

# Using historical and weather data for marketing and category management in eCommerce

The experience of EW-Shopp

David Čreslovník  
Ceneje d.o.o.  
Ljubljana, Slovenia  
david@ceneje.si

Aljaž Košmerlj  
Jožef Stefan Institute  
Ljubljana, Slovenia  
aljaz.kosmerlj@ijs.si

Michele Ciavotta  
University of Milan-Bicocca  
Milan, Italy  
michele.ciavotta@unimib.it

## ABSTRACT

Contemporary marketing business in the European Union is currently dominated by large global companies, such as Google and Facebook, which have access to large amounts of data about the consumer journey. Small and medium-sized enterprises (SMEs), instead, only have access to a fraction of those data and thus their decisions are based on intuition, experiments and data analysis in the small. This paper proposes a way of merging the data across consumer journey from various SMEs and, eventually, using big data analytics along with weather data to provide SMEs with valuable market insights to help them with marketing activities and category management to be able to compete with global players.

## KEYWORDS

Big Data Analysis, eCommerce Marketing, eCommerce Category Management, Sales Prediction

### ACM Reference Format:

David Čreslovník, Aljaž Košmerlj, and Michele Ciavotta. 2019. Using historical and weather data for marketing and category management in eCommerce: The experience of EW-Shopp. In *XXXXXXX*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Present-day business is facing a profound transformation of business dynamics. Putting aside slow and old-fashioned "small" data analysis and market experiments backed up by management intuitive capabilities, future businesses need to focus on harnessing large databases to support decision-making and predictive models. Industries are disrupted by entities from all around, but we can spot one common trait of successful newcomers: their supreme usage of the available data that support business decisions in terms of speed and insights, eventually enable them to succeed on the market. Meanwhile, small and medium-sized enterprises (SMEs) often do not have access to trusted data sources and to the knowledge of how to enrich the business knowledge with heterogeneous external information. They do not even have access to those valuable data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

XXX, XXXX, XXXX

© 2019 Association for Computing Machinery.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00  
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Two companies from Slovenia, *Ceneje*<sup>1</sup> and *Big Bang*<sup>2</sup> (the former is e-commerce search engine and a comparison shopping platform in the Adriatic region and the latter is an omnichannel retailer) have joined forces to recreate the typical SME environment in EU. By sharing data and enriching it with weather forecast data, an integrated data-as-a-service platform has been developed, which can then be used by other market players to optimize their processes, predict activities or to engage with target clients more efficiently.

The proposed platform for data sharing, enrichment and analytics has been realized in the frame of EW-Shopp<sup>3</sup>. This is a H2020 EU project aiming at supporting the e-commerce, retail, and marketing industries in improving their efficiency and competitiveness through enabling them to perform predictive and prescriptive analytics over integrated, enriched, and extended large data sets using open and flexible solutions. This paper describes the analytics and prediction subsystem of the EW-Shopp integrated platform for the Ceneje and Big Bang business case. The aim of the case is to develop data-driven services in the following key areas:

- Data-driven user information add-ons for real-time informing and individual purchase decision-making. The services target individual users across the web in a freemium model.
- Category management and marketing management optimization services that target retailers and manufacturers as a key target group.

The paper is structured as follows. In Section 2 the results of an analysis aimed at identifying and presenting to the user interesting information for purchases are shown whereas the category relevant data enrichment and presentation are discussed in Section 3. Finally, Section 4 concludes the document.

## 2 FOSTERING USER ENGAGEMENT IN E-COMMERCE (B2C)

In this section the output of a service that aims at creating additional value in the online search process. In a nutshell, some decision-relevant information that could speed up the purchase process and create a sense of urgency or satisfaction are collected, manipulated, and presented. The service design is shown in Figure 1.

So far, we have identified three types of categories:

