A Work Project, presented as part of the requirements for the Award of a Master Degree in Management from the NOVA – School of Business and Economics.

FIELD LAB - CUSTOMER RELATIONSHIP MANAGEMENT AT SPORT LISBOA E BENFICA:

EXAMINING THE POTENTIAL OF MACHINE LEARNING TO CUSTOMER RELATIONSHIP MANAGEMENT IN FOOTBALL

João André Calado Gouveia, 3522

A Project carried out on the Master in Management Program, under the supervision of:

Elizabete Cardoso

3rd January 2018

Abstract

This report studies the potential application of Machine Learning algorithms on SL Benfica, as part of the Customer Relationship Management Field Lab. For this purpose, topics regarding the concepts of Big Data and Machine Learning were addressed as well as the several applications of the former in CRM practices in conventional industries and more specifically, in the sports industry. Such research, along with a members' data analysis enabled general recommendations and identify specific applications of these algorithms on SL Benfica processes, aiming for the optimization and enhancement of the relationship with the fans and members.

Keywords: machine learning, Benfica, CRM, data

Table of Content

Introduction

This project was proposed and oriented by Prof. Elizabete Cardoso, in the context of the Customer Relationship Management Field Lab, and counted with the participation of six master students. Our client was SL Benfica, a Portuguese football club that plays in the top level of the Portuguese football league system. In fact, SL Benfica is considered to be one of top three football clubs in the country as well as the one with the highest number of members (around 200.000). Regarding the European context, the club ranked the 9th position in UEFA Ranking during the 2016/17 season, being the Portuguese club with the highest score.

The consultancy project consisted in three major topics: the internationalization of SL Benfica, understanding paid members' churn and supporters' opinions of Benfica's digital communication through qualitative interviews, and the analysis of the club's members data. Our goal with member data analysis was to identify the segments of members sharing similar behavior, to highlight the most profitable members for the club and, ultimately, to uncover the drivers that can lead members to churn. In addition, a spin-off stream of this topic was to infer the potential applications of Machine Learning on the club's activities to enhance and optimize its relationship with the members and supporters. As such, this report will start by examining existing literature in the field of Machine Learning, moving on to its applications in Customer Relationship Management and ending with a focus on the sports industry. The final section uses this knowledge and the information gathered through SL Benfica during our Field Lab to propose potential applications of Machine Learning for the club.

Literature Review

Big Data & Machine Learning

Nowadays, society is experiencing a shift in the way it works, due to the increasing evolution and application of Artificial Intelligence in the most diverse areas (Gazit, 2017). Indeed, AI is penetrating everything an individual consumes, creates and owns, to the extent that it has been considered as a new industrial revolution (Gazit, 2017). At the same time, the growing generation rate of data had led to a "data flood era" in which modern society lives in. Consequently, the term Big Data was introduced and, with it, the necessity to infer from such extensive datasets towards better decision making process (Patel, 2017).

Under this scenario, within the field of AI and due to the urgency of big data analysis, Machine Learning (ML) started to gain some relevance, since its algorithms could provide to software or machines the capability to improve its performance through the exposition and consequent learning from data (Tarwani, Saudagar, & Misalkar, 2015), aiming to identify complex patterns and to automate output, whether that is clustered data or predictions based on historical relationships (Patel, 2017). Moreover, the process of integrating ML models on datasets can be divided in three stages further explained: (1) data collection and preparation, (2) model building and (3) prediction (Patel, 2017).

Along with world digitalization, new data sources appear every day, as well as enormous quantities of (1) data collected from them (Kitchin, 2015). This increasing variety of sources extends from social networks sites and software applications until mobile devices and IoT devices. At this stage, it seems important to highlight that there are three types of data: structured, semi-structured and unstructured (Tarwani et al., 2015). As such, there must occur a data preprocessing phase in which the semi-structured and unstructured data are transformed into an understandable format, whose output will consist on the final training set used for the model building stage (Patel, 2017).

