borrower or both without any traditional intermediation.P2P lending platforms, like eBay and Amazon in retail industry, connect borrowers on the demand side and lenders on the supply side directly and sometimes with a third party instance (Hassett et al., 2011; Herrero-Lopez, 2009; H. Wang & Greiner, 2011).

2.2.1 Direct P2P Lending Models

In direct P2P lending model, borrowers and lenders are connected directly without any support from a bank or other financial intermediary. As an alternative way of applying for a loan from a bank or taking a cash advance from a credit card company at their fixed rate, a borrower can post a loan listing on the platform with an asking interest rate where strangers as lenders declare the amount of money to fund the loan request even at the interest rate lower under the bidding system. The money, after fully funded, gets transferred to borrower's bank account and the borrower repays the loan at the rate settled by the bids (Collier & Hampshire, 2010; Persson, 2012; H. Wang & Greiner, 2011). The Zopa (December 2005), Prosper (February 2006), and LendingClub (May 2007) are the three largest platforms operating in UK and US for profit. There are many other P2P lending platforms continue to grow in US, Europe, and even in Asia with some variations in operation (Frerichs & Schumann, 2008; Hassett et al., 2011).

2.2.2 Indirect P2P Lending Models

Indirect P2P lending model allows borrowers and lenders to connect through local intermediary or field partner who manage borrowers locally and get them connected to lenders via the platform. A borrower goes to the field partner or local financial intermediary and requests for loan. The field partner makes a thorough evaluation of the borrower's business to make sure he/she can repay the loan with cost. Then the loan request uploads (after or before the loan disbursement from field partner's own capital) to the platform so that lenders can read about the borrower and can consider the loan application for funding. A lender can fund the loan individually or in team, and finally the borrower repays the loan to the lenders via field partner and the platform. Lender receives only loan principal or both principal amount and interest depending on the specific model's business strategy (Uddin, Vizzari, & Bandini, 2015b). The Kiva (October 2005), MyC4 (May 2006), and DEKI (2008) are the three leading platforms operating globally.

2.2.3 Lending Operations and Stakeholders

Direct P2P lending platforms generate their revenue from service fees paid by borrowers as well as lenders. Fees for loan closure, late or failed payments are charged to the borrowers. Fees for lending are charged to the lenders on the amount funded (Klafft, 2008). For indirect P2P platforms, their revenues are generated from donations, optional lender fees and the interest income generated from the instalment money they hold for the time being (Flannery, 2007).

The stakeholders are grouped into internal and external perspectives. Internal stakeholders include management, employees and owners who run the platform whereas external stakeholders consist of lenders, borrowers, communities, partner banks/microfinance institutions, credit bureaus, regulatory Authorities who use and support the platform from their own capacities (Bachmann et al., 2011).

2.2.4 Leading Local and Global Models

Although there is no complete list of such models to refer, some compilations have been found to get an idea how this innovation is growing. The weblog P2P-Banking.com names 48 different platforms worldwide in February 2016. Bachmann et al., (2011)find a list of top 10 P2P lending platforms based on loan volume (Table 2.1). Also, Hassett et al. (2011) identity 24 Indirect P2P lending participants, and Garman et al. (2008) record 24 platforms existing worldwide.

The following table shows the ten largest lending companies relating to the total volume of loans created.

Position	Company	Country	Vol.(Million) in US$
1	Virginmoney	USA	390.0
2	Prosper	USA	178.0
3	Kiva	USA	57.9
4	Zopa UK	UK	45.6
5	Lending Club	USA	26.9
6	MyC4	DK	9.0
7	Smava	DE	8.6
8	Moneyauction	JP	7.8
9	Zopa IT	IT	5.9
10	Boober	NL	3.3

Table 2.1 *Loan volume of P2P lending companies(Bachmann et al., 2011)*

Till now most of the available P2P lending sites operate within the country because of the compliance variations across the countries or regulatory requirements(Berger & Gleisner, 2009). Although the platforms vary in type and operating styles depending on the purpose and other compliance issues, following on our study purpose and importantly based on the wide advertised value proposition, these platforms can be divided into basic two types[11]: direct and indirect (Hassett et al., 2011). The direct P2P platforms like Prosper, Lending Club, Zopa often run the business on a national level while the indirect platforms like Kiva, MyC4 usually work globally. Again, another classification can be made as commercial/for-profit and prosocial/non-profit platforms. While commercial platforms run for profit in general are limited to national markets, non-profit platforms with philanthropic or social purpose often operate globally. Lenders in commercial platforms get return on investment while lenders provide loan in non-profit sites at free of interest and only get back principal amount of loan.

Among the indirect P2P platforms in Hassett et al.(2011) study, 11 sites operate with local intermediaries globally and the rest run their businesses with regional or national focus. Of the 24 Indirect P2P Platforms identified in (Hassett et al., 2011) study, 19 are not-for-profit, 4 are for-profit and 1 is hybrid (mixed) entity. Among the non-profit platforms, Kiva is the largest one which captures 90% of the market lending in this type. As of March 2011 approximately $233 million had been raised through Indirect P2P platforms. Kiva is the first Indirect P2P online marketplace founded in 2005 as non-profit. MyC4 is one of the other Indirect P2P platforms established in 2006 as for-profit. Wokaiis another platform focuses on a single geographic area- China. Hassett et al. (2011) found both types of models (lending for-profit and not-for-profit) in his list of Indirect P2P platforms. As mentioned earlier, the majority of the platforms operate not-for-profit (provide no financial return to their lenders) and only few platforms operate for-profit. Among the for-profit platforms, lenders, for instance at MyC4, can receive financial return from their socially-motivated investment capital following "Dutch auction" on the sites. Over 90% of funds provided through Indirect platforms led by Kiva as pioneer model provide social capital where lenders make their investment at free of cost for filed partners[12].

Wang et al.(2009)provided an overview of P2P lending in the U.S., considering the motive for lending and degree of separation among participants (see Figure 2.3).

[11]Again, different variations of P2P lending platforms are seen within the two basic categories in terms of interest rate determination, interest groups, degree of separation between lenders and borrowers etc.

[12]Field partners charge interest to their borrowers and recover the loan with interest. However, they only return the principal amount of the loan to the online lenders through P2P models like Kiva.

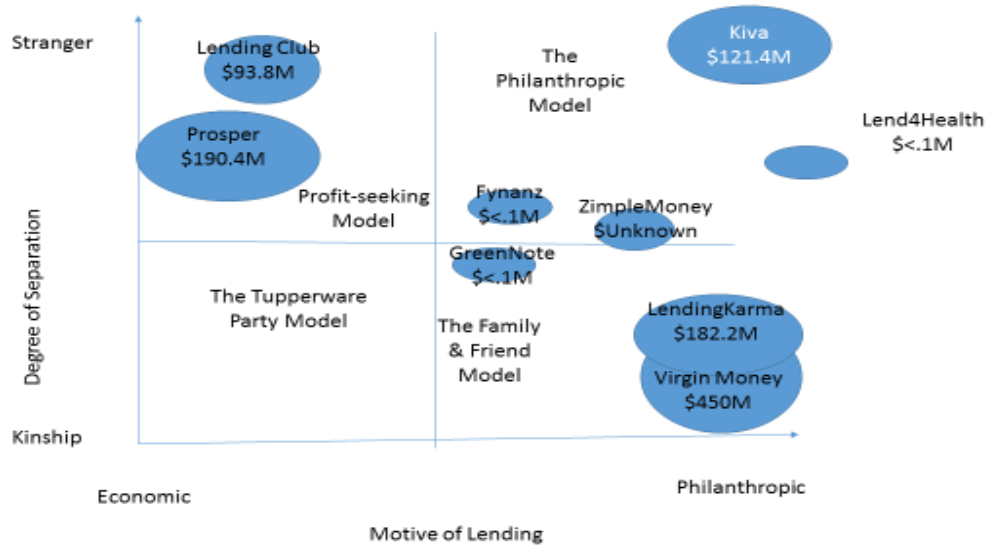


Figure 2.3*P2P Lending Marketplaces in the US*

Considering the above citations made earlier in different studies, we can provide a list based on the current visibility with our focus on those indirect P2P platforms who operate their business globally following indirect models with the motivation as pro-social capital providers (Table 2.2).

Online Indirect P2P Platforms	Operation Status & No. of countries (Cs)	Focus	Motivation	Loans (in Millions)	No. of Lenders
Kiva	Global; 82 Cs	Microfinance	Non-profit	$950.80	1.60 M
Zidisha	Global; 9 Cs	Microfinance	Non-profit	$7.95	21,619
MyC4	Global; 7 Cs	Microfinance	For-profit	€24.35	7,523
MicroWorld	Global; 9 Cs	Microfinance	Social business	€0.43	23,995
Deki	Global; 5 Cs	Microfinance	Non-profit	£1.05	2,863
Lendwithcare	Global; 7Cs	Microfinance	Non-profit	NA	34,595
Veecus	Global; 4 Cs	Microfinance	For-profit	NA	NA
myELEN	Global; 5 Cs	Microfinance	Social business	NA	NA

Table 2.2 *List of online indirect P2P lending platforms operated globally*

Our focus with those indirect P2P platforms who capture global operations in microfinance in the following diagram (Figure 2.4):

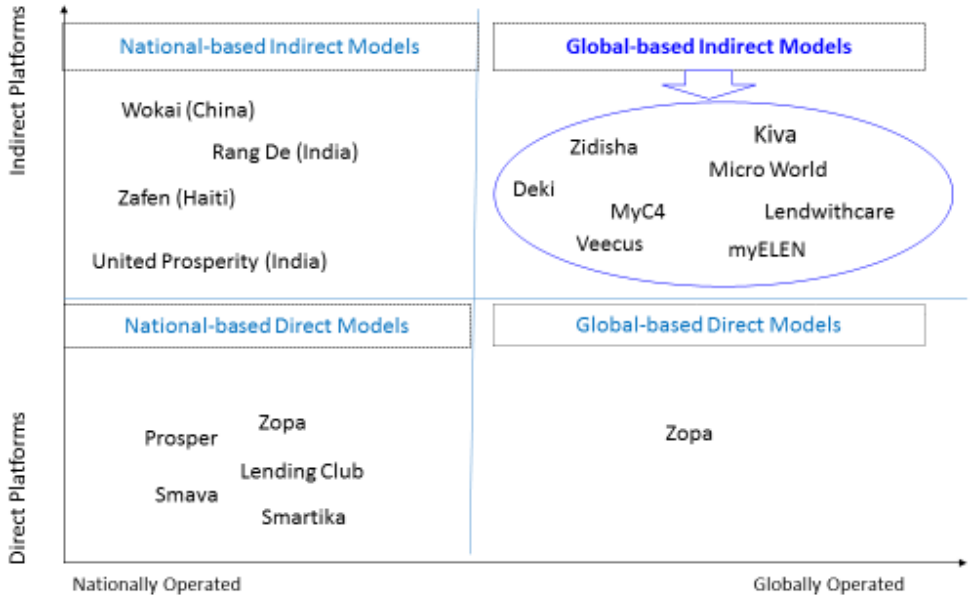


Figure 2.4*Lending Platforms with the focus on indirect globally operated models*

2.3 Borrower Selection in online P2P Platforms and Lender's Decision Making

Today's information and communication technology (ICT), especially Internet changes many things including the finance industry providing individuals to participate online microlending globally. One such opportunity is Kiva.org, the first and largest indirect P2P platforms in the world, to meet the need for entrepreneurial supports by connecting individual lenders from developed countries to the low-income entrepreneurs in developing countries as well as in different cities in the United States via local partners(R. Chen, Chen, Liu, & Mei, 2014).With its many successes, Kiva faces challenges like lenders' motivation to continue their lending after registration on the site or their first loans. Most of the members who have registered as lenders give few or no loans although membership of Kiva and its total number of loans and volume have increased remarkably(Liu, Chen, Chen, Mei, & Salib, 2012). Premal Shah, the president of Kiva, comments that many Kiva lenders remain inactive after their first loans although their loans have been repaid and thus they could make another loan at no additional cost[13]. This is not only the significant comments from the Kiva management but also a serious thinking about the strategies, here lending in team as new strategy, to make the platform effective and sustainable.

Some recent studies investigate different factors in online microfinance like lender motivations(Liu et al., 2012), biases (Jenq et al., 2012; Riggins & Weber, 2012) and sensitivity to transaction costs(Meer & Rigbi, 2012) as the challenges in online microfinance. These studies, focusing on Kiva model, identify some key factors for granting loan to the borrowers. For example, on one hand, lenders favor more attractive, lighter-skinned and less obese borrowers. On the other hand, they prefer the borrowers who they perceive to be needy, honest and creditworthy (Jenq et al., 2012).Team lenders are found more capable in

[13]https://www.youtube.com/watch?v=aEC3OwKWgfc

repayment than the individual borrowers(R. Chen et al., 2014). Due to the nature of risk and the characteristics of borrowers, for P2P lenders, it is difficult to judge the quality of the deal offered beforehand(Heng et al., 2007). According to Tan & Thoen (2000),*"pseudonymous online environments are characterized by information asymmetries, which make the exploitation of lenders particularly easy for the borrowers (opportunistic behaviour)"*. Lenders' willingness to place bids is identified as the factor on which the long-term success of the online platforms rely since *"rational, risk-neutral and profit-oriented lenders will only do so if they obtain at least as good a return as in comparable alternative investments"*(Klafft, 2008).Such online platforms, e.g., Prosper.com, were found risky to many lenders since they were unable to generate acceptable returns with their investments due to a high number of loan defaults. This situation led to unexpected and wrathful discussions in several online communities including users' threats to quit or boycott the platform[14]. However, contrary to the fact whether P2P lending is suitable at all for pseudonymous online environment, lending online is assumed to have a chance for long term success, "*if the platform actively addresses the issue of bad investments and low loan performance. Several such measures have already been taken, including offering webinars (= web-based seminars) to lenders to raise their problem awareness, or making additional verified borrower information available online"*(Klafft, 2008).Since a large portion (70% of the $291 billion) of charitable giving market consists of individual giving, it is considered as one of the important sources of capital for non-profit and social causes (USA, 2012). Hence, it is important to know the reasons of individual donations and if systematic biases affect charitable decision making. However, no relationship is found between loan performances and attributes of objective and subjective borrower as well as between borrower attributes and loan funding times. Considering these facts, it is suggested that *the observed lender biases are unlikely to be driven by statistical discrimination where lenders favor borrower attributes that are correlated with loan performance or borrower enterprise performance. The findings appear more consistent with lenders exhibiting bias, implicit or explicit, in their lending decisions* (Jenq et al., 2012; Riggins & Weber, 2012).It may be the presence of information asymmetry where relevant information are not available. So, lenders may behave rational in the case of available information on which they can decide. However, they may be bias in the case of asymmetrical information. We found the evidence of such different behaviors in the study of Jenq et al. (2012). In this study, the authors argued that despite the limitation of the paper in providing direct evidence on the extent to which observed bias is attributable to explicit or implicit discrimination, they are able to show that lenders with more experience on Kiva are less likely to fund loans in a pattern consistent with lender bias on physical characteristics. This argument lead them to interpret this as indirect evidence that *greater lender familiarity with the choice problem reduces the lender's tendency to rely on implicit mental processes – although it could also be evidence that more committed lenders simply have a different type of preference.* Clarifying the argument, the authors assume Kiva lenders face two potential considerations: *first, they are likely to care about the social impact of their loan and, all else equal, we may expect them to prefer borrowers who would maximize social impact, such as borrowers who appear more needy than others and second, while Kiva lenders are really donors, recovery of the loan principal is important since a recovered principal allows the 're-gift' of the principal to a new borrower, promoting an additional charitable goal.* Hence, it is needed to pay attention to borrower profitability and default risk by the Kiva lenders. Besides, the author also analyze the speed at which loans are

[14] https://blog.p2pfoundation.net/the-prosper-lender-rebellion-and-the-us-creditborrowing-black-hole/2007/08/16

funded as a proxy for the relative attractiveness of a given loan that since virtually all the loans on Kiva eventually receive full funding.
According to Galak et al. (2011), the context of microfinance decision making constitutes a new hybrid decision form which is called pro-social lending. This is hybrid since it consists both financial and pro-social characteristics. On one hand, from financial perspective, it shares many characteristics with conventional financial decision making (e.g., likelihood of repayment, repayment terms, etc.). On the other hand, from pro-social perspective, its stated purpose is to help others. All of these features could compel the lenders to treat the decision in a more calculative manner which could make psychological or emotional drivers ineffective(see Small, Loewenstein, & Slovic, 2007).
As lender's decision is both financial in nature as well as pro-social, risk assessment of borrower might help lender to assess the borrower more efficiently from financial perspective (Baklouti Ibtissem, 2013).

2.4 Borrower Indebtedness, Delinquency Rate and Credit Information System

As we have mentioned earlier that P2P lending platforms, characterized by information asymmetries (Tan & Thoen, 2000), operate in a pseudonymous online environments where individual lenders meet strangers (borrowers from across the world in our focused platforms) and make lending decisions without experience (Heng et al., 2007). The context poses a threat to this innovative online lending platforms for borrowers' indebtedness, particularly delinquency and default rates(H. Wang et al., 2009).

2.4.1 Borrower Indebtedness and Delinquency Rate

In many developing countries, the microfinance revolution has brought about a new type of competition in credit market between the lenders that resulted in a number of new and unexpected challenges (Baklouti Ibtissem, 2013; Luoto, McIntosh, & Wydick, 2004). Borrower over-indebtedness, reduced loan repayment incentives, and growing arrears for Microfinance Institutions (MFIs) in competitive environments are the outcomes of such competitions between the lenders(Campion, 2001; McIntosh & Wydick, 2005). Moreover, in Bangladesh, overlapping loan problems among major MFIs and borrowers has emerged as a crucial problem in the credit market. The study of Yuge (2011) mentioned two main reasons for this remarkable increase of indebtedness. One is poor people have more options in choosing MFIs to borrow money, and the other is the number of people who use multiple loans from various MFIs has been increasing. This phenomenon of increased indebted people poses threat to MFIs and to the microfinance industry since repayment among overlapping borrowers has become more and more irregular. There are different reasons behind the over-indebtedness of the borrowers found in different studies: borrowers' experience, high credit limit, levels of indebtedness (Khandker, Faruqee, & Samad, 2014; Pytkowska & Spannuth, 2011; Schicks, 2011; Soman & Cheema, 2002). Borrowers may wrongly view the size of their credit limits i.e., the limit of credit that they can really afford. Consequently, high credit limits may encourage them to borrow beyond their affordability. Moreover, the level of indebtedness increases with the number of active loan contracts. Clients with a single loan are insolvent compared to the clients who have two or more loans. Also, the share of the clients facing a critical situation and those at risk increases significantly with the number of loan.

Moreover, over-indebtedness is more often seen among experienced clients than that of the inexperienced clients. As it is mentioned in the study of Yuge (2011), overlapping is considered as an emerging problem in the credit market of Bangladesh like other developing countries, for example, Bolivia has suffered from overlapping problems in 1990s when microfinance and consumer credit confronted social turmoil due to the protests by the borrowers who asked for debt remission or waiver. In order to mitigate such kind of situation, Bolivia has been trying to introduce an effective Credit Information System (CIS). Moreover, such trend of worsening indebtedness in microfinance may push MFIs into a trap of accumulating non-performing loans which might cause their sustainability.

The same case has been witnessed in the credit market of online P2P lending platforms. Wang found high default rates with Prosper and Lending Club (both models are for-profit) accompanied with considerable delinquency rates. The author found different default rates based on Prosper's proprietary risk model (maximum 9% with good grade borrowers and 43% with poor grade borrowers). These rates are likely to be higher (more than 15% in good grade and about 60% in poor grade) if delinquency rates are considered. Philanthropic or non-profit models in P2P lending marketplaces emphasize on lending money to improve borrowers' living conditions. The usual target group of these P2P lending marketplaces are the borrower groups who have some particular needs. In these cases, Faynanz emphasizes on the need of the education loans, Lend4health emphasizes needs for health loans, Kiva prioritizes need for business loan in developing countries. In case of such loans, lenders' appetites for risk and interest rate for this model can be speculated. Although the focus in non-profit models has been given more on social welfare or giving than risk of lending, the borrowers, however, are not free from the risk of over-indebtedness and finally the risk of default to the lenders on the platforms. Because the borrowers are charged the interest by the local partners (mostly MFIs) and they need to repay the loan with interest which are taken by local partners to cover their operation cost and profit to sustain. Therefore, the repayment of principal amount or loan given to the borrowers is in the same risk as off-line microfinance or traditional microfinance.

2.4.2 Credit Information System

Conventional CIS have been functioning around the world for decades basically in the form of public and private credit bureaus for commercial lending markets. These systems are the oldest and most vigorous in the countries like US, UK, Germany, Japan, Sweden and Switzerland, in other words, in the developed countries (Jappelli & Pagano, 2000). Some factors like strong legal infrastructure, high lending volumes, advanced communication technology, borrower mobility and heterogeneity of credit events and economic activities facilitate an environment in these countries that encourage and support such a vigorous system (Luoto et al., 2004). Besides, CIS have been operating for many years at a smaller, less comprehensive scale in countries like Argentina, Brazil, Finland, the Netherlands and Australia. However, there is no or little existence of CIS in many countries in Latin America, Asia and Africa that only share negative information mainly in the form of blacklists (Rozycki, 2006).

Information Sharing System (ISS) is always an important consideration in preventing overlapping borrowing among microfinance borrowers. This importance is well established in different research works (Akerlof, 1970; Stiglitz & Weiss, 1981). In order to check overlapping borrowing among microfinance borrowers, some of the South American countries established effective ISS in the 1990s and 2000s. Since it is possible to establish an efficient information system *"which would create a screening effect that improves risk*

assessment of loan applicants" with the introduction of a credit bureau, establishing a "Credit Bureau" is one of the most effective measures to prevent overlapping (Yuge, 2011).
The paper authored by Luoto et al. (2004) argued that the weakening performance of microfinance in competitive environments is due, in part, to the absence of information sharing in these markets. CIS can help to increase the transparency of credit markets where MFIs are endeavouring to address the problem of asymmetric information between borrowers and lenders to overcome the effects of adverse selection and moral hazard. Despite the undisputed importance of CIS in credit market, in many developing countries CIS are still in their infancy and information sharing between lenders remains insignificant. Since, competition in microfinance lending exaggerates in the developing countries, borrower information is considered as the most important for this market.
Considering the crucial issues like over-indebtedness of microcredit borrowers and high rates of delinquency that threaten MFIs sustainability, it requires a system that can inform the lenders about the credit-worthiness or reputation of potential borrowers (Khandker et al., 2014; Rozycki, 2006). Otherwise, these challenges may cause a serious sustainability question like violent debtor uprisings in some cases for MFIs in global microfinance sector.

2.5 Credit Risk Management Practices and Credit Scoring

In the financial sector, microfinance has turned into a booming industry in the period 1998 to 2008 due to its growth both in MFIs and number of customers. Within this period, the number of MFIs grew by 474% while the number of customers grew by 1048%. This phenomenon has brought about changes in the nature of business of commercial banks that start to operate in the microfinance sector on one hand; on the other hand, it has created the competition between the players in this industry. However, it causes increase in the operational cost of the MFIs and poses threat to their survival in the long run. Hence, it is essential for MFIs to "*increase their efficiency in all their processes, minimize their costs and control their credit risk if they want to survival a long-term."*(Cubiles-De-La-Vega, Blanco-Oliver, Pino-Mejías, & Lara-Rubio, 2013). The main challenge of microcredit is to check and control the risk associated with a client due to his behaviour (Baklouti; Ibtissem & Bouri, 2013). For example, a borrower is assumed to be risky when he does not pay the loan or pay late or does not return for repeat loans.

2.5.1 Traditional Risk Management Practices

Basically, joint-liability groups and careful investigation of an individual's business and his character by skilled loan officers are considered as the most vital two innovations as the bases of microcredit in microfinance sector that reduce the cost of managing risk(Schreiner, 2005). In order to assess the credit risk of borrowers, usually most of the financial institutions (FIs) relied mainly on subjective analysis system or banker expert system. In so doing, generally some basic information about the borrowers [*various borrower characteristics like borrower's character (reputation), capital (leverage), capacity (earnings' volatility), collateral, and condition (macroeconomic cycle)*] are used by the bank loan officers (Vaish, Kumar, & Bhat, 2011).The borrowers and their projects are needed to be assessed with regard to credit-worthiness and business risk so that the credit risk or default rate can be reduced. This assessment is done based on the quantitative parameters or by subjective appraisal by the lender. Between the loan delivery mechanisms of group-lending and individual-lending,

the latter is more risky due to the dependency on the credit-worthiness and ability of the sole borrowers. Hence, more attention is required in selecting borrowers in the case of individual lending. Usually, risks associated with individual lending are sought to be minimized with the available collateral. Generally, potential loss of collateral motivates the individual borrower to repay the loan in time or to behave properly and this helps the borrower to have a reputation of a solvent debtor. Since microfinance borrowers lack collateral, therefore, group-lending is preferred to individual-lending (Pellegrina and Masciandaro, 2006 in Vaish et al., 2011). Group-lending has both advantages and disadvantages over individual-lending. On the one hand, they allow a member whose project yields very high returns to pay-off the loan of a group-member whose project does very badly. On the other hand, a moderately successful borrower may default on her own repayment because of the burden of having her partner's loan (Ghatak & Guinnane, 1999).Basically, the main objective of most models for focusing on explaining joint-liability group-lending and its implications is to reduce information asymmetries. Despite, there are lots of theoretical literatures on whether and how microfinance helps to reduce existing information asymmetries, there are only a few studies on investigating its empirical part(Hermes & Lensink, 2007).

2.5.2 Credit Scoring

In West (2000), *an appropriate automatic evaluation* of the credit applicants is considered as a tool that offers several important advantages like reduced credit analysis cost, improved cash flow, faster credit decisions, reduced losses, a closer monitoring of existing accounts, and prioritizing collections. As known, finance in general, and microcredit in particular, is all about managing risk. Apart from joint-liability groups and careful evaluations of an individual applicant's business and his characteristics, scoring is taken into consideration as a third risk-management innovation (to microcredit) to judge repayment risk. Scoring is regularly used in the developed countries in order to rationalize decision-making and increase profits(Schreiner, 2002, 2005). It helps to detect historical links between repayment performance and the quantified characteristics of loan applications. It also assumes those links will persist through time. Then, it forecasts future repayment risk based on the characteristics of current applications. In high-income countries, scoring (through credit cards) has been the biggest breakthrough ever in terms of providing millions of people of modest means with access to small, short, unsecured, low-transaction-cost loans. Research shows that scoring increases not only profits but also the number of clients and the number of poor clients. In general, scoring improves risk management, leading to a cascade of benefits (Schreiner, 2005). It is found favourable and profitable for both small and large micro-lenders. For small micro-lenders "...*scoring can indeed expand the efficiency frontier and so improve both poverty outreach and organizational sustainability"* since it does not only reduce time spent for collecting overdue payments from delinquent borrowers by the loan officers (a typical loan officer might save about two days per month), they can also then use some of their new-found time to search for more good borrowers. For large micro-lenders, scoring can also be profitable since it helps to reject the riskiest disbursed loan. For example, one test with historical data in Bolivia suggested that rejecting the riskiest 12 percent of loans disbursed in 2000 would have reduced the number of loans that reached 30 days overdue by 28 percent (Schreiner, 2001 in Schreiner, 2002).

2.5.2.1 Credit Scoring in Microfinance

Scoring does not have a long track record in microcredit in wealthy countries as well. Though there are pilots and proof-of-concept tests with past data, in practice, there are no long term uses of scoring in microcredit. Basically, dependence on the limited funding available from social-minded donors has undersized wide scale development of microcredit. Two reasons have been identified why profit-minded investors have been biding their time: on one hand, it is risky to investigate in a new industry since microcredit returns are too low to compensate for the risk; on the other hand, there is also uncertainty about 'unknown risk' of investing in microcredit(Schreiner, 2005).However, bankers and investors understand lending that are based on scoring is better than the evaluation technique of group lending and individual lending. In this case, scoring is treated as a technique that can help to reduce the uncertainty about the risk of investment since scoring helps in centralized decision-making and give non-specialist investors more confidence that they can maintain effective control. When scoring is used for evaluating investment risk, an amateur investor in microcredit can evaluate investment risk with more confidence because scoring process is familiar with the investor and it helps to quantify the risk of the microlender's loan portfolio (Ayayi, 2012; Schreiner, 2005). Therefore, microlenders can adapt credit scoring to leverage the benefits though it is less powerful in microcredit in developing countries than the consumer credit in wealthy countries. However, it will be complementary to the existing approaches (Schreiner, 2002; Serrano-Cinca, Gutierrez-Nieto, & Reyes, 2013) and will add a final hurdle, detecting some high-risk cases that slipped through standard screens. Scoring cannot approve applications, but it can reject applications that would otherwise have been approved or flag applications for additional analysis and possible modifications to the loan contract(Schreiner, 2005).
The progress of credit scoring models in the microfinance sector is very insignificant. Table 2.3 gives an overview on credit scoring models in microfinance.

Author (Date, Country)	**Institution type**	**Sample size**	**Number of (included) inputs**	**Description (Technique(s) & Variables)**
Vigano (1993, Burkina Faso)	Microfinance (individual)	100	53 (13)	Discriminant Analysis
Sharma and Zeller (1997, Bangladesh)	Microfinance (group)	868	18 (5)	TOBIT Maximum Likelihood Estimation
Zeller (1998, Madagascar)	Microfinance (group)	146	19 (7)	TOBIT Maximum Likelihood Estimation
Reinke (1998, South Africa)	Microfinance (individual)	1641	8 (8)	Probit Regression
Schreiner (1999, Bolivia)	Microfinance (individual)	39 956	9 (9)	Logistic Regression
Vogelgesang (2003, Bolivia)	Microfinance (individual)	8002	28 (12)	Multinomial Logistic Regression Random Utility Model
Vogelgesang (2003, Bolivia)	Microfinance	5956	30 (13)	Random Utility Model
Diallo (2006, Mali)	Microfinance (individual)	269	17 (5)	Logistic Regression, Discriminant Analysis
Deininger and Liu (2009, India)	Microfinance (group)	3350	15	Tobit Regression
Van Gool et al.	Microfinance	6722	16	Logistic Regression

(2012, Bosnia)	(individual)			
Serran-Cinca et al. (2013, Colombia)	Microfinance (individual)	1	26	Multiple-attribute Utility Theory (MAUT); Multiple-attribute Value Theory (MAVT); Analytic Hierarchy Process (AHP)
Blanco et al.(2013, Peru)	Microfinance (individual)	5500	39	Neural Networks
Cubiles-Di-La-Vega et al. (2013, Peru)	Microfinance (individual)	5451	39	LDA, QDA, LR, CART, MLP, Bagging, Boosting, SVM, RF
Gutiérrez-Nieto et al. (2016, Colombia)	Microfinance (individual)	1	26	Multiple-attribute Utility Theory (MAUT); Multiple-attribute Value Theory (MAVT); Analytic Hierarchy Process (AHP)

Table 2.3 *Credit Scoring Models in Microfinance*

2.5.2.2 Parametric Statistical Techniques

Despite the fact that the non-parametric methodologies are considered as the best as the classical statistical models, the existing models are based on parametric statistical techniques, mainly linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR)(Lee & Chen, 2005; West, 2000). In order to improve the performance of credit scoring models the microfinance industry has not yet received the benefits of the advantages of non-parametric techniques which results in a failure in competing on equal terms with the international commercial banks who are considered as their new competitors. This phenomenon is an outcome of lack of literature on credit scoring model designed for the microfinance industry applies a non-parametric methodology (Cubiles-De-La-Vega et al., 2013). Similar evidences have also found in other literatures that parametric methodologies like LDA and LR have been used in the development of credit scoring models for MFIs (Dinh & Kleimeier, 2007; Gool, Verbeke, & Baesens, 2012; Kinda & Achonu, 2012; Rayo, Lara, & Camino, 2010; Schreiner, 1999; Sharma & Zeller, 1997; Vigano, 1993; Vogelgesang, 2003; Zeller, 1998). Basically, the strict assumptions, like linearity, normality and independence among predictor variables, of these statistical models limit their application in the real world, along with the pre-existing functional form that relates response variables to predictor variables. It has been identified that, in case of applying to credit scoring problems, two basic assumptions of LDA are often violated: *"(a) the independent variables included in the model are multivariate and normally distributed, (b) the group dispersion matrices (or variance–covariance matrices) are equal across the failing and the non-failing groups"* (Eisenbeis, 1978). Apart the fact that LDA is reported to be a more robust and precise technique, in the cases where the covariance matrices of the two populations are unequal, theoretically, QDA is recommended to be adopted(Dillon & Goldstein, 1984 in Blanco et al., 2013). Like LDA, LR is found most favourable under the assumption of multivariate normal distributions with equal covariance matrices. Moreover, it is also found that LR remains most favourable in a wider variety of situations. Yet, in order to obtain stable results LR requires larger data sets. Besides, interactions between predictor variables are needed to be formulated. Moreover, to obtain stable results complex non-linear

relations between the dependent and independent variables could be incorporated through appropriate but not evident transformations. Our review also found that, for these reasons recently non-parametric statistical models have been successfully applied to credit scoring problems.

2.5.2.3 Non-Parametric Statistical Techniques

Among the non-parametric statistical models, the k-nearest neighbour (KNN) algorithm, support vector machines (SVM), decision tree (DT) models, and neural network (NN) models are most significant (Vapnik, 1999). Artificial neural networks (ANNs) comprise one of the most powerful tools of these for pattern classification for their non-linear and non-parametric adaptive-learning properties. Since the default prediction accuracies of ANNs are better than those using classical LDA and LR, we have found many studies that compared ANNs with other classification techniques in the field of credit scoring models(Blanco, Pino-Mejías, Lara, & Rayo, 2013; Che, Wang, & Chuang, 2010; Desai, Conway, Crook, & Overstreet, 1997; Desai, Crook, & Overstreet, 1996; Hand & Henley, 1997; T. S. Lee, Chiu, Lu, & Chen, 2002; T.-S. Lee & Chen, 2005; Malhotra & Malhotra, 2002; Piramuthu, 1999; West, 2000). Despite the fact that ANNs yielded satisfactory results in the field of credit scoring models, it has some disadvantages as well, for example its black box nature and the long training process involved in the design of the optimal network topology(Chung & Gray, 1999 in Blanco et al., 2013).