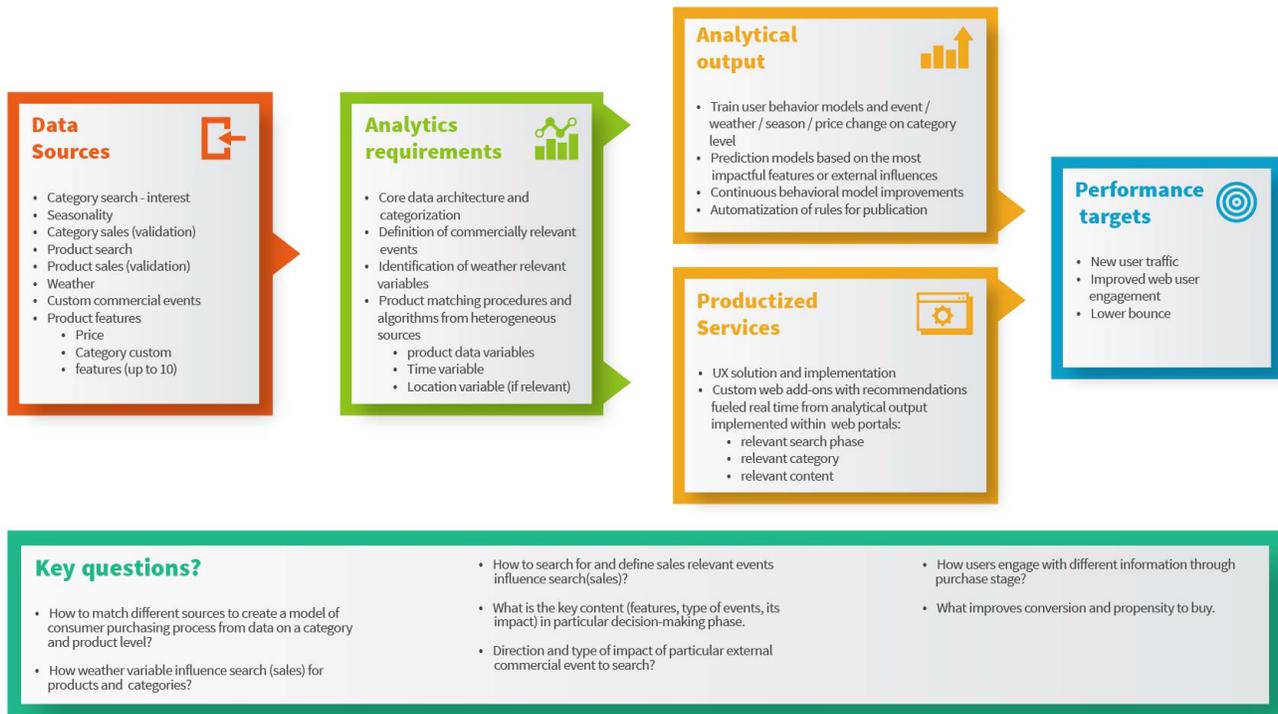
- sales in a category depending on weather,
- sales in a category depending on the season,
- no clear pattern is seen.

<sup>1</sup><https://www.ceneje.si/>

<sup>2</sup><https://www.bigbang.si/>

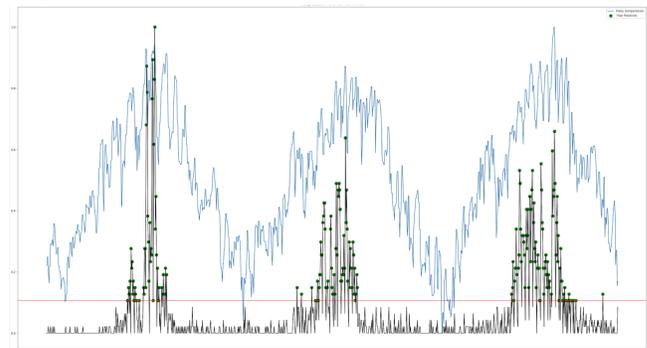
<sup>3</sup><http://www.ew-shopp.eu/>

**Business case 1 /**  
**Pilot I. Enrichment of purchase information for web platforms (B2C)**



**Figure 1: B2C data enrichment process**

As an example of decision-relevant information consider the following scenario: a series of hot days in the summer strikes and consequently the air conditioning device sales increase dramatically. Using weather forecasts, it is possible to predict a sudden spike in sales and inform the user that the stock of air conditioning (AC) appliances might decrease and that the prices will probably rise. To observe this relation, we have enriched historic data of AC appliance daily sales from Big Bang in years 2015-2017 with temperature data. Sales data comes from the online store as well as from physical stores from all over Slovenia. All weather data in our analyses were obtained from the Meteorological Archival and Retrieval System (MARS) repository maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF)<sup>4</sup>. We have aggregated the weather data over the entire country by averaging the measurements over the regions with the highest population density. The relation between daily temperature and AC sales is shown in Figure 2. Domain experts from Big Bang consider any day with more than 10 units sold as a *spike* in sales. This threshold

