DATA MINING AND BIG DATA PROCESSING FOR MARKETING

Ciprian-Marcel POP [1], Eniko Elisabeta BODEA [2]

[1, 2] *Babes-Bolyai University*

Abstract
This papers approaches marketing through the eyes of informatics, how this mix (between marketing and informatics) does works out for consumers and marketers. We look in both direction and try to find out who is winning more: consumers or marketing specialists. We look specifically into concept like Data Mining and Big Data or algorithm in Marketing. This being the consequence of the fact that in our days data is stored in data bases and, in order to obtain information out of it, some tools and methods are needed, and we think that this two (data mining and big data) are used widely in marketing to target customers profiles and to discover and predict trends. We are looking to answer the questions: Who needs more informatics and marketing: Consumers or Marketing Specialists? Are we using informatics in marketing or marketing in informatics? Some twisted questions that challenged us.

1. Introduction

The technology is something we use frequently; it is also the time when technology and the online web technology are massively used. It is something we use at work, in our social life, when shopping and having fun at the same time. Digital means that customers are now connected all the time through their smartphones, tablets, smart TV, services and channels through different devices, and this represents valuable information to marketers. They need to take advantage of them and transform data into information.

---

[1] *Professor, Faculty of Economic Sciences and Business Administration, Cluj-Napoca, Romania, marcel.pop@econ.ubbcluj.ro*

[2] *Ph.D. Faculty of Economic Sciences and Business Administration, Cluj-Napoca, Romania, eniko.bodea@econ.ubbcluj.ro*

Marketing is such a complex field and it is developing even more with the help of IT (Information Technology) tools, and develops through its levers. We talk particularly about data mining and big data concepts. These concepts now (marketing, data mining and big data) are working as one and the central piece of them, is the consumer.

How is data mining and big data, influencing marketing and marketing decisions in particular? Well, it is complicated, because some of the facts we expose cannot be proved 100% because companies are not willing to share their intern marketing management and that is why we bring them up.

We argue that airline companies use algorithms when selling tickets that, depending on the information and number of searches (retain cache), can change the price of a ticket to the detriment of the customer. They say that they are not interested in this and that they modify prices in such a short time due to the high occupancy of the seats or to occupy the seats of a flight as soon as possible. Nevertheless, when we are stubborn to make a small test, as it will be seen in the following result is not on their side. On the other hand, another similar but different approach, is when the monitored product's price is "fluctuating" and the consumer does not know when it is the right time to buy it. If we are wise and use price tracking tools we are not going to be fooled by big sales when there is just big price and not what they advertise.

Data management is going to be a more sensitive topic, because starting with 26 May things are about to change due to GDPR (General Data Protection Regulation). It is an EU Regulation 2016/679 by which the European Parliament, the Council of the European Union, and the European Commission intend to strengthen and unify data protection for all individuals within the European Union (EU). It also addresses the export of personal data outside the EU. The GDPR aims primarily to give control back to citizens and residents over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU (ST, 2015).

We take a brief look at the literature and how the concepts are defined and used, then apply a little test by our own and present the conclusions of then and our opinion on the aspects that we decided to discuss.

2. Literature review

Data mining is the process of extracting hidden valuable information from the data in given data sets (A. Mushtaq, H. Kanth, p. 985 - 991). Big Data is as a

big deal for marketing. The big data category most familiar to marketing may include behavioral, attitudinal and transactional metrics from such sources as marketing campaigns, points of sale, websites, customer surveys, social media, online communities and loyalty programs (SAS, 2018).

Data mining is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data and data mining as the extraction of patterns or models from observed data (Berzal et. all, 2001).

Data mining is an essential process where intelligent methods are applied in order to extract data pattern (Han J., Kamber, M., 2007). Data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules (Berry, M. J., Linoff, G. S., 2008).

It is all about information and customers, and the fact that marketing is now using more and more informatics to be able to develop and find new ways of discovering and predict trends. For that, it is necessary to use information, and to rely to data. Because data is the main source to the answers that consumers and marketing specialists are looking for, tools are needed to help them find what they need. Data can be seen as the universe and the information needed as a star, if the tool, measurements and those who search for it, are not at the same page they cannot find what are they searching for.

Consumer analytics is at the epicenter of a Big Data revolution. Technology helps capture rich and plentiful data on consumer phenomena in real time. Thus, unprecedented volume, velocity, and variety of primary data, Big Data, are available from individual consumers (Sunil et all, p. 897-904).