2.5.2.4 Statistical Learning Techniques

According to current literature, in non-microfinance environments a wide range of supervised classification algorithms have been successfully applied for credit scoring (Cubiles-De-La-Vega et al., 2013; Hens & Tiwari, 2012). There are many papers providing empirical evidences supporting these alternative algorithms in credit scoring (Ince & Aktan, 2009; Kim & Sohn, 2010; T. S. Lee et al., 2002; Malhotra & Malhotra, 2003; West, 2000). However, similar works in the microfinance field are still expected to be done. By using neural network, Blanco et al. (2013) developed credit scoring models for the microfinance industry. In this paper, the authors construct several non-parametric credit scoring models based on the multilayer perceptron approach (MLP) and uses their performance as yardsticks against other models which employ the traditional LDA, QDA, and LR techniques. The results reveal that neural network models outperform the other three classic techniques both in terms of area under the receiver-operating characteristic curve (AUC) and as misclassification costs.
In the work of Cubiles-De-La-Vega et al. (2013), the authors developed credit scoring models for MFIs based on statistical learning techniques (LDA and QDA, LR, MLP, SVM, classification trees (CT), and ensemble methods based on bagging and boosting algorithm). They claimed that there is a lack in developing credit scoring using classical statistical method which is surprising since "*the implementation of credit scoring based on supervised classification algorithms should contribute towards the efficiency of microfinance institutions, thereby improving their competitiveness in an increasingly constrained environment"*. They explored an extensive list of statistical learning techniques as microfinance credit scoring tools from an empirical viewpoint. In their work, they considered a data set of microcredit belonging to a Peruvian Microfinance Institution. They applied their models to decide between default and non-default credits, in other words, they used the models in LDA and QDA, LR, MLP,SVM, CT, and ensemble methods based on bagging and boosting algorithm and found that, with the implementation of this MLP-based model, the

MFIs´ misclassification costs could be reduced with respect to the application of other classic models.
It is also found that credit scoring algorithms in microfinance sector have been mainly based on statistical techniques mainly LDA, QDA, Probit regression (PR), and LR (Deininger & Liu, 2009; Dinh & Kleimeier, 2007; Rayo et al., 2010; Sharma & Zeller, 1997; Vigano, 1993; Vogelgesang, 2003; Zeller, 1998) which are considered as less fitted to credit scoring problems due to the violations of the assumptions LDA and QDA (Karels and Prakash, 1987 in Cubiles-De-La-Vega et al., 2013).The mixed nature of quantitative and qualitative data and the high non-linearity in the association between the target variable and the predictors are generally appeared as problems in credit scoring data sets. These problems can be faced with statistical learning algorithms, which is a framework for machine learning with a strong statistical basis. In this case, data mining is considered as an important element of statistical learning(Hastie, Tibshirani, & Friedman, 2001). Knowledge Discovery from Data (KDD) is a process that contains both statistical learning and data mining and which is oriented to identify patterns in data sets(Fayyad, Piatetsky-Shapiro, & Smyth, 1996).
Taking into consideration the importance of effectiveness and competency of management and control of credit risk, credit scoring is considered as one of the most significant uses of technology that may influence management of MFIs. Other authors also claim that implementation of credit scoring not only improves the judgment of credit risk and helps in cutting costs of MFIs (Schreiner, 2005), but also incorporates social parameters to evaluate social aspects of this lending (Gutierrez-Nieto, Serrano-Cinca, & Camon-Cala, 2016).

2.6 Conclusion

The meta-analysis of literature review gives insights on how the P2P platforms got success in pro-social lending and how platform opened the access to the borrowers to avail the loan without the affiliation of any group. The review finds lenders always face challenges in choosing a borrower among many candidates on such platforms, particularly for individual lenders who are not expert in lending. Moreover, lenders are provided with little information, which lack the details of the financial aspects, particularly risk assessment of the loan applicants and eventually they are confronted with judging the worthiness of applicants for which making their lending-decisions is really a tough job. Different risk management tools are practiced in the sector but most of them are for group borrowers. Most importantly, risk rating of borrowers is not provided to the lenders on indirect P2P platforms. This lack of risk rating of borrower being embedded to P2P is surprising since credit scoring could help the online P2P model's lenders to evaluate the loan applicants more efficiently and thereby enable lenders to match their lending risk perception with the degree of risk associated with a particular loan applicant.

Chapter 3

The Methodological Approach on Risk Rating for Microfinance

The scope of this research is in the arena of online indirect P2P lending models that facilitate lending service in microfinance globally. Such microlending platforms are Kiva, Zidisha, MyC4, Microworld, Deki, Lendwithcare etc. that connect people through lending to alleviate poverty. Among the platforms, Kiva model is the largest and leading one (Hassett et al., 2011)which has been chosen for this study. Kiva allows researchers to its open-access database of large number of borrowers that can meet all the necessary requirements for fulfilling the objectives of this research. The population size of this database represents true attributes of diversified set of more than 2.3 million borrowers' loan history from 82 countries in 8 zones with 302 field partners from 2006 to till date across the world (www.kiva.org).

3.1 Research Design

Considering the problem undertaken for this research study and its context, Case-Based Reasoning (CBR) has been chosen as a method which fits best as one of the successful techniques of Artificial Intelligence (AI). It has been chosen as prime methodology to assess borrower risk in online indirect P2P models, here *Kiva* is the particular model, sinceno other statistical models fit well with the unique nature of borrower profiles in microcredit where nature of borrower characteristics demands for special knowledge and adequate relevant data other than financial performance data exist in online indirect P2P lending platforms for global borrowers. Hence, CBR system works as a statistical model to improve the results (risk ratingor prediction) ofjudgmental/expert rules through bootstrapping process (Bunn & Wright, 1991, p.505).
Unlike corporate finance, opaque microfinance borrower varies from one another based on his/her intention of borrowing and capacity to repay the borrowed money without guarantee/collateral security and formal documentation of financial reporting. Moreover, CBR technique has been used in corporate finance to predict the market (Oh & Kim, 2007) and in retail banking for consumer credit loans, credit card loans through credit scoring of borrowers to assess their possibility to repay the loan. Such CBR-based credit scoring has been done in mostly developed country like Australia, Germany where borrowers database are available for tracing their history (Vukovic, Delibasic, Uzelac, & Suknovic, 2012). However, no work, using CBR, has been done yet to represent the borrower's profile from developing nations/regions like Africa and Asia to the lenders in P2P microfinance lending platforms (like kiva). Hence, CBR technique is more suitable than other approaches due to the lack of generally accepted credit decision models in P2P microfinance platform except the basis of borrower's raw/unstructured profile that contains business information, biographic information as well as field partner's and country's information(see details in Chapter 4).

The main issue with the application of this approach to the present problem is certainly not the lack of data. In fact, Kiva makes available all the information associated to past loan requests and to the actual repayments made by the borrowers. All the information necessary to define a case description is available (see details in Chapter 5), and also the final outcome is known (the information about the actual repayments), but of course no actual risk rating is present, in Figure 3.1, and therefore all cases would be missing the solution part. To solve

this cold boot problem, it has been decided to adopt a strategy to select a reasonable number of past loans that are sufficiently representative of all the countries, economical sectors for the funded activities, kind of borrowers, and actually rate them (filling thus the solution part of the case) employing expert rules for rating the risk associated to loan requests in developing countries, coded into a spreadsheet. This activity cannot, as of this moment, be completely automated due to the need to interpret elements of the borrower description written in natural language and not structured in fields of a database. Moreover, the above mentioned rules are not completely formalized and the experts sometimes manually modify the results of their direct application to define the risk rating.

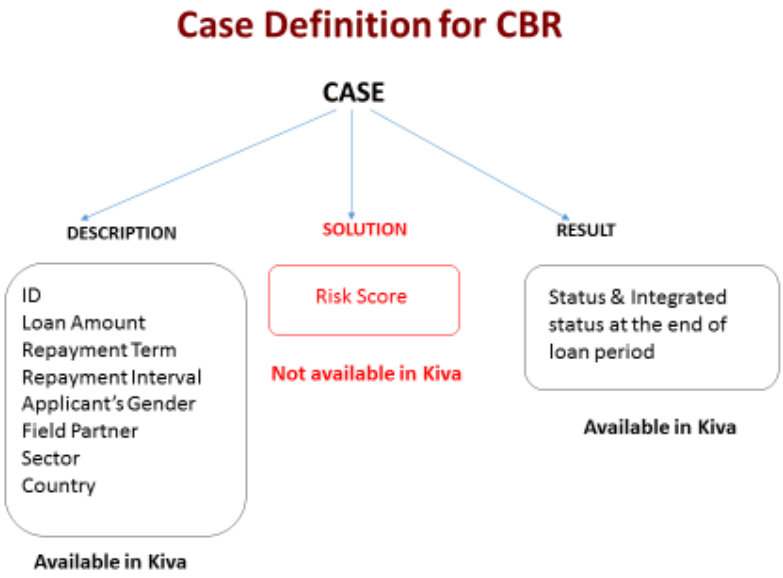


Figure 3.1*Case structure in CBR cycle*

Expert rules have been chosen for using knowledge-based rating (Baklouti; Ibtissem & Bouri, 2013) for providing the missing part *solution (risk scoring)* to make the loan cases complete to use in CBR system/approach. Because knowledge/judgement-based rating works well where opacity problem and little data exist(Bunn & Wright, 1991, p.505). The context of microcredit lending in online P2P platforms especially in developing countries conforms both opacity issue and little data availability for which expert rules are justified. Therefore, a constrained expert modelor integrated model have finally been chosen combing expert-based manual model (expert-judgment approach or knowledge-based approach) with automated statistical model (CBR approach) in Figure 3.2.

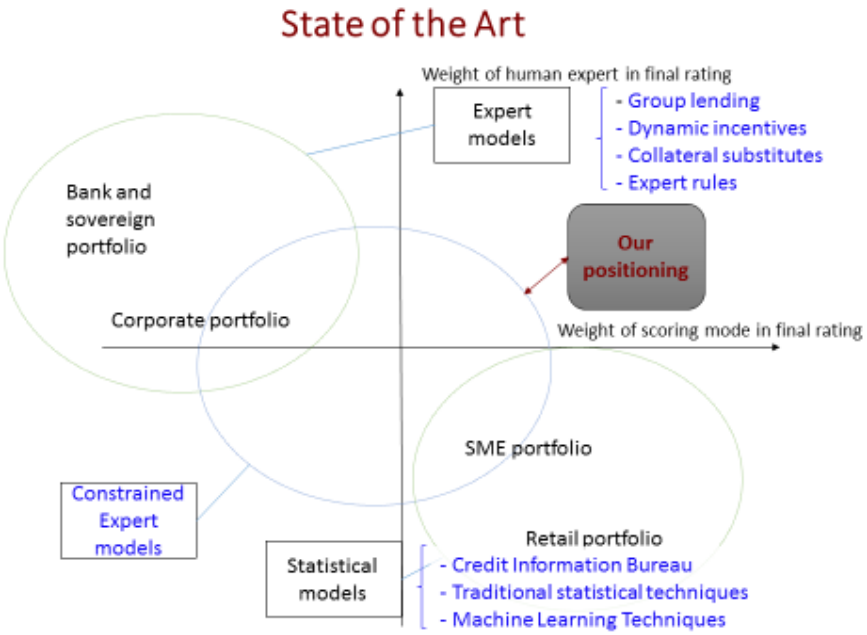


Figure 3.2*Current bank practices: rating system (Balthazar, 2006: From Basel 1 to Basel 3, p.118*

Other models like group lending approach, dynamic incentives, or collateral substitutes that work well for borrowers in group lending but do not fit with the risk assessment of individual lending in microcredit system, particularly in P2P lending platforms(Armendáriz de Aghion & Morduch, 2005; Kono & Takahashi, 2010).

3.2 Data Collection

Kiva has the scope to access its open source data for the study. Therefore, Kiva XML datahave been recovered, in Figure 3.3, using XQuery to organize an adhoc database for past loans of individual borrowers (unit of analysis) with representative numbers and then examination method has been used for identifying relevant and readily extractable features for the sample of past individual borrower loans.

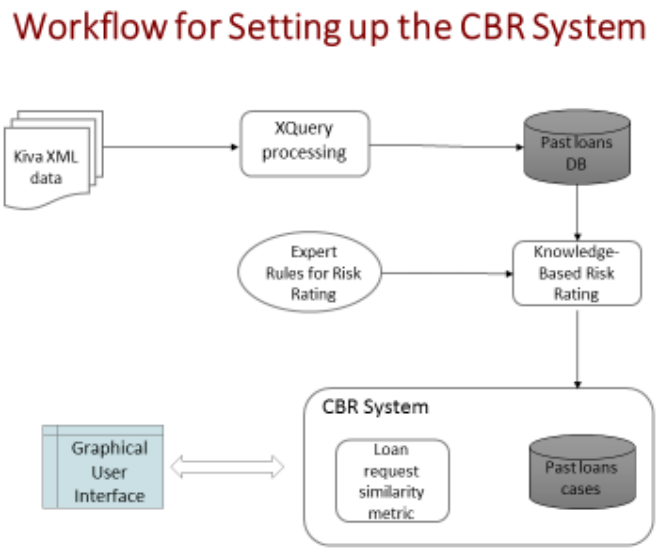


Figure 3.3*Workflow for CBR-based borrower risk assessment in P2P lending platforms.*

In this study, secondary datahave been usedfrom Kiva database since the research objective (borrower risk rating using statistical model) demands for historical data on which an adhoc new *casebase* can be developed. The database of Kiva is large enough to qualify the requirements of large size database for CBR application. From open source database of Kiva, only African and Asian zoneshave been chosen with 45 countries which count more than 50% coverage of Kiva's original database in terms of country covers (82 countries). The reason for choosing these two zones are the homogeneity in terms of loan size and nature of borrower's activities. Only individual borrower's loan data have been chosen as unit of analysis skipping group borrower data for selected variables.

After examining the Kiva XML data, expert ruleshave been used to select variables, assign values and weights on the selected variables based on five expert opinions. This has been done using a spreadsheet coding which was redesigned by adopting a framework used in previous similar works (Gutierrez-Nieto, Serrano-Cinca, & Camon-Cala, 2016; Point, n.d.; PKSF) and finally risk scoring has been developed based on 13 selected variables.

These selected variables represent uniqueness in this model in terms of multi-country data, data availability in practice in most of the online P2P lending models, and common and general data across the countries in the world. Moreover, from experts' perspectives, the type and number of features have been selected seem reasonable to build a parsimony, but a predictive model for giving a light on the level of risk associated with the borrowers in online P2P lending platforms. It is noticeable that there exists *missing data* for borrower's financial data concerning business which have been mitigated by the data used from field partner as proxy. The five selected variables of field partner deem reasonable as these data are being used as proxy to represent the riskiness of the borrowers as comprehensive metrics. The selection of variables has been done based on the distribution of each variable. Using this expert-based models *credit scoring* has been done for a set of representative cases of 107 loans and then they have been used in CBR system as complete loan cases to run the system as bootstrap problem for assessing new loan applicants or new borrowers. Finally, the CBR

rating has been tested with a set of test loan cases (75 cases from holdout sample from 2014) for evaluating its predictive power.

3.3 CBR System

The CBR-based prototypical solution provides borrower risk assessment to the users in online P2P lending platforms. The users with the support of Graphical User Interface (GUI) can access the CBR application that links with a proper set of initial relevant past loan cases using a similarity function to get the most similar case with its solution. The solution of the most similar case is reused to assess the risk of new loan applicant or new borrower by which the users or lenders can get an idea about the level of risk of the borrower whom they wish to lend by matching the degree of risk to their risk tolerance attitudes/perceptions. For example, high risk borrower might be fitted with risk aggressive lenders and low risk borrower might be chosen by risk avert lenders as informed decision taken by themselves. If the reused solution (risk score) does not fit with the new problem description of the applicant (as the most similar past loan case may not be perfectly or cent percent similar to the new borrower or loan applicant), then this score can be revised by the users to adapt it perfectly. Finally, this application gives the opportunity to retain the new solution with the case description of the loan applicant or new borrower which is incremental learning or new learning to the system that generates autometrically and improve the borrower risk prediction in the system as bootstrapping problem.
This CBR-based prototypical solution or system is composed of three major components: the database, CBR application, and a GUI. A SQL-based ad hoc database has been created and populated with the open source data from Kiva model (recovered through XQuery language) for a representative set of relevant past loan cases with description and results. Then, the missing element (solution or risk score or risk prediction) of a case structure has been fulfilled by using expert-based borrower risk scoring to make the cases complete in the database. The CBR application has been done based on jCOLIBRI platform (a java-based application framework (API) for supporting or implementing CBR system) where similarity algorithm was done following nearest neighbor method with the weights taken from expert-based model. Finally, the GUI has been developed based on the GWT web application (a development toolkit of Google Web Toolkit for browser based applications)(see details in Chapter 5).

3.4 Database

An ad hoc database is being used for hosting data from Kiva model for the loan history of past borrowers to use as past cases in CBR system. The specific use of this database is to hold the past loan cases as previous borrower history to find the solution for the new loan applicant or new borrower from the similar borrowers in the past. The main reason for creating an ad hoc database is to get a *casebase* for previous loan history with the only relevant features of the borrower to be linked with the CBR system. As Kiva open source database (consists of both relevant and other additional features of the borrower and more importantly there is missing data for solution of the borrower case) does not fit directly with the purpose of the research, it became essential to have an ad hoc own database for a proper set of relevant previous borrower data with selected features from Kiva database and the data for the solution part of the case structure from expert-based scoring model.
This database has been developed based on MySQL Workbench (a tool for designing, development and administration of database) and has been populated with the open source data from Kiva model (recovered through XQuery language) for a representative set of

relevant past loan cases with description and results. The Entity Relationship (ER) diagram helped the *database* to relate the selected variables each other and store them orderly. For considering the past loans of borrower history as complete cases, the missing element (solution or risk score) has been gathered from expert-model and has finally been updated the *casebase* with all required and relevant data (see details in Chapter 5).

3.5 Borrower Risk Scoring

Borrower risk scoring has been done in spreadsheet coding based on expert rules that are the expression of credit experts in objective way for their subjective judgments and opinions. The main reason for the use of expert-based borrower risk scoring is to provide the solution (here, risk scoring) to make the cases as complete in the case base of CBR system. As an existing dominant system for borrower risk assessment, expert-based rating works well in traditional brick-and-mortar models (models of MFIs). However, it does not fit with the web-based delivery platforms (P2P models) where thousands of loan applicants or borrowers need to be assessed daily. The most significant limitation of this expert system is that it is not fully automated for which it demands for high user time (requirements) resulting higher operation cost. Some other limitations of this system are: not applicable on large scale, needs maintenance, no learning, and computationally expensive (Figure 3.4). Therefore, for overcoming these limitations of this effective system, CBR approach has been chosen which is completely automated and has incremental learning. In CBR system, a proper set of initial relevant cases has been used by taking solution and similarity function from expert model.

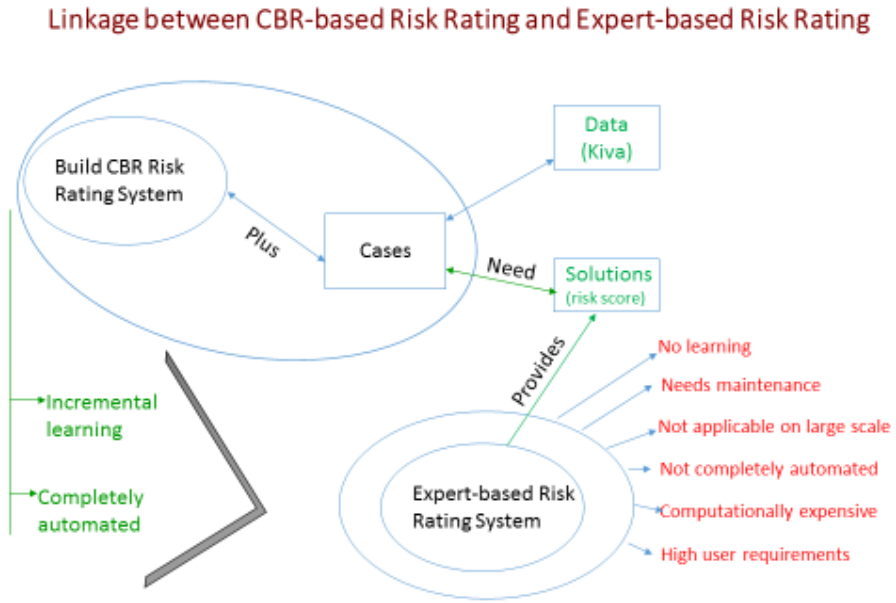


Figure 3.4*CBR-based risk model vs Expert-based risk model.*

In expert-based model, the assessment framework has been designed based on the similar framework used in previous studies(Gutierrez-Nieto, Serrano-Cinca, & Camon-Cala, 2016; Point, n.d.; PKSF).In this model, variable selection, value assignment (variable translation into a scale or value statement) and weights giving on the selected variables have been done based on five expert opinions. The initial assigned weights of selected variables for scoring

and similarity function have been refixed based on the distribution of the sample data set. Finally, the obtained credit risk score has been validated in holdout sample taken from the period of 2014 (see details in Chapter 6).

3.6 Summary

The area of this research is online indirect P2P lending models that facilitate lending service in microfinance globally. The CBR approach has been chosen as a method which fits best as one of the successful techniques of AI. It has been chosen as prime methodology to assess borrower risk in online indirect P2P models and expert rules have been chosen for using knowledge-based rating for providing the missing part *solution (risk scoring)* to make the loan cases complete to use in CBR system. Therefore, a constrained expert modelor integrated model have finally been chosen combing expert-based manual model with automated statistical model (CBR approach).In this study, secondary data have been used from Kiva database since the research objective demands for historical data on which an adhoc new *casebase* can be developed. The database of Kiva is large enough to qualify the requirements of large size database for CBR application. From open source database of Kiva, only African and Asian zoneshave been chosen with 45 countries which count more than 50% coverage of Kiva's original database in terms of country covers (82 countries). Only individual borrower's loan data have been chosen as unit of analysis skipping group borrower data for selected variables.After examining the Kiva XML data, expert ruleshave been used to select variables, assign values and weights on the selected variables based on five expert opinions. This has been done using a spreadsheet coding which was redesigned by adopting a framework used in previous similar works and finally risk scoring has been developed based on 13 selected variables. It is noticeable that there exists *missing data* for borrower's financial data concerning business which have been mitigated by the data used from field partner as proxy. Using this expert-based models *credit scoring* has been done for a set of representative cases of 107 loans and then they have been used in CBR system as complete loan cases to run the system as bootstrap problem for assessing new loan applicants or new borrowers. Finally, the CBR rating has been tested with a set of test loan cases (75 cases from holdout sample from 2014) for evaluating its predictive power.

Chapter 4

Case-Based Reasoning System in Microfinance

Case-Based Reasoning (CBR[15]) is a paradigm or an approach to solve a new problem using the solution from an old similar situation or case. The standard delineation of CBR was first devised by Riesbeck and Schank (I. Watson, 1999): "A case-based reasoner solves problems by using or adapting solutions to old problems". From this perspective, CBR can be described by an example in the work of Aamodt & Plaza (1994). The authors viewed CBR as a means or system to solve a new problem by remembering an old similar situation to adopt directly or adapt information and knowledge of that situation. For instance, "a financial consultant working on a difficult credit decision task, uses a reminding to a previous case, which involved a company in similar trouble as the current one, to recommend that the loan application should be refused".

As a problem solving paradigm, CBR is fundamentally different from other AI approaches. While other AI approaches are relied on general knowledge of a problem domain, CBR is based on special or expert knowledge gained through previous similar situations. In addition, unlike other AI approaches, CBR is a dynamic one which is able to utilize sustained learning by retaining the new solution in the domain (Aamodt & Plaza, 1994).

The basic ideas of CBR were coined by the desire to understand how a human being usually solves a problem and what is the process of recalling any previous similar problem solved with its solutions (specific information or knowledge)(I. Watson, 1999). In the study of Slade (1991), the process of remembering the similar problem in the past is described clearly. Here, past episodes (cases) are the driving forces that represent experience of an expert (human being or knowledge domain) to solve a new problem by recalling similar case (successful or failure). Also, it requires the know-how to modify the recalled case to fit a new situation. For doing this, CBR is a general paradigm for reasoning from experience. This paradigm, based on memory model[16], runs on a scientific cognitive model for the representation of episodic knowledge, memory organization, indexing, case modification, and learning. Besides, computer-based CBR improves knowledge acquisition and robustness by addressing many of the technological limits of standard rule-based expert systems.

Since the inception of CBR idea, it has been applied to divergent domains from its initial specific and insulated research area (Aamodt & Plaza, 1994). Several studies have shown empirically the role of specific, past experienced situations in human problem solving (Ross, 1989 in Aamodt, 1994). Schank (1982) developed a theory of learning and remembering assuming the retention of previous experience in a dynamic, evolving memory[17] structure. Anderson (1983) found the use of previous situations as models when people try to solve

[15]In some cases, CBR is considered as an artificial intelligence (AI) technology like rule-based reasoning, neural networks or genetic algorithms. However, in this case, CBR is used as a methodology (Watson, 1999). In Watson study Checkland and Scholes (1990) describe a methodology as: "an organised set of principles which guide action in trying to 'manage' (in the broad sense) real-world problem situations."

[16]Memory model represents, indexes, and organizes past cases and processes the model for retrieving and modifying old cases and assimilating new ones.

[17] The term 'memory' is often used to refer to the storage structure that holds the existing cases, i.e. to the case base. A memory, thus, refers to what is remembered from previous experiences. Correspondingly, a reminding is a pointer structure to some part of memory.

problems, particularly in the early learning. The same findings have been evidenced by (Kolodner, 1988 in Aamodt, 1994). Other research studies like analogical[18] research showed the frequent use of previous similar cases in solving new and different problems(Carbonell, 1986; Gentner, 1983).Even the idea of CBR has been enriched from theories of concept formation, problem solving and experimental learning within philosophy and psychology(Smith & Medin, 1981; Tulving, 1972).

The first CBR was the CYRUS system, developed by Kolodner (1983) at Yale University and then the next system was developed by Porter & Bareiss(1986) at the University of Texas, Austin. At the beginning, machine learning problem of concept learning was the prime issue for classification tasks. This led to the development of the PROTOS system by Bareiss (1989) focusing on integrating general domain knowledge and specific case knowledge into a unified representation structure. With the continuation of this trend across the world (see e.g., DARPA-1991; IEEE-1992; EWCBR-1993), the increased number of publications on CBR are found available in any AI journal (Aamodt & Plaza, 1994).

Following the trend, a lot of help desk applications initially exists for a more general coupling of CBR- and AI in general -to information systems. The practice of case application to human interaction-based decision making accelerates the attention in intelligent computer-aided learning, training, and teaching. Both human-computer interaction within flexible control environment and the motivation towards total inter-activeness of systems favors a case-based approach to intelligent computer assistance, since CBR systems are able to continually learn from, and evolve through, the capturing and retaining past experiences (Aamodt & Plaza, 1994).

CBR methods have been successfully applied to realize knowledge-based decision support systems in airline industry for optimizing heat treatment of composite materials (Hannessy & Hinkle, 1992)and in shipping line industry for solving non-conformances frequent problems of selecting appropriate mechanical equipment (Brown & Lewis, 1991). Moreover, other applications of CBR system are continuously in test or regular use. Among them many applications are rapidly growing as “Tool” or “help desk systems” (Kolodner, 1992) utilizing indexing and retrieval methods to retrieve cases for information purpose as a first step towards a more full-fledged CBR system (Aamodt & Plaza, 1994). The ReMind from Cognitive Systems Inc., CBR Express/ART-IM from Inference Corporation, Esteem from Esteem Software Inc., and Induce-it (later renamed to CasePower) from Inductive Solutions Inc. are such few CBR application tools (Harmon, 1992 in Aamodt, 1994).

4.1 CBR in Microfinance

CBR has many applications in finance from different perspectives. Among the divergent applications, forecasting and monitoring are the prime focuses in different branches of finance like corporate (Chun & Park, 2006; Oh & Kim, 2007), SMEs (Moon & Sohn, 2008) and even credit scoring in banking(Chuang & Lin, 2009; T.-S. Lee & Chen, 2005; Vukovic et al., 2012; G. Wang, Ma, Huang, & Xu, 2012; Yap, Ong, & Husain, 2011). In addition, CBR applications are found in predicting business failure (Li & Sun, 2011) and bankruptcy (Jo, Han, & Lee, 1997; Min & Lee, 2008; Shin & Han, 2001; Ye, Yan, Wang, Wang, & Miao, 2011).Most of the applications have been found in corporate sector relating to accounting, portfolio management, decision support and associated areas (Mechitov, Moshkovich, Olson,

[18] CBR and analogy are sometimes used as synonyms having a different focus: CBR is intra-domain and analogical research is based on across domains (Carbonell, 1986).

&Killingsworth, 1995; O'Roarty, Patterson, McGreal, & Adair, 1997). The use of CBR methods is widely appreciated in credit scoring(T.-S. Lee & Chen, 2005) for corporate clients (Chuang & Lin, 2009).

Risk assessment, a fundamental activity in lending, is done mostly using statistical tools like logistic regression along with data mining tools like neural network (NN), k-nearest neighbor (KNN), CBR. Although all the above tools have been widely applied for assessing lending risk in corporate finance (large scale, better structured), and even risk scoring for consumer loan as well as credit card customers in developed countries, a few of them has been employed in microfinance (small scale, unstructured) in developing countries. Unlike corporate finance, no specific rule can be applied to borrower selection in microcredit[19]. It may be due to lack of available data and its proper structure. Such non-availability of structured database in microfinance might be the result of feasibility issues among other prime causes. Moreover, microcredit borrower varies from one another based on borrower's personal information, loan requirements, & repayment features {data on borrower's profile} for which special knowledge is required or useful. The prime typical risk assessment system is the use of special knowledge of loan officers who gain experience through long service period (Schreiner, 2005). In most of the cases, Financial Institutions (FIs)usually relied on subjective analysis or expert system to evaluate borrower's credit risk. Credit experts used information on borrower's business and personal characteristics like borrower character (reputation), capital (leverage), capacity (earnings volatility), collateral(security), and condition (macroeconomic cycle) (Nair et al., 2011).Beyond experience, loan officers' diverse ability is matter to sense bad risks and they may take time to learn the riggings and hone their sixth sense (Schreiner, 1999). The theory using skilled credit officers' subjective judgment and joint-liability model are only the methods under group-lending approach in microfinance to solve the problem of borrower-selection based on risk. However, there is no existing reliable risk modeling tool in online Peer-to-Peer (P2P) microcredit lending platforms except the reliance of field partner's risk rating and other pieces of advisory information. Such information on P2P platforms cover lending portfolio diversification through choosing different countries, field partners, or sectors by online microcredit lenders (Uddin et al., 2015a). Fortunately, recently there exist available open access database (build.kiva.org) of microcredit borrowers in developing countries which gives an opportunity to exploit data mining approaches to tackle the risk of borrowers in online platforms. Therefore, considering the overall environment of special knowledge-based microcredit system with the availability of large volume open access database, CBR-based risk assessment approach can give better result than other approaches.

4.2 CBR Process

The prime functions of CBR process are to get the current problem as a new case, find a previous problem or situation in the case base similar to the new one, use the solution of that old case to propose a solution to the current problem, evaluate the suggested solution, and update the process by learning from this experience. It is based on the reasoning by analogy method (i.e., similar problems have similar solutions). In Aamodt & Plaza (1994), the CBR approach is described through a cycle of four activities which is well-known as 4R's cycle. The activities are *Retrieve, Reuse, Revise* and *Retain* which are illustrated in Figure 4.1.

[19] Microcredit is a loan delivery product in microfinance sector which also includes other products or services like microsavings, microinsurance, money transfer etc.

The authors described the process in a cyclical form where a new problem is solved by retrieving one or more previously experienced cases, reusing the case in one way or another, revising the solution based on reusing a previous case, and retaining the new experience by incorporating it into the existing case-base (knowledge-base).

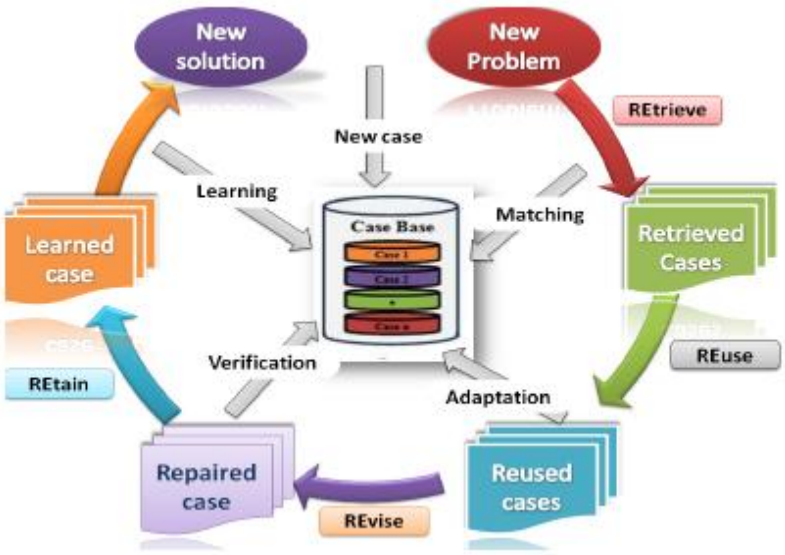


Figure 4.1*The CBR Cycle*

Here, a new case is defined by an initial description of a problem (top of the Figure 1). Then, this new case is compared to the description of similar problems already solved and stored in the case base according to similarity algorithm. The most similar problem description is then *Retrieved* and its solution is *Reused* as a first attempt to solve the new problem without starting from scratch. This solution is tested for success through the *Revise* process. Revise phase can be done in various ways: this case can be applied to the real world environment or it can be evaluated by a teacher. In case it is failed, it can be repaired. Finally, useful experience (new problem description and its solution) is retained for future reuse during *Retain* phase, and the case base is updated by a new learned case, or by modification of some existing cases.

In this cycle, general knowledge usually plays a significant role by supporting the CBR process. Based on the type of CBR method, the degree and the nature of the support of general knowledge may vary from very weak (or none) to very strong. Here general knowledge refers to *general domain-dependent knowledge, as opposed to specific knowledge embodied by cases.* The author exemplified this situation with the diagnosis situation of a new patient considering the case of an old patient. *In diagnosing a patient by retrieving and reusing the case of a previous patient, a model of anatomy together with causal relationships between pathological states may constitute the general knowledge used by a CBR system.* In this case, a set of rules may have the same role (Aamodt & Plaza, 1994).

4.3 How CBR Methodology Can be Applied to Solve Problem in Microcredit

Borrower risk assessment, particularly credit scoring process needs special knowledge based on experience. It has been found that loan officers, here experts use their previous experience to assess the credit risk of borrowers. Bank loan officers use information on various borrower characteristics and evaluate the new borrower (applicant) with the characteristics of the similar borrowers previously they had and take the decision based on the similar cases (successful or default borrowers). In this decision making process, loan officers or experts need to know which loans have been repaid successfully and which have been defaulted or failed from past similar cases. They also need to know how to modify an old best similar case to fit the new problem perfectly. We found the use of CBR model by Vukovic et al. (2012)) using preference theory and a genetic algorithm (GA) to decide whether to grant a credit to new applicants in credit card and consumer loans using credit scoring. The authors used the dataset from Australia and Germany which are different from the aspects in developing regions like African and Asian countries. In this study, the authors mentioned the challenges of each phase of CBR cycle on which its performance depends. The performance of *retrieval phase* is affected by case representation, case indexing and similarity metric (Buta, 1994 in Vukovic et al., 2012). For a successful CBR system, it is important to retrieve relevant previous cases to propose a solution to the new situation and ignore those previous cases that are irrelevant(Montazemi & Gupta, 1997). In this regard, special knowledge of the domain represented by different features of the case in structured way is very significant and highly recommended in modeling a successful CBR system(Park & Han, 2002).
In order to solve a problem, CBR approach involves designing/getting a proper case description, computing analogy or relevancy of the current situation to the old cases stored in a database with their known solutions, retrieving similar cases and attempting to reuse the solution of one of best similar retrieved cases directly or modifying the nearest one to adapt to the current situation if necessary, and finally, storing the new case description and its proposed solution in the database as new knowledge or learning(MÁNTARAS et al., 2005). Kolodner (1993) described CBR system in four steps including case representation, case indexing, case retrieval and case adaptation. Case representation represents the features associated with a past case; case indexing intends to facilitate the search and retrieval of similar cases; case retrieval retrieves the cases most similar to the studied case from the database; and case adaptation is a process of modifying an existing case or building a new one if all the retrieved cases do not comply with the case encountered(Y. K. Chen, Wang, & Feng, 2010). However, the same system has been described by Aamodt & Plaza (1994) with well-known 4 REs (retrieve, reuse, revise and retain) which again being extended by (Reinartz, Iglezakis, & Roth–Berghofer, 2001) including two new steps (review and restore) (Vukovic et al., 2012).