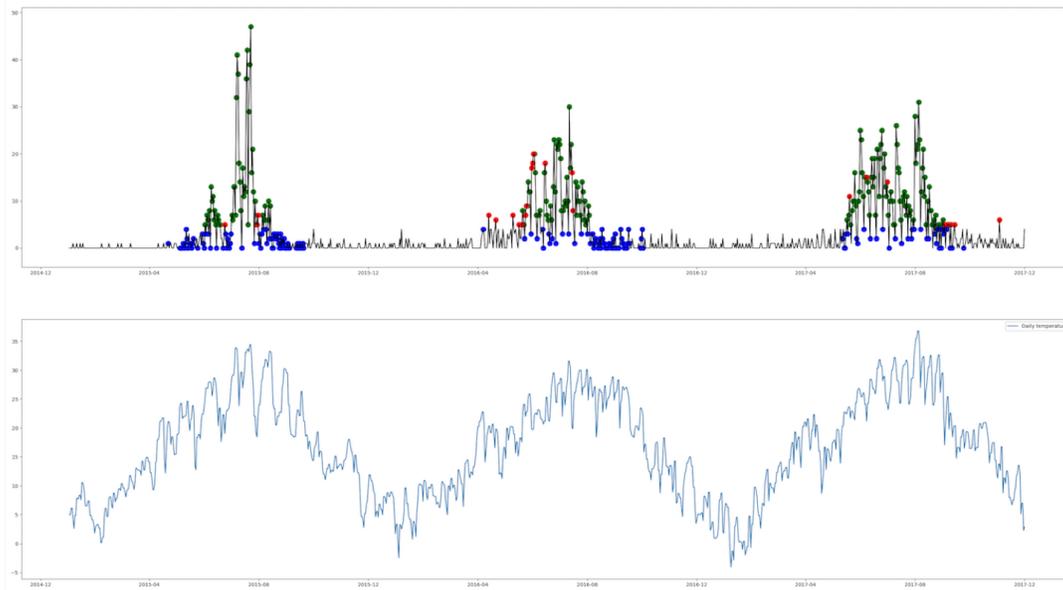


**Figure 2: Sellout in correlation with temperature**

is denoted in the plot by the red line and spikes are marked by the green dots.

The plot indicates there is a correlation between temperature and AC sales. We validated this by building a classification model that predicts whether a spike will occur on a day with some weather

<sup>4</sup><http://www.ecmwf.int>



**Figure 3: Prediction of spikes in correlation with temperature forecast**

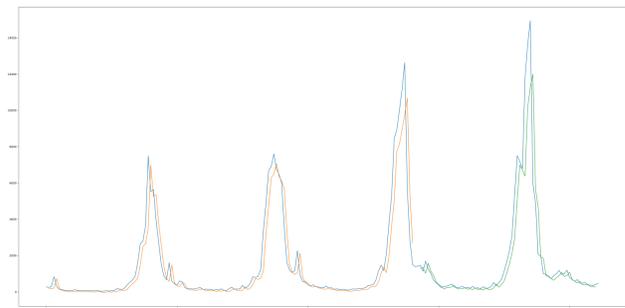
conditions. We extended the weather data with additional measurements such as precipitation volume, wind speed, relative humidity, cloud cover and surface solar radiation. Based on the distribution seen in Figure 2 we have taken into account only the time range from May to September and built an SVM model [2] with a linear kernel. By using actual historical data, the model achieves 81.6% precision and 68% recall whereas if we use weather forecast data for three days in advance, the performance drops to 78.7% precision and 54% recall. A visualization of prediction results using actual weather data is presented in Figure 3. Using this model, we can inform the user of an oncoming sales spike. The user, once informed, should show an increase in the inclination to purchase the product to avoid the sold out or just a delay in delivery.

Another representative example of a season dependent category is winter tires (in Slovenia winter tires are obligatory from 15. 11. until 15. 3. every year). In this case a Long Short-Term Memory neural network model [1] (particularly suitable for time series) has been applied for predicting future sales of tires – the prediction can be seen in Figure 4. Again, we can notify the user about coming spikes in sales providing an added-value service.

### 3 CATEGORY AND MARKETING OPTIMIZATION (B2B)

The Internet allows customers to make purchase decisions that are becoming more and more informed. For retailers and brand manufacturers, it is, therefore, vital to understand and monitor how and when customers go from getting informed about performing the purchase. It becomes of paramount importance to be able to define variables that trigger this momentum to optimize the business performance.