Both data mining and big data, help marketing professional to improve their understandings about the behavior and actions that the customer is making and will make. Data mining technology allows learning more about their customers and making smart marketing decisions, Big data on the other hand affords the opportunity to dig deeper and deeper into the data, peeling back layers to reveal richer insights then used to make the perfect marketing decision. And if we take a look above we conclude that both are tools used to take the smartest and best marketing decisions that a marketer needs. It is all about turning information into data and then using it at its best.

For marketers when discussing online marketing information is about personal informations as well and not just name or surname, but IP- address, biometric data, cookies and web location for example. They all use them and search for them in order to create profiles, customer profiles and more. As well as by the

example that tested bellow in order to make decisions they use web location and cookie data and RFID tags to change prices for products.

Moreover, if the assumption is that all of this is beneficial for marketing department in firms and corporation, is there something beneficial for the customers? The benefits of using them are both for companies and customers or it goes in just one way? And we are about to find this out by using examples of tools used on the other hand by marketers and tool used by consumers and see where the balance scales.

3. How informatics influences Marketing decisions

To be able to do more, and by more we understand influence marketing decision, they use Analysis Algorithms. These algorithms are used to track prices online and lower or rise them according to what the code specifies. Analysis Algorithms used to track or adapt prices online raise competition concerns, according to a recent submission to the OECD (Organization for Economic Co-operation and Development) by the European Union. If pricing practices are illegal when implemented offline, they are very likely to be illegal when implemented online as well. Firms involved in illegal pricing practices cannot avoid liability on the grounds that their prices were determined by algorithms (OECD, 2017).

Semantic search is the process of searching in natural language terms and keywords that are the most used on a platform. Big data and machine learning make it easier for search engines to understand the user and to create a search profile based on what they access and search on the web, and smart marketers are using it to their advantage to improve the user experience for their visitors. However this is about to change when GDPR is going to be applied. GDPR stand for General Data Protection Regulation which was adopted on April 2016 and will come into effect beginning with 26 May 2018.

Why is that important, you may ask? Well, starting with May 2018, GDPR will affect virtually any company that either is based in Europe, or has any customers in Europe. Since ecommerce is making it easier than ever to sell to the world, that might be your store, too (GDPR, 2018).

GDPR affects companies of all sizes, not equally, there are some differences, for example, certain record-keeping requirements applies differently to companies with a larger number of employees, nevertheless, from one employee to thousand's, even if that company is not situated in Europe, if it

handles data about Europeans, then GDPR applies to that company. Some of the key features of GDPR that businesses in all sectors, that takes orders in online and process vast amounts of personal data, need to follow are (GDPR, 2018):

•   It specifically gives people the right to access, correct, delete, and restrict processing of their data, and sets out strict guidelines about how you need to get customers to agree that you can use their data (aka, consent). This is especially important if you're using your customers: data for purposes beyond simply filling orders, like for marketing or advertising.

•   Only collect data that you need. The heart of GDPR compliance is protecting people's data. You can limit your exposure by not collecting data that you don't need. Being transparent and implementing best practices, you won't face the massive fines that come with GDPR

•   Make everything really clear: the privacy policy of the store should be easy to access, the key is transparency.

•   Allow access to data: The GDPR also gives data subjects the right to access any information that's held on them. This means that organizations must store information in a way that allows them to access information quickly. They will also need to offer any data for download where possible.

For non-compliant organizations to GDPR rules, supervisory authorities will have the ability to fine organizations 20 million or 4% of their annual global turnover - whichever is greater.

Some of the internet biggest actors, Google and Facebook, have taken big steps in order to be GDPR compliance. And all the other firms/companies are now dealing with this problem and trying to find solution.

To better understand the impact of Big Data on various marketing activities and how do these algorithms work, we take the example of a person who searches to buy plane tickets. The information that caught our attention was, that if more searches are made and repeatedly, the entered information is the same, the ticket price will increase. Somehow, they advertise this by a message, which says something like: "Hurry up: The prices can change at every moment". So we wanted to test this and made the followings:

We had the following pattern: To fly from Cluj-Napoca to Paris - 2 Adults and one Infant (Children under the age of two). Period Departure 11.02.2018-Arrival 18.02.2018 in the first attempt and to fly from Cluj-Napoca to Paris - 2

Adults and one Infant (Children under the age of two). Period Departure 14.03.2018- Arrival 18.03.2018 in the second one.

This is how algorithms work, and just by feeding the web page with information, it can modify its content without the need of a person to do that. If, for example, we are in a hurry and in need to buy the tickets, at the second attempt well, we spend more money than if we had bought it for the first time. The company wins. If we are aware of the practices and how the informatics work we win and not the company.