The (2) ML models can be divided into three categories: Supervised, Unsupervised and Reinforcement (Bishop, 2006). Regarding Supervised Learning, a training data set (whose inputs have pre-defined labels) is introduced to an algorithm that learns the pattern from input to output (label). It is commonly used where historical data is available and, in general, it suits for Classification (where the available algorithms are Neural Network, Support Vector Machine, Decision Trees, Naïve Bayes, etc) and Regression (either linear or non-linear) (Bishop, 2006). For the unsupervised learning, in contrast with the supervised one, there is no need for historical data since the algorithm's task is to map within the data set. So it includes Cluster Analysis, in which the observations are grouped based on similarity by algorithms that can be from Self-Organizing Maps until K-Means (Tarwani et al., 2015) (Exhibit 1). Finally, the reinforcement learning consists on just providing a hint and not the fully labeled data to the algorithm, which will automatically establish ideal behavior to achieve the maximum efficiency through trial and error (Bishop, 2006).

Once the model is built and tested, new and unseen data can be introduced in order to (3) get predictions (Patel, 2017). However, the success of ML techniques is intrinsically related with the quality of these predictions, which depends on several factors, such as the relevance of the input data to answer the problem and its reliability (Tarwani et al., 2015). These algorithms have been widely used in many fields, including supply chain, medical diagnosis and behavior analysis, being the former one essential for the way companies, and more specifically marketers, address and relate with their current and potential customers (Florez-Lopez & Ramon-Jeronimo, 2009).

Applications in Customer Relationship Management

Customer Relationship Management (from this point forward called CRM) is an strategic approach used by companies that aims to build, manage and strengthen loyal and long-lasting customer relationships (Vafeiadis, Diamantaras, Sarigiannidis, & Chatzisavvas, 2015) by learning more about their interests and behaviors (Farquad, Ravi, & Bapi Raju, 2014). In recent years, due to an increasingly competitive business environment and to the development of new methods to contact customers, there was a major shift in the way they behave, being more heterogeneous. Under this scenario, CRM programs have been gaining significant importance, given the fact that they allow companies to tailor their service or product offering to each customer's needs, in order to increase their satisfaction and to achieve higher profits and market share (Florez-Lopez & Ramon-Jeronimo, 2009). For this effect, the CRM strategy must be designed upon the foundation of customer information, which can be obtained through sales (purchasing history), marketing (campaign response), financial (payment history) and service data (Farquad et al., 2014).

But for that purpose, the company should begin by conducting an initial segmentation of clients and this is where ML can start to have an impact (Florez-Lopez & Ramon-Jeronimo, 2009). Basic cluster analysis combined with discriminant ones and logistic regressions have been the statistical techniques traditionally used in the development of segmentation models, but the existence of big datasets was a barrier to their effectiveness and robustness. Hence the importance of ML techniques, since they do not consider any *a priori* restriction about data and, in fact, based on behavioral trends, these algorithms are able to predict future customer decisions (Florez-Lopez & Ramon-Jeronimo, 2009). Moreover, decision trees (algorithms based on sets of *IF-THEN* rules) are considered to be one of the most widely used ML models for market segmentation and, given the fact that it is a supervised learning technique, takes into account historical data from customers (Dullaghan & Rozaki, 2017). Still regarding segmentation, there is a pressing need for companies to develop a dynamic customer management program in order to match with the constantly changing heterogeneous behavior from them (Florez-Lopez & Ramon-Jeronimo, 2009). For instance, some shopping malls are applying ML techniques to predict clients' preferences based on their real-time location and trajectory and, consequently, adapting the discount depth according to their proximity to the store (Spann, Molitor, & Daurer, 2016).

However, ML is not only interfering with the segmentation process, but it also has a role on the way companies interact and communicate with the customer. On one hand, by combining natural language processing and ML algorithms, firms are personalizing customer service by developing chatbots that are continuously learning from previous interactions (Wellers, Elliott, & Noga, 2017). On the other hand, ML algorithms (namely Artificial Neural Network and Bayesian Network) have been also supporting marketers in modelling consumer responses from direct marketing by identifying who are the ones most likely to answer as well as selecting the right channels (Cui, Wong, & Lui, 2006).

Another common application of ML in the CRM process occurs on the analysis of customer satisfaction. Indeed, with a Bayesian approach, it is possible for companies to model this aspect using unstructured data from customers' reviews, by transforming it beforehand into a semi-structured form and frequency counting the positive, negative and neutral sentiments (Dullaghan & Rozaki, 2017).