In our study, the aim is to develop a risk assessment tool, here credit scoring system, to measure the risk of individual microcredit borrowers in developing countries seeking loans in online P2P lending platforms. As we accept CBR as generally quite simple to implement and can often handle complex and unstructured decisions very effectively (Ahn, Kim, & Han, 2007), it fits to our research context rationally. Because we have already mentioned earlier that unlike corporate finance, no specific rule, except skilled credit officers' subjective judgment, can be applied to borrower selection in microcredit. We intend to see the CBR as a methodology which is based on the reasoning by analogy method (i.e. similar problems have similar solutions) (I. Watson, 1999). Also, we follow CBR approach in the way summarized in the well-known 4R's cycle by Aamodt & Plaza (1994). Of course this approach requires

the definition of (i) a case structure, comprising a description of the situation, an adopted solution and an outcome, and (ii) a proper similarity metric supporting the retrieval of cases that are relevant to the one at hand. This problem-solving paradigm is suitable to deal with domains whose problem solving methods have not been fully understood and modeled, but in which experiential and episodic knowledge is instead present. In fact, within this paradigm, it is not necessary to elicit and to represent the knowledge required for constructing a solution from the description of the current problem, but it is rather necessary to have an idea of how to compare two situations, two cases, and rate their degree of similarity (Uddin et al., 2015a).

Provided that the number of past cases is sufficiently covering the range of possibilities, it is plausible to think that the solution to a past situation sufficiently similar to the one at hand will be a useful support to the definition of a line of work for the current problem. Knowledge elicitation and representation phases in the definition, design and implementation of a CBR system are therefore focused on the definition of a proper structure for the case description (as suggested above, composed of a description, solution and outcome parts) and also of a proper similarity metric. The most knowledge intensive phase of the CBR cycle is about the adaptation of the past case solution to the present situation (Manzoni, Sartori, & Vizzari, 2007): it is not unusual that this phase is actually delegated to the human expert (the so-called null adaptation approach) due to the lack of sufficient knowledge to systematically perform this kind of activity (Uddin et al., 2015a).

We intend to use other relevant technologies in this CBR approach to achieve the central goal- credit scoring for individual microcredit borrowers in online P2P lending platforms. For data collection, XQuery has been applied to retrieve the data from open sourced database of an online P2P lending platform (Kiva.org: having enough data volume to support the requirements of large sample size for CBR system) and then MySQL Workbench (a database technology or tool for designing, development and administration of databases) has been used to create a database with relevant features of the target cases- here, microcredit borrowers. For similarity metric, nearest neighbor (a popular and widely used technology for classification problem) with weighted average of the results has been used. Finally, to get the missing solution (credit scores) of previous cases in the database, expert-rule technique (spreadsheet-based objective assessment of subjective judgment of experts in the domain) has been deployed.

4.4 Summary

This chapter has introduced Case-Based Reasoning (CBR) as an approach and described the CBR approach detailing its process and its application in finance for credit scoring. Also, this chapter has given a clear idea how the CBR approach can be applied to borrower risk assessment in online P2P microfinance platforms.

Chapter 5

Case-Based Reasoning to Support Microcredit Systems: The Prototypical Solution

The growth of dedicated web-based platforms[20] guides this project in the development of a first prototype able to provide lenders an estimate of risk, based on similar past cases, for a new loan request from a person who needs loan. This prototypical solution has been developed based on *Case-Based Reasoning (CBR)* approach. It has defined the *Case Structure*[21]by analyzing the information and data available to the leading platform-Kiva[22]. Then, recovery of such information from the open source database provided by Kiva with the use of the XQuery and importing them in specially created *Database Tables (DB)* have been done while the domain expert has defined a risk function to calculate the risk score for loans. After creating an own database with risk score, the CBR application has been implemented. Finally, the web-based *User Interface (UI)* has been developed to interact with the CBR system.

In what follows, the current scenario of microcredit system (web-based P2P lending) and the scope of data access (Kiva open source data) have been described. Then, the technologies (XQuery language, CBR approach, and Google Web Toolkit-GWT) that have a crucial role both in development and in implementation of the system have been discussed.

After discussing the technological issues, the development and implementation aspects (DB creation, Queries & Data import, Developments of CBR application, and GWT application online) of the CBR system have been presented and explained. Finally, the implications of the system and an overall assessment of the effectiveness of the score created, and the suggestions for possible future developments to improve and to expand the project have been discussed.

5.1 Trend of Microcredit System towards Web-based Peer-to-Peer Lending Platforms and Scope of Data Access

5.1.1 Microcredit

Microcredit is a small loan amount to those individuals who are in poverty and also who, because of their social and economic status, do not have access to the services provided by the formal financial sector. The common critical factors that undermine this access to a

[20]Also referred to as 'Internet-based platform' and 'online platform'. The term 'Web-based platform' will be used in this paper. Three of such web-based microcredit platforms are Kiva, Zidisha and MyC4. These platforms help those marginalized people who cannot make the use of traditional lending channels for accessing credit or loans they need.

[21] Case Structure consists of three components: Problem description, Solution, and Result.

[22]For reasons of transparency, Kiva.org provides a large amount data regarding loan applications, applicants and payments made.

traditional bank loan are lack of collateral, unsteady employment, poor verifiable credit records, and small size of business or entrepreneurial activities. Most of the people who require loan are living in developing countries and they are often called 'micro-entrepreneurs' because of small scale or size of business activities (Milana & Ashta, 2012).In recent years, microcredit has spread not only among the families living in a subsistence economy but also in developed economy where consumers as well as small businessmen cannot make use of the traditional loans. In this scenario, Internet has brought an opportunity to provide the financial services, particularly loans to the target group based on the principles of micro credit system. The network of lenders[23]using web-based platforms is completely dedicated to connecting people through loan in order to alleviate poverty in the world. This is the mission of Kiva.org, the web-based Peer-To-Peer (P2P)[24] platform that will be further explained in the next sub-section.

5.1.2 Kiva Microfunds

Kiva Microfunds[25] is a nonprofit organization dedicated to microcredit lending to alleviate poverty globally by leveraging the Internet and a global network of Micro Finance Institutions (MFIs). Kiva online platform allows anyone to lend money to entrepreneurs and people who do not have access to traditional banking systems in 82 countries worldwide.

5.1.2.1 Operation of Kiva.org

Kiva operates its lending activities as indirect online lending platform with the global network of filed partners (MFIs, Schools, NGOs, and Nonprofit Associations) who support for loan management locally. The filed partners help Kiva to select borrowers, disburse loans, monitor the activities, and collect the payments of loan installments from borrowers. In Figure 5.1Kiva lending operations have been described in 3 steps.

[23] Mostly individual lenders are connected across the world to the system.

[24] Also referred to as Person-to-Person lending, People-to-People lending, social lending, or P2P lending. We will use peer-to-peer lending and P2P lending interchangeably. P2P lending is an Internet-based platform of financial transactions where borrowers place requests for loans online and private lenders fund them directly or indirectly.

[25] Kiva Microfunds is Kiva.org web platform: www.kiva.org

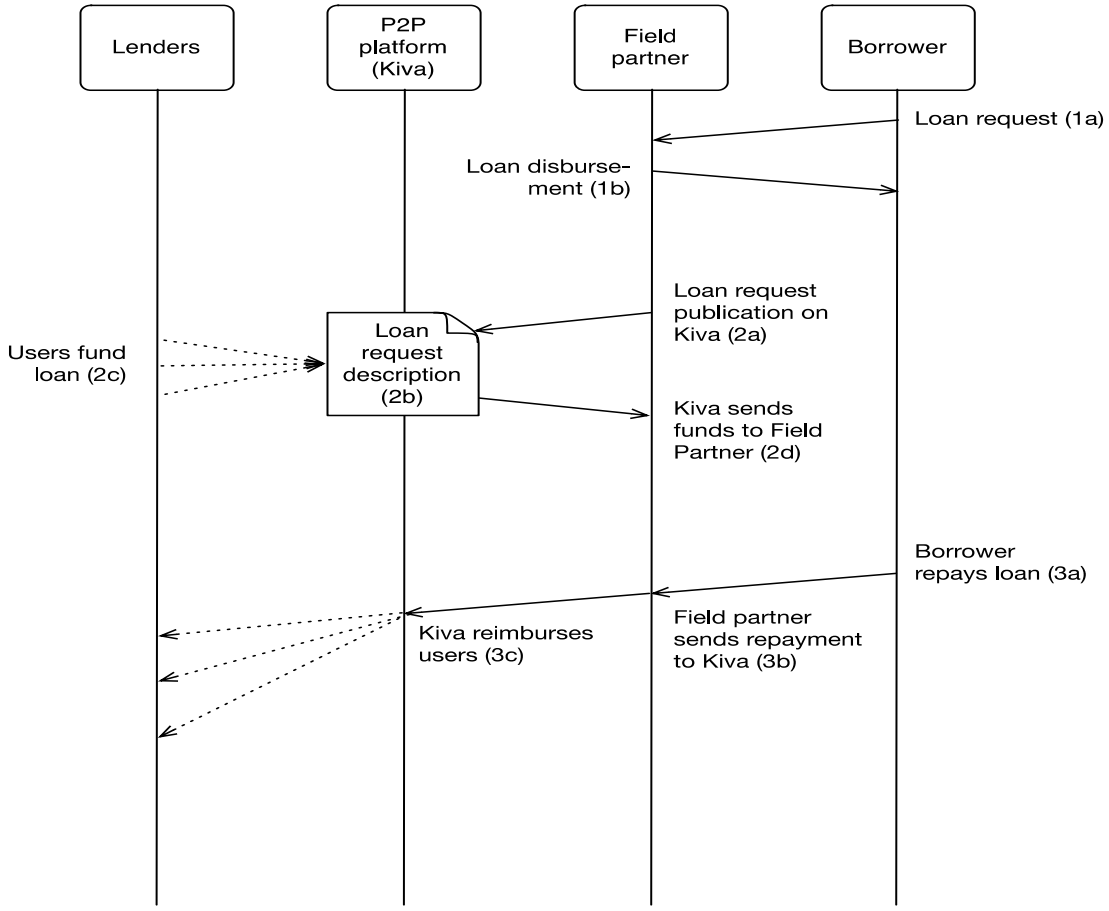


Figure 5.1*Life cycle of a loan*

It follows, Figure 5.1, the following steps: (1a) the borrower meets with the Field Partner and requests a loan. The Field Partner, if certain criteria are met, disburses (pre-disbursement[26]) a loan to the borrower so that the applicant is satisfied immediately without waiting until the loan is fully funded on the web platform (1b).After loan disbursement, the Field Partner uploads the loan request to Kiva (2a), it's reviewed by a team of volunteer editors and translators and then published on Kiva.org (2b). Kiva lenders[27] fund the loan request assuming the risk of the loan that they have decided to support (2c). When the loan is fully funded Kiva sends the loan safely through bank transfer to the Field Partner (2d). The borrower, later on, makes repayments (3a) and the Field Partner[28] sends funds owed to Kiva (3b). Kiva repays the principal amount to its lenders (3c). The lenders can make another loan,

[26] On the contrary, in case of post-disbursement, loan amount is disbursed to the borrower after the loan request uploaded to the Kiva web platform and then funded fully by the lenders on the site.

[27] The loan amount can be funded by a lender with full requested amount or can be funded with the partial amount, starting from $25 and multiple, where full loan will be funded by a number of lenders in group/pool at interest free.

[28] Field partner charges interest to the borrower and collects the repayments from borrowers from which they keep interest portions to cover their operating costs and profit margin, and then finally send the principal portion to Kiva.

donate to Kiva[29], or withdraw the money to their PayPal account (Choo, Lee, Lee, & Park, 2014).

In this funding operation, the lenders[30] viewing loan request descriptions have an indication of the risk rating associated to the Field Partner, based on historical data of loans managed by that organization. However, no indication on the risk associated to the *specific* loan request is provided, as shown in Figure 5.2. Selecting a borrower is a challenging task to online microcredit lenders as individual borrowers' profiles do not provide any risk rating on the platform except the microfinance intermediaries' aggregate risk indicators (depicted on the right bottom corner in Figure 5.2) based on the actual repayment of previous borrowers managed by the same field partner. This information can surely suggest good assessment and management capabilities of the field partner, but it is essentially unrelated to the current loan request.

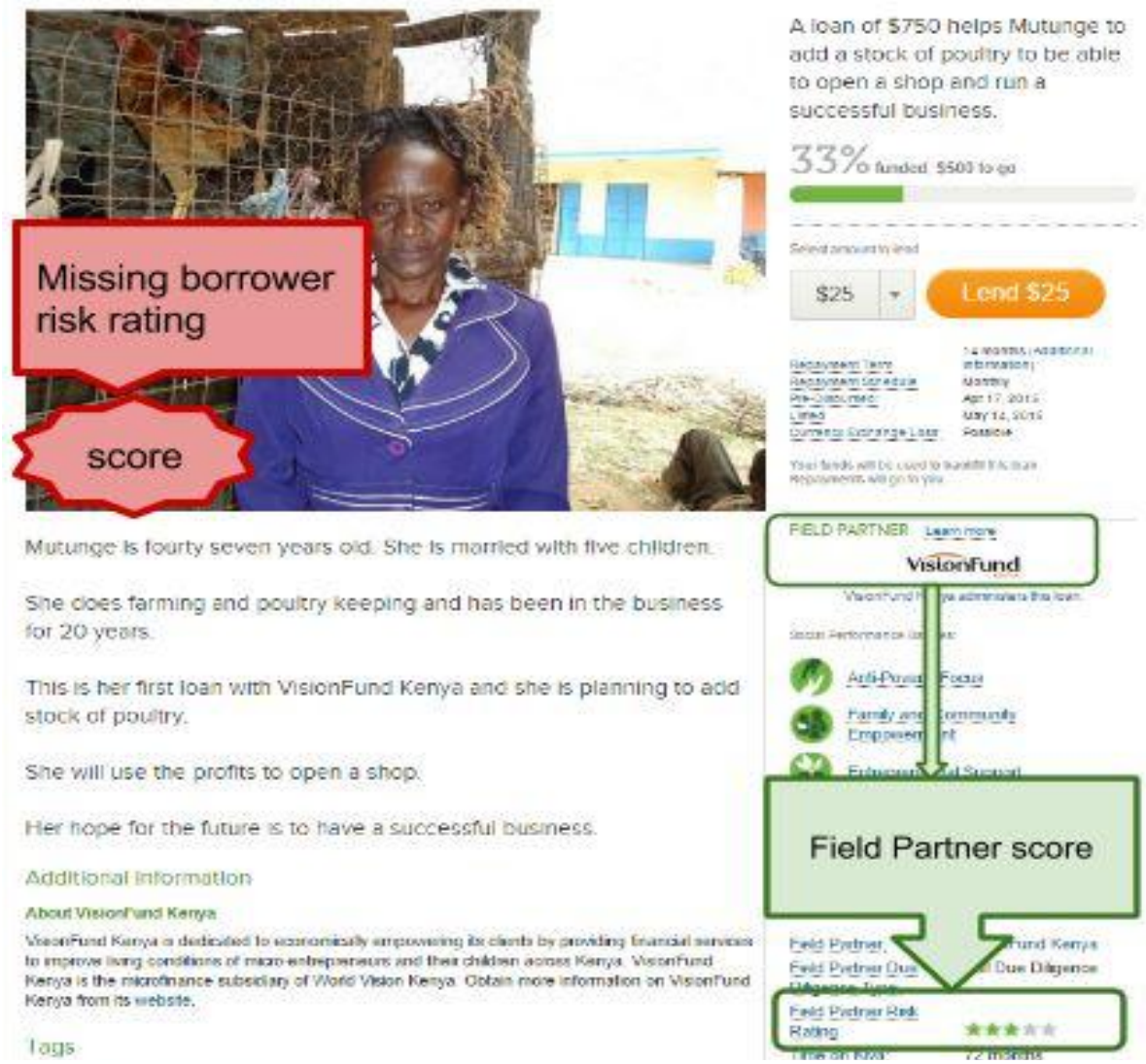


Figure 5.2*Profile of an applicant and published on its loan kiva.org. Note the presence of a score of the Local Partner (not very useful to the creditor in the judgment of the loan) and the lack of an appropriate assessment that refers to the loan risk.*

[29] Kiva generates its revenues from donations, optional lender fees and the interest income generated from the instalment money they hold for the time being (for details see Flannery, 2007).

[30] Lenders are also known as users.

Moreover, the platforms merely keep typical advices for lenders or end users to diversify their portfolios through lending to more than one borrower via different field partners as well as in different countries and/or sectors. However, an indication of the borrower's risk, which is missing on the models (indicated in Figure 5.2), remains critical to the aggregate or individual lenders in the sites (Uddin et al., 2015).

5.1.2.2 Data Access

API [31] created by Kiva, based on REST architecture, allow researchers to get all the information on applicants, loans, lenders and obviously in full compliance with privacy by making a request at a time and having the data related to available data in 4 different formats: HTML, JSON, XML, and RSS. For example, if one wants to retrieve the loan details with id = 495172, the seeker should send the following HTTP request: api.kivaws.org/v1/loans/495172.xml, which will provide data in XML format. For more insights on the methods provided by Kiva API and on available tools for developers, please visit the build.kiva.org website. For those who, as in our case, need a larger amount of data, we are provided so-called "data snapshots", a historical archive of loans, available in JSON format or XML and updated frequently, which contain a number of information comparable to thousands of requests to Kiva API.

5.1.2.3 Analysis of a "data snapshot"

The database we have used contains "data snapshots" in XML format, more human readable compared to JSON format. A snapshot contains 500 loans where each loan records, identified by the tag <loan>,root node to the data of the relevant applicant or group of applicants, represent data on applicant's personal profile, business information, loan requirements and its repayment structure, history with status, and many other intuitive information assigned to sub nodes that exist within the root node. Figure 5.3 shows an extract from an XML document, among the thousands contained in the archive.

The meanings of different dates given, not disclosing so excessively, will only be explained for each loan and who have a close relationship with the cycle of life already shown in Figure 5.1.

[31] Kiva API: build.kiva.org/api

Figure 5.3*Extract of a "snapshot date" in XML format*

In chronological order:

- Disbursed date: the date on which the local partner finances in cash the applicant and how, already mentioned in Section 2.2.1,it canbe pre-disbursal or post-disbursal.
- Posted date: the date on which the profile of the applicant is posted onKiva platform.
- Funded date: the date on which the loan is fully funded by lenders on Kiva, in fact, before this date, the loan status is set as "Fundraising", while later, it turns into "funded", followed by "in_repayment".
- Scheduled payment dates: the date appears several times within nodes as "scheduled_payments". Each date indicates the date when the lenders are expected to receive a portion (installment) of the amount lent.
- Local payment dates: the date appears several times within nodes as "local_payments". Each date indicates the scheduled date on which the applicant should make a payment (installment) to the local partner.
- Processed dates: the date of a payment made by the applicant to the local partner (for each installment, if it is installment loan where payments are made more than one time).

- Settlement dates: the date on which the local partner repays the installment to lenders (a portion of the total loan); the transaction is handled by Kiva to credit the repayment tothe respective accounts of lenders.
- Paid dates: may coincide with the date of the last repayment ("settlement date") in Kiva account of lenders and indicates the end of the loan.

The "data snapshots" provided all are well-structured and consistent to the same model without errors or omissions that might be expected in the oldest data. Those affecting the last two years of loans, one may notice additional use, but not constant, of the nodes <tag> and <theme>, probably due to the implementation of categories to filter search results on the platform. This is not, however, important for the project.

5.1.3 Goal of Prototypical Solution

As previously mentioned in the description of the Kiva platform's operation that once the local partner post the profile of the borrower, the loan is ready to be funded by lenders. The choice of a lender for helping borrowers on the platform is based on social and psychological aspects, which fall outside this discussion, but it lacks a reliable method to alert the lender for the risk that may be incurred to fund that particular loan[32].In fact, as shown in Figure 5.2, there is already a score of 1 to 5 scale (5-star) for assessing the risk of the local partner, but that score does not have a close or direct relationship with the current loan. Most importantly, it lacks a risk assessment specific to each published loan or loan applicant.

The purpose of developing this CBR system is to specify for each loan or applicant with a score from 0 to 10, where 0 means "very risky" and 10 means "low-risk" based on the information of previous borrowers or loans. Hence, it is needed to create the score for a specific number of previous or historical loans in order to create an adequate knowledge base so that it helps to search the assessed loans, similar to the new loan application.

For this phase of recalling and the next steps, it uses an approach, CBR, discussed in Section 3.2.

5.2 Technologies

We now describe the technologies that have played an important role in the realization of the system, such as XQuery to retrieve open data in XML format provided by Kiva, CBR for the type of approach used in the application itself, and GWT for the realization of the web-based UI.

5.2.1 XQuery

XQuery 1.1[33] is a functional programming language for querying collections of structured or unstructured data. It helps to extract and manipulate the data in XML documents or anything that might be seen as XML.The language was developed by the working group "XML Query" W3C and is closely related to the XPath 2.0 standard[34], briefly described below, from

[32] It is mentioned on the platforms that lending risk is always on lenders. For instance, on Kiva site, *"Lending through Kiva involves risk of principal loss. Kiva does not guarantee repayment nor we offer a financial return on your loan"*. See more at https://www.kiva.org/

[33]XQuery 1.1: http://www.w3.org/TR/2009/WD-xquery-11-20091215/

[34]XPath 2.0: www.w3.org/TR/xpath20/

which it inherits the model data defined in the XDM[35,] the syntax to navigate in an XML document, consisting of "path expressions"(path symbols), and supporting the same functions and operators. Every expression in Query, working on sequences, is ordered in the list of objects. Each object can be a node, which represents a component of the XML document, or an atomic value, which is an instance of a basic type of XML Schema, such as an xs: integer, xs: string, xs: date etc. Each node can belong to one of the 7 types: document node (root), element, attribute, text node, namespace, Processing Instruction, and comment. An atomic value, however, is a node with no child or parent nodes. XQuery is able to navigate the nodes of the tree. It characterizes the structure of a XML document which becomes possible due to the syntax inherited from XPath. The nodes are labeled with the parent endorsements (parent), children (son), sibling (brother), ancestor (ancestor), descendant (descending), depending on the relationships which exist between them. Each element and attribute node has a parent node and the parent node, hierarchically higher XML tree, is the root (root) of the document. On the contrary, a parent node may have zero, one or more children nodes below it. All element nodes that have the same parent are named with the sibling relationship. One can guess that the ancestor and descendant reports resulted from the involvement of at least one parent node, in the first case, and children, in the second. Some examples are shown in Figure 5.5

```
<loan_set>
    <loan>
        <id>6532</id>
        <borrower>
            <name>Mark</name>
            <gender>M</gender>
            <age>43</age>
        </borrower>
        <amount>650</amount>
        <repayment_term>6</repayment_term>
    </loan>
</loan_set>
```

Figure 5.4 *Relations between nodes.*

Various examples:

i. The element 'loan' is the children of the element 'loan_set', ii. The elements 'name', 'gender' and 'age' share sibling relationship each other and have parental relationship with the element 'borrower' as the parent node, iii. The element 'name' is the descendant of the loan node and also the borrower node, iv. The element 'amount' has its ancestor both elements 'loan' and 'loan_set'.

[35] XDM - XQuery and XPath Data Model: www.w3.org/TR/xpath-datamodel-30/

The symbols used to specify a path with the aim to navigate and select a node in the document are shown in Table5.1

Symbol	Description
nodename	Select all nodes with the name "nodename".
/	Select the root node.
//	Select the nodes in the document starting from the current node that matches the selection no matter where they are.
.	Selects the current node.
..	Select the parent of the current node
@	Select attributes

Table 5.1 *Path expressions frequently used to select a node or attribute. It comes from w3schools.com*

In addition to the above symbols, they are still available so-called predicates, always indicated in square brackets as used to select a specific node or a node that contains a specific value.

Some examples are illustrated in Table 5.2

Paths with predicates	Description
/parentnode/child[1]	Select the first child element, a child of the parentnode.
/parentnode/child[last()]	Select the last child element, child of the parentnode.
/parentnode/child[position() < 3]	Select the first two child elements, children of parentnode.
//child[@attr = 'val']	Select all child elements that have an attribute "attr" with a value "val".
/ancestor/parent[cond > 35]/child	Select all child elements of the parent elements of ancestor elements that have an element "cond" with a value> 35.

Table 5.2 *Some examples of routes with specific predicates. Taken from w3schools.com*

New XML documents can be created using elements and attributes retrieved by others XML documents, or taking advantage of the features available to build new ones. In addition to more than 100 built-in functions available in XQuery, the user can create their own or import the online libraries available. Moreover, the language is designed to use a syntax similar to SQL queries, call "FLWOR expression", shown in Figure 5.5, can join the operation between various documents and to interrogate a collection of documents, usually included in the same working directory, replacing fn: doc with the fn: collection.

```
distinct-values(
    for $c in doc('loans.xml')/loans/loan/location
    order by $c/country
    return concat($c/country_code, ''', ''', $c/country))
```

Figure 5-5 *Example of a query that uses the FLWOR syntax. The query result is the concatenation of code and corresponding name of all countries in the loans.xml document, alphabetical order (order by keywords), and returned no more than once (fn: distinct values).*

FLWOR, derived from the five keywords used to create the query: For, Let, Where, Order by, Return, but, in fact, with the release of XQuery 3.0[36] has also introduced a new keyword named *group by*, is used to group nodes in the XML document according to their name or to the one of the possible attributes, in addition to the keyword *count* that allows, for example, to filter the result of a query by returning only a certain number of tuples. For the full change log, visit www.w3.org/TR/xquery30/#idrevisionlog website.

5.2.2 Case-Based Reasoning

Case-Based Reasoning (CBR) is a recent approach that aims to combine the resolution and learning problems. It is based on the idea that a problem tends to occur more than once, then, a new problem can be solved by recalling a similar case in the past, reusing information and knowledge associated with that situation; this technique is frequently used by humans to find the solution to a new problem. In CBR, a case usually denotes a problematic situation. A situation experienced in the past has been learned and stored so that it can be reused in solving future problems and is referred to as the event passed, if stored or preserved. On the other hand, a new event or an unsolved case is the description of a new problem to be solved. An important feature of CBR is, precisely, the learning phase- the resolution of the problem had success and remembering the experience to solve future problems. That means, it has performed updating the knowledge base to check whether it has failed to remember the reason and to avoid making the same mistake again. CBR is, therefore, a cyclical process to solve a problem, learn from that experience, to solve a new problem, and so on (Aamodt & Plaza, 1994; Slade, 1991).

5.2.2.1 CBR Cycle

The CBR cycle may be described by dividing it into four main steps (the four REs):

i. RETRIEVE: the most similar case or cases to the new problem with a similarity function created ad hoc.

ii. REUSE: the information and knowledge in that past event or case for solving the new problem.

iii. REVISE: the proposed solution, and

iv. RETAIN: the new experience in the knowledge base (Case-base) that will be useful in the future to solve another problem (Aamodt & Plaza, 1994).

[36] XQuery 3.0: www.w3.org/TR/xquery-30/

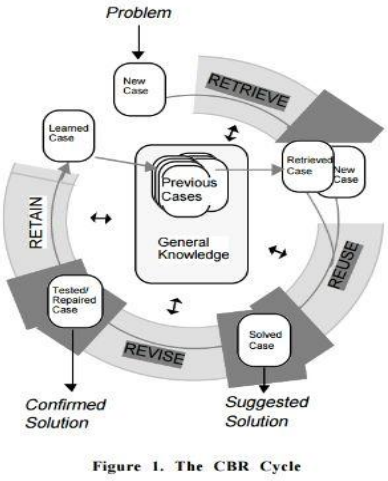


Figure 1. The CBR Cycle

Figure 5-6 *CBR cycle (from jCOLIBRI Tutorial).*

Figure 5.6 illustrates the cycle. The description of a problem (top of the figure) defines a new event which must be solved. This new case is used to retrieve one or more similar cases from the collection of previous cases contained in the knowledge base, through a similarity function. The case recovered is reused to suggest a solution to the current problem. Through the adaptation process, this solution is tested or evaluated if it's succeed in the current problem and modified in case of failure. During the storage phase, the useful experience is retained for re-use in future and the knowledge base is updated by inserting the new case, or learned by modifying some existing cases. As shown in the figure, the knowledge base plays an important role in the CBR cycle, supporting the CBR processes. By general knowledge, it refers to a knowledge-based dependent on the application domain (expert knowledge), as opposed to specific knowledge embodies by cases.

5.2.2.2 Critical issues

In this section, we give an overview of the major sensitive issues in developing CBR application. As we mentioned earlier, the CBR reasoner is heavily dependent on the structure and content of the collection of past cases. It is important, therefore, to represent and to describe properly the structure of a problem to keep a consistency of data in the knowledge base, and to decide how it should be organized and indexed for a proper recovery, reuse and storage cases. An example from jCOLIBRI Tutorial exhibits CBR reasoner as follows:

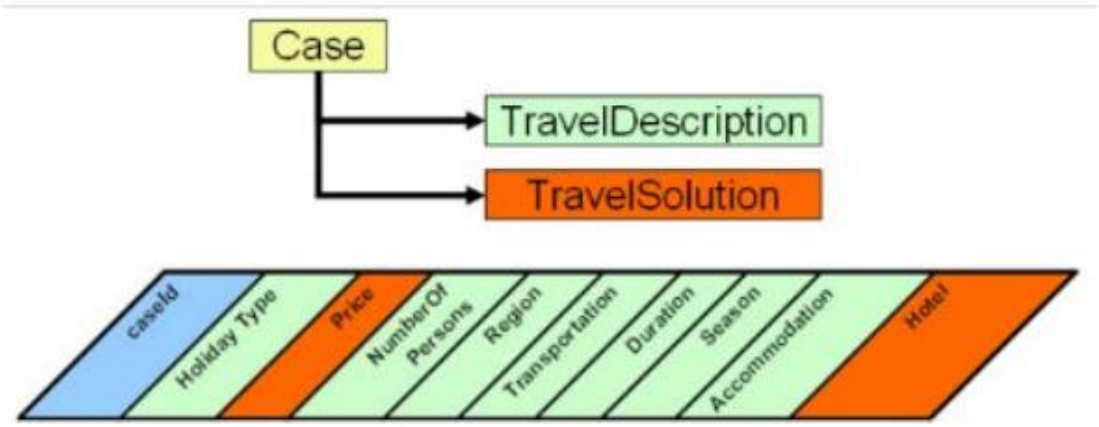


Figure 5.7 *Example of a case frame in an application Travel Recommendation.*

The highlighted attributes that arein green make up the description of the case, those in orange are of the solution and that in blue- the "caseId" attribute shared by both. The "Region" attribute can be deployed as a compound attribute. This structure ***has no result*** (Recio-García, Díaz-Agudo, & González-Calero, 2008).

The cases are represented as a collection of objects. The structure of an object, in object oriented representations such as Java, is described by a class having one for each of the objects need to describe (description, result and solution, see Figure 5.7). Each of which can consist of simple attributes, i.e., attribute-value pairs of a Java primitive type, or from compounds attributes, namely in connection with other classes of objects.

The recovery phase, having a query as input, i.e., the description of a new problem, has the aim of finding a set of cases sufficiently similar to it, carrying out measures of similarity. We talk about the global similarities when comparing two objects, i.e., the query and the description of a past event. For each simple attribute belonging to the two descriptions, local similarity is calculated between the two values of the attributes, while for each compound attribute is calculated a new global similarity which in turn compares the simple attributes of the associated classes. The overall similarity between query and description of the past event, is determined by aggregating, in our case with a weighted average, the local similarity values resulting from comparisons. It is possible that the values of simple attributes, to be compared, are not always numerical, but rather character strings and that it is not useful to compare the equality between them, but rather determine whether a string has or has not a certain property in common with another string, creating functions that associate a higher value to strings with similar properties and lowest in strings with different properties. This is possible through ontologies that, in this case, the use of membership classes is remarkable, grouped into each of the objects with the same properties (see Section 4.3.2 Recovery, reuse and adaptation).

So, the global function, which has the task of creating the similarity ratings for past cases, it is the second critical aspect in a 'CBR application and can affect the proper recovery cases. These issues will be covered in Section 5.3 ‘System Design’.

5.2.3 Google Web Toolkit

GWT[37] is a development tool for creating AJAX applications cross-browser using the favorite IDE, such as Eclipse, through Java APIs and Widgets. These allow the developer to create a graphical client-side interfaces by writing code easily in Java, then translated by the GWT compiler into highly optimized JavaScript code that will be executed in the user's web browser.

The tool provides two modes of performance for the phase of creating, editing and debugging involves the development mode, which lets the user run Java code in the JVM, without going through the compilation in JavaScript, and using the GWT Developer Plugin which allows viewing content in the most popular browsers (e.g., Firefox v.27 or smaller). Once the application is running in development mode, it is useful to run in production mode, then translated into JavaScript, in order to test the performance and appearance that could take in different browsers. The graphical user interface also needs to interact with the backend server side. One can communicate with a Java Servlet through GWT Remote Procedure Call (RPC) over HTTP asynchronously and transparently, leaving the GWT to serialize objects task Java, or use custom HTTP requests using the HTTP library included in GWT. In any case, the server side code is not compiled into JavaScript, unlike the client side. The peculiarity of developing an AJAX application is to be able to move the workload created from the graphical interface on the client side, reducing in this way the bandwidth used, as well as to achieve a more fluid and responsive system, in this case, the asynchronous calls are recalled, avoiding to download an HTML page for each user request as in traditional web application.

5.3 CBR System Design

The system consists of three major components as shown in Figure 5.8: the database, newly created and populated with the open source data provided by Kiva (recovered through XQuery), the application based on the concept of CBR that interacts with it, and a graphical interface developed in GWT for a general approach of the web-based system.