Table 1. Tested patterns

First Attempt	Departure 11.02.2018 From Cluj-Napoca to Paris 702 RON	Arrival 18.02.2018 From Paris to Cluj-Napoca 522 RON	Under 3 searches
	Departure 11.02.2018 From Cluj-Napoca to Paris 1122 RON	Arrival 18.02.2018 From Paris to Cluj-Napoca 972 RON	After the fift search
	Departure 11.02.2018 From Cluj-Napoca to Paris 1292 RON	Arrival 18.02.2018 From Paris to Cluj-Napoca 972 RON	After 12 Hours and other 3 searches
Second Attempt	Departure 15.03.2018 From Cluj-Napoca to Paris 392 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 522 RON	First search
	Departure 15.03.2018 From Cluj-Napoca to Paris 412 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 522 RON	After the second search
	Departure 15.03.2018 From Cluj-Napoca to Paris 482 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 572 RON	After 10 searches
	Departure 15.03.2018 From Cluj-Napoca to Paris 522 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 572 RON	After 12 hours and other 2 searches
	Departure 15.03.2018 From Cluj-Napoca to Paris 442 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 482 RON	After a 2 days break of searching
	Departure 15.03.2018 From Cluj-Napoca to Paris 522 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 572 RON	After 4 searches in a row
	Departure 15.03.2018 From Cluj-Napoca to Paris 442 RON	Arrival 18.03.2018 From Paris to Cluj-Napoca 482 RON	After another 24 hours break of searching

Let us look at another type of mix, between marketing and informatics. Dynamic pricing or real-time pricing as others refer to it, demand pricing, or

time-based pricing is a pricing strategy in which businesses set flexible prices for products or service based on market demands at some point. This is the process or the algorithms that take into account competitor pricing, supply and demand, and other factors from the market, and adjust the price of the product according to the specified information. Dynamic pricing is a common practice in several industries such as travel, entertainment, and in online shopping. We are somehow aware of it, but each company has its own strategy and approach when reprising their products. Retailers, and in online in particular, adjust the price of their products according to traffic on the page, number of visitor accessing in the same time the product, unique visitors and multiple searches for a product, accessing and storing the product as favourite. The aim of dynamic pricing is to increase revenue and profit. There are different ways to do this. Some align the price of their product based on the price that the direct competitors have, some drop the price if the demand is low or increase price when the demand is high, some advertise the product as "promotion" even if some times this is not necessary true. In addition, we have some tools, programs that can help us as consumers to not fall into the trap of false promotions. Keepa, Pricy, Camel Camel Camel are some of these tools that track and display product history, showing us the price history of a certain product, to be more precise.

Keepa is a modern Amazon Price Tracker; elegant, efficient and easy to use (Keppa, 2018), the same goes with Pricy or Camel Camel Camel. This Chrome extension can be added and are meant to help you to take the best buying decision. The information that this adds offer are:sales rank, other best deals for the same product but in a different "shop", graphics that display price history for the last month, three months, six months or a year back for example.

The fast growing of online e-commerce led to the appearance of the price comparisons/tracker extensions which you can add them to your browser, where a customer can use them to filter and compare products based on price, features, reviews and other criteria. In many cases, these price comparisons can reduce the time of searching for a product, and give the best price for the product from different merchants leaving the decision to buy in the hands of the consumer. Some of these price comparison extensions are pricy.ro, camel camel camel, keepa and many other. We shall gave a short presentation of the three mentioned above.

Pricy.ro (Pricy, 2018) is a extension for chrome and firefox that can be used by consumers from Romania to make the best decision regarding the product they intend to purchase, and also a website in which there is a search engine that gives information about the products, including the history price. It monitors all major stores in Romania: Emag, Altex, Evomag, Pcgarage, F64, Flanco, Cel, Mediagalaxy and more than 200 stores, and finds the lowest price for the desired product and shows you the price evolution.

We used Pricy to track two products and the result was that:

•   Both of the products have a fake promotion; one had 50% of the price, the second 45%. But if we searched the history of them, we never found the product to be sold at its full price. In the first case we checked the history for the past 2 years. The product was never sold for the full price (299.99 Ron), even more the highest price to be sold at, was less than it. So not much of a deal if the product has a similar or even lower price.

•   Actually the heist price was on June 2017 and it was 259.99, so less than the full price they display now. Also the product was sold for less the 149.99 if we look to the chart meaning that if we are not in a hurry this is not the right time to buy it.

•   In the second case we have 45% of, but three months ago the price was even lower, meaning that again the real price is the one advertised as the promotional one.