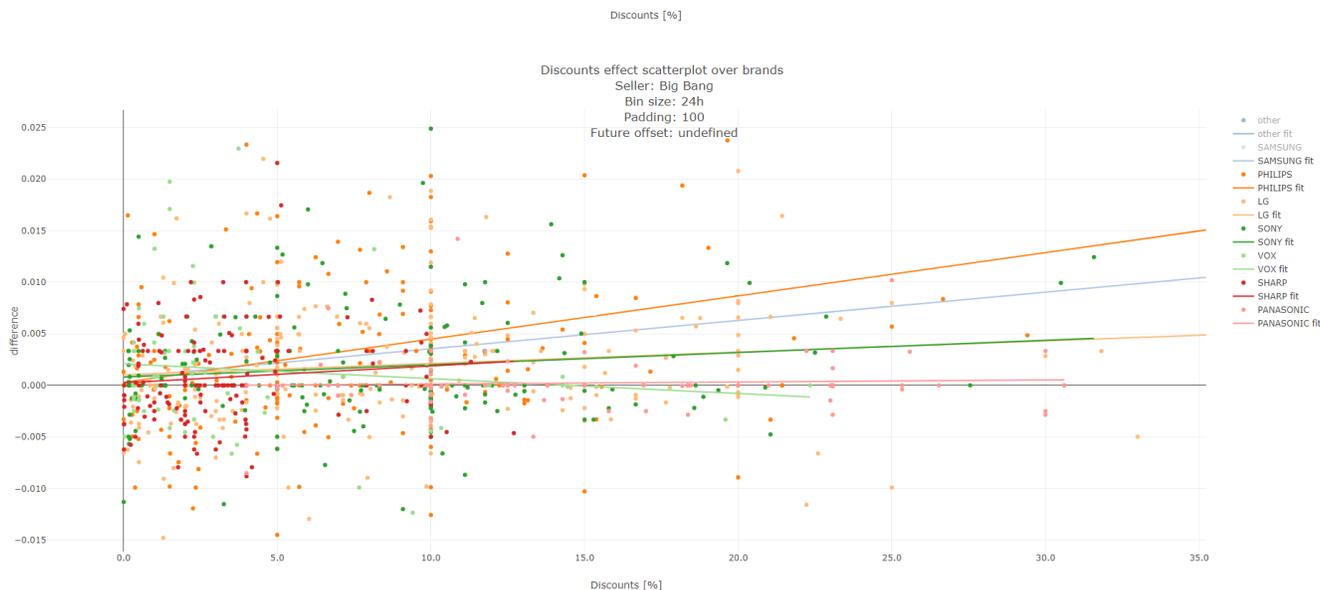
The core idea, here is that, by applying multivariate predictive models based on search data, retail and brand manufacturers could



**Figure 4: Winter tyres sellout and prediction**

improve efficiency and margin as well as shorten decisions cycle in category and procurement management. It would also offer new possibilities to optimize marketing management and increase efficiency. The same steps to predicting category sellout can be applied here as in the B2C case with a difference that a retailer has some additional means to increase sales: marketing activities and discounts. The predictive model must be able to handle such variables.

To measure the effect of discounts we have observed the change in average daily user interactions (i.e. clicks) on the Ceneje website. We have collected all discounts (i.e. changes of price from higher to lower) in years 2015 to 2017 in the category of television products and compared the average number of daily clicks in the time interval when the product had the higher price with the average daily interactions for the first seven days with the discounted price. This gives us a measure of the immediate markdown effect on consumer interest. Since some products have very little interactions (less than 10 per day on average), small changes there could lead to spurious values with little informative value. To ensure stable



**Figure 5: Effect on sales by applying discounts**

trends we have pad both average values with 100. The plot of these discount effects with their trend lines is shown in Figure 5. We can see that discount size positively correlates with the effect of some brands (PHILIPS) and negatively for others (VOX). Note that this does not mean necessarily that bigger discounts lead to lower sales, but it indicates that a particular brand generates more immediate interest if associated with smaller discounts as compared to larger ones. We'd like to predict high/low discount effects by modeling the effect of external factors such as weather. The modeling work on this topic is still ongoing. So far we can predict that an effect will be above threshold 0.005 with recall 0.623 and precision 0.342 by building a linear SVM model using weather information over the first seven days of the discount.

#### 4 CONCLUSIONS

In this paper, we have presented a proof of concept on how to join multiple players across the consumer user journey. Each player can benefit greatly by analyzing the joined data, whether by providing the user with some additional useful information with the aim of creating engagement or by gaining some valuable insights into category future sales and being able to plan marketing activities accordingly.

#### ACKNOWLEDGMENTS

The work in this paper is partly supported by the H2020 project EW-Shopp (Grant number: 732590).

#### REFERENCES

[1] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Comput.* 9, 8 (Nov. 1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

[2] Ingo Steinwart and Andreas Christmann. 2008. *Support Vector Machines* (1st ed.). Springer Publishing Company, Incorporated.