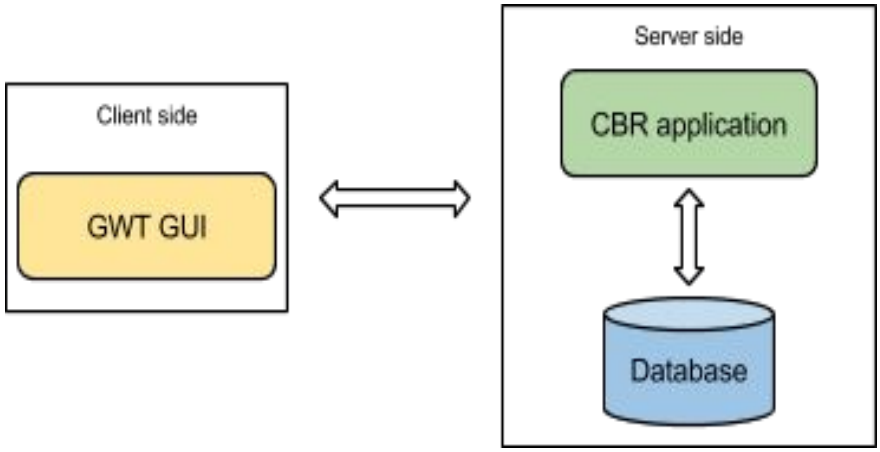


Figure 5.8 *Architecture of the CBR System*

[37] GWT - Google Web Toolkit: www.gwtproject.org/

5.3.1 SQL Database Creation

It was necessary to create an ad hoc database with which the CBR application could interact properly, retrieving relevant information, i.e., the attributes that make description, solution and result, for subsequent use as input and output in calculating the similarity between the concerned loans. The development begins with the realization of an ER diagram shown in Figure 5.9, that conceptually represents the database; an entity is an object or concept in the real world which can be described by the attribute, while a relationship is an association that binds two or more entities, having four types of possible cardinality (one to one, one to many, many to one, many to many).The heart of the model is the entity Loan Request (loan application) that includes many attributes fundamental to its representation, from the identifier unique; from here, four different relationships branch out to related entities. In particular, a loan application is supported by one and only one Field Partner (local partner), which may present from 0 (at the time of its insertion) to n requests loan; it was decided, for the sake of simplicity, that a loan request may be made by only one borrower (or applicant) (although in Kiva there is the possibility that a group of people requires a collective loan) and the applicant may have, over time, more than a loan; in turn the applicant is in one to one relationship with country, that may reside in a single country; the loan application is also related with sector and it might have only one economic sector involved. Finally, the loan request could have from 0 (when it is still in the financing stage or has expired) to n payments associated, each of which is uniquely identified.

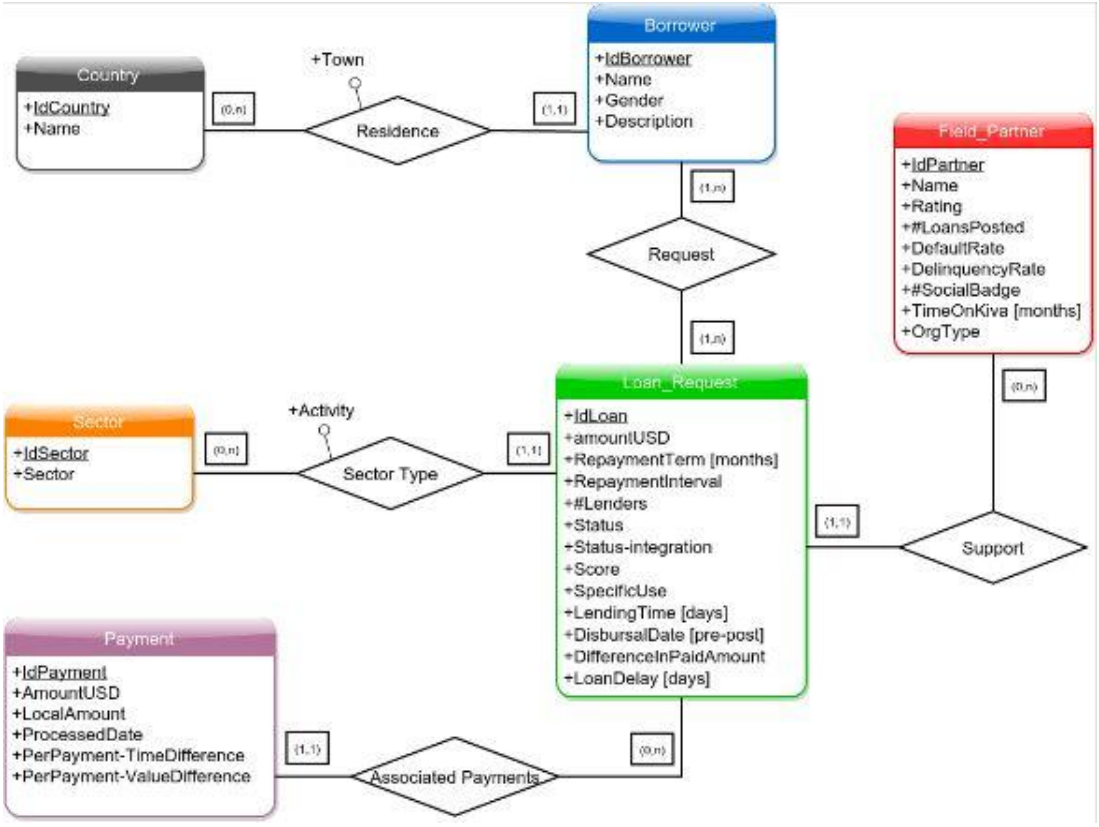


Figure 5-9 *Conceptual representation of the database creating an ER diagram.*

From the ER diagram, it is brought about the associated relational model, which represents the real own database schema, depicted in Figure 5.10.Unlike the ER diagram, here, the entities are called tables, attributes become table fields and using their primary keys to express the relationship, allowing a developer to maintain the concept of referential integrity, one of the cornerstones of relational model.

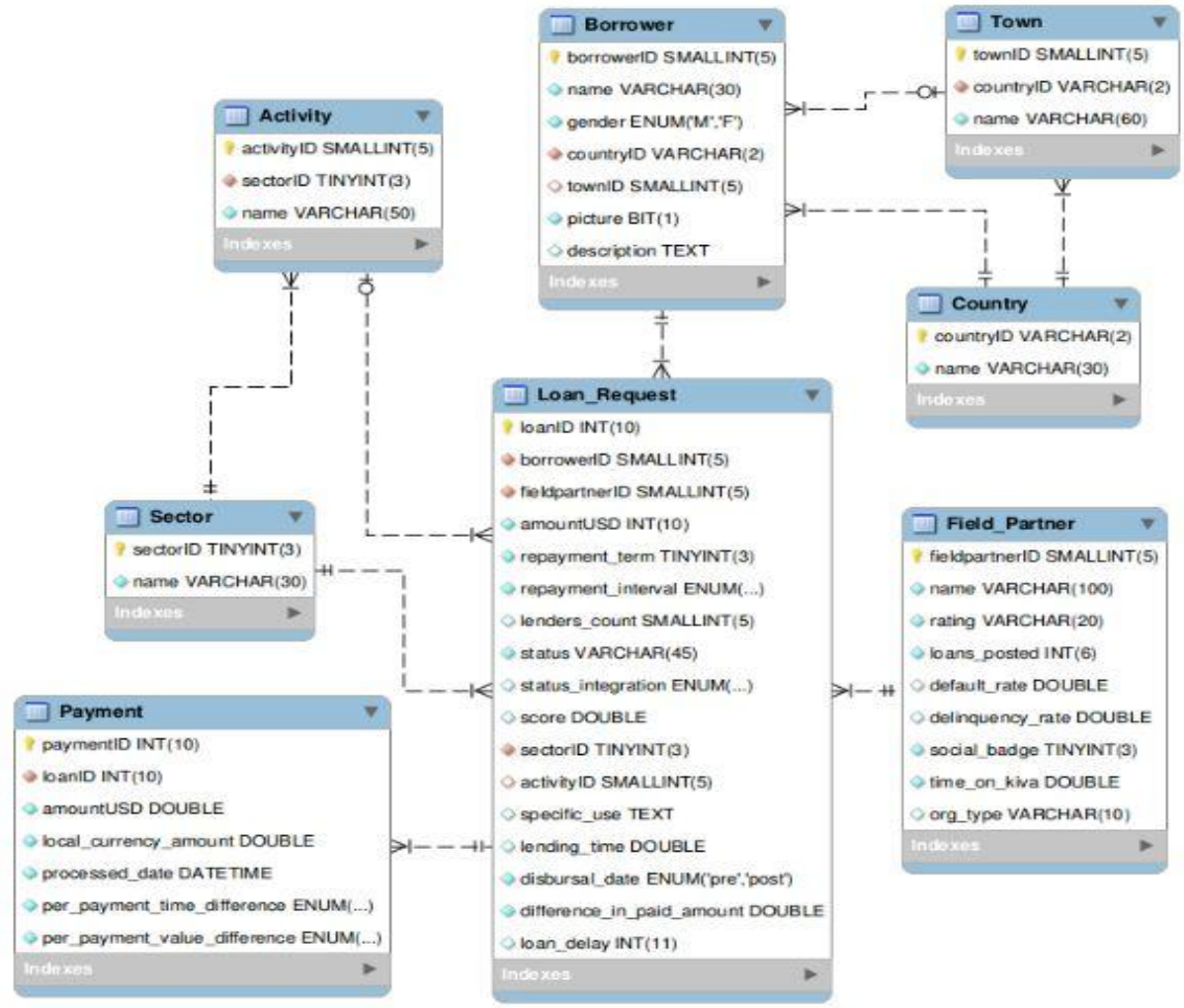


Figure 5.10 *Relationship model database schema in MySQL Workbench.*

The database was implemented in MySQL Workbench[38], a tool for designing, development and administration of databases, allowing easy deployment; in this case, the option of forward engineering starting from the relationship model is helpful.

5.3.2 Queries and Data Import

Between the two snapshot date formats provided by Kiva, namely JSON and XML, it has decided for the latter, extrapolating atomic values in the XML nodes through creation of a

[38]MySQL Workbench: www.mysql.com/products/workbench

Java program, Eclipse environment Luna[39] Java EE, and leveraging queries in XQuery. After trying a couple of XQuery processors with unsatisfactory results, it was decided a great product offered by open source Saxonica[40]. The package Saxon 9.6 HE[41] that includes a collection of tools to process XML sources, including support to XPath 2.0, XPath 3.0, and XQuery 3.0 especially critical and lacking in processors previously tested. The XQuery processor is invoked in a Java environment with masses API available by XQJ (XQuery API for Java, also known as JSR 225) and implemented entirely in Saxon.

```
1 import java.io.*;
2 import javax.xml.xquery.*;
3 import com.saxonica.xqj.*;
4
5 public class Main {
6
7
8     public static void main(String[] args) throws XQException {
9
10         XQDataSource datasource = new SaxonXQDataSource();
11
12         // Establishing a Session
13         XQConnection connection = datasource.getConnection();
14
15         // XQExpression object allows to execute XQuery expressions.
16         XQExpression expression = connection.createExpression();
17         try{
18             File sqlfile = new File("/home/andre/Desktop/Stage/
   loansDB/SQLQuery/Script.sql");
19             if (!sqlfile.exists()){
20                 sqlfile.createNewFile();
21             }
22
23             BufferedWriter bw = new BufferedWriter(new FileWriter
   (sqlfile.getAbsoluteFile()));
24
25             Country.readAndWrite(expression, bw);
26             Sector.readAndWrite(expression, bw);
27             FieldPartner.readAndWrite(expression, bw);
28             Loan.readAndWrite(expression,  bw);
```

Figure 5.11 *Main code extracted that shows the invocation of the Saxon processor.*

As depicted in Figure 5.11, the system creates a new *SaxonXQDataSource*() object. From the latter, called a *getConnection*() to get a connection from createExpression connection, can invoke () method to create an object *XQExpression* that, in turn, it becomes possible due to its *executeQuery()* method which allows the query, contained in a separate file, to be evaluated. The result of the query evaluation is a *XQSequence* object, shown in Figure 5.12, which behaves like an iterator, in fact, by its next () method, one can scroll through sequence, with *getItem*() method, retrieve the item to the current location. The result of *getItem*() is a *XQItem* object through its methods allows to determine the type of item and convert it to a value or Java object that is suitable in our case of a String type for writing to a file named "script.sql" used then to import the data in the database tables. Each query inquires the XML source and

[39] Eclipse IDE: eclipse.org
[40] Saxonica: www.saxonica.com/welcome
[41] Saxon Documentation: www.saxonica.com/documentation/index.html

extracts the necessary information using 1 to 5 seconds depending on the complexity of the query, the type (local path or URL) and to the number of sources.

```
1 import java.io.*;
2 import javax.xml.xquery.*;
3
4 public class Loan{
5
6     public static void readAndWrite(XQExpression expression,
  BufferedWriter bw) throws IOException, XQException{
7
8         System.out.print("Retrieving the loans info and creating
  the SQL Script...   ");
9         InputStream query;
10        query = new FileInputStream("Loan.xquery");
11
12        // It executes the query to retrieve the countries
13        XQSequence sequence = expression.executeQuery(query);
14        bw.write("\nINSERT INTO Loan_Request (loanID,
  fieldpartnerID, amountUSD, repayment_term, repayment_interval,
  lenders_count, status, activity_name, specific_use, lending_time,
  disbursal_date, difference_in_paid_amount, loan_delay) VALUES\n");
15
16        while (sequence.next()) {
17            bw.write("(" + sequence.getAtomicValue() + "'),\n");
18        }
19
20        System.out.println("***Done.");
21    }
22 }
```

Figure 5.12 *Loan class that implements the readAndWrite () method invoked by the Main class. You may notice the reading of the query from Loan.xquery files and writing the result query, managed by the object passed to the method as BufferedWriter input.*

It has created a class for each of the popular tables so that the object *XQExpression*, passed as input to *readAndWrite ()* method, can handle a query at a time overwriting the previous one. Figure 5.13 shows an extract of the SQL script the developer just created that will run for populate database tables. A note to consider is that the applicants, as the sectors, activities and cities do not have an identifier provided by Kiva; therefore, it has been created one for each of them, so as to maintain proper referential integrity.

```
211 INSERT INTO Loan_Request (loanID, fieldpartnerID, amountUSD, repayment_term, repayment_interval, lenders_count, status, activity_name,
    specific_use, lending_time, disbursal_date, difference_in_paid_amount, loan_delay, countryID) VALUES
212 (618873, 215, 950, 3, 'At_end_of_term', 9, 'paid', 1, 'Agriculture', 'to buy seed and fertilizers', 0.1, 'pre', 0, -8, 'TJ'),
213 (618820, 42, 2400, 5, 'At_end_of_term', 75, 'paid', 9, 'Personal Housing Expenses', 'to buy insulation materials', 0.4, 'pre', 0, -13, 'MN'),
214 (610024, 332, 20275, 5, 'At_end_of_term', 613, 'paid', 15, 'Goods Distribution', 'To buy solar lamps that he will sell to cotton and sesame
    farmers.', 3, 'post', 0, -18, 'ML'),
215 (591350, 225, 2850, 5, 'At_end_of_term', 109, 'paid', 2, 'Crafts', 'to invest in the purchase of natural gemstones and silver to prepare new
    designs and increase his stock.', 2.3, 'post', 0, -36, 'ID'),
216 (579082, 203, 250, 3, 'At_end_of_term', 10, 'paid', 13, 'Auto Repair', 'to buy rivets and sprays', 0.2, 'pre', 0, 165, 'KE'),
217 (569003, 42, 4900, 5, 'At_end_of_term', 91, 'paid', 9, 'Personal Housing Expenses', 'to purchase building materials', 2.4, 'pre', 239.77,
    -3, 'MN'),
218 (548798, 161, 1400, 5, 'At_end_of_term', 51, 'paid', 1, 'Farming', 'To buy fertilizers for use in the cultivation of maize', 0.1, 'pre', 0,
    -56, 'RW'),
219 (545657, 170, 800, 3, 'At_end_of_term', 28, 'paid', 1, 'Cattle', 'to buy a cow for resale.', 0.3, 'pre', 0, 269, 'RW'),
220 (544045, 215, 275, 6, 'At_end_of_term', 11, 'paid', 11, 'Personal Purchases', 'to buy a washing machine', 0.1, 'pre', 0, -11, 'TJ'),
221 (539869, 276, 15450, 6, 'At_end_of_term', 452, 'paid', 12, 'Renewable Energy Products', 'to pay for solar energy systems and offer financing
    plans to lower-income farmers', 0.4, 'post', 0, -29, 'KE'),
222 (526936, 225, 800, 5, 'At_end_of_term', 23, 'paid', 2, 'Crafts', 'to invest in the purchase of larger amounts of wood', 0.3, 'post', 0, -19,
    'ID'),
223 (536696, 123, 1250, 6, 'At_end_of_term', 49, 'paid', 1, 'Pigs', 'to buy feeds, vitamins, vaccines and additional piglets to raise.', 0.9,
    'pre', 0, -7, 'PH'),
224
225 (618371, 106, 375, 10, 'At_end_of_term', 15, 'paid', 1, 'Farm Supplies', 'to buy a few sacks of fertilizers and a water pump.', 0.1, 'pre',
    0, -163, 'KH'),
```

Figure 5.13 *Extract of the SQL script that will be run to populate the table Loan_Request.*

A significant support was given from the library FunctX[42], in which it is drawn for several built-in functions are not included in the package of XQuery.

5.3.2.1 Creating Risk Score

The associated risk score to each loan has been created by a domain expert so that it could indicate which attributes and what their respective weights would come into the consideration in the function to calculate this score. The score should be calculated for all loans in the database so that the CBR application can draw on a vast collection of cases, but was accomplished a targeted selection of 107 cases, with the purpose of having a set of the more heterogeneous results. Because scoring was performed manually based on a spreadsheet framework. The rating can vary in a range from 0 to 10, including the extreme values; this interval was then divided into bands and for each of them was assigned a degree of risk with respective explanation, depicted in Table 5.3.

Grade Type	Score (10-point)	Explanation
UG1	8.6 - 10.0	Excellent Grade
UG2	7.6 - 8.5	Very Good Grade
UG3	6.6 - 7.5	Good Grade
AG	5.6 - 6.5	Marginal/Average Grade
LG1	4.6 - 5.5	Lowest in High Risk Grade
LG2	3.1 - 4.5	Moderate in High Risk Grade
LG3	1.6 - 3.0	Higher in High Risk Grade
LG4	0 - 1.5	Worst/Absolute High Risk Grade

Table 5.3 *Creating a degree of risk associated with each interval class along with the explanation.*

[42] FunctX XQuery Functions Library: www.xqueryfunctions.com/xq

5.3.2.2 Creating Derived Attributes

A fundamental work performed during this phase was to create some new attributes, resulting from the comparison of existing information, of considerable importance in subsequent development of Solution and Result for the CBR application, in-depth the next paragraphs.

New attributes in question are:

- lending_time: is the time difference in days between funded_dateandposted_date.Therefore, it indicates the length of time taken for the loan has been fully funded. More time indicates the loan offer is not so attractive or is more risky than the one with less time taken for being funded. It can consider as the proxy of social perception (mental or psychological judgement) of the lenders who evaluate the description (in text) of the loan applicant along with other financial aspects in their lending decision process.
- loan_delay: is the time difference in days between the last payment processed_date performed and disbursal_date + repayment_term – grace period[43].So, a positive result indicates a delay in payment on the loan.
- difference_in_paid_amount: is the difference between the sum of all payments andfunded_amount. Therefore, a loss of more than 5% of the amount financed (loss due to currency exchange) indicates that the loan has not been successful.
- per_payment_time_difference: for each payment (payment made), month and year of theprocessed_dateis compared with the month and year of the respective due_datefor local_payments (local payment) to ensure whether the payment is made in time (within the scheduled month). Mismatching indicates that the payment is not made in scheduled month and it is risky. [Possible Results: match&mismatch].
- payment_value_difference: for payment(s), the amount paid in local currency is compared with the amount planned in the scheduled month forlocal_payment. If the result is less than 20% of the amount planned, the payment is not regular [possible Results: paid_full installment on time, paid_less,not_paid, &unknown].
- status_integration: is an integration of the loan repayment status on time and value of the installments performed during loan period. It considers the above five latent factors (three factors are associated with time and the rest two are associated with value) derived from other attributes to evaluate the status (historical payment status = paid, default, in-repayment) critically. The paid status of any historical loan falls into the degree of lending risk if any of the five factors evaluates it for risky loan in terms of their respective scale mentioned in each factor category. For instance, the status ‘paid’ in historical result is not out of risk if mismatching was found for per_payment_time_difference that indicates the degree of risk associated with the loan made. Following the same rules, it may fall into the risky category of loan for other four factors that test/evaluate the loan for any risk belong. [Possible Results: regularly, not_totally_regular, in_delay, default_condition, serious_default_condition].

The result of hard work now deepens the creation of two attributes for payment_value_difference and per_payment_time_difference. It was necessary to create two

[43] Grace period is the initial time allowed to the field partner to start the repayment to the lenders after availing the loan as funded. It is usually 1 month.

new XML derived sources, then to compare: the first attribute that contains all the identifiers of the loans and each loan grouping local_payment (scheduled payments by the applicant) by month and year of due_date (scheduled dates) with relative amounts, added together if more than one in month; the second attribute that contains again all the identifiers of loans, but for every loan, this time, the identification of payments (payments) grouped according to month and year of the relevant processed_date (date of payment) with corresponding amounts. In this way, it was possible to compare payments with payments due, taking into account of all the possible variants, for example, only one payment made to compare with more payments due in the same month and so on. In Figure 5.14a depicts an extract of query XQuery code to make the first described operation, note the use of the group by keywords provided by support XQuery 3.0 by Saxon, without which it would have been much more difficult to make a complicated job.

```
for $loan in collection('file:/home/andre/eclipse/j_workspace/XQue
return
  <loan>
  { $loan/id }
  {$loan/status}
  <local_payments>{
    for $payment in $loan/terms/local_payments/local_payment
    let $year-month := year-from-dateTime($payment/due_date) ||
                       '-' || functx:pad-integer-to-length(month-
    group by $year-month
    order by $payment[1]/due_date
    return if ( count($payment/amount) = 1 ) then
     <local_payment>
    { $payment/due_date }
    <year-month>{ $year-month }</year-month>
    { $payment/amount }
    </local_payment>

    else <local_payment>
    { $payment/due_date }
    <year-month>{ $year-month }</year-month>
    { $payment/amount }
    <total_amount>{ round-half-to-even(sum($payment/amount), 2) }<
    </local_payment>

  }</local_payments></loan>
```

Figure 5.14 *aCode extract query to group the dates of the payments.*

```
3046 </loan><loan>
3047     <id>495338</id>
3048     <status>paid</status>
3049     <local_payments>
3050         <local_payment>
3051             <due_date>2012-10-29T07:00:00Z</due_date>
3052             <year-month>2012-10</year-month>
3053             <amount>1333.33</amount>
3054         </local_payment>
3055         <local_payment>
3056             <due_date>2012-11-05T08:00:00Z</due_date>
3057             <due_date>2012-11-12T08:00:00Z</due_date>
3058             <due_date>2012-11-19T08:00:00Z</due_date>
3059             <due_date>2012-11-26T08:00:00Z</due_date>
3060             <year-month>2012-11</year-month>
3061             <amount>1333.33</amount>
3062             <amount>1333.33</amount>
3063             <amount>1333.33</amount>
3064             <amount>1333.33</amount>
3065             <total_amount>5333.32</total_amount>
3066         </local_payment>
3067         <local_payment>
3068             <due_date>2012-12-03T08:00:00Z</due_date>
3069             <due_date>2012-12-10T08:00:00Z</due_date>
3070             <due_date>2012-12-17T08:00:00Z</due_date>
```

Figure 5.14b*To the left, extract resulting XML document.*

After creating the two new XML documents, it was necessary to create two queries: the first one to compare the amounts of payments made by the amounts of payments due, creating a SQL script that would update the attribute per_payment_value_difference for each payment identifier; the second one to compare the dates and update each attribute per_payment_ time_difference for each payment identifier, as shown in Figure 5.15.

```
for  $payment in $loan/payments/payment,
     $local in $due/local_payments

let $year-month := year-from-dateTime($payment/processed_date) ||'-' || functx:pad-integer-to-lengt
    $firstDate := $local/local_payment[1]/year-month,
    $lastDate := $local/local_payment[last()]/year-month,
    $processed_date := xs:date(xs:dateTime($payment/processed_date)) || ' ' || functx:time(hours-fro
                       minutes-from-dateTime($payment/processed_date),seconds-from-dateTime($paymen

return if ( functx:is-value-in-sequence($year-month, $local/local_payment/year-month) )
           then 'UPDATE loansDB.Payment SET per_payment_time_difference=''Match'' WHERE paymentID='

       else if ( (substring($year-month,1,4) < substring($firstDate,1,4)) or (substring($year-month
           then 'UPDATE loansDB.Payment SET per_payment_time_difference=''In_Advance'' WHERE paymen

     else if ( (substring($year-month,1,4) > substring($lastDate,1,4)) or (substring($year-month,1,
       then 'UPDATE loansDB.Payment SET per_payment_time_difference=''In_Delay'' WHERE paymentID='

     else if (  (substring($year-month,1,4) > substring($firstDate,1,4)) or (substring($year-month,
                (substring($year-month,1,4) < substring($lastDate,1,4)) or (substring($year-mont
       then 'UPDATE loansDB.Payment SET per_payment_time_difference=''Mismatch_but_in_time'' WHERE

     else 'UPDATE loansDB.Payment SET per_payment_time_difference=''UNKNOWN'' WHERE paymentID=' ||
```

Figure 5.15 *Extract the query code that compares the dates just grouped payments due to Figure 5.14b with the dates of payments made and returns a SQL statement to update per_payment_time _difference the attribute in the table Payment.*

5.3.3 COLIBRI2 Platform for CBR System

jCOLIBRI2[44] is one of the reference platforms for application development with an approach, CBR. The CBR application was implemented in Eclipse Luna[11] Java library for importing jCOLIBRI2, and the developer will have to implement the "cbrapplications.StandardCBRApplication" interface to divide its behavior in 4 steps:

- configure (): is a configuration method to set the application- base case, connectors, ontology etc.
- precycle (): is typically performed when the cases are read and organized in a database of cases, as in our case. The method returns the database of cases with cases stored. It is performed only once.
- cycle (): runs the CBR cycle with the given query. It can be run multiple times.
- postcycle (): executes the code to terminate the application. Typically closes connector.

[44] Framework jCOLIBRI2: gaia.fdi.ucm.es/research/colibri/jcolibri

5.3.3.1 Application Configuration

First of all, the application must be able to interact with the underlying database. jCOLIBRI divides the problem of managing the base case into two separate but related concepts: persistence mechanism and memory organization. Persistence is built around the connectors, objects that know how to access and retrieve case and return them to CBR system. Among the three types of connectors it has been chosen to use *jcolibri.connectors.DataBaseConnector* that organizes the persistence of cases in the database using Hibernate[45] library internally, middleware that interfaces directly with the database and creates SQL queries automatically. The connectors are configured through the databaseconfig.xml configuration file, shown in Figure 5.16.

```
1 <DataBaseConfiguration>
2     <HibernateConfigFile>jcolibri/myapplication/hibernate.cfg.xml</HibernateConfigFile>
3     <DescriptionMappingFile>jcolibri/myapplication/Description.hbm.xml</DescriptionMappingFile>
4     <DescriptionClassName>jcolibri.myapplication.Description</DescriptionClassName>
5     <SolutionMappingFile>jcolibri/myapplication/Solution.hbm.xml</SolutionMappingFile>
6     <SolutionClassName>jcolibri.myapplication.Solution</SolutionClassName>
7     <ResultMappingFile>jcolibri/myapplication/Result.hbm.xml</ResultMappingFile>
8     <ResultClassName>jcolibri.myapplication.Result</ResultClassName>
9 </DataBaseConfiguration>
```

Figure 5.16 *The connector configuration file. As one can see where it explicitly states find the configuration file for Hibernate mappings with a description, result and solution and the path to the classes that represent them.*

The second layer of the database management of cases is the data structure used to organize after the cases are loaded into memory. It was decided to use that *jcolibri.casebase.LinealCaseBase* stores the cases in a list object. Before describing the *Hibernate* configuration, it has been discussed the representation of cases jCOLIBRI.jCOLIBRI represents cases using Java Beans, which is a class that has a *get()* method and *set()* for each attribute its audience. *Hibernate* uses its Java Beans to interact with the database. In addition, each case must implement the interface *jcolibri.cbrcore.CaseComponent*, which binds every class to have a unique attribute that identifies each component. In short, cases and queries consist *CaseComponents*. A case will be divided into four components: description of problem, solution to the problem, result of the application of the solution, and justification of the solution (optional- why that solution was chosen).Each query always has a description, defined in *jcolibri.cbrcore.CBRQuery* class. This class represents the defining query as a description of the problem or case. The query is then a case without solution or outcome. In our application, a case is a query plus a solution and a result. In fact, *jcolibri.cbrcore.CBRCase* extends *jcolibri.cbrcore.CBRQuery* class by adding the *get()* and *set()* for each component's justification, result and solution. At this point, must configure *Hibernate* so that it can interact appropriately and automatically with the database already created. Its configuration file, *hibernate.cfg.xml*, is simple and

[45]Hibernate Middleware: www.hibernate.org

requires to enter the class of MySQL driver, URL for the connection, username and password. Of course, much more articulated items are the mapping file, fundamental for mapping a Java Bean in a table of the database. One has to define which table is used to store the Bean, then configure which column of the table contains each attribute of the Bean.

```
1 <?xml version="1.0"?>
2 <!DOCTYPE hibernate-mapping PUBLIC "-//Hibernate/Hibernate Mapping D
  mapping-3.0.dtd">
3 <hibernate-mapping default-lazy="false">
4 <class name="jcolibri.myapplication.Description" table="Loan_Request
5     <id name="loanID" column="loanID">
6       <generator class="native"/>
7     </id>
8     <many-to-one name="fieldPartner" column="fieldpartnerID" not-nul
9     <many-to-one name="borrower" column="borrowerID" not-null="true"
10    <property name="amountUSD" column="amountUSD"/>
11    <many-to-one name="sector" column="sectorID" not-null="true" cas
12    <property name="repaymentTerm" column="repayment_term"/>
13    <property name="repaymentInterval" column="repayment_interval">
14        <type name="jcolibri.connector.databaseutils.EnumUserType">
15            <param name="enumClassName">jcolibri.myapplication.Descr
16        </type>
17    </property>
18    <property name="disbursalDate" column="disbursal_date">
19        <type name="jcolibri.connector.databaseutils.EnumUserType">
20            <param name="enumClassName">jcolibri.myapplication.Descr
21        </type>
22    </property>
23 </class>
24 <class name="jcolibri.myapplication.FieldPartner" table="Field_Partn
25    <id name="fieldPartnerID" column="fieldpartnerID"/>
26    <property name="name" column="name"/>
27    <property name="rating" column="rating"/>
28 </class>
```

Figure 5.17 *Hibernate mapping file for the class in the Description of Loan_Request table.*

In our case, we are going to map attributes of the solution and result in the same class table, i.e.,Loan_Request. The *Description* class, however, will be mapped over that Loan_Request on the table. Also on its compound attributes of the tables, each of which has a Java Bean with an ID and other simple attributes. The mapping documents are in .hbm.xml format and different XML tags must be used for the configuration, depicted in Figure 5.17.The tag<*class*> is the class that must be stored in which table. The loanID attribute of the class is mapped in the Description column of the table loanIDLoan_Request. The tag <*id*> indicates that the attribute is the primary key of the table. The other simple attributes are mapped to the respective columns of the table using the tag <*Property*>.For compound attributes, such as *fieldPartner* used the tag<*many-to-one*> and the attribute is mapped to the primary key of the table Field_Partner, and then declared in a new tag <*class*>. The tag <*type*>, nested in the <property>, is used when attributes are not of a built-in type Java. For example, user-defined types or enumerations.

5.3.3.2 Recovery, Reuse and Adaptation

The *configure()* and *precycle()* are completed. Now it is the time to invoke the *cycle()* method that deals with the recovery phase in which it obtains the most cases similar to the query date. The primary method to perform recovery is in class *jcolibri.method.retrieve.NNretrieval.NNScoringMethod*. This class provides the method *Nearest Neighbor* to compare the attributes. It uses a global similarity function for compounds attributes (CaseComponents) and a local similarity function to compare the simple attributes. In our case, *NNScoringMethod* calculates the similarity for each simple attribute and then the global similarity by performing the weighted average of the results of these local similarity for each case in the database. If the overall result is 0, no similarity, and 1, complete correspondence. The functions of these configurations are stored by the object *jcolibri.method.retrieve.NNretrieval.NNConfig* (see Figure 5.18), in which we find:

- global similarity function for the description.
- global similarity function for each attribute compound (except the Case Component-the description).
- local similarity functions for each simple attribute.
- weight for each attribute.