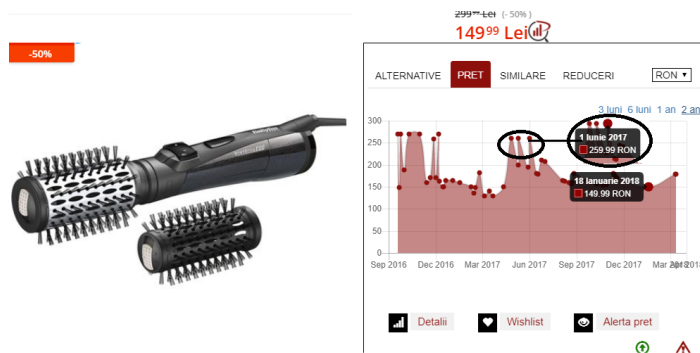**Figure 1. First product advertise (capture from an anonymized online shop)**

Figure 2. Figure 1. Second product advertise (capture from an anonymized online shop)

More known price trackers are CamelCamelCamel and Keepa (Camel, Keppa, 2018), both used for the world's biggest online retailer, AMAZON. They have Price Chart History, Price Drop Availability Alerts and other features that help customers to make the right buying decision.

Camel has a Watch Management Interface accessible at the top of every page that makes a list of prices for the products you track in a table format where you get the name and picture of the product with a small description; your desired price as well as the price at which you began tracking the product, and the current price. Once you install Keepa, price history graph will be displayed directly on each Amazon product page. Both have as a main advantage, the huge and constantly updated product database.

Using these tools, the consumer can be sure to buy the products he wants at the best price at the moment. The difference between the two Amazon tools are: for instance Keepa needs Sing Up, Camel does not, Keepa offers Notifications on Facebook, Twitter and Email, Camel just for the last 2, Camel has the possibility to import wishlist, Keepa does not and the applicabilty in different countries, Keepa (Brazil, Canada, China, France, Germany, India, Italy, Japan, Mexico, Spain, UK, US), CamelCamelCamel (Canada, China, France, Germany, Italy, Japan, Spain, UK, US).

4. Conclusion

Both, big data and data mining, are now a must for the Marketing field and, that is because without them the amount of information cannot be transformed and transposed into data. Without the right information, marketers cannot sell, and more important, cannot profile and predict trends. Along with the two concepts applied in marketing, we have also algorithms that we saw that are used in the first case for the benefits of the company, and in the second case in the benefit of consumers. Data mining and big data are now essential to marketers to predict trends, to analyze consumer behavior and to create sales profiles. Without them, the amount of time and effort would be massive, and could not keep up with the market that is in constant change.

We conclude that marketers advertise and promote products at their best, in the case of both flying tickets and online shopping as well, and that is the consumer's decision if the price is right or not.

It is a topic that in the near future will change massively due to the EU regulation and another similar test should reveal changes and improvement for consumers. Marketing department will invest more and be more careful with personal information and will be not able to use it as they desire. The practice of dynamic pricing should disappear as well in the online market as it is now in the offline one.

5.   References

- Aiman Mushtaq, Hina Kanth, Data Mining For Marketing, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Volume: 3 Issue: 3 Pages 985 – 991
- Berzal, Fernando, Cubero, Carlos, J., Nicolas, M., Serrano, et al. (2001). An Efficient Method for Association Rule Mining in Relational Databases. 47-64
- Berry, M. J., & Linoff, G. S. (2008). Data Mining Techniques For Marketing Sales and Customer Relationships Management (2nd ed.). Wiley Publishing,Inc
- Han, J., & Kamber, M. (2007). Data Mining Concepts and Techniques (2nd ed.). Morgan Kaufmann Publishers.
- Sunil Erevelles, Nobuyuki Fukawa, Linda Swayne, Big Data consumer analytics and the transformation of marketing, Journal of Business Research, Volume 69, Issue 2, February 2016, Pages 897-904
- https://www.sas.com/en_us/insights/big-data/big-data-marketing.html , Accesed at 2:08 AM on 04.02.2018
- Http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DAF/COMP/WD(2017)2&docLanguage=En , Accesed at 5:58 AM on 06.02.2018
- https://keepa.com/#! Accesed at 2:17 AM on 12.02.2018

- https://pricyro.wordpress.com/2014/02/17/tutorial-extensie-chrome/   Accesed at 2:17 AM on 12.02.2018
- https://camelcamelcamel.com/features Accesed at 2:47 AM on 12.03.2018
- https://keepa.com/#!features Accesed at 2:17 AM on 12.03.2018
- http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf Accesed at 5:58 AM on 06.02.2018