In *jcolibri.method.retrieve.retrieval.similarity* package, there are different functions for similarities including the *GlobalSimilarityFunction* and *LocalSimilarityFunction* interface implemented in the corresponding similarity functions. The widely used global similarity is jcolibri.method.retrieve.NNretrieval.similarity.global.Average which calculates a weighted average of the similarity of its sub-attributes, be they simple or compounds. Local similarities are as follows:

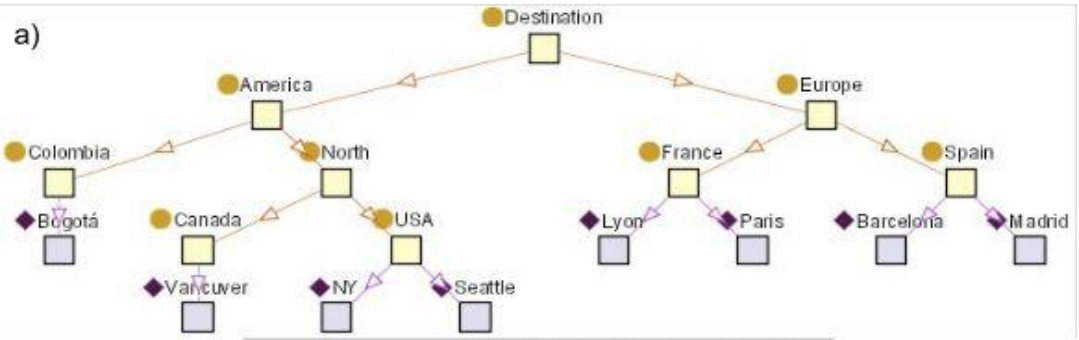
- jcolibri.method.retrieve.retrieval.similarity.local.Equal, which returns 1 if the two simple attributes are equal, otherwise 0.
- jcolibri.method.retrieve.NNretrieval.similarity.local.Interval which returns the similarity between two numbers within an intervalsim(x,y) = 1 - (|x - y | ÷ interval).
- jcolibri.method.retrieve.NNretrieval.similarity.local.ontology.OntCosine which calculates the cosine similarity (see Figure 5.19b) on ontologies, described below.

```
// Obtain configuration for KNN
final NNConfig simConfig = new NNConfig();
simConfig.setDescriptionSimFunction(new Average());

simConfig.addMapping(new Attribute("amountUSD", Description.class), new Interval(900));
simConfig.setWeight(new Attribute("amountUSD", Description.class), 0.15);

simConfig.addMapping(new Attribute("repaymentTerm", Description.class), new Interval(10));
simConfig.setWeight(new Attribute("repaymentTerm", Description.class), 0.10);

simConfig.addMapping(new Attribute("repaymentInterval", Description.class), new Equal());
simConfig.setWeight(new Attribute("repaymentInterval", Description.class), 0.05);

simConfig.addMapping(new Attribute("disbursalDate", Description.class), new Equal());
simConfig.setWeight(new Attribute("disbursalDate", Description.class), 0.05);

simConfig.addMapping(new Attribute("borrower", Description.class), new Average());
simConfig.setWeight(new Attribute("borrower", Description.class), 0.15);
simConfig.addMapping(new Attribute("gender", Borrower.class), new Equal());

// Configure the OntCosine() function for the similarity of Country
simConfig.addMapping(new Attribute("countryID", Borrower.class), new OntCosine());
```

Figure 5.18 *Code extract showing the configuration of NNConfig (). Note the mapping the attributes both simple compounds, in which the type of similarity is set corresponding.*

The application uses the ontologies to represent country and sector, creating four classes based on the risk attributed to each of them: *maximum_risk, high_risk, intermediate_risk* and *low_risk.* The countries and areas are defined in a file and follow roughly owl the tree hierarchy shown in the example of Figure 5.19a. In this way, the result of the cosine function applied to two countries (or sectors) will be equal to 1 if they are in the same risk class, instead it will be close to 0 if they belong respectively to the minimum and maximum risk class. For the moment, the function compares the following attributes: *amount, repayment_term, repayment_interval, disbursal_date, gender, field_partner_rating, country* and *sector*, to which the following weights are associated: *amount* 0.15 *repayment_term* 0.10 *repayment_interval* 0.05 *disbursal_date* 0.05, *borrower* 0.10, *field_partner* 0.35, *country*0.10 *and Sector* 0.10.Note that for the attribute-field_partner_rating, the weight is set to the respective compound attributes. After evaluating the similarity with the query, it must be selected only the case with the highest rating, using a process called kNN retrieval equipped in jcolibri.method.retrieve.selection.SelectCases class. At this point, it passes to the phase of re-use (or adaptation), which adapts the solution of the case retrieved at the request of the query. It is a very dependent transition from the domination of application, in our case, a method of basic adaptation is used, i.e., perform the direct copy of the query attributes, replacing the ones of the case just selected. Next, the user, expert of the application domain, can change the solution better adapt to the newly copied attributes.

a)



	fdeep_basic	fdeep	cosine	detail
Sim(Mad,Bcn)	3/4	1	1	5/6
Sim(Mad,Paris)	1/2	1/2	2/3	3/4
Sim(NY,Seattle	1	1	1	7/8
Sim(Seattle, Vanc)	3/4	3/4	3/4	5/6
Sim(Bcn,Bogotá)	0	0	$1/\sqrt{12}$	1/2
Sim(Mad,Bogotá)	0	0	$1/\sqrt{12}$	1/2
Sim(Vanc,Bogotá)	1/2	1/2	1/2	3/4

b)

$$\text{cosine}(i_1, i_2) =$$

$$\text{sim}(\text{t}(i_1), \text{t}(i_2)) = \frac{\left|\left(\bigcup_{d_i \in \text{t}(i_1)} (\text{super}(d_i, CN))\right) \cap \left(\bigcup_{d_i \in \text{t}(i_2)} (\text{super}(d_i, CN))\right)\right|}{\sqrt{\left|\bigcup_{d_i \in \text{t}(i_1)} (\text{super}(d_i, CN))\right|} \cdot \sqrt{\left|\bigcup_{d_i \in \text{t}(i_2)} (\text{super}(d_i, CN))\right|}}$$

Figure 5.19 *Images and descriptions taken from jCOLIBRI Tutorial. a. As the result of various functions of ontologies that represent some cities in the world. b. The cosine function which, in our case, was more appropriate than the other, in which we find: CN: is the 'set of all concepts in the knowledge base, super (c, C) is the subset of concepts in C that are super concepts of c, and t (i) is the set of concepts of which the individual i is the instance.*

5.3.3.3 Retention

In the last phase, the case just adopted is stored in the case base (or DB) with a new ID. It was decided not to use Hibernate, internal jCOLIBRI to save the event, but MySQL Connector / J[46], the JDBC driver to communicate with the MySQL server, saw the approach web-based application. This creates a new connection to the MySQL server, then recognizing to the method executeUpdate () that can execute an INSERT INTO statement to add a new Loan_Request record in the table, by entering the values of the appropriate attributes; the method returns a ResultSet object that contains the new unique ID of the loan, created automatically with the option of auto increment of loanID field. As shown in Figure 5.20, the developer must create a new record in the Borrower table before proceeding with the inclusion of the new loan.

[46]MySQL Connector/J: dev.mysql.com/doc/connector-j/en/index.html

```
    conn = DriverManager.getConnection("jdbc:mysql://localhost:3306/loansDB", "root", "password");

} catch (SQLException ex) {
    // handle any errors
    System.out.println("SQLException: " + ex.getMessage());
    System.out.println("SQLState: " + ex.getSQLState());
    System.out.println("VendorError: " + ex.getErrorCode());
}

try {

    stmtBorrower = conn.createStatement();
    stmtBorrower.executeUpdate("INSERT INTO Borrower (gender, countryID) "
            + "VALUES (" + gender + "," + countryID +");", Statement.RETURN_GENERATED_KEYS);

    rsBorrower = stmtBorrower.getGeneratedKeys();
    if (rsBorrower.next()) {
        borrowerID = rsBorrower.getInt(1);
    }
    stmtLoan = conn.createStatement();
    stmtLoan.executeUpdate("INSERT INTO Loan_Request_min (borrowerID,fieldPartnerID,amountUSD,repaym
            + "repayment_interval,status,score,sectorID,disbursal_date) "
            + "VALUES ("+borrowerID+","+ FPID+","+ amount+","+repTerm+",'"+repInt+"','recorded',"+ri
            +","+sectorID+",'"+ disbDate+"');", Statement.RETURN_GENERATED_KEYS);


    rsLoan = stmtLoan.getGeneratedKeys();

    if (rsLoan.next()) {
        loanID = rsLoan.getInt(1);
    }
```

Figure 5.20 *Code excerpt regarding the storage of the new case in database.*

Attribute values are those arising from the previous phase of recovery and adaptation, except for the status of the loan, it is associated with a "recorded" value (registered) and for remaining attributes in Loan_Request that is automatically assigned null, because one can enter a value only after the conclusion of the loan, with success or failure.

5.3.4 GWT Application Online

The GWT web application will allow the user, in Figure 5.21, to do two things: i) run the CBR application that will return the most similar loan to the request sent with the possibility to adapt its solution and store the new case in database, providing the user's ID that will serve in the future to enter the result, and ii. enter the result of a completed loan, stored in the database with the first operation, searching for it using its identifier.

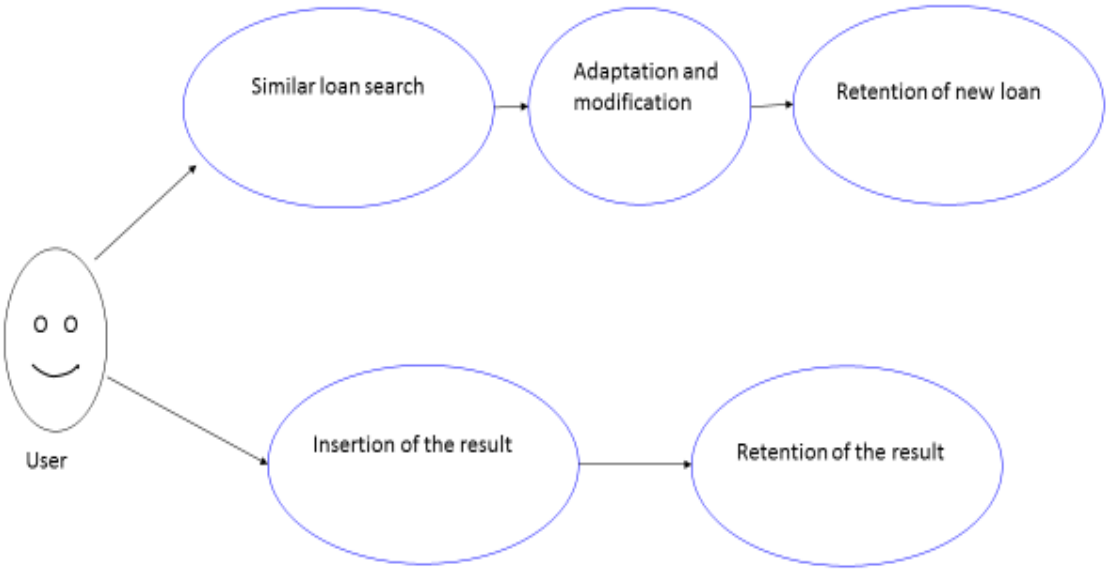


Figure 5.21 *Application usage.*

In terms of design, the application is structured in panels, one for each task. How long it concerns the search for a similar case in the database a panel is shown for the step of submitting the query where the user is expected to indicate the amount of the loan, the term of repayment, the repayment period, the type of financing, the applicant's gender, the field partner's rating, the business sector, and finally the country. After sending the request, in addition to the just mentioned associated attributes, the more similar loan selected by the CBR system is displayed on a new panel, including its outcome and risk score. Following a panel shows the adaptation of the solution to request made with the possibility of manual modification of the risk score and finally a panel for storing the new event in the database is shown in which it is asked to indicate the name of the field partner. For the operation result of the insertion, instead, a panel is shown in which the user inserts the identifier of a loan. A request sent is shown a field for the insertion of the loan result, only in the case where it is not present.

5.3.4.1 Implementation

GWT provides the Google Plug in for Eclipse[47]available that divides the project into different packages, depending on whether the Java source code belongs to the client-side, the server-side or is shared by both, as shown in Figure 5.22.Another important file is the XML document called GWTCBR.gwt.xml containing the path to the Java class used as the starting point of the application, which implements obligatorily the EntryPoint interface and its onModulLoad () method, and also declares definitions of GWT modules inherited. That is, the path of files that define each library used in the class entry points (see Figure 5.23).

[47]Google Plugin per Eclipse: developers.google.com/eclipse/?hl=en

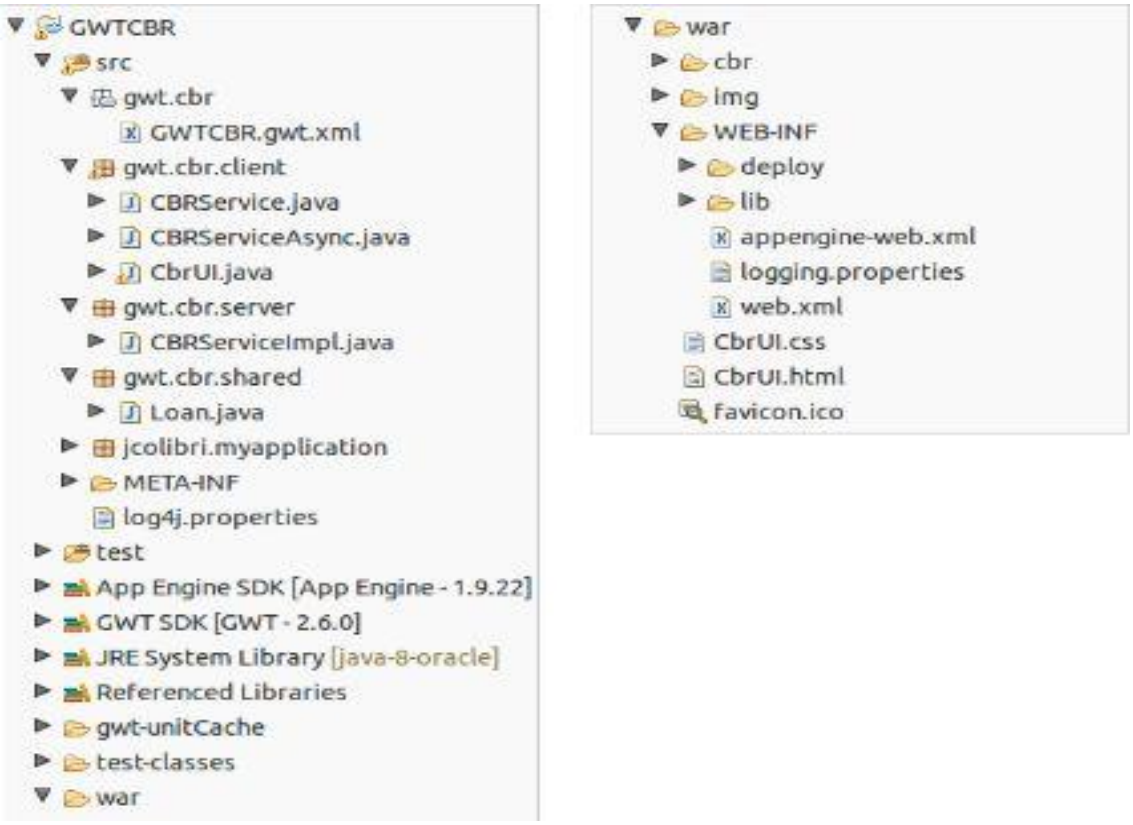


Figure 5.22 *Organization of the GWT project in Eclipse for a web based approach.*

In client-side package we include:

- the Java class CRUI, just the starting point, just mentioned, the application that it contains the implementation in Java language UI, translated later in JavaScript by GWT compiler.
- the CBRService interface, which defines the RPC service, inheriting the GWT interface RemoteService and defining the RPC methods sendLoanDescription (), which accepts the new loan attributes entered by the user and returns a Loan object (the most similar loan), and retainLoan () accepting the new loan attributes with the solution adapted and returns the new identity of the loan.
- theCBRServiceAsync interface, which defines the same methods in CBRService with the addition of the AsyncCallback parameter to perform asynchronous calls from client.

In the package of the server-side, instead we find the class CBR ServiceImpl that the implementation CBR Service interface, which also inherits the GWT class RemoteServiceServlet with task to de-serialize incoming requests arising from the client and serialize the answers output from the server. There are also the war folder containing static resources such as style sheets, images and so-called host HTML page to be served, the directory war / WEBINF containing Java web application files, and related libraries are included in the folder war/WEBINF/lib. The host page, CbrUI.html, is the HTML page where the code for the Web application is executed, and refers to both the application's style sheet, CbrUI.css, both the code path JavaScript source, generated by the GWT compiler, responsible for the dynamic elements the page, declared mandatory as an identifier of an

HTML used to position the content created dynamically. The host page can contain, of course, also static elements.

```
<!DOCTYPE module PUBLIC "-//Google Inc.//DTD Google Web Toolkit
  "http://google-web-toolkit.googlecode.com/svn/tags/2.6.0/dist
<module rename-to='cbr'>
  <!-- Inherit the core Web Toolkit stuff.
  <inherits name='com.google.gwt.user.User'/>

  <!-- Inherit the default GWT style sheet.  You can change
  <!-- the theme of your GWT application by uncommenting
  <!-- any one of the following lines.
  <inherits name='com.google.gwt.user.theme.clean.Clean'/>
  <!-- <inherits name='com.google.gwt.user.theme.standard.Stand
  <!-- <inherits name='com.google.gwt.user.theme.chrome.Chrome'
  <!-- <inherits name='com.google.gwt.user.theme.dark.Dark'/>

  <!-- Other module inherits
  <inherits name='eu.maydu.gwt.validation.ValidationLibrary'/>
  <inherits name='eu.nextstreet.gwt.components.IntoGwt'/>

  <!-- Specify the app entry point class.
  <entry-point class='gwt.cbr.client.CbrUI'/>
```

Figure 5.23 *Extract of the GWT module configuration file.*

5.3.4.2 Compilation and Deployment on a Web Server

After filling the application in JavaScript, different products files will automatically be in the output folder / war / CBR, depicted in Figure 5.24, named with GUIDs (Unique identification number), which contain implementations Java script application, one for each supported web browser. Now just upload these files, in addition to static resources in/war described earlier, in the output folder of any web server that supports static web pages.

Figure 5.24 *Files generated by the GWT compiler in Eclipse.*

5.3.4.3 User Demo

The simple starting page, shown in Figure 5.25, allows the user to choose between the two options already described, or try the case more similar to its request or enter the result in a loan that does not have it.

Figure 5.25 *Initial web application page.*

By clicking on the first option ("Record a new loan") it will be displayed to the user in the panel in which it will be possible to insert the required attributes to the CBR application for the calculation of similarity, shown in Figure 5.26. Note the client-side validation performed on Loan amount, Repayment term, and Country; for first attribute, the system will display the label "OK!" only in the case in which the user must type a integer between 1 and 10,000 (1 and 100 for the term of repayment), in case instead you will see the error label that prompts the user to the correct input insert, as shown on the row for Repayment term, and will be applied to an edge red to associated text field with the aim to capture the user's attention; the field of the country, however, allows the user to enter the attribute in two different ways: by clicking the arrow on the far right, the user will see the list of all available countries, or start typing, the user will see a list of only those countries that correspond to the letters type (highlighted in green in the names in the list), up auto completion of the word in when there are no more possible alternatives. Even in this case, if the word typed does not match any country in the list, you will see the label and the edge error. The other attributes will be inserted through the selection, by the user, a voice from a drop down menu, then the user will not need the validation. The Reset button allows the user to clear all fields, setting them as after the first loading of the panel.

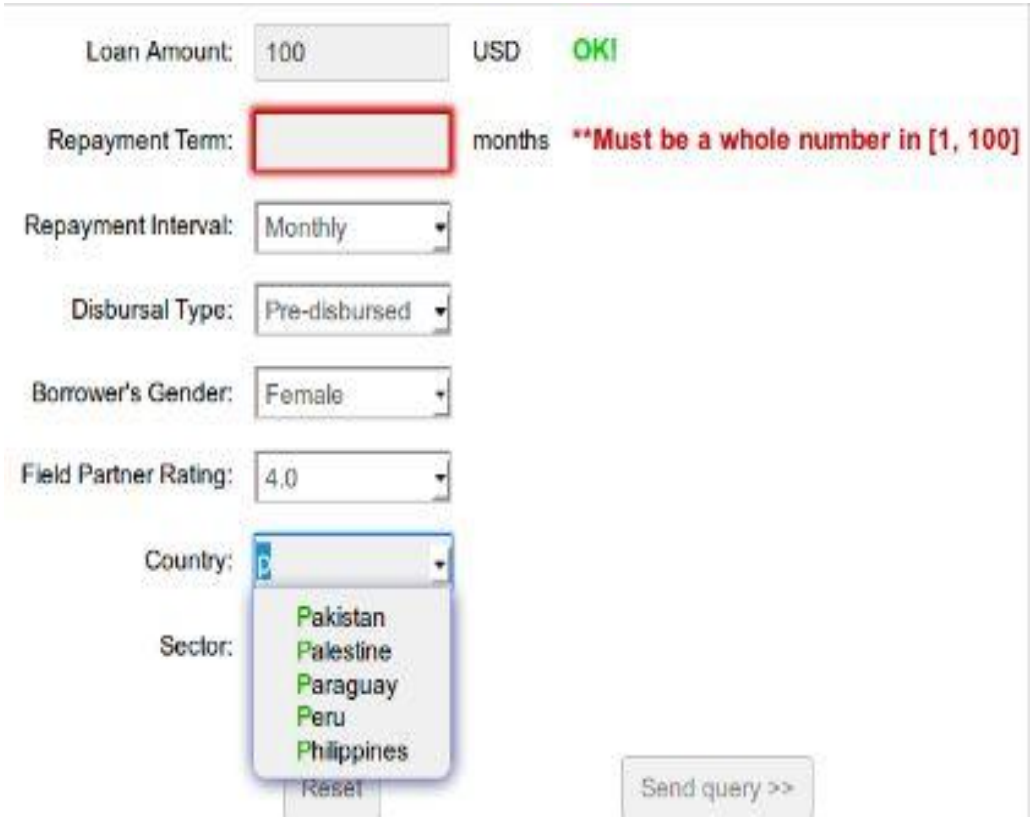


Figure 5.26 *Panel for the insertion of the required attributes. The image also shows the validation mode for the three labels Loan amount, Repayment Term and Country.*

By submitting the request, the CBR application selects the case most similar to it and will return its attributes, shown in a new panel shown in Figure 5.27, with the addition of LoanID (unique loan identifier), percentage of similarity with the request, Loan status (as Result for

CBR) and associated Risk Score (as Solution for CBR).The user is given the option to continue with the adaptation phase by clicking on Button Proceed to revise, or can return to the previous screen by clicking the Back button. In the panel on the adaptation phase, shown in Figure 5.28, is displayed the solution (associated Risk Score) of the recovered loan, re-using the attributes of the request, with the option to adjust manually by entering a numeric value between 0 and 10 including the extreme value. Even in this case, error label will be shown next to the text box, and the edge of red color if the input is not correct. By clicking on Proceed to retain button, it will show the last option panel Record a new loan that indicates the user to enter the missing information on name of the Field Partner (local partner), mandatory for the proper storage of new loan in the database. Now, retain by clicking the button, the system will register the loan with a status equals recorded (registered) and the user will be displayed its new identity, essential for the operation Insert the result of a recorded loan. At this point, an enable button allows the user to return to the home page Figure 5.25.

Figure 5.27 *Panel in which case the attributes retrieved from the application are displayed CBR. It tips the user to continue.*

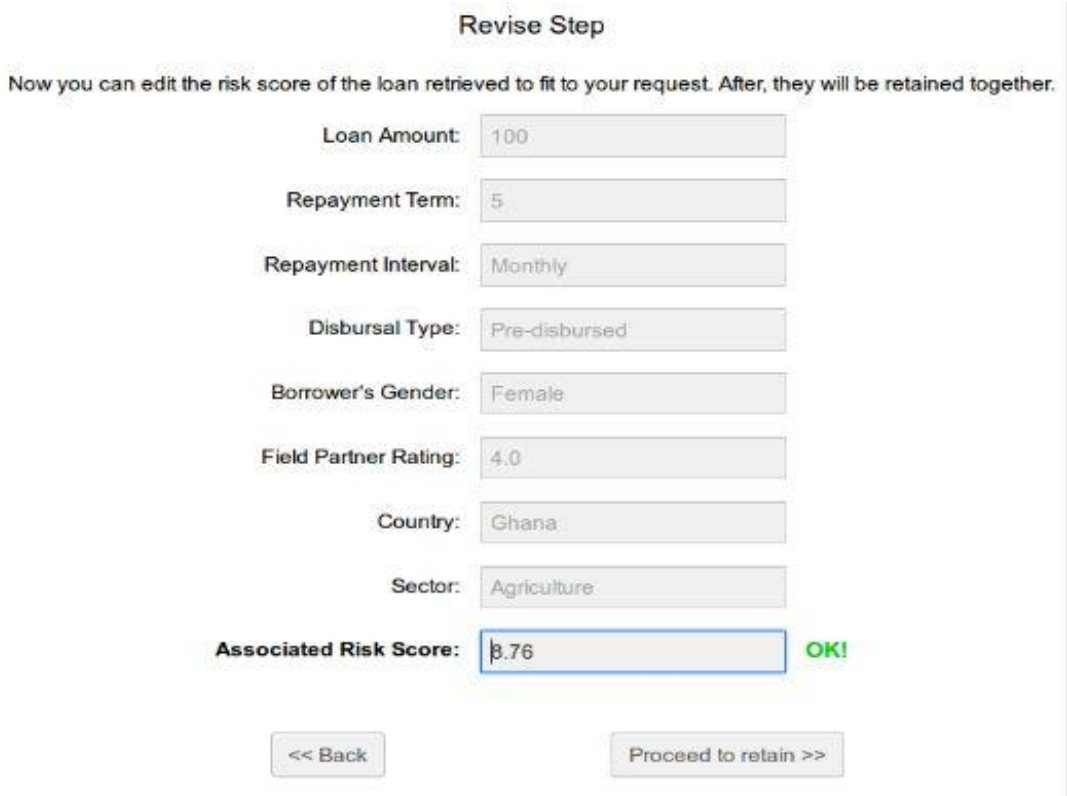


Figure 5.28 *Panel concerning the adaptation phase, client side validation of the text box.*

The second operation possible, insert the result of a recorded loan, start with display a new panel that displays a drop-down menu, in which they are listed only the identification of the loans without result, and a Search button. After clicking the button, the user will see a new panel with a summary of loan attributes and a drop down menu from which the user can choose the loan results from the following alternatives: Regularly paid, paid not totally regular, paid in delay, expired, defaulted. Retain the result by clicking on the button, the result will be stored and added to the user will be redirected to the home page.

5.4 Implications of the CBR System and Effectiveness of the Credit Score

The prototypical solution has met the expectation of achieving an application that can be used by a domain expert user, albeit with a mainly demonstrative purpose with a manual creation of a risk score for only 107 loans: a too small number to establish a clear assessment of the effectiveness of such a score in relation to the effective result of the loans, but still acceptable for this first prototype. The most important work was definitely the design of a knowledge base adequate both for the CBR application, which uses the stored information on the calculating the similarity between the cases, both for the creation of function of the risk score, not yet automated and likely future development, which also exploits all derived attributes described in Section 4.2, the object of hard work with XQuery which lasted for most of the internship.

At first, in the database they were included 21 cases which, however, have proved little heterogeneous in general, so it was decided, with the support of a domain expert, the add others more wisely reaching 107 cases stored. The database was conceived with an eye to the future, including additional information such as the city of residence of the applicant and the activity associated to the specific sector of the loan that were not taken into account because they are considered too detailed in this stage. In addition, one can expand it by inserting new records of countries and sectors, for now, they are about half of those visible on the Kiva platform, and also including loan requests resulting from groups of more people.

The similarity function and that for the calculation of the risk score are related under a certain point of view, in the sense that all of the first attributes are also involved in second function, so it is reasonable to think that one can improve both adding new parameters or modifying the weight of each of them, analyzing them more thoroughly. The advantage and the disadvantage of the results arising from the second mentioned feature, it is calculated manually, so it was possible to integrate more parameters but produce a few scores. Instead, bearing in mind the training period, the function of similarity is much simpler, but it has been already found how to improve it, for example, extracting, automated, targeted information from the text description of the applicant already in the database for this purpose, such as age, marital status, state of employment of family, number of children and loan past experiences and going to enhance local similarity functions used for the comparison of simple attributes, both numerical that for ontologies.

Analyzing the adaptation phase of the solution, one realizes that, in fact, the first is a null adaptation and the task of testing and change the solution is entirely up to the domain expert, without a minimal support from the CBR application. This could be an additional proposition to create a function able to adapt the based solution to the confrontation between the user input parameters in the initial stage and those of if selected by the system.

The main objective of the application created with GWT was never to be visually impact, but the realization of web-based interface, albeit minimal, that allows the user to easily avail the system. Here too, it was decided to grant the user only the two basic steps, but it will extend the choices and the user control in the near future, for example, with the possibility, in the case of the new phase of storage, to insert a Field Partner not present in the list, effectively creating a new record in the homonymous table. In short, this first prototype is a good starting point to develop and refine, as well as to automate aspects just described which, for the moment, are still manual and require an extreme amount of work to the expert.

It offered the developer the prototype project was both lucrative and challenging from an academic point of view. The interdisciplinary approach tied to microcredit has made the developer closer to a really unknown topic and think about the possibilities that this system can only be the beginning of a more concrete system that can truly help creditors, suggesting the risk they will go in against, and applicants, giving them the advantage of being more visible assigning a positive score, fills the researcher with satisfaction.

5.5 Summary

This chapter has introduced the trends of microcredit systems towards web-based P2P lending platforms. For supporting the microcredit, it has designed and developed the CBR System and proprietary database along with technical details. Before the development, this chapter has also discussed the relevant technologies and finally, it has ended up with the implications of the CBR system and effectiveness of the credit score.

Chapter 6

Borrower Risk Scoring

Credit Scoring[48] is one of the innovative risk assessment (in broader term, management) tools in finance that assesses the credit[49] risk of borrowers. Credit risk or the risk of default by the borrower has been identified as one of the main risks in the Basel II framework and therefore an internal rating-based approach has been proposed as one of the methodologies that allows and encourages banking institutions to develop their own internal measures for the assessment of credit risk capital. Basel II explicitly mentions credit scoring as a possible technique to determine drivers of credit risk (Gool et al., 2012). Typically bank loan officers do this assessment subjectively based on their experience and used borrower's various information like character (reputation), capital (leverage), capacity (volatility of earnings), collateral (guarantee), and condition (macroeconomic cycle) in a score card(Vaish et al., 2011). In microcredit, credit scoring is considered as a third risk management innovation after two initial tools: joint-liability groups and skilled loan officer's careful evaluation to judge credit repayment risk of an individual applicant based on information available with respect to personal, social, business and chattel (Schreiner, 2005).

Credit scoring quantifies repayment risk of a borrower through establishing the historical link between repayment performance and the quantified characteristics of a loan applicant. The prime purpose of the use of credit scoring is to rationalize decision-making of lending. A loan is basically known as debt which is provided by any organization or individual to another organization or individual at an interest rate. In this case, it is evidenced mainly by a promissory note that states three basic information: amount of money borrowed, the interest rate the lender is charging, and date of repayment. According to Investopedia, a loan is defined essentially as a promise, and a credit rating determines the probability that the borrower will pay back a loan within the date of repayment. A high credit rating specifies a higher likelihood of paying back the loan in its entirety without any issues; on the contrary, a poor credit rating implies the fact that the borrower has a previous history of default loan and might follow the same pattern in the future. Hence, the credit rating helps the lenders to take decision in approving a loan or fixing terms for a particular applicant for loan [http://www.investopedia.com/terms/c/creditrating.asp].

Credit scoring does not have a long track record in microcredit even in developed countries. Dinh & Kleimeier (2007) found that in 1996, 97% of all US banks used credit scores for credit card applications and 70% of them used the same for small business loans. They also found the wide spread of the usage of credit scores across the world to banking sectors in other developed countries. In web-based P2P lending, which is the focus of this study, most of the direct P2P lending models[50] like Prosper (USA), Zopa (UK), Smartika (Italy) operate

[48]Alternative terms used in different sections of the study are borrower risk rating or scoring, credit risk grading, lending risk rating.

[49]There are alternative terms of 'credit': loan, advance etc. These terms have been used synonymously based on the context in the respective section.

[50] Direct P2P Model allows borrowers and lenders to connect directly, eliminating conventional intermediaries (bank or other financial intermediary), to provide for greater access to credit at a lower cost; Indirect P2P Model typically provides capital to developing markets by connecting borrowers and lenders through local intermediaries or field partners (Hassett et al., 2011).

nationally and capture the retail market for consumer loans and credit card loans across the world, more particularly in the US (Weib et al., 2010). Although the use of credit scores for consumer and credit card loans is evidenced widely in developed countries, there exists a limited number of empirical evidence and scientific studies on credit scoring for microlenders in developing countries. The main challenge of the studies in developing countries is contextual in general, and is the lack of data adequacy on borrower characteristics and their credit histories in particular(Dinh & Kleimeier, 2007; Vogelgesang, 2003). Many of the studies including the first work by Vigano (1993) on the application of the credit scoring models for microfinance used organizational or national data capturing Latin American, East European-Central Asian, and African markets as their sample countries(Clarke, Cull, Peria, & Sánchez, 2005; Gool et al., 2012). Besides, there are criticisms against the methodologies used in credit scoring models for microfinance. On the one hand, existing credit scoring is done mostly on qualitative judgment(Gool et al., 2012), on the other hand, the performance of credit scoring models for microfinance is usually described using percentage correctly classified (PCC) or (pseudo-) R^2, which is considered methodologically wrong, and it does not allow to compare different scoring models in a best way(Baesens et al., 2003). More interestingly, to the best of our knowledge, no studies have been focused on the risk scoring of borrowers in online P2P microcredit lending platforms[51] operating globally. Regarding lending risk, most of the P2P lending platforms (such as kiva.org) declare the risk of lending is absolutely on the lenders who decide to fund the borrowers. The platforms merely keep typical advices for lenders to diversify their portfolios through lending to more than one borrower via different field partners as well as in different countries and/or sectors. The platforms also provide field partners' risk rating which proxies the level of risk for all borrowers from a particular field partner, but not a direct credit risk score of individual borrowers who are the target to the P2P platforms for funding.

This study, therefore, aims to extend the knowledge on credit scoring for microfinance by developing credit scoring models using data from a P2P lending platform, Kiva.org[52].The Case-Based Reasoning (CBR) approach is adopted for rating the borrowers in online P2P lending platform. To solve the cold boot problem in CBR, the following strategic position has been adopted in Figure 6.1.

[51] Risk rating with credit score is available in most of the direct P2P platforms who operate nationally like Prosper, Zopa which are out of this study (Ceyhan et al., 2011; Slavin, 2007; H. Wang & Greiner, 2011). However,this study focuses only on the indirect P2P lending models who operate globally like Kiva, Zidisha(Hassett et al., 2011).

[52] Kiva is the leading platform among the indirect P2P lending models which allows researcher accessing data from its open source database (https://build.kiva.org/). But the other models like Zidisha, MyC4 have no access to data though they maintain the similar database.

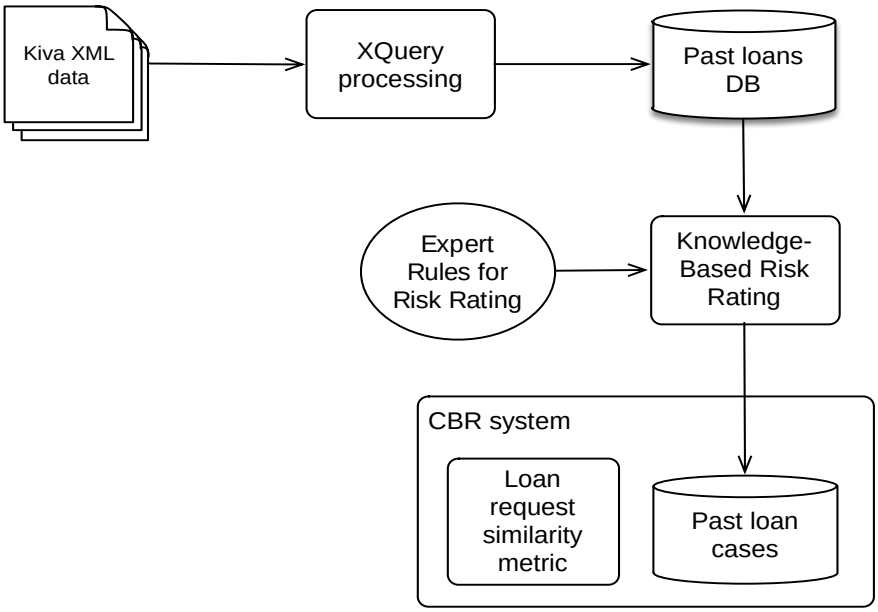


Figure 6.1 *Workflow for setting up the CBR system from initial Kiva data (Uddin at el., 2015)*

The main issue with the application of this approach to the present problem is certainly not the lack of data. In fact, Kiva makes available all the information associated to past loan requests and to the actual repayments made by the borrowers. All the necessary information including borrower description, outcome or payment history are available except risk rating. For resolving the missing component, expert rules have been employed for rating the risk associated to loan requests in developing countries of African and Asian zones, coded into a spreadsheet. Some features are not readily usable due to the need to interpret elements of the borrower description like age, experience, marital status etc. written in natural language and not structured in fields of a database. Moreover, the above mentioned rules are not completely formalized and the experts sometimes manually modify the results of their direct application to define the risk rating. Expert rules are based on their opinion regarding objective assessment of subjective judgment in order to arrive at the credit risk score for rating of borrower's risk. The obtained credit risk scores will be required to be validated in holdout sample. The knowledge elicitation activity carried out in order to define these rules was characterized by five interviews with experts in risk assessing in the microcredit context. These experts are actually proficient in the usage of spreadsheets(Uddin et al., 2015).

In what follows, the following section starts with an overview of literature review, and provides a thorough analysis of the credit scoring for microfinance. Then, the process of developing a spreadsheet-based credit scoring system for microfinance has been described. The performance of the applied scoring models has been done using the holdout samples in CBR system.

6.1 Credit Scoring Models in Microfinance

This section traces the origins of credit scoring and its evolution in order to identify the latest trend in credit scoring in microfinance and to justify the methodology in developing a credit scoring application. In Gool et al. (2012), the authors provided an overview of eight credit scoring models for developing countries published till 2012(see in Appendix A).

The progress of credit scoring models in the microfinance sector is very insignificant. Table 6.1 gives an overview on credit scoring models in microfinance.

Author (Date, Country)	**Institution type**	**Sample size**	**Number of (included) inputs**	**Description (Technique(s) & Variables)**
Vigano (1993, Burkina Faso)	Microfinance (individual)	100	53 (13)	Discriminant Analysis [applicant characteristics, business characteristics & loan characteristics]
Sharma and Zeller (1997, Bangladesh)	Microfinance (group)	868	18 (5)	TOBIT Maximum Likelihood Estimation [group characteristics (people, lands), program /loan characteristics & community characteristics]
Zeller (1998, Madagascar)	Microfinance (group)	146	19 (7)	TOBIT Maximum Likelihood Estimation [group characteristics, microcredit program characteristics, & community characteristics]
Reinke (1998, South Africa)	Microfinance (individual)	1641	8 (8)	Probit Regression [applicant characteristics, business characteristics & MFI branch characteristics]
Schreiner (1999, Bolivia)	Microfinance (individual)	39 956	9 (9)	Logistic Regression [loan characteristics, applicant characteristics, & credit officer experience]
Vogelgesang (2003, Bolivia)	Microfinance (individual)	8002	28 (12)	Multinomial Logistic Regression Random Utility Model [applicant characteristics, business characteristics, loan characteristics & environmental characteristics]
Vogelgesang (2003, Bolivia)	Microfinance (individual)	5956	30 (13)	Random Utility Model
Diallo (2006, Mali)	Microfinance (individual)	269	17 (5)	Logistic Regression, Discriminant Analysis [credit history, applicant characteristics, business characteristics, & credit officer experience]
Berger et al. (2007, New Mixico& USA)	Microfinance (individual)	500	21	Logistic Regression [business characteristics, borrower's profile, payment and credit histories]
Dinh&Kleimeier (2007, Vietnam)	Retail Bank (individual)	56 037	22 (17)	Logistic Regression [loan characteristics, applicant characteristics & the applicant's relationship with the MFI]

Deininger and Liu (2009, India)	Microfinance (group)	3350	15	Tobit Regression [loan characteristics, applicant characteristics & business practices of community organizations]
Che et al. (2010, Taiwan)	SME Loans (individual)	30 (22)	11	Fuzzy Analytical Hierarchy Process (FAHP) & Data Envelopment Analysis (DEA) [FAHP for variable selection & DEA for solving decision problem. Data on solvency, management & risk of the applicant]
Kinda and Achonu (2012, Senegal)	Retail Loans (individual)	30	14(13)	Logistic Regression [applicant socio-economic characteristics, loan characteristics & credit officer's experience]
Van Gool et al. (2012, Bosnia)	Microfinance (individual)	6722	16	Logistic Regression [applicant characteristics, loan characteristics & branch and credit officer characteristics]
Serran-Cinca et al. (2013, Colombia)	Microfinance (individual)	1	26	Multiple-attribute Utility Theory (MAUT); Multiple-attribute Value Theory (MAVT); Analytic Hierarchy Process (AHP) [credit history, applicant characteristics & loan characteristics]
Blanco et al.(2013, Peru)	Microfinance (individual)	5500	39	Neural Networks [historic data, collateral, applicant characteristics, business characteristics & macroeconomic variables]
Cubiles-Di-La-Vega et al. (2013, Peru)	Microfinance (individual)	5451	39	LDA, QDA, LR, CART, MLP, Bagging, Boosting, SVM, RF [applicant characteristics, business characteristics, loan characteristics, macroeconomic context]
Gutiérrez-Nieto et al. (2016, Colombia)	Microfinance (individual)	1	26	Multiple-attribute Utility Theory (MAUT); Multiple-attribute Value Theory (MAVT); Analytic Hierarchy Process (AHP) [credit history, applicant characteristics & loan characteristics]

Table 6.1 *Credit Scoring Models in Microfinance.*

6.2 Scoring Model Development

The prime goal of borrower risk rating is to help the lenders (financial institutions, individuals) to understand various dimensions of risk involved in a credit. The aggregation of such rating across the borrower's profile, business activities and the lines of business can provide better assessment of the quality of credit portfolio of a lender. The credit risk rating system is vital to take decisions both at the pre-sanction and post-sanction stages.

At the pre-sanction stage, credit rating helps the sanctioning authority (decision-maker) to decide: whether to lend or not to lend, what should be the loan price, what should be the extent of exposure, what should be the appropriate credit facility, what are the various facilities, and what are the various risk mitigation tools to put a cap on the risk level.At the post-sanction stage, the lender can decide: about the depth of the review or renewal, frequency of review, periodicity of the grading, and other precautions to be taken.Having considered the significance of credit risk rating, it becomes imperative for the lending system to carefully develop a credit risk rating model which meets the objective outlined above.

The credit risk rating model is not new to the banking system. Rather such proprietary models are used as internal control system within the lending institutions. Some models are used across the industry obliged by regulatory authority such as CRG (Credit Risk Grading) model in Bangladesh[53]. In 1993, the central bank of Bangladesh (Bangladesh Bank) introduced the Lending Risk Analysis (LRA) and made it for mandatory use in practice by the banks and financial institutions for the loan size above a certain limit. Later in 2003, the Bangladesh Bank introduced the Risk Grade Score Card for risk assessment of credit applications in order to overcome the problems of subjectivity and vagueness of LRA model. Finally, the regulatory authority came up with CRG model to make the previous versions a need-based simplified and user friendly model for application by the banks and financial institutions in processing credit decisions and evaluating the magnitude of risk involved therein.

Now a days, credit risk grading system is being developed and practiced in microfinance sector for assessing and evaluating a borrower. Gutierrez-Nieto et al. (2016) developed a credit score system for socially responsible lending incorporating financial and social aspects of the borrower. In this study, Multi-Criteria Decision Making (MCDM) models have been used in incorporating both financial and social issues with balanced weights by the specialists in a given subject.

In both cases, CRG model and DCDM-based credit score system, a tailor-made spreadsheet program has been used to develop a credit score of the borrower. In our study, we have adopted this framework and made necessary modifications based on the features we have selected and used to manipulate our data to come up with a score. The Figure 6.2 shows the steps done in the study and the indication of future work for its validation.

[53]http://www.assignmentpoint.com/business/report-on-credit-risk-grading-manual.html

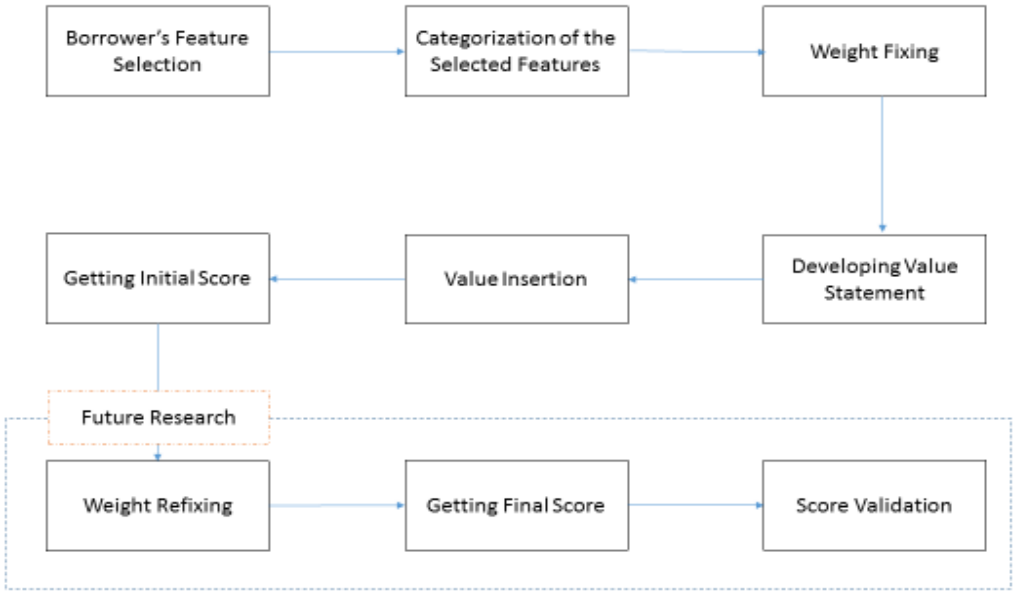


Figure 6.2 *Borrower Risk Scoring Diagram.*

6.2.1 Definition of Borrower Risk Scoring (BRS)

Borrower risk scoring is a spreadsheet based grading model based on the pre-specified scale or value of the selected factors or features of microcredit borrower to assess and evaluate the underlying credit-risk for a given borrower. It deploys a number (ranging from 0 to 10) as a primary summary indicator of risks associated with a credit exposure and then translates this number into a particular grade (from excellent grade to worst grade for eight grades in total).

6.2.2 Functions and Use of BRS

Well-managed credit risk grading system promotes lenders or lending institutions safety and soundness by facilitating informed decision-making. Grading system measures credit risk and differentiate individual credits and groups of credits by the risk they pose. This allows credit management and examiners to monitor changes and trends in risk levels. The CRG matrix allows application of uniform standards to credits to ensure a common standardized approach to assess the quality of individual obligor, credit portfolio of a unit, line of business, the financial institution as a whole. As evident, the BRS outputs would be relevant for individual credit selection, wherein a borrower is rated. The other decisions would be related to pricing (credit-spread) and specific features of the credit facility. These would largely constitute obligor level analysis. Risk grading would also be relevant for surveillance and monitoring, internal MIS and assessing the aggregate risk profile of a lending institution.

6.2.3 Grading Structure and Scale of BRS

The grading structure and scale or value has been shown in Table 6.2.

Number	Grade	Short Name	Score (10-point)
1	Excellent Grade	UG1	8.6 - 10.0
2	Very Good Grade	UG2	7.6 - 8.5
3	Good Grade	UG3	6.6 - 7.5
4	Marginal/Average Grade	AG	5.6 - 6.5
5	Lowest in High Risk Grade	LG1	4.6 - 5.5
6	Moderate in High Risk Grade	LG2	3.1 - 4.5
7	Higher in High Risk Grade	LG3	1.6 - 3.0
8	Worst in/Absolute High Risk Grade	LG4	0.0 - 1.5

Table 6.2 *Grading structure and scale.*

6.2.3.1 Excellent Grade (UG1)

UG1 is in the tier of low risk class. It is the superior tier among the tiers in low risk class.

- The borrower is reliable or committed
- Repayment is regular
- Field Partner has strong reputation
- Sector is safe and stable
- Country is favorable for funding

6.2.3.2 Very Good Grade (UG2)

UG2 is in the tier of low risk class. It is the 2nd best tier among the tiers in low risk class.

- The borrower is reliable or committed
- Repayment is expected to be regular
- Field Partner's reputation is good
- Sector is safe and stable
- Country is favorable for funding

6.2.3.3 Good Grade (UG3)

UG3 is in the tier of low risk class. It is the 3rdbest tier among the tiers in low risk class.

- The borrower is reliable or committed
- Repayment may beirregular
- Field Partner's reputation is acceptable
- Sector is safe and stable
- Country is favorable for funding

6.2.3.4 Marginal or Average Grade (AG)

AG is the tier between low and high risk classes. It is marginal tier which indicates risk neutrality.

- The borrower is reliable or committed
- Repayment may be irregular
- Field Partner's reputation is marginal
- Sector is less vulnerable
- Country is favorable for funding

6.2.3.5 Lowest in High Risk Grade (LG1)

LG1 is in the tier of high risk class. It is the best tier among the tiers in high risk class.

- The borrower is reliable or committed
- Repayment is in risk
- Field Partner's reputation is below the average
- Sector is vulnerable
- Country is not favorable for funding

6.2.3.6 Moderate in High Risk Grade (LG2)

LG2 is in the tier of high risk class. It is the 2nd best tier among the tiers in high risk class.

- The borrower is reliable or committed
- Repayment is doubtful
- Field Partner's reputation is doubtful
- Sector is vulnerable
- Country is unfavorable for funding

6.2.3.7 Higher in High Risk Grade (LG3)

LG3 is in the tier of high risk class. It is the 2^{nd} most risky tier among the tiers in high risk class.

- The borrower is reliable or committed
- Repayment is in great trouble
- Field Partner's reputation is not good
- Sector is vulnerable
- Country is unfavorable for funding

6.2.3.8 Worst in or Absolute High Risk Grade (LG4)

LG4 is in the worst tier among all the risk grades.

- The borrower's capacity is in question clearly
- Repayment is most uncertain
- Field Partner's has no reputation

- Sector is vulnerable
- Country is absolutely unfavorable for funding

6.2.4 Process of Developing BRS

The following step-wise activities outline the detail process for arriving at BRS.

6.2.4.1 Variables or Factors Selection

In Gool et al. (2012), once data had been prepared and a statistical model had been chosen, the next step was to decide on the treatment of the explanatory variables. Afterwards, the explanatory variable selection process was considered. Both categorical and continuous variables have been used in different scoring models in microfinance. Some recent works show the details, in Table 6.3, of the variables used in different contexts.

Author (Date, Country)	Technique(s)	Number of variables	Variables
Vigano (1993, Burkina Faso)	Discriminant Analysis	13	13 factors are composed of 63 variables related to borrower's personal, business, loan and lender in factor loading models
Schreiner (2004, Bolivia)	Logit regression	9	Disbursement date, amount disbursed, guarantee type, branch, loan officer, gender, sector, number of spells of arrears, length of the longest spell of arrears.
Baklouti (2013, Tunisia)	Binary Logistic regression	10	Socio-demographic variables(5): age, gender, education, job experience, marital status; Loan characteristic variables(3): amount, purpose, sector; Behavioral variables(2): number of previous loans, previous loan default;
Dukic et al (2011, Croatia)	Logistic regression	10	Socio-demographic variables(5): gender, age, education level, marital status, members of household; Financial indicators (5): salary, other income, expenditures, debts, account balance;
Serrano-Cinca et al (2013, Spain)	Social net present value (SNPV)	26	Variables include qualitative and quantitative, social and financial variables in different scales in past, present and future contexts.
Blanco et al (2013, Peru)	multilayer perceptron approach (MLP)	39	Variables related to borrowers, loan and lenders in financial, non-financial and macroeconomic categories.
Gool et al (2012, Bosnia)	Logistic Regression (logit model)	16	Borrower characteristics (8): age, job experience, net earnings of business, business capital, business register, net earnings of household, household capital, other debt; Loan characteristics (6): purpose, amount, requested duration, cycles (how many times loan taken), beginning month(cyclical effect), year of initiation; Lender characteristics (2): branch (rural has more social control), loan officer (experience)

Table 6.3 *Variables used in different credit scoring models for developing countries.*

The studies used spreadsheet program (the CRG model in Bangladesh and the DCDM-based credit score system in Spain) used both quantitative and qualitative information which are completely outlined in Table 6.4:

Model	Number of Variables used	Risk Category	Weight	Risk Factors	Score
CRG	20	Financial	50%	-Leverage-15% -Liquidity-15% -Profitability-15% -Coverage-5%	0-15 0-15 0-15 0-5
		Business/ Industry	18%	-Size of business-5% -Age of business-3% -Business outlook-3% -Industry growth-3% -Market competition-2% -Entry/Exit barriers-2%	0-5 0-3 0-3 0-3 0-2 0-2
		Management	12%	-Experience-5% -Second line/Succession-4% -Team work-3%	0-5 0-4 0-3
		Security	10%	-Security (primary)-4% -Collateral (property)-4% -Support (guarantee)-2%	0-4 0-4 0-2
		Relationship	10%	-Account conduct-5% -Limit utilization-2% - Covenants compliance-2% -Personal deposit-1%	0-5 0-2 0-2 0-1
DCDM	26	Repayment history	29.4%	Repayment behavior with: -lender -Other lenders -Suppliers/customers	0-10
		Company	50.2%	**Accounting information (59.05%):** -Business growth -Profitability, efficiency, productivity -Short term liquidity -Long term liquidity **Intangibles (40.95%):** -Management board -Staff -Labor responsibility -Vision and values -Processes and technology -Innovation -Customers -Social image -Networks -Transparency	0-10 0-10
		Loan	20.4%	**Financial (32.47%):** -Profitability -Risks -Liquidity **Social (67.53%):** -Impact on employment -Impact on education -Diversity & equal opportunity -Community outreach -Impact on health -Impact on environment	0-10 0-10

Table 6.4 *Details of variables used in spreadsheet-based credit scoring models.*

In our sample database, there are about 37 variables representing borrower's personal and business details, loan information and repayment history along with filed partner's reference. Among the available data, some are not directly relevant to our purpose (credit rating) and some are not readily extractable due to the structure of the data in the database. Considering the availability, relevancy and computational complexity we have finally selected 13 variables under 5 risk categories for our model. The selected variables represent borrower's personal information, business, industry and loan information that are essentially representative. Additionally, like other credit scoring models, we have 5 variables to assess the local intermediary (field partner) which are (among other information) are currently considered by the P2P platforms as proxy to the borrower risk. Moreover, we have added a variable "country" that represents country risk. As most of the indirect P2P lending models operate globally this variable gives an indication on national or macro level risk and it is unique to our model. What significant data missing here are financial data concerning borrower's business or project. However, as almost all the P2P lending platforms that we are considering in our study focus on not only typical loan funding but also social aspects of the borrowers where qualitative variables get more significance. From this perspective, the type and number of variables, Table 6.5, we have selected seem reasonable to build a simple but predictive model for giving a light on the level of risk associated with the borrowers in online P2P lending platforms that operate non-profit models globally.

Model	**Number of Variables used**	**Risk Category**	**Risk Factors**
BRS	13	Borrowers' profile	Gender
		Loan profile	Amount or size, purpose,disbursal mode, repayment term, repayment interval
		Field Partner's Reputation	Risk rating, default rate, delinquency rate, social badge, borrower volume rate
		Sector	Sector of the firm
		Country	Country of the borrower

Table 6.5 *Selected variables in BRS.*

6.2.4.2 Categorization of the Selected Features

Credit risk for borrower arises from an aggregation of five risk categories: borrower's profile risk, borrower's loan risk, field partner's reputational risk, sectoral risk and countryrisk.Each of the key risk areas requires to be evaluated and aggregated to arrive at an overall risk grading measure.

- Borrower's Profile Risk: It is the risk that the borrower may default due to the lack of commitment made by the person who is taking the loan and his/her personal capabilities, experience as well as the family background. Here, gender, age, experience, marital status and or spouse employment, number of family members and their occupations play the vital role in repaying the loan taken. However, due to the computational complexities only gender has been considered as the risk factor where female is considered more committed than the male borrower.

- Borrower's Loan Risk:It is the risk that the borrower may fail to repay the loan due to the lack of suitability and or inappropriate use of the loan. Here, size of loan, purpose of use, disbursal mode, repayment terms and type of installment have been considered. There is a high risk of default if the loan size is large and the loan usage is for the purposes other than productive activities. Also, high risk exposes or reveals if the repayment mode is ballooning payment or one time and at the end of the loan term granted.
- Field Partner's Reputational Risk: It is the risk that the loan may fail due to the reputational lackingof the local financial intermediary who is solely responsible to screen and select the borrower. It is the proxy risk category by which borrower risk can be measured. Because, in practice, till now most of the cases, microcredit borrowers are selected and managed by the group-lending technology and by the help of experience orexpertise of the loan officers of the microfinance institutions (MFIs). Even in the case of individual microenterprise loans, the loan officers follow the expert rules in subjective manner. Therefore, the risk of field partner's reputational exposure has been considered by risk rating (done by online P2P platform), social performance (in terms of social badge), default and delinquency rates, number of borrowers served using the P2P platform. High default and delinquency rates, low risk rating, social badges and borrower volume may put the loan in high risk and vice versa.
- Sectoral Risk:It is the risk that the loan may fail due to the poor performance of a specific sector as macro factor. Usually sectoral performance is measured by the productivity or return from the sector. However, the data shows such return are not available in the database. Therefore, the risk of sector has been considered empirically by the high default loans belong to the sectors categorized in online P2P lending platform. The higher default loans exist in a particular sector the more risk of defaulting the loan in that sector in future.
- Country Risk:It is the risk that some countries globally inherent high level of risk to fail not only the loan in microcredit but also other activities there. Like sector, it is also a macro factor which may impact on any cluster of the economy. It may be due to political issues, economic recession, natural disasters, or even social problems.

6.2.4.3 Fixing Weights

Initially weights for both categorical risk components and factor risk components are given by value judgment. Among the five categorical risk components, borrower's loan risk and field partner's reputational risk have been given with more weights than the rest: borrower's profile risk, sectoral risk and country risk. Later, the initial assigned weights will be re-fixed based on the distribution of the sample.
According to the importance of risk profile, the following weightages are proposed for corresponding categorical risks in Table 6.6.

Categorical Risk Components	**Weight**
Borrower's Profile Risk	10%
Borrower's Loan Risk	35%
Field Partner's Reputational Risk	35%
Sectoral Risk	10%
Country Risk	10%

Table 6.6 *Categorical risk with initial weights.*

- Borrower's Profile Risk: 10%
  As the borrower's profile risk is measured by only single factor 'gender', it gets 10% weight given to the category.

- Borrower's Loan Risk: 35%
  The weight assigned to this category is sub-allocated to the following risk factors:

  i. Loan Amount-40%
  ii. Loan Use- 15%
  iii. Disbursal Mode - 15%
  iv. Repayment Term-15%
  v. Repayment Interval-15%

- Field Partner's Reputational Risk-35%
  This category is composed of the following factors along with the assigned risk percentage within the category:

  i. Risk Rating-25%
  ii. Default Risk-15%
  iii. Delinquency Risk-10%
  iv. Social Badge Risk-25%
  v. Borrower Volume Risk-25%

- Sectoral Risk-10%
  Like borrower's risk profile this category of risk is also measured by single factor 'sector' with the assigned percentage of risk -10%.

- Country Risk-10%
  This risk category is measured by single factor, country risk along with the assigned risk of 10%.

6.2.4.4 Developing Value Statement of the Selected Features or Factors

The features or factors associated with a loan request in online P2P lending are translatedinto a scale or value statement following expert opinions as follows:

- Borrower's Profile Risk

Gender

Score	**Attributes (based on literature)**
10.0	Female
7.0	Male

- Borrower's Loan Risk

Loan Amount

Score	**Attributes (based on loan size distribution applicable for Africa and Asia Zones; modal value)**
10.0	If loan amount <=$300
7.5	If loan amount >$300 but <=$500
5.0	If loan amount >$500 but <=$700
2.5	If loan amount >$700 but <=$1200
0.0	If loan amount >$1200

Loan Use

Score	Attributes (based on the analysis of 500 loan cases from a selectedcomprehensive file from kiva database)
10.0	If the loan is used for working capital or raw materials/products purchase (WC)
7.5	If the loan is used for operating assets in business (OPA)
5.0	If the loan is used for non-earning assets in business (NEA)
2.5	If the loan is used for personal assets (PA)
0.0	If the loan is used for personal non-assets purposes (PU)

Disbursal Mode (Type)

Score	Attributes (based on existing modes of disbursal in kiva lending system)
10.0	If the disbursement is made before posting the loan to the Kiva website (pre-disbursal)
5.0	If the disbursement is made after posting the loan to the Kiva website (post-disbursal)

Repayment Term

Score	Attributes (base 'repayment term' is determined using average &/mode)
10.0	if the repayment term <=14 months
7.5	if the repayment term >14 months but <=20
5.0	if the repayment term >20 months but <=26
2.5	if the repayment term >26 months but <=60
0.0	if the repayment term >60 months

Repayment Interval

Score	Attributes (based on existing types of repayment interval in kiva lending system)
10.0	If repayment interval <=1 month (monthly)
7.5	Irregularly (within the repayment term but not in regular interval)
5.0	At the end of the term (one ballooning payment)

- Field Partner's Reputational Risk

Risk Rating

Score	Attributes (based on distribution of Kiva's risk rating for their field partners in Africa and Asia zones)
10.0	Between 4 and 5 stars (3.5, 4.0, 4.5 & 5.0)
7.5	Rating with 3 stars (2.5 & 3.0)
5.0	Rating with 2 stars (1.5 & 2.0)
2.5	Upto 1 star (0.5 & 1.0)
0.0	No risk rating (experimental, paused, inactive)

Default Risk*

Score	Attributes (based on distribution of default rate of the field

	partners in Africa and Asia zones at Kiva platform)
10.0	Upto 1.00% of the amount of ended loan
7.5	Between 1.01% to 2.00% of the amount of ended loan
5.0	Between 2.01% to 5.00% of the amount of ended loan
2.5	Between 5.01% to 20.00% of the amount of ended loan
0.0	Above 20% of the amount of ended loan

* Defined by Kiva. Default Rate: Percentage of Ended Loans (no longer paying back) which have failed to repay (measured in dollar volume, not units). How this is calculated: Amount of Ended Loans Defaulted / Amount of Ended Loans; Amount of ended loans are total amount of loans raised and disbursed which are no longer in the process of being paid back by an entrepreneur. This excludes refunded loans (total amount of loans refunded to lenders due to an error).

Delinquency Risk*

Score	Attributes (based on distribution of delinquency rate of the field partners in Africa and Asia zones at Kiva platform)
10.0	Upto 5% of outstanding principal balance
7.5	Between 5.00% to 10.00% of outstanding principal balance
5.0	Between 10.00% to 20.00% of outstanding principal balance
2.5	Between 20.00% to 50.00% of outstanding principal balance
0.0	Above 50% of outstanding principal balance

*Kiva defines the Delinquency (Arrears) Rate as the amount of late payments divided by the total outstanding principal balance Kiva has with the Field Partner. Arrears can result from late repayments from Kiva borrowers as well as delayed payments from the Field Partner. Delinquency (Arrears) Rate = Amount of Paying Back Loans Delinquent / Amount Outstanding

Social Badge Risk*

Score	Attributes (based on distribution of social badges of the field partners in Africa and Asia zones at Kiva platform)
10.0	6 badges and above
7.5	between 4 and 5 badges
5.0	between 2 and 3 badges
2.5	1 badge
0.0	no badges/nil

*Considered the degree of social dimension based on social badge developed & published by Kiva at its site. The more a field partner has social badges the less the risk to be a non-performer/defaulter to run its business.

Borrower Volume Risk*

Score	Attributes (based on distribution of kiva borrowers per month of each field partner in Africa and Asia zones at Kiva platform)
10.0	Borrowers 200 and above per month
7.5	Borrowers between 100 to 199 per month
5.0	Borrowers between 50 to 99 per month
2.5	Borrowers between 10 to 49 per month
0.0	Borrowers less than 10 per month

*Kiva borrowers per month risk= number of borrowers associated with kiva's funding of a field partner/ number of months a field partner is with Kiva.

- Sectoral Risk

Sector

Score	Attributes (based on 125 default cases in both African and Asian zones for individual loan cases)
10.0	Manufacturing, Wholesale, Arts, Education, Entertainment, Health, Housing, Personal Use, Transportation
7.5	Agriculture, Construction
5.0	Retail, Services
2.5	Clothing, Food

- Country Risk

Country

Score	Attributes (based on country risk ratingby S&P, Moody's, and Fitch)
10.0	China
7.5	Azerbaijan, Botswana, India, Indonesia, Namibia, Philippines, South Africa, Thailand
5.0	Nigeria, Tunisia
2.5	Benin, Burkina Faso, Cambodia, Cameroon, Ethiopia, Ghana, Kenya, Mongolia, Mozambique, Pakistan, Rwanda, Senegal, Uganda, Vietnam, Zambia
0.0	Burundi, Congo (Dem. Rep.), Congo (Rep.), Cote D'Ivoire, Kyrgyzstan, Lao PDR, Liberia, Madagascar, Malawi, Mali, Mauritania, Myanmar (Burma), Nepal, Sierra Leone, Somalia, South Sudan, Tajikistan, Tanzania, Togo, Zimbabwe

The assignment of weights and scores of the categorical risk components and individual factor risk components has been shown comprehensively in Table 6.7 as follows:

Model	Number of Variables used	Risk Category	Weight	Risk Factors	Score
BRS	13	Borrower's Profile Risk	10%	-Gender -10%	7-10
		Borrower's Loan Risk	35%	-Loan Amount - 40% -Loan Use -15% -Disbursal Mode - 15% -Repayment Term - 15% -Repayment Interval - 15%	0-10 0-10 5-10 0-10 5-10
		Field Partner's Reputational Risk	35%	-Risk Rating - 25% -Default Risk - 15% -Delinquency Risk -10% -Social Badge Risk - 25% -Borrower Volume Risk - 25%	0-10 0-10 0-10 0-10 0-10
		Sectoral Risk	10%	-Sector - 10%	0-10
		Country Risk	10%	-Country - 10%	0-10

Table 6.7 *Comprehensive Structure of BRS*

6.2.4.5 Value Insertion

After factoridentification,categorization, weightage and value assignment process (as mentioned above), the next stepis to input the value in the score sheet to arrive at the scores corresponding to the actual features.This model provides a well programmed MS Excel based credit risk scoring sheet to arrive at a total score on each borrower. The excel program requires inputting data accurately in particular cells for input and will automatically calculate the risk grade for a particular borrower based on the total score obtained. The following steps are to be followed while using the MS Excel program.

- Open the MS XL file named, BRS_SCORE_SHEET
- The entire XL sheet named, BRS is protected except the particular cells to input data.
- Input data accurately in the cells which are BORDERED & are colored YELLOW.
- All the cells provided for input must be filled in order to arrive at accurate risk grade.

The following is the proposed Credit Risk Grade matrix, Table 6.8, based on the total score obtained by an obligor or borrower or loan applicant.

Number of Grade	Risk Grading	Short Name	Score
1	Excellent Grade	UG1	8.6 - 10.0
2	Very Good Grade	UG2	7.6 - 8.5
3	Good Grade	UG3	6.6 - 7.5
4	Marginal/Average Grade	AG	5.6 - 6.5
5	Lowest in High Risk Grade	LG1	4.6 - 5.5
6	Moderate in High Risk Grade	LG2	3.1 - 4.5
7	Higher in High Risk Grade	LG3	1.6 - 3.0
8	Worst/Absolute High Risk Grade	LG4	0.0 - 1.5

Table 6.8 *Credit Risk Grading.*

6.2.4.6 Getting Initial Score

The following Figure 6.3 shows how a score is computed to assess the risk of a particular borrower and label him a grade through our tool.

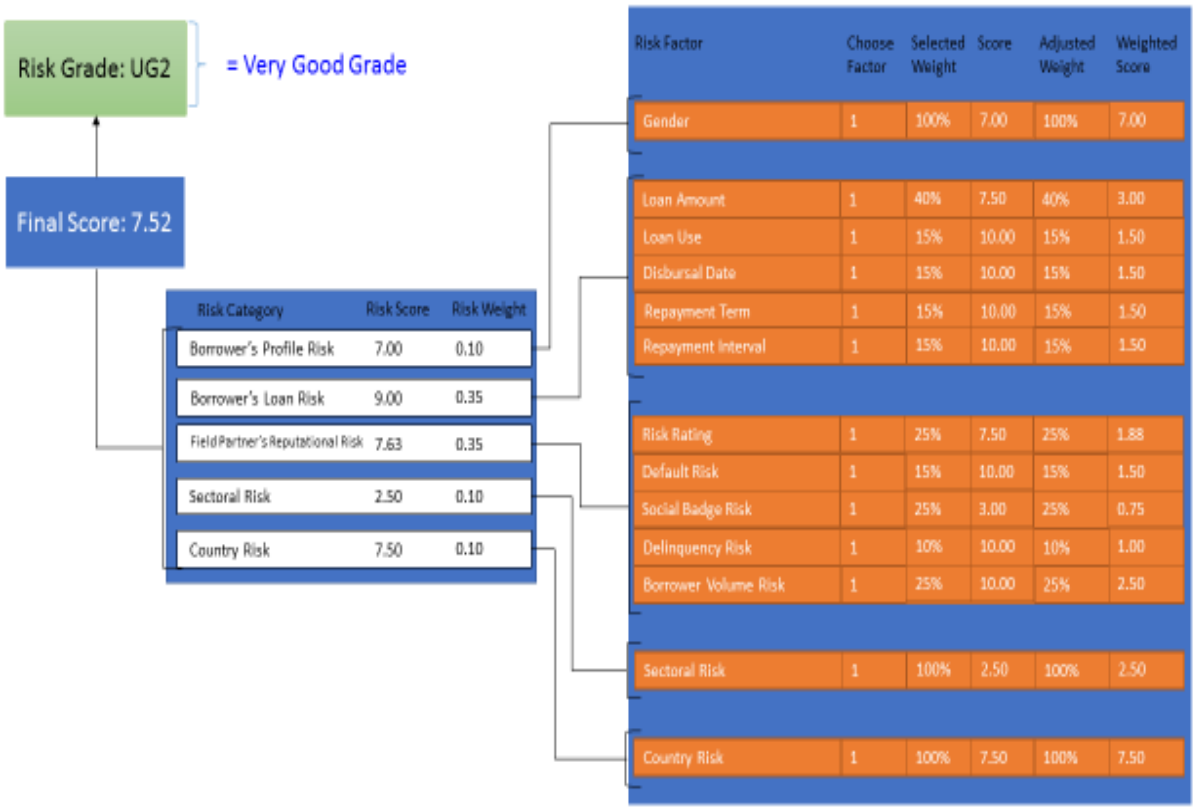


Figure 6.3 *Borrower Risk Scoring.*

6.3 Summary

This study developed BRS system based on the risk factors that are associated with individual borrowers; i.e., borrower's profile risk, borrower's loan risk, field partner's reputational risk, sectoral risk and country risk. The study, to some aspects, has also been developed on the same framework used in Gutierrez-Nieto et al. (2016) and a report on CRG score sheet of private commercial banks. Though Gutierrez-Nieto et al. (2016) evaluated the social and financial aspects of the borrowers and tested the model in a single country context, findings of the three papers have some similarities. First, all three papers used financial and nonfinancial parameters. Second, all three papers used 'weight' for scoring. Third, all three papers considered developing country context with respect to credit scoring for microfinance.

Despite the similarities mentioned above, our study is different from referred works. First, this study has developed credit scoring tool based on the variables to light on the level of risk associated with the borrowers in online indirect P2P lending platforms that operate non-profit

models globally. Second, the present study has used multiple country data from Africa and Asia zones for which country risk variable has been added as a new feature of the model. Third and finally, this model has been developed with the variables that are most common and universal across the countries in the world. For this reason, this model can be used as a common tool in any online indirect P2P microcredit lending platforms and hence the online microcredit lenders can get an idea about the level of risk of the borrower whom they wish to lend by matching the degree of risk to their risk tolerance attitudes or perceptions. For example, high risk borrower might be fitted with risk aggressive lenders and low risk borrower might be chosen by risk avert lenders. In this regard, this work is a first contribution in developing a credit scoring tool for identifying the borrower's risk in developing country context, in particular for online indirect P2P microcredit lending model.

Chapter 7

Analysis and Evaluation of the Results of the CBR Integrated System

In this chapter we have presented howborrower risk assessements have done with the detail elaborations of assigned weights and factor values. It has also shown how sample data probability distributions were used in the intergration of initial weights of risk factors to make the weights assignment more objective or scientific. Then the results of four versions of two sets of expert models have been shown and explained for using the best score in the CBR integrated system. Finally, the evaluation of the CBR integrated system has been made with the test scores of the loan cases taken from hold out sample or test set.

7.1 Analysis of Outcome of Expert Model (EM) and It's Predictive Power

Borrower risk scoring, here expert model, is a spreadsheet based grading model based on the pre-specified scale or value (given by experts) of the selected factors of microcredit borrower to assess and evaluate the underlying credit-risk for a given borrower. It deploys a number (ranging from 0 to 10) as a primary summary indicator of risks associated with a credit exposure and then translates this number into a particular grade from excellent to worst for eight grades in total (details in Chapter 6).

There are four versions of two sets of expert models with different assumptions (Tables 7.1 to 7.4). The purpose of having these four versions is to compare the results of initial training set of relevant cases of these models with the empirical results(outcomes of the cases) and to choose that version with highest predictive power of sensitivity and specificity, among the four versions. Then the results (credit scores or solutions of the cases) of that version of the expert model will be used to classify and complete the cases of the training set. With the complete representative cases, the case base of the CBR system will be updated.

The general functions of expert model are as follows:

$$Risk\ Score = \sum_{c \in Categories} w_c \cdot CategoryScore$$

$$CategoryScore = \sum_{f \in Factors} w_f \cdot FactorMark$$

Risk score of a loan case is the summation of the products of category weight and category score. Again, category score is defined as the summation of the products of individual factor weight and factor mark based on score on value judgement of expert.

7.1.1 Two versions of Expert Model-1 (EM01a & EM01b)

7.1.1.1 Assumption: Both the weights of categorical risk components and factor risk components are given by experts.

7.1.1.2 First version of Expert Model-1 (EM01a) has been developed with initial weights for both categories and factors in Table 7.1.

Categorical Risk Component	Category Weights	Factor Risk Component	Factor Weight
Borrower's profile risk	0.10	Gender	1.00
Borrower's loan risk	0.35	Loan amount	0.40
		Loan use	0.15
		Disbursal mode	0.15
		Repayment term	0.15
		Repayment interval	0.15
Field partner's reputational risk	0.35	Risk rating	0.25
		Default risk	0.15
		Delinquency risk	0.10
		Social badge risk	0.25
		Borrower's volume risk	0.25
Sectoral risk	0.10	Sector	1.00
Country risk	0.10	Country	1.00

Table 7.1 *Expert model 1 for version a (EM01a).*

7.1.1.3 Second version of Expert Model-1 (EM-01b) has been designed with different sets of weights for both categories and factors in Table 7.2.

Categorical Risk Component	Category Weights	Factor Risk Component	Factor Weight
Borrower's profile risk	0.05	Gender	1.00
Borrower's loan risk	0.35	Loan amount	0.60
		Loan use	0.10
		Disbursal mode	0.05
		Repayment term	0.20
		Repayment interval	0.05
Field partner's reputational risk	0.35	Risk rating	0.35
		Default risk	0.15
		Delinquency risk	0.10
		Social badge risk	0.20
		Borrower's volume risk	0.20
Sectoral risk	0.10	Sector	1.00
Country risk	0.15	Country	1.00

Table 7.2 *Expert model 1 with only expert weights for version b (EM01b).*

7.1.2 Two versions of Expert Model-2 (EM02a & EM02b)

7.1.2.1 Assumptions:

- Both the initial weights of categorical risk components and factor risk components given by experts are reweighted by the probability distribution of large sample data.
- New category defined as Macro risk with the merger of two factors of sector & country risk has been made to consider the effects of large sample data distribution in the model.
- Different set of initial weights have been considered for two different versions.

7.1.2.2 First version of Expert Model-2 (EM02a) has been developed in Table 7.3.

Categorical Risk Component	Category Weights	Factor Risk Component	Initial Factor Weight	Sample Distribution[54]
Borrower's profile risk	0.10	Gender	1.00	69%, 31%
Borrower's loan risk	0.40	Loan amount	0.40	24%,22%,15%,26%,13%
		Loan use	0.15	79%,10%,4%,4%,3%
		Disbursal mode	0.15	99%,1%
		Repayment term	0.15	73%,17%,8%,2%,0%
		Repayment interval	0.15	89%,5%,6%
Field partner's reputational risk	0.40	Risk rating	0.25	40%,25%,11%,3%,21%
		Default risk	0.15	78%,6%,7%,9%,0%
		Delinquency risk	0.10	76%,2%,4%,1%,17%
		Social badge risk	0.25	4%,39%,45%,7%,5%
		Borrower's volume risk	0.25	38%,26%,23%,11%,2%
Macro risk	0.10	Sector	0.50	9%,22%,9%,56%,4%
		Country	0.50	0%,33%,12%,41%,14%

Table 7.3 *Expert model 2 for version a (EM02a).*

- An instance of the distribution effects for a loan case is shown in Figure 7.1:

Risk Category (1)	Risk Factor (2)	Initial Factor Weight (3)	Factor Mark (4)	Disb.[55] Weight (5)	Adj. Weight {6=(3*5)/100}	Re-weighted {7=(6/c_sum)*100}	Weighted Mark {8=(4*7)/100}	Category Score {9=c_sum(8)}
Client Name								
Borrower's Profile Risk	Gender	100	10.0	100	100	100	10.00	10.00
Borrower's Loan Risk	Loan Amount	40	5.0	15	6.0	13.5	0.68	9.28
	Loan Use	15	10.0	79	11.9	26.7	2.67	
	Disbursal Date	15	10.0	99	14.9	33.4	3.34	

[54] Distribution is based on 6436 loan cases. Sample distribution has been shown on the different scale of a particular factor. The details of the scale have been discussed in chapter 6.

[55] Distribution weight was taken from the right column of figure 3. For example, the weight of 15 for loan amount is taken from the 3rd value of 5-point scale in the distribution column of figure 3.

	Repayment Term	15	10.0	73	11.0	24.7	2.47	
	Repayment Interval	15	7.5	5	0.8	1.7	0.13	
					44.4	100.0		
Field Partner's Reputational Risk	Risk Rating	25	10.0	40	10.0	20.6	2.06	9.50
	Default Risk	15	10.0	78	11.7	24.1	2.41	
	Delinquency Risk	10	10.0	76	7.6	15.7	1.57	
	Social Badge Risk	25	7.5	39	9.8	20.1	1.51	
	Borrower's Volume Risk	25	10.0	38	9.5	19.6	1.96	
					48.6	100.0		
Macro Risk	Sectoral Risk	50	2.5	56	28.0	57.7	1.44	2.50
	Country Risk	50	2.5	41	20.5	42.3	1.06	
					48.5	100.0		

Figure 7.1 *An instance of reweights of factors with sample distribution effects.*

7.1.2.3 Second version of Expert Model-2 (EM02b) has been designed in Table 7.4.

Categorical Risk Component	**Category Weights**	**Factor Risk Component**	**Initial Factor Weight**	**Sample Distribution**[56]
Borrower's profile risk	0.05	Gender	1.00	69%, 31%
Borrower's loan risk	0.35	Loan amount	0.60	24%,22%,15%,26%,13%
		Loan use	0.10	79%,10%,4%,4%,3%
		Disbursal mode	0.05	99%,1%
		Repayment term	0.20	73%,17%,8%,2%,0%
		Repayment interval	0.05	89%,5%,6%
Field partner's reputational risk	0.35	Risk rating	0.35	40%,25%,11%,3%,21%
		Default risk	0.25	78%,6%,7%,9%,0%
		Delinquency risk		76%,2%,4%,1%,17%
		Social badge risk	0.20	4%,39%,45%,7%,5%
		Borrower's volume risk	0.20	38%,26%,23%,11%,2%
Macro risk	0.25	Sector	0.50	9%,22%,9%,56%,4%
		Country	0.50	0%,33%,12%,41%,14%

Table 7.4 *Expert model 2 for version b (EM02b).*

[56] Same as note 54.

- An instance of the distribution effects for a loan case is shown in Figure 7.2.

Risk Category (1)	Risk Factor (2)	Initial Factor Weight (3)	Factor Mark (4)	Disb[57]. Weight (5)	Adj. Weight {6=(3*5)/100}	Re-weighted {7=(6/c_sum)*100	Weighted Mark {8=(4*7)/100}	Category Score {9=c_sum(8)}
Client Name								
Borrower's Profile Risk	Gender	100	10.0	100	100	**100**	10.00	10.00
Borrower's Loan Risk	Loan Amount	60	5.0	15	9.0	**24.5**	1.23	8.76
	Loan Use	10	10.0	79	7.9	**21.5**	2.15	
	Disbursal Date	5	10.0	99	5.0	**13.5**	1.35	
	Repayment Term	20	10.0	73	14.6	**39.8**	3.98	
	Repayment Interval	5	7.5	5	0.3	**0.7**	0.05	
		100			36.7	**100.0**		
Field Partner's Reputational Risk	Risk Rating	35	10.0	40	14.0	**28.9**	2.89	9.60
	Default Risk	25	10.0	78	19.0	**39.3**	3.93	
	Delinquency Risk		10.0	76				
	Social Badge Risk	20	7.5	39	7.8	**16.1**	1.21	
	Borrower's Volume Risk	20	10.0	38	7.6	**15.7**	1.57	
		100			48.4	**100.0**		
Macro Risk	Sectoral Risk	50	2.5	56	28.0	**57.7**	1.44	2.50
	Country Risk	50	2.5	41	20.5	**42.3**	1.06	
		100			48.5	**100.0**		

Figure 7.2 *An instance of reweights of factors with sample distribution effects.*

7.1.3 Results (Credit Score or Performance) of different versions of Expert Model

The classification table of 107 loan cases based on computed initial credit score of 107 loan cases being predicted with the versions of expert model has been presented sequentially in four Tables (7.5 to 7.8). There are total eight grades of the scoring scale. Success cases are those which fall in the upper four grades (UG3, UG2, UG1, AG) of the scale. Failure cases are those which fall in the lower four grades (LG1, LG2, LG3, LG4) of the scale.

[57]Same as note 55.

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	With Only Expert Weights (EM01a) Predicted Successful	Predicted Failure	Total
Successful loan cases	93 (100%)	79 (85%)	14 (15%)	93 (100%)
Failure loan cases	14 (100%)	11 (79%)3 (21%)		14 (100%)
Total	107 (100%)			

Table 7.5 *Classification Table of Prediction results of EM01a.*

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	With Only Expert Weights (EM01b) Predicted Successful	Predicted Failure	Total
Successful loan cases	93 (100%)	69 (74%)	24 (26%)	93 (100%)
Failure loan cases	14 (100%)	8 (57%)6 (43%)		14 (100%)
Total	107 (100%)	77 ()	30 ()	

Table 7.6 *Classification Table of Prediction results of EM01b.*

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	Reweights With Probability Distribution of Large Sample Data (EM02a) Predicted Successful	Predicted Failure	Total
Successful loan cases	93 (100%)	92 (99%)	1 (1%)	93 (100%)
Failure loan cases	14 (100%)	14 (100%)	0 (00%)	14 (100%)
Total	107 (100%)	106 ()	1 ()	

Table 7.7 *Classification Table of Prediction results of EM02a.*

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	Reweights With Probability Distribution of Large Sample Data (EM02b) Predicted Successful	Predicted Failure	Total
Successful loan cases	93 (100%)	87 (94%)	6 (6%)	93 (100%)
Failure loan cases	14 (100%)	12 (86%)2 (14%)		14 (100%)
Total	107 (100%)	90 ()	17 ()	

Table 7.8 *Classification Table of Prediction results of EM02b.*

Comparison of the results of sensitivity and specificity of four versions of expert models has been shown in Table 7.9 with empirical results (outcomes) for choosing the best model.

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	With Only Expert Weights		With Probability Distribution of Large Sample Data Effects (Reweights)	
		EM01a	**EM01b**	**EM02a**	EM02b
Sensitivity (True successful loan cases)	93 (100%)	79 (85%)	**69 (74%)**	**92 (99%)**	87 (94%)
Specificity (True failure loan cases)	14 (100%)	3 (21%)	**6 (43%)**	**0 (0%)**	2 (14%)
	107 (100%)				

Table 7.9 *Results of Expert Models.*

7.1.4 Analysis of the Results of Expert Model

As per the definitions of successful loan cases and failure loan cases mentioned above the empirical results or outcomes for success (paid) and failure (default) of 107 loan cases are 87% and 13% respectively.Comparing with the empirical resultsin Table 7.9, we have got the highest sensitivity rate (99%) in EM02a and the lowest rate (74%) in EM01b. Further, we have found the highest specificity rate (43%) in EM01b and the lowest rate (0%) in EM02a. Here, we have found both the models as the best performers in two separate criteria: sensitivity and specificity. As we are concerned about the borrower risk assessment (less risky or safe borrowers and high risky or problematic borrowers) aiming to give an indication of the degree of borrower's risk to lenders, with conservative policy we have chosen the model 'EM01b' which has shown higher percentage of specificity. This model has the power to identify the highest number of unsuccessful or default borrowers (6 of 14) among the versions of the expert models. Therefore, we have used the credit scores of 107 loan cases from the model-EM01b with only expert weights to complete the initial relevant cases in the case base of the CBR integrated system.

7.1.5 Discussion of the Results of Expert Models

The accuracy of the results of expert models depends on several issues. Some of them are under the scope of this study and the others are out of the scope of this research. Under this scope, it depends on the definitions of successful and failure loan cases where the expert jury was straight forward to make the border line for the grades. The upper grades including the average grade they have considered as good borrowers (to be predictive as successful or paid borrowers) and the lower four grades which are below to the average grade as bad borrowers (to be predicted as unsuccessful or failure or problematic borrowers). Another issue is the distribution of sample data. The results of the expert model with the integrated weights for large sample distribution effect were poor to the results of the expert model with only expert

weights. The reason for the worse results may be outcome of either non-normality of the sample distribution or ineffectiveness of the distribution at all. Because almost half of the factors is qualitative and in nominal scale where parametric distribution is assumed to be wrong. Out of 13 selected factors, 6 are categorical and the rest 7 are numerical. For the numerical factors, the distribution is skewed to the right. Therefore, the integration of distribution effects with initial expert weights is not effective or fruitful. Rather, expert weights work well in the performance of expert models. It may be the reason that still loan officer's evaluation or expert knowledge performs best in the assessment of microcredit borrowers or loan applicants in microfinance sector and the use of expert models is a common practice in the industry (Bunn & Wright, 1991; Gool et al., 2012; Schreiner, 2005). Other issues which are out of this research scope are missing and non-extractable data that might be relevant for borrower risk assessment. In the data set of selected factors, financial performance data of borrowers are not available. However, this type of numerical data still dominates the scoring model in loan market ranging from corporate loans to consumer or credit card loans(Crook, Edelman, & Thomas, 2007; Dinh & Kleimeier, 2007; Hand & Henley, 1997; Schreiner, 1999). In this research, five proxy factors have been considered from field partners to mitigate the problem of financial data missing. Another issue is non-extractable data of personal profile. In Kiva database, some personal data like borrower's age, marital status, family size, spouse employment status, business or loan experience etc. are available. But these data are not readily extractable at this moment although they might improve the results of credit rating.

7.1.6 Database Update with the Results of Expert Model

Among the results of four versions of two sets of expert models, the second version of expert model with only expert weights (EM01b) has shown the best predictive power. The case base of CBR system has been updated with the results of EM01b for 107 loan cases.

7.2 Evaluation of the results of the CBR system test set for testing it's predictive power

With the test set or hold out sample of the period from year 2014, the predictive power of the CBR system has been performed for 75 loan cases. Like expert model, there are eight grades in the scoring scales which fall on zero (0.0) score to full score of ten (10.0). For success loan cases, the scores range from 5.60 to 10.0 or upper four grades (AG, UG3,UG2 and UG1). For failure loan cases, the scores range from 0.00 to 5.50 or lower four grades (LG4, LG3,LG2,&LG1). Two types of loan cases of success are:

- Upper grades (Risk Score 5.6-10.0)- Successful loans (paid)
- Lower grades (Risk Score 0.0-5.5)- Failure loans (defaulted, in repayment, & expired)

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	CBR Rating		Total
		Predicted Successful	Predicted Failure	
Successful loan cases	**60 (100%)**	**46 (77%)**	14 (23%)	**60 (100%)**
Failure loan cases	**15 (100)**	6 (40%)	**9 (60%)**	**15 (100%)**
Total	**75 (100%)**			

Table 7.10 *Classification Table of Prediction results of CBR system.*

Actual Loan Status (Successful & Failure Cases)	Empirical Results (Outcomes)	CBR Rating (Prediction)
Sensitivity (True successful loan cases)	60 (100%)	46 (77%)
Specificity (True failure loan cases)	15 (100%)	9 (60%)
	75 (100%)	

Table7.11 *Prediction Power of CBR Rating.*

In Table 7.11, out of 75 loan cases, 52 loan cases or 69% are predicted or rated to be less risky or safe borrowers who will repay the loan timely if they are funded and 23 loan cases or 31% are rated to be more risky or doubtful borrowers who may not repay the loan if they are granted. However, in the real outcome (empirical result) of those 75 loan cases, 60 borrowers or 80.00% have been found successfully repaid and 15 borrowers or 20.00% have been found unsuccessful in the forms of fully defaulted, in repayment status, and expired loans. Comparing with the real outcomes or empirical results, the CBR system has predicted 46 loan cases or 77% of the successful (less risky) borrowers as perfectly and 9 loan cases or 60% of the unsuccessful borrowers correctly as high risk or problematic borrowers (Table 7.10; see in Appendix B for details). Therefore, the sensitivity rate of 77% and the specificity rate of 60% are quite encouraging and better than the predictive power of the expert model EM01b for the same indicators.

7.2.1 Discussions of the Results

There are several reasons or justifications for getting such rating performance of the CBR system. Most of them are associated with the poor performance of identifying the high risk loan applicants or borrowers who are the critical factors for the survival of the microfinance industry. Some of the reasons are related to data and the methodology and the rest are concerned with macro issues and industry trend of credit scoring in microcredit. Among the reasons, few important issues have been discussed as follows:

7.2.1.1 Data Related:

- The representative number of loan cases in the case base of the CBR system seems not sufficient. Because only 107 loan cases have been considered as complete loan cases with solution in big data environment (about 2.0 million loan cases). It has been observed in the results (risk scoring) that upto 61.78% similarity level was found as best similar case that infers the weakness of the representativeness of the case base to find the best similar case as the initial solution of the new loan cases.
- Besides the small number of representative loan cases in the case base, the number of test cases is only 75 which is also insufficient in compare to large number of test cases for 0.26 million cases in the hold out sample.
- Also, there exists data missing for financial performace of loan applicants or microcredit borrowers which are usually considered as the prime data group to any credit scoring in consumer loan or credit card industry. However, here proxy data from field partners have been used as the mitigation policy.

7.2.1.2 Method Related:

- The definitions of successful loan cases and unsuccessful loan cases have been made straight forward having the top four grades (AG, UG3, UG2 & UG1) or the scores range from 5.60 to 10.0 as successful loan cases and the lower four grades (LG4, LG3, LG2, & LG1) or the scores range from 0.00 to 5.50 as unsuccessful loan cases. If any grade is shifted from either group, then the results of the CBR system will be affected. As the first attempt, the boundary of both definitions seem reasonable although there are lots of room for manipulation.
- Both categorical weights and individual factor weights have been finalized based on the comparison of two different sets of expert models with two versions of each (the details in sub-section 1.1 and 1.2). As the distributional effects of sample data have no positive impact on the improvement of the performance, the weights assigned by the jury of expert model have been considered as final to use in similarity function of the CBR system. Still it has scope to improve by using discriminant analysis or logit regression.
- In Kiva database, some personal profile data like borrower's age, marital status, family size, spouse employment status, business or loan experience etc. are available. According to the jury of credit experts, these data may improve the performance of the CBR system. Because the group of data represents the degree of personal background of a loan applicant which are highly regarded as vital indicators in today's social lending system. So, these data will be considered in future research although these data are not readily extractable at this moment.

7.2.1.3 Macro Issues Related:

Some socio-economic, natural and political factors seem reasonable to impact the performance of the CBR system. For example, the customers of micro credit are living in the locations across the world where political unrest exists. Vulnerable groups of poor society in the world are found as micro credit borrowers in different developing countries in Asia-Africa zone (sample zones in this research) who are affected by natural disaster, social and economic reasons.

7.2.1.4 Industry Trend of Credit Scoring:

The industry trend of credit scoring in micro credit is less powerful than the impact of credit scoring in consumer loans (mortage loans, home loans, car loans etc.) and credit card loans in developed or wealthy countries(Schreiner, 2005). The reasons behind this trend may be that weak or even no database of the clients in developing regions in compare to the rich database in the developed worlds. Moreover, microcredit borrowers vary from one another where identifying common factors for credit scoring is a challenging job due to complex nature of factors associated with natural, social, economic and political contexts.

7.3 Credibility of the results of the CBR System

The credibility of the findings of any research study depends on the proper design of this research. Reliability and validity are the two particular concerns about the research design. Reliability concerns about data collection technique or analysis procedure that yields the consistent findings in terms of same results of other occasions, similar observations by others, and the transparency in how sense was made by raw data. The data from Kiva open source database (build.kiva.org) were retrieved in XML format by XQuery programming language and then the snapshots of data in XML format were critically examined by credit experts (domain experts) to review the data availability. After the proper examination of the sample data snapshots (all data are in the same structure in this big database), the credit experts have selected the factors that are relevant and most common in the line of similar works(Blanco et al., 2013; Dukiü, Dukiü, & Kvesiü, 2011; Gool et al., 2012; Baklouti; Ibtissem & Bouri, 2013; Schreiner, 2004; Serrano-Cinca et al., 2013; Vigano, 1993). With the selected factors, an ad hoc SQL database has been created to interact with CBR system. Because all the data in Kiva database are not relevant and the structure is not suitable or complete as cases (solution is missing in Kiva database) to fit with the CBR system adopted in this study. Here, CBR approach has been taken as prime methodology to assess borrower risk in online indirect P2P lending models since no other statistical models fit well with this unique nature of borrower profiles in microcredit where nature of borrower characteristics demands for special knowledge (Bunn & Wright, 1991). In this system, spreadsheet based expert model for completing the missing component of loan case ‘risk score’ has been considered as an integral part of the CBR system. This expert model has contributed to the cold boot problem of the CBR system(Uddin et al., 2015).

7.4 Summary

This chapter has presented four different results of expert models with different sets of assumptions and has selected the model ‘EM01b’ considering the best comparative predictive results (borrower risk scores). With the results of the best expert model (EM01b), the case base has been updated and finally the evaluation of the CBR integrated system has been made with 75 loan cases from hold out sample or test set.

Chapter 8

Main Contributions and Future Research Directions

This chapter presents the conclusions of the research presented in the previous chapters. A summary of the whole research is provided in Section 8.1. Then, Section 8.2 specifies the contributions made in this research. The limitations of the study are discussed in Section 8.3 that is followed by the directions of our future research and presented in Section 8.4.

8.1 Research Summary

Lenders in online indirect P2P microcredit lending platforms always face challenges in choosing a borrower from many candidates on such platforms, particularly for individual lenders who are not expert in lending. An inherent risk exists in a pseudonymous online environment of P2P lending where most of the individual lenders are not professional investors which causes serious information asymmetry problems. In this context, lenders are provided with little information, which lack the details of the financial aspects, particularly risk assessment of the loan applicants and eventually they are confronted with judging the worthiness of applicants for which making their lending-decisions is really a tough job. Therefore, loan default and loan fraud would be the most fundamental concern, among others, with lending money unsecured to complete strangers over the Internet. In this case, borrower information and its accuracy are critical for lenders to assess a borrower's credit risk. However, obtaining and verifying borrower information would increase the operation cost considerably. It is more acute in online indirect P2P lending platforms that are serving globally in general, developing countries in particular. In addition, being a new innovative business model, online P2P lending platform is under the most influential challenges to overcome the regulatory issues as well as to replicate the social network learned from off-line solidarity lending. Among the challenges, the problem with the borrower's or loan applicant's information is critical to the web-based lenders to remain active in such platforms and to sustain them in the long term in the promotion of novel goal, reducing global poverty. Moreover, it is serious because no individual credit risk rating is provided directly or indirectly by the field partners or by such lending platforms resulting bearing the default risk lies absolutely with the lenders who ultimately refinance the field partners. Selecting borrower is the challenging task to online microcredit lenders as individual borrower's profile does not provide any risk rating on the platform except the microfinance intermediaries' aggregate risk indicators and the information that these intermediaries screen/assess the borrowers before being posted and made available to the lenders. Moreover, the platforms merely keep typical advices for lenders to diversify their portfolios through lending to more than one borrower via different field partners as well as in different countries and/or sectors. Different risk management tools are practiced in the sector, particularly in off-line brick-and-mortar models, but most of them are for group borrowers and risk rating of borrowers is not provided to the lenders on indirect P2P platforms[58]. This lack of missing information on borrower risk assessment is surprising since credit scoring could help the online indirect P2P

[58]Risk rating with credit score is available in most of the direct P2P platforms who operate nationally like Prosper, Zopa which are out of this study (Ceyhan et al., 2011; Slavin, 2007; H. Wang & Greiner, 2011). However, this study focuses only on the indirect P2P lending models who operate globally like Kiva,Zidisha(Hassett et al., 2011).

model's lenders to evaluate the loan applicants more efficiently and thereby could match their lending risk perception with the degree of risk associated with a particular loan applicant.
Historical data about loan requests, actual loans, successful repayments, delays and delinquency situations are in fact made available by one of these platforms and this allows applying Artificial Intelligence techniques to support the evaluation of new loan requests. This large amount of data represents an asset that can be exploited to develop a support system exploiting, at the same time, available expert knowledge and historical data: the latter contains description of loan episodes and final outcomes of actual loans, but it lacks actual indication of what should have been the suggested risk rating associated to the loan request. Therefore, the main objective of this research is to build a Case Based Reasoning (CBR) system for borrower risk assessment in online indirect P2P microfinance platforms and to suggest how risk assessment, especially credit scoring, can be useful to online P2P micro-lenders. To achieve this goal a proper case base is needed: the loan episodes include a case description and outcome, but they lack a solution that, in our context, is represented by the risk rating. Therefore, solving the problem of achieving the missing solution part in case structure another sub-objective is to develop an expert-based risk rating model. This model has been used to effectively bootstrap the CBR system by producing a set of representative complete cases.
The CBR approach has been chosen for this research since it allows considering the unique nature of borrower profiles in microcredit, where specific characteristics demands for special knowledge. The proposed CBR system works as a statistical, incremental learning model to improve the results (risk rating/prediction) of judgmental or expert rules, that were employed in the bootstrapping process of initial cases definition, and that however are instrumental in the definition of the similarity function guiding the retrieval of past cases relevant to the present one.
The Kiva model has been chosen as the largest and leading one as a case study which allows to access its open source data for study. The Kiva XML data have been recovered using XQuery to organize a database for past loans of individual borrowers (unit of analysis) with representative numbers and then examination method has been used for identifying relevant and readily extractable features for the sample of past individual borrower loans. The database of Kiva is large enough to qualify the requirements of large size database for CBR application. From the database of Kiva, only African and Asian zones have been chosen (for a total of 45 countries) accounting for more than 50% coverage of the whole database (83 countries). The reason for choosing these two zones is the homogeneity in terms of loan size and nature of borrower's activities. Only individual borrower's loan data have been chosen as a unit of analysis, not considering group borrower data for selected variables. All the information necessary to define a case description are available, in addition to the final outcome (the information about the actual repayments), but no actual risk rating is present and therefore all cases would be missing the solution part. To solve this "cold boot" problem, a strategy is adopted to select a reasonable number of past loans that are sufficiently representative of all the selected countries, economic sectors for the funded activities, kind of borrowers, and actually rate them (filling thus the solution part of the case) employing expert rules for rating the risk associated with loan requests in developing countries, coded into a spreadsheet. Therefore, a constrained expert model or integrated model has finally been chosen combining expert-based manual model (expert-judgment approach or knowledge-based approach) with automated statistical model (CBR approach).
Using this expert-based models *credit scoring* has been carried out for a set of representative cases of loans, and then they have been used in the CBR system as complete loan cases to run the system for assessing new loan applicants or new borrowers. Finally, the CBR rating has been tested with a set of test loan cases for evaluating its predictive power. The CBR system

considered as low risk borrow requests 60 of them, 87% of which were correctly repaid; the system turned out to be quite conservative, since requests considered risky often turned out to be correctly repaid, but in general results are encouraging.

8.2 Main Contributions

The more general contribution of this research is to have investigated an understudied field in Borrower Risk Assessment in Peer-to-Peer (P2P) Web based Microfinance Platforms that result in building a CBR system for the same and in developing an Expert-based Risk Rating Model.

Hence, this research work has brought to two contributions: expert-based risk rating and CBR-based risk rating. As a dominant risk rating approach in microfinance, till now, expert-based risk rating approaches could assess borrower risk in microfinance very well. However, despite its good practice in traditional brick-and-mortar models, it does not fit well online P2P situations due to its inherent limitations like its static nature (no learning), which leads to the need of maintenance, often heavy computational costs, high user requirements and sometimes just partial automation, and in general it is not applicable to large scale operations. Thus, the expert-based rating (risk scoring) has been developed and used, as shown in Figure 8.1, for providing the solution to a proper set of representative relevant cases to use in CBR-based risk rating which is instead completely automated, with an incremental learning, and in general more suited to an online P2P microcredit setting.

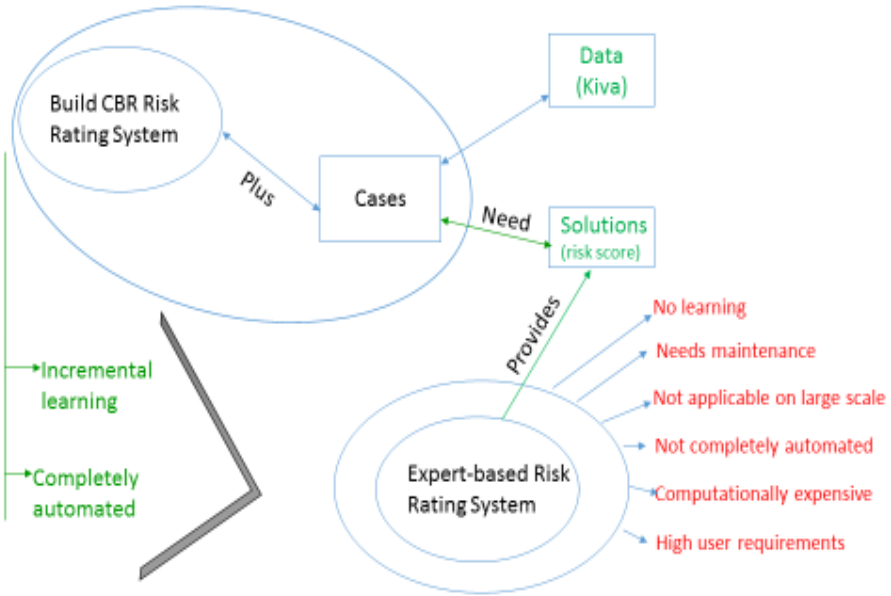


Figure 8.1 *Research contributions.*

In one hand, A CBR system for improving microcredit system (for example, Kiva), in particular for providing a loan request risk rating based on past loans that are most similar to the new one to be published on the microcredit web site. The system has been developed and a strategy to solve the cold boot problem has been devised and implemented: as of this

moment, the case base is being populated to better cover the variety of the potential loan requests, and then proceed with a quantitative evaluation of the CBR system effectiveness.
This tool can be deployed to filed partners or alternatively to Kiva Systems. The field partners can use this tool to assess/rate the new applicants (who will make loan requests to the field partners) and based on the rating they can also provide suggestions to the applicants for how to make their businesses more appealing, competitive for the loans and also, hopefully, more successful. In case of Kiva Systems, they can adopt/align this tool in their existing systems and thereby incorporate the rating in borrower's description space. Such kind of incorporation will definitely help the end users/lenders understand the risk category of the borrowers. As a result, the lenders will be able to diversify the lending risk of their lending portfolios.
On the other hand, borrower risk scoring i.e. Expert Model, is a spreadsheet based grading model based on the pre-specified scale or value (given by experts) of the selected factors of micro credit borrower to assess and evaluate the underlying credit-risk for a given borrower.

8.3 Limitation of the Study

Despite the contributions illustrated in the previous section, this research has one main limitation. This limitation concerns the reliability and validity of the system, it requires few more trials which will help to confirm the reliability and validity issues in future. Now, it is a newly developed system based on CBR approach which is introduced as a tool for assessing borrower risk in online indirect P2P lending platforms in microfinance industry. Due to time and resource constraints, we used this CBR system in Kiva only. If the system could have been used in similar platforms too, a more comprehensive and generalized systems would have been developed.

8.4 Future Research Directions

The limitations of this study provide direction for new research for investigating the usefulness and potential of a CBR system in P2P online platform. In this research setting, data have been used from an open source data base of the leading model- Kiva.org in online indirect P2P lending platforms in microfinance industry. Although the data have been collected from one organization, similar data exist in other models in the industry (online indirect P2P lending platforms), although few other models like Zidisha, MyC4, Deki maintain and support the access to this kind of data from their online lending operations. Therefore, the developed CBR system for assessing borrower risk can be used with in other online indirect P2P microcredit lending initiatives.
Our findings clearly indicate that a system based on CBR approach is indisputably helpful to the lenders in assessing borrowers risk in online platform. Hence, more rigorous work for investigating its potential will help to improve microcredit initiatives in a broader scope.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59. doi:10.1.1.56.4481

Ahn, H., Kim, K. jae, & Han, I. (2007). A case-based reasoning system with the two-dimensional reduction technique for customer classification. *Expert Systems with Applications*, *32*(4), 1011–1019. doi:10.1016/j.eswa.2006.02.021

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, *84*(3), 488–500. doi:10.2307/1879431

Anderson, J. R., (1983). The architecture of cognition. Harvard University Press, Cambridge in Aamodt,A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Armendáriz de Aghion, B., & Morduch, J. (2005). *The Economics of Microfinance* (Vol. 31). doi:10.1086/523604

Ashta, A., & Assadi, D. (2010). An Analysis of European Online micro-lending Websites. *Innovative Marketing*, *6*(2), 7–17. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1518637

Assadi, D., & Ashta, A. (2010). Should online Micro-lending be for profit or for philanthropy? *Economic Issues*, (November 2009). doi:10.3917/jie.006.0123

Ayayi, A. G. (2012). Credit risk assessment in the microfinance industry. *Economics of Transition*, *20*(1), 37–72. doi:10.1111/j.1468-0351.2011.00429.x

Bachmann, A. ;, Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., & Tiburtius, P. (2011). Online Peer-to-Peer Lending: A Literature Review. *Journal of Internet Banking and Commerce*, *16*(2).

Baesens, A. B., Gestel, T. Van, Viaene, S., Stepanova, M., Suykens, J., Baesensl, B., … Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627–635.

Bareiss, R. (1989). Exemplar-based knowledge acquisition: A unified approach to concept representation, classification, and learning. Boston, Academic Press in Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Bauchet, J., Marshall, C., Starita, L., Thomas, J., & Yalouris, A. (2011). *Latest Findings from Randomized Evaluations of Microfinance*. *October*. Retrieved from http://www.cgap.org/p/site/c/template.rc/1.9.55766/

Berger, S. C., & Gleisner, F. (2009). Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending. *BuR - Business Research*, *2*(1), 39–65. doi:10.1007/BF03343528

Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications*, *40*(1), 356–364. doi:10.1016/j.eswa.2012.07.051

Brown, S. J., & Lewis, L. M. (1991). A Case-Based Reasoning Solution to the Problem of Redundant Resolutions of Nonconformances in Large-Scale Manufacturing.

Bruett, T. (2007). Cows, Kiva, and Prosper.com: How Disintermediation and the Internet are Changing Microfinance. *Community Development Investment Review*, *3*(2), 45–50.

Bunn, D., & Wright, G. (1991). Interaction of Judgemental and Statistical forecasting methods: issues and analysis. *Management Science*, *37*(5).

Buta, P. (1994). Mining for financial knowledge with CBR. AI Expert, 9(2), 34–41 in Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*, *39*(9), 8389–8395.

Campion, A. (2001). Client Information Sharing in Bolivia. *Journal of Microfinance*, *3*(1).

Carbonell, J. G. (1986). Derivational Analogy: A Theory of Reconstructive Problem Solving and Expertise Acquisition. *Machine Learning An Artificial Intelligence Approach*, *2*(CMU-CS-85-115), 371–392.

Ceyhan, S., Shi, X., & Leskovec, J. (2011). Dynamics of bidding in a P2P lending service. *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, 547. doi:10.1145/1963405.1963483

Che, Z. H., Wang, H. S., & Chuang, C. L. (2010). A fuzzy AHP and DEA approach for making bank loan decisions for small and medium enterprises in Taiwan. *Expert Systems with Applications*, *37*(10), 7189–7199. doi:10.1016/j.eswa.2010.04.010

Chen, R., Chen, Y., Liu, Y., & Mei, Q. (2014). Does Team Competition Increase Pro-Social Lending ? Evidence from Online Microfinance, 1–54.

Chen, Y. K., Wang, C. Y., & Feng, Y. Y. (2010). Application of a 3NN+1 based CBR system to segmentation of the notebook computers market. *Expert Systems with Applications*, *37*(1), 276–281. doi:10.1016/j.eswa.2009.05.002

Choo, J., Lee, C., Lee, D., & Park, H. (2014). Understanding and Promoting Micro-Finance Activities in Kiva . org.

Chuang, C. L., & Lin, R. H. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications*, *36*(2), 1685–1694. doi:10.1016/j.eswa.2007.11.067

Chun, S.-H., & Park, Y.-J. (2006). A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting. *Expert Systems with Applications*, *31*, 329–336. doi:10.1016/j.eswa.2005.09.053

Clarke, G., Cull, R., Peria, M., & Sánchez, S. (2005). Bank lending to small businesses in Latin America: does bank origin matter? *Journal of Money, Credit and Banking*, *37*(1), 83–118. doi:10.2307/3838938

Coleman, R. W. (2007). *Is the Future of the Microfinance Movement to be Found on the Internet ?*

Collier, B., & Hampshire, R. (2010). Sending Mixed Signals: Multilevel Reputation Effects in Peer-to-Peer Lending Markets. *ACM Conference on Computer Supported Cooperative Work*, 1–10. doi:10.1145/1718918.1718955

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, *183*, 1447–1465. doi:10.1016/j.ejor.2006.09.100

Cubiles-De-La-Vega, M.-D., Blanco-Oliver, A., Pino-Mejías, R., & Lara-Rubio, J. (2013). Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert Systems with Applications*, *40*(17), 6910–6917. doi:10.1016/j.eswa.2013.06.031

Deininger, K., & Liu, Y. (2009). Determinants of Repayment Performance in Indian Micro-Credit Groups, (March), 2–11.

Desai, V. S., Conway, D. G., Crook, J. N., & Overstreet, G. A. (1997). Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. *IMA Journal of Management Mathematics*, *8*, 323 –346. doi:10.1093/imaman/8.4.323

Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, *95*(1), 24–37. doi:10.1016/0377-2217(95)00246-4

Dinh, T. H. T., & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, *16*(5), 471–495. doi:10.1016/j.irfa.2007.06.001

Dukiü, D., Dukiü, G., & Kvesiü, L. (2011). A Credit Scoring Decision Support System, 391–396.

Eisenbeis, R. A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal of Banking and Finance*, *2*(3), 205–219. doi:10.1016/0378-4266(78)90012-2

Everett, C. R. (2015). Group membership, relationship banking and loan default risk: the case of online social lending. *Banking and Finance Review*, (November), 1–40. doi:10.2139/ssrn.1114428

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 37–54. doi:10.1145/240455.240463

Flannery, M. (2007). Kiva and the Birth of Person-to-Person Microfinance. *Innovations*, (winter & spring), 31–56.

Freedman, S., & Jin, G. Z. (2008). Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper. com. *Social Networks*, (December 2007), 63. doi:http://dx.doi.org/10.2139/ssrn.1304138

Frerichs, A., & Schumann, M. (2008). Peer to Peer Banking – State of the Art. *Institut Für Wirtschaftsinformatik Der Georg-August-Universität Göttingen, Arbeitsbericht*, (02), 80.

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance Decision Making: A Field Study of Prosocial Lending. *Journal of Marketing Research*, *48*, S130–S137. doi:10.1509/jmkr.48.SPL.S130

Garman, S., Hampshire, R., & Krishnan, R. (2008). A search theoretic model of person-to-person lending. *Working Paper Available at http://www.heinz.cmu.edu/research/244full.pdf*, 1–32.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi:10.1016/S0364-0213(83)80009-3

Ghatak, M., & Guinnane, T. W. (1999). The economics of lending with joint liability: Theory and practice. *Journal of Development Economics*, *60*(May), 195–228. doi:10.1016/S0304-3878(99)00041-3

Gool, J. V., Verbeke, W., & Baesens, B. (2012). Credit Scoring for Microfinance: Is it Worth IT? *International Journal of Finance & Economics*, *17*(2), 103–123. doi:10.1002/ijfe

Gutierrez-Nieto, B., Serrano-Cinca, C., & Camon-Cala, J. (2016). A Credit Score System for Socially Responsible Lending. *Journal of Business Ethics*, 691–701. doi:10.1007/s10551-014-2448-5

Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: a Review. *Royal Statistical Society*, 523–541. doi:10.1111/j.1467-985X.1997.00078.x

Hannessy, D., & Hinkle, D. (1992). Case-Based Reasonfng. *Ieee Expert Intelligent Systems And Their Applications*, *IEEE Exper*.

Harmon, P. (1992). Case-based reasoning III, Intelligent Software Strategies, VIII (1) in Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Hassett, T., Bergeron, J. D., Kreger, M., & Looft, M. (2011). Indirect P2P Platforms. *2011 Global Microcredit Summit*, 1–32.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). Model assessment and selection. *The Elements of Statistical Learning*, (X), 219–259. doi:10.1007/b94608

Hawkins, R., Mansell, R., & Steinmueller, W. (1999). Toward digital intermediation in the information society. *Journal of Economic Issues*, *33*(2), 383–391. doi:10.1080/00213624.1999.11506169

Heng, S., Meyer, T., & Stobbe, A. (2007). Implications of Web 2.0 for financial institutions: Be a driver, not a passenger. *Munich Personal RePEc Archive*, *2007*, 0–11. Retrieved from http://mpra.ub.uni-muenchen.de/4316/

Hens, A. B., & Tiwari, M. K. (2012). Computational time reduction for credit scoring: An integrated approach based on support vector machine and stratified sampling method. *Expert Systems with Applications*, *39*(8), 6774–6781. doi:10.1016/j.eswa.2011.12.057

Hermes, N., & Lensink, R. (2007). The empirics of microfinance: What do we know? *Economic Journal*, *117*, 1–10. doi:10.1111/j.1468-0297.2007.02013.x

Herrero-Lopez, S. (2009). Social interactions in P2P lending. *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, *09*, 1–8. Retrieved from http://delivery.acm.org/10.1145/1740000/1731014/a3-herrero-lopez.pdf

Ibtissem, B. (2013). Determinants of Microcredit Repayment : The Case of Tunisian Micro finance Bank, *25*(3), 370–382.

Ibtissem, B., & Bouri, A. (2013). Credit Risk Management in Microfinance : the conceptual framework, *2*(1), 9–24.

Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management*, *10*(February 2015), 233–240. doi:10.3846/1611-1699.2009.10.233-240

Jappelli, T., & Pagano, M. (2000). *Information Sharing in Credit Markets: A Survey* (Vol. 5). Retrieved from www.csef.it/WP/wp35.pdf

Jenq, C., Pan, J., & Theseira, W. (2012). What Do Donors Discriminate On ? Evidence from Kiva . org.

Jeong, G., Lee, E., & Lee, B. (2012). Does Borrowers' Information Renewal Change Lenders' Decision in P2P Lending? An Empirical Investigation, (February), 83–86.

Jo, H., Han, I., & Lee, H. (1997). Bankruptcy prediction using case-based reasoning, neural networks,

and discriminant analysis. *Expert Systems with Applications*, *13*(2), 97–108. doi:10.1016/S0957-4174(97)00011-0

Kauffman, R. J., & Riggins, F. J. (2012). Information and communication technology and the sustainability of microfinance. *Electronic Commerce Research and Applications*, *11*(5), 450–468. doi:10.1016/j.elerap.2012.03.001

Khandker, S. R., Faruqee, R., & Samad, H. A. (2014). *Are Microcredit Borrowers in Bangladesh Over-indebted?*

Kim, H. S., & Sohn, S. Y. (2010). Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, *201*(3), 838–846. doi:10.1016/j.ejor.2009.03.036

Kinda, O., & Achonu, A. (2012). Building A Credit Scoring Model For The Savings And Credit Mutual Of The Potou Zone (MECZOP)/Senegal. *Consilience: The Journal of Sustainable Development*, *7*(1), 17–32.

Klafft, M. (2008). Online Peer-to-Peer Lending: A Lenders' Perspective. *SSRN Electronic Journal*. doi:10.2139/ssrn.1352352

Kolodner, J. L. (1983). Maintaining organization in a dynamic long-term memory. *Cognitive Science*, *7*(4), 243–280. doi:10.1016/S0364-0213(83)80001-9

Kolodner, J. (1988). Retrieving events from case memory: A parallel implementation. In: Proceedings from the Case-based Reasoning Workshop, DARPA, Clearwater Beach, 1988, pp. 233-249 in Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*, 3–34. doi:10.1007/BF00155578

Kolodner, J. L. (1993). Case-based reasoning. San Mateo, CA: Morgan in Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*, *39*(9), 8389–8395.

Kono, H., & Takahashi, K. (2010). Microfinance revolution: Its effects, innovations, and challenges. *Developing Economies*, *48*(1), 15–73. doi:10.1111/j.1746-1049.2010.00098.x

Lee, T. S., Chiu, C. C., Lu, C. J., & Chen, I. F. (2002). Credit scoring using the hybrid neural discriminant technique. *Expert Systems with Applications*, *23*(3), 245–254. doi:10.1016/S0957-4174(02)00044-1

Lee, T.-S., & Chen, I.-F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, *28*(4), 743–752. doi:10.1016/j.eswa.2004.12.031

Li, H., & Sun, J. (2011). On performance of case-based reasoning in Chinese business failure prediction from sensitivity, specificity, positive and negative values. *Applied Soft Computing*, *11*, 460–467. doi:10.1016/j.asoc.2009.12.005

Light, J. (2012). Is "Peer-to-peer" lending worth the risk? *The Wall Street Journal*.

Liu, Y., Chen, R., Chen, Y., Mei, Q., & Salib, S. (2012). " I Loan Because ...": Understanding Motivations for Pro-Social Lending. *Proceedings of the Fifth ACM International Conference on Web Search*

and Data Mining (2012), 503–512. doi:10.1145/2124295.2124356

Luoto, J., McIntosh, C., & Wydick, B. (2004). Credit Information Systems in Less-Developed Countries : Recent History and a Test, 1–31.

Magee, J. R. (2011). Peer-to-Peer Lending in the United States: Surviving After Dodd-Frank. *North Carolina Banking Institute*, *15*. doi:10.1525/sp.2007.54.1.23.

Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. *European Journal of Operational Research*, *136*(1), 190–211. doi:10.1016/S0377-2217(01)00052-2

Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega*, *31*(2), 83–96. doi:10.1016/S0305-0483(03)00016-1

MÁNTARAS, R. L. DE, MCSHERRY, D., BRIDGE, D., LEAKE, D., SMYTH, B., CRAW, S., … WATSON, I. (2005). Retrieval, reuse, revision, and retention in case- based reasoning. *The Knowledge Engineering Review*, *000*, 1–31. doi:10.1017/S000000000000000

Manzoni, S., Sartori, F., & Vizzari, G. (2007). Substitutional Adaptation in Case-Based Reasoning: a General Framework Applied To P-Truck Curing. *Applied Artificial Intelligence*, *21*(October 2014), 427–442. doi:10.1080/08839510701253641

McIntosh, C., & Wydick, B. (2005). Competition and microfinance. *Journal of Development Economics*, *78*(2), 271–298. doi:10.1016/j.jdeveco.2004.11.008

Mechitov, A. I., Moshkovich, H. M., Olson, D. L., & Killingsworth, B. (1995). Knowledge acquisition tool for case-based reasoning systems.

Meer, J., & Rigbi, O. (2012). Transaction costs and social distance in philanthropy: evidence from a field experiment.

Milana, C., & Ashta, A. (2012). Developing microfinance: A survey of the literature. *Strategic Change*, *21*(7-8), 299–330. doi:10.1002/jsc.1911

Min, J. H., & Lee, Y.-C. (2008). A practical approach to credit scoring. *Expert Systems with Applications*, *35*, 1762–1770. doi:10.1016/j.eswa.2007.08.070

Montazemi, A. R., & Gupta, K. M. (1997). A framework for retrieval in case-based reasoning systems. *Annals of Operations Research*, *72*, 51–73. Retrieved from http://www.springerlink.com/index/m04085818k38p361.pdf

Moon, T. H., & Sohn, S. Y. (2008). CASE-BASED REASONING FOR PREDICTING MULTIPERIOD FINANCIAL PERFORMANCES OF TECHNOLOGY-BASED SMEs. *Applied Artificial Intelligence*, *22*, 602–615. doi:10.1080/08839510701734285

Nair, L., Mehrotra, R., Vaish, A. K., In, A., Indian, T. H. E., Scenario, B., & Kushwaha, S. (2011). CO - ORDINATOR Dean (Academics), Tecnia Institute of Advanced Studies , Delhi CO - EDITOR, *1*(1041).

O'Roarty, B., Patterson, D., McGreal, S., & Adair, A. (1997). A case-based reasoning approach to the selection of comparable evidence for retail rent determination. *Expert Systems with Applications*, *12*(4), 417–428. doi:10.1016/S0957-4174(97)83769-4

Oh, K. J., & Kim, T. Y. (2007). Financial market monitoring by case-based reasoning. *Expert Systems with Applications*, *32*, 789–800. doi:10.1016/j.eswa.2006.01.044

Park, C. S., & Han, I. (2002). A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. *Expert Systems with Applications*, *23*(3), 255–264. doi:10.1016/S0957-4174(02)00045-3

Pellegrina L. D. and Masciandaro D. (2006.). Informal Credit and Group Lending: Modelling the Choice. Finance India, XX: 491-511 in Vaish, A. K., Kumar, A., & Bhat, A. (2011). Need for Credit Scoring in Micro-Finance: Literature Review, *1*(1041).

Persson, A. (2012). Microfinance 2.0: Can a crowdsourced model save microfinance? *Human Rights Studies*, 1–47.

Peters, M. J., Howard, K., & Sharp, M. J. A. (2012). *The management of a student research project*. Gower Publishing, Ltd.

Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. *European Journal of Operational Research*, *112*(2), 310–321. doi:10.1016/s0377-2217(97)00398-6

Point, A. (n.d.). *Report on Credit Risk Grading Manual*.

Porter, B., & Bareiss, R. (1986). {PROTOS}: An Experiment in Knowledge Acquisition for Heuristic classification tasks. *First International Meeting on Advances in Learning (IMAL)*.

Pytkowska, J., & Spannuth, S. (2011). Indebtedness of Microcredit Clients in Kosovo Results from a comprehensive field study, (May).

Rayo, S., Lara, J., & Camino, D. (2010). A Credit Scoring Model for Institutions of Microfinance under the Basel II Normative. *Journal of Economics, Finance and Administrative Science*, *15*(28), 89–124. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=51381543&lang=es&site=ehost-live&scope=site

Recio-García, J. a, Díaz-Agudo, B., & González-Calero, P. (2008). *jCOLIBRI2 Tutorial*. *Tutorial*.

Reinartz, T., Iglezakis, I., & Roth–Berghofer, T. (2001). Review and Restore for Case-Base Maintenance. *Computational Intelligence*, *17*(2), 214–234. doi:10.1111/0824-7935.00141

Riggins, F. J., & Weber, D. M. (2012). A model of peer-to-peer (P2P) social lending in the presence of identification bias. *Proceedings of the 13th International Conference on Electronic Commerce - ICEC '11*, 1–8. doi:10.1145/2378104.2378127

Ross B.H. (1989). Some psychological results on case-based reasoning. Case-Based Reasoning Workshop, DARPA 1989. Pensacola Beach. Morgan Kaufmann,pp. 144-147) in Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Rozycki, V. (2006). *Credit Information Systems for Microfinance: A foundation for further innovation*.

Saunders, M. N., Saunders, M., Lewis, P., & Thornhill, A. (2011). *Research methods for business students, 5/e*. Pearson Education India.

Schank R. (1982). Dynamic memory; a theory of reminding and learning in computers and people. Cambridge University Pressin Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, *7*, 39–59.

Schicks, J. (2011). *Microfinance Over-Indebtedness : Understanding its drivers and challenging the*

common myths Microfinance Over-Indebtedness : and challenging the common myths. *Centre Emile Bernheim* (Vol. 32). Retrieved from http://www.rb-cms.nl/classes/FCKeditor/upload/65/File/overindebtness_cermi_2010.pdf

Schreiner, M. (1999). A Scoring Model of the Risk of Costly Arrears at a Microfinance Lender in Bolivia. *Development and Comp Systems*. Retrieved from http://ideas.repec.org/p/wpa/wuwpdc/0109005.html

Schreiner (2001). Un Sistema de Scoring del Riesgo de Créditos de FIE en Bolivia, report to Fomento de Iniciativas Económicas, La Paz in Schreiner, M. (2002). Scoring : The Next Breakthrough in Microcredit?

Schreiner, M. (2002). Scoring : The Next Breakthrough in Microcredit ?

Schreiner, M. (2004). Scoring Arrears at a Microlender in Bolivia. *Journal of Microfinance*, *6*, 65. Retrieved from https://journals.lib.byu.edu/spc/index.php/ESR/article/view/1456/1417

Schreiner, M. (2005). Can Scoring Help Attract Profit-Minded Investors to Microcredit ?

Serrano-Cinca, C., Gutierrez-Nieto, B., & Reyes, N. M. (2013). *A Social Approach to Microfinance Credit Scoring* (Vol. 32).

Sharma, M., & Zeller, M. (1997). Repayment performance in group-based credit programs in Bangladesh: An empirical analysis. *World Development*, *25*(10), 1731–1742. doi:10.1016/S0305-750X(97)00063-6

Shin, K., & Han, I. (2001). A case-based approach using inductive indexing for corporate bond rating. *Decision Support Systems*, *32*, 41–52. doi:10.1016/S0167-9236(01)00099-9

Slade, S. (1991). Case-Based Reasoning : A Research Paradigm. *AI Magazine*, *12*(1), 41–55.

Slavin, B. (2007). Peer-to-peer lending–An Industry Insight. In *Retrieved from bradslavin. com* (pp. 1–15). Retrieved from http://scholar.google.com/scholar?

Small, D. a., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153. doi:10.1016/j.obhdp.2006.01.005

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. *Cognitive Science Series*. doi:10.2307/414206

Soman, D., & Cheema, A. (2002). The Effect of Credit on Spending Decisions: The Role of the Credit Limit and Credibility. *Marketing Science*, *21*(1), 32–53. doi:10.1287/mksc.21.1.32.155

Steelmann, A. (2006). Bypassing Banks, Region Focus, Federal Reserve Bank of Richmond, 10 (3): 37-40 in Berger, S. C., & Gleisner, F. (2009). Emergence of Financial Intermediaries in Electronic Markets:The Case of Online P2P Lending. *BuR - Business Research*, *2*(1), 39–65.

Stiglitz, J. E. ., & Weiss, A. (1981). Credit Rationing in Market with imperfect inoformation. *The Amercian Economic Review*, *71*(3), 393–410. Retrieved from http://pascal.iseg.utl.pt/~aafonso/eif/pdf/crrinf81.pdf

Tan, Y. H., & Thoen, W. (2000). Toward a Generic Model of Trust for Electronic Commerce. *International Journal of Electronic Commerce*, *5*(2), 61–74. doi:10.1080/10864415.2000.11044201

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*. doi:10.1017/S0140525X00047257

Uddin, M. J., Vizzari, G., & Bandini, S. (2015a). CASE BASED REASONING AS A TOOL TO IMPROVE MICROCREDIT. *IC3K Conference Proceedings*.

Uddin, M. J., Vizzari, G., & Bandini, S. (2015b). Case Based Reasoning As a Tool To Improve Microcredit. *IC3K Conference Proceedings*, *3*(Ic3k), 466–473.

USA, G. (2012). *Giving USA*.

Vaish, A. K., Kumar, A., & Bhat, A. (2011). Need for Credit Scoring in Micro-Finance: Literature Review, *1*(1041).

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988–99. doi:10.1109/72.788640

Vigano, L. (1993). A credit scoring model for development banks: An African case study. *Savings and Development*, *XVII*(4), 441–482.

Vogelgesang, U. (2003). Microfinance in times of crisis: The effects of competition, rising indebtedness, and economic crisis on repayment behavior. *World Development*, *31*(12), 2085–2114. doi:10.1016/j.worlddev.2003.09.004

Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications*, *39*(9), 8389–8395. doi:10.1016/j.eswa.2012.01.181

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. *Knowledge-Based Systems*, *26*, 61–68. doi:10.1016/j.knosys.2011.06.020

Wang, H., Greiner, M., & Aronson, J. E. (2009). People-to-people lending: The emerging E-Eommerce transformation of a financial market. *Springer*, *36 LNBIP*, 182–195. doi:10.1007/978-3-642-03132-8_15

Wang, H., & Greiner, M. E. (2011). Prosper: The eBay for money in lending 2.0. *Communications of the Association for Information Systems*, *29*(1), 243–258.

Watson, I. (1999). Case-based reasoning is a methodology not a technology. *Knowledge-Based Systems*, *12*(December 1998), 303–308. doi:10.1016/S0950-7051(99)00020-9

Watson, J. W. R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2).

Weib, G. N. F. ., Pelger, K., & Horsch, A. (2010). Mitigating Adverse Selection in P2P Lending: Empirical Evidence From Prosper.com.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, *27*(11-12), 1131–1152. doi:10.1016/S0305-0548(99)00149-5

Yap, B. W., Ong, S. H., & Husain, N. H. M. (2011). Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Systems with Applications*, *38*(10), 13274–13283. doi:10.1016/j.eswa.2011.04.147

Ye, K., Yan, J., Wang, S., Wang, H., & Miao, B. (2011). Knowledge level modeling for systemic risk management in financial institutions. *Expert Systems with Applications*, *38*(4), 3528–3538.

doi:10.1016/j.eswa.2010.08.141

Yuge, Y. (2011). The Current Situation of Microfinance in Bangladesh : A Growing Concern about Overlapping Loan Problems – From a Field Visit to Rajshahi and Comilla. *Center for Emmerging Markets Enterprises, Student Research Series*, *March 2011*.

Yum, H., Lee, B., & Chae, M. (2012). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Electronic Commerce Research and Applications*, *11*(5), 469–483. doi:10.1016/j.elerap.2012.05.003

Zeller, M. (1998). Determinants of repayment performance in credit group: The role of program design, intragroup risk pooling and social cohesion. *Economic Development and Cultural Change*, *46*(3), 599–620. doi:10.1086/452360

Appendices

Appendix A

Overview of published credit scoring models for developing countries (Gool et al., 2012)

Author (Date, Country)	Institution type	Sample size	Number of (included) inputs	Technique(s)	Performance metrics
Vigano (1993, Burkina Faso)	Microfinance	100	53 (13)	Discriminant Analysis	PCC, R^2
Sharma and Zeller (1997, Bangladesh)	Microfinance	868	18 (5)	TOBIT Maximum Likelihood Estimation	N/A
Zeller (1998, Madagascar)	Microfinance	168	19 (7)	TOBIT Maximum Likelihood Estimation	N/A
Reinke (1998, South Africa)	Microfinance	1641	8 (8)	Probit Regression	N/A
Schreiner (1999, Bolivia)	Microfinance	39 956	9 (9)	Logistic Regression	PCC
Vogelgesang (2003, Bolivia)	Microfinance	8002	28 (12)	Random Utility Model	PCC, Pseudo-R^2
Vogelgesang (2003, Bolivia)	Microfinance	5956	30 (13)	Random Utility Model	PCC, Pseudo-R^2
Diallo (2006, Mali)	Microfinance	269	17 (5)	Logistic Regression, Discriminant Analysis	PCC, R^2
Kleimeier et al. (2007, Vietnam)	Retail Bank	56 037	22 (17)	Logistic Regression	PCC, SENS, SPEC
Gool et al. (2012, Bosnia)	Microfinance	6722	16	Logistic Regression	AUC

Sample size is total number of observations used, combining training and test sets. Number of inputs is the total number of inputs available. Number of included variables is the number of selected inputs in the final model. If known, a 5% significance level is employed as selection criterium. Dummy variables or transformations belonging to one (categorical) variable are counted as one variable. PCC stands for Percentage Correctly Classified, SENS for sensitivity and SPEC for specificity. Vogelgesang (2001) published multiple models in her study; the two models reviewed in this table are illustrative for the other models.

Appendix B

sl_no	L_ID	L_amount	RT	RI	dis_mode	Gender	FP_rating	Country	Sector	CBR_Score	Sim_%	Grade	status
1	646028	450	6	m	pre	F	closed	Ghana	Retail	5.24	94.84%	LG1	defaulted
2	645923	650	8	m	pre	F	closed	Ghana	Retail	5.24	93.20%	LG1	defaulted
3	640856	350	14	m	pre	F	2.0	Kenya	Agriculture	6.79	99.03%	UG3	paid
4	657477	125	12	m	pre	F	1.5	Liberia	Clothing	6.26	99.03%	AG	paid in repayment
5	656352	425	14	ir	pre	F	4.5	Kyrgyzstan	Agriculture	7.86	92.15%	UG2	paid
6	645924	350	11	m	pre	F	3.5	Kenya	Retail	7.6	92.74%	UG2	paid
7	656496	550	14	ir	pre	M	4.0	Tajikistan	Agriculture	6.44	95.28%	AG	expired
8	656500	325	12	ir	pre	M	3.5	Tajikistan	Agriculture	8.38	92.56%	UG2	paid
9	657401	150	8	m	pre	F	2.0	Philippines	Agriculture	8.61	92.82%	UG1	paid
10	657404	500	7	m	pre	F	3.0	Philippines	Services	7.17	88.19%	UG3	paid
11	657326	150	8	m	pre	F	2.0	Philippines	Food	8.61	92.82%	UG1	paid
12	657329	325	8	m	pre	F	2.0	Philippines	Food	6.09	93.71%	AG	paid
13	645764	425	7	m	pre	F	3.0	Philippines	Services	7.69	90.30%	UG2	paid
14	657330	675	7	m	pre	F	3.0	Philippines	Food	7.14	96.93%	UG3	paid
15	657416	350	9	m	pre	M	2.0	Philippines	Agriculture	7.03	92.21%	UG3	paid
16	645974	2050	14	m	pre	M	-	Azerbaijan	Food	3.15	69.10%	LG2	paid
17	646084	250	14	m	pre	F	3.5	Kenya	Retail	7.6	100.00 %	UG2	paid
18	657393	450	8	m	pre	F	4.0	Philippines	Food	8.93	96.93%	UG1	paid
19	657345	575	8	m	pre	F	3.0	Philippines	Services	7.17	89.98%	UG3	paid
20	645775	400	10	m	pre	F	3.0	Uganda	Retail	5.5	93.12%	LG1	paid
21	645796	1375	5	m	pre	F	4.0	Philippines	Retail	7.13	74.72%	UG3	paid
22	645818	1300	20	m	pre	M	-	Azerbaijan	Transportation	2.04	92.73%	LG3	paid
23	645847	1925	14	m	pre	F	1.0	Azerbaijan	Food	4.64	61.78%	LG1	paid
24	645864	200	6	m	pre	M	3.0	Uganda	Agriculture	8.72	89.72%	UG1	paid
25	645883	925	8	m	pre	F	2.0	Uganda	Services	7.17	89.98%	UG3	paid
26	645912	125	11	m	pre	F	1.5	Liberia	Food	6.26	92.23%	AG	paid in repayment
27	645931	200	14	m	pre	F	4.0	Pakistan	Arts	8.38	86.58%	UG2	paid
28	645944	850	14	m	pre	F	4.5	Tajikistan	Transportation	6.35	80.64%	AG	paid
29	645961	300	14	m	pre	F	4.0	Pakistan	Agriculture	8.38	93.36%	UG2	paid
30	645981	1800	18	m	pre	M	-	Azerbaijan	Services	2.04	67.37%	LG3	paid
31	646008	225	6	m	pre	F	closed	Ghana	Food	5.24	86.09%	LG1	defaulted
32	646042	1175	10	m	pre	F	1.0	Sierra Leone	Retail	4.12	92.08%	LG2	paid

33	646119	1175	8	m	pre	F	1.0	Sierra Leone	Retail	4.12	94.33%	LG2	paid
34	657310	675	14	m	pre	F	3.0	Philippines	Retail	7.14	91.31%	UG3	paid
35	657355	350	17	m	pre	M	2.0	Kenya	Agriculture	6.13	90.07%	AG	paid
36	657392	475	12	m	pre	F	2.5	Pakistan	Manufacturing	5.81	92.83%	AG	paid
37	657417	1125	8	m	pre	F	3.0	Philippines	Retail	6.82	86.81%	UG3	paid
38	657464	125	12	m	pre	F	1.5	Liberia	Retail	6.26	93.36%	AG	paid in repayment
39	657409	2575	20	m	pre	M	4.5	Tajikistan	Food	5.13	83.35%	LG1	expired
40	645954	2050	14	m	pre	M	4.5	Tajikistan	Agriculture	5.13	72.97%	LG1	expired
41	646010	1175	10	m	pre	F	1.0	Sierra Leone	Retail	4.12	92.08%	LG2	defaulted
42	645095	725	12	m	pre	F	closed	Vietnam	Retail	4.8	89.15%	LG1	paid
43	645163	100	8	m	pre	F	2.0	Uganda	Clothing	8.61	90.14%	UG1	paid
44	645186	1700	14	ir	pre	M	4.5	Tajikistan	Agriculture	5.83	73.96%	AG	expired
45	645207	1925	17	m	pre	M	1.0	Azerbaijan	Food	6.83	64.24%	UG3	paid
46	645221	650	14	m	pre	F	4.5	Tajikistan	Clothing	6.35	94.33%	AG	paid
47	645237	250	43	eot	pre	F	3.0	India	Agriculture	8.72	84.03%	UG1	paid in repayment
48	645250	175	8	m	pre	F	2.5	Kenya	Education	6.53	80.23%	UG3	paid
49	645271	1225	14	m	pre	M	paused	Kenya	Agriculture	2.04	78.96%	LG3	paid in repayment
50	645290	350	6	m	pre	F	closed	Ghana	Food	5.24	90.95%	LG1	defaulted
51	645317	125	8	m	pre	F	2.5	Kenya	Food	7.47	89.19%	UG3	paid
52	645338	375	14	m	pre	F	4.0	Pakistan	Agriculture	8.18	91.77%	UG2	paid
53	645359	450	6	m	pre	F	closed	Ghana	Retail	5.24	94.84%	LG1	defaulted
54	655019	825	12	m	pre	M	experimental	Indonesia	Retail	4.8	84.78%	LG1	paid
55	656316	1025	14	ir	pre	F	4.5	Kyrgyzstan	Agriculture	6.09	87.02%	AG	paid
56	656644	600	12	m	pre	M	3.0	Uganda	Retail	7.78	89.97%	UG2	paid
57	656675	325	15	m	pre	F	2.5	Cameroon	Retail	5.81	97.08%	AG	paid
58	656694	250	11	m	pre	M	3.5	Kenya	Services	8.38	89.01%	UG2	paid
59	656861	350	11	m	pre	F	3.0	Philippines	Food	5.5	94.99%	LG1	paid
60	656885	350	8	ir	pre	M	3.0	Philippines	Education	7.37	92.29%	UG3	paid
61	656911	1000	20	ir	pre	F	4.5	Cambodia	Housing	5.51	90.28%	AG	paid
62	656925	175	8	m	pre	F	3.0	Philippines	Agriculture	8.89	97.08%	UG1	paid
63	644543	1450	10	eot	pre	M	2.5	Rwanda	Agriculture	5.41	98.06%	LG1	paid
64	644344	500	8	m	pre	F	2.5	Zimbabwe	Agriculture	5.81	87.76%	AG	paid
65	644623	1375	10	eot	pre	M	2.5	Rwanda	Agriculture	5.41	99.03%	LG1	paid
66	656338	650	15	m	pre	F	2.0	Senegal	Services	5.76	88.35%	AG	paid
67	645956	1275	14	m	pre	F	0.5	Togo	Food	4.1	73.20%	LG2	paid
68	644741	500	8	m	pre	F	2.5	Zimbabwe	Agriculture	5.81	87.76%	AG	paid
69	655577	325	8	m	pre	F	2	Mali	Food	6.09	92.74%	AG	paid
70	656043	425	15	m	pre	F	2.5	Cameroon	Food	5.81	99.03%	AG	paid
71	644500	125	8	m	pre	F	4.0	Philippines	Retail	8.18	89.15%	UG2	paid
72	644515	125	8	m	pre	F	3.0	Philippines	Agriculture	8.89	95.14%	UG1	paid

73	644593	300	13	eot	pre	M	3.0	Philippines	Transportation	8.07	85.76%	UG2	paid
74	644646	1450	10	eot	pre	M	2.5	Rwanda	Agriculture	5.41	98.06%	LG1	paid
75	644502	200	8	m	pre	F	3.0	Philippines	Services	8.89	92.39%	UG1	paid