Furthermore, one of CRM main objectives lies on customer retention, since the cost of their retention is generally lower when compared to the cost of acquisition (Vafeiadis et al., 2015). As such, companies feel the urgency of developing models that are capable to predict the churn phenomenon or, in other words, predict which current customers are going to switch their "loyalty" to another service provider (Farquad et al., 2014). But to be able to model that behavior, companies need to understand what are the drivers for customer churn and, with this in mind, several supervised ML algorithms (such as Decision Trees, Support Vector Machines, Artificial Neural Networks, among others) have been used to predict which customers are more likely to switch between providers, based on data from previous relationships (Vafeiadis et al., 2015). Regarding the changing customer behavior towards the company, it is possible for the former to develop more tailored loyalty reward programs and, consequently, build stronger relationships (Farquad et al., 2014).

Companies constantly ought to answer one important question about CRM policy: who are the most valuable customers? With the increasingly capability of ML techniques to aggregate customers into clusters and predict future behavior, it is becoming easier for companies to compute with relatively accuracy the customer lifetime value and, consequently, measure the success of CRM (Florez-Lopez & Ramon-Jeronimo, 2009).

Applications in the Sports Industry

Along with the continuously growing application of ML techniques over several industries, the sports one is not an exception. However, even though there is vast and extensive research about the application of ML algorithms on clubs' sportive performance, there is a surprising lack of studies regarding their use on the areas of marketing and CRM towards the sports fans. Although CRM practices have been a concern for marketers for decades, the sports industry was a later adopter regarding its implementation, since their customers' loyalty levels are much higher when compared to the ones from conventional industries. In fact, this belief had led sports entities (namely clubs) to take their fans for granted and, as a consequence, ignoring their needs (Adamson & Tapp, 2005).

Like in any other industry, in order to enhance the relationship with the fans, sports clubs need to uncover the distinct segments that, beyond mere demographic characteristics, share similar psychological reasons driving motivation and attachment (Rohm, Milne, & McDonald, 2006). Once the fan data is gathered and prepared, clubs are able to run unsupervised ML algorithms (like Self-Organizing Maps or K-Means) to group the fans according to traits of their behavior (Tsiotsou & Alexandris, 2012).

Meanwhile, ML techniques have also been present in the churn prediction by sports clubs. Nevertheless, it seems important to highlight that "churn" in the sports industry does not literally follow the concept like in conventional industries, since for clubs, this phenomenon occurs when a supporter does not renew the season ticket or when a member stops paying his/her fees. In point of fact, the NBA's Portland Trail Blazers have been using a predictive model through Microsoft Azure Machine Learning platform (Exhibit 2) that is able to identify customers most likely to change the season ticket purchase plan as well as to compute a single game purchase propensity for each fan (Microsoft Ignite, 2017). With such analytics techniques, the Blazers better identified customers for targeted marketing campaigns, which led to a significant increase in terms of conversion rate (Microsoft Ignite, 2017).

Overall, clubs share the same goal as other conventional firms: find the most profitable customers and strengthen the relationship with them. ML algorithms can have a decisive role in this aspect, since they are able to predict with relative accuracy future behavior from the fans and ultimately, model fan lifetime value (Adamson & Tapp, 2005; Florez-Lopez & Ramon-Jeronimo, 2009).

Even though a lot can be said about Machine Learning in the context of Customer Relationship Management, and some research already examines applications in the sports industry, the next sections will propose specific applications in SL Benfica's strategy, hopefully contributing to further research in this area. On the other hand, the conjunction of both the reviewed knowledge and SL Benfica's information allows to turn academic insights into business recommendations likely to lead to enhancement and optimization of SL Benfica's customers management base.

Conclusions and Recommendations

Taking into consideration the research related with this field and all the insights from the Field Lab, it is possible to infer how and with which purpose SL Benfica can benefit from the application of ML in its CRM process. As such, this final section of the report will start by providing general recommendations that will allow the club to adopt and consolidate ML techniques on its database. Finally, it will focus on more specific applications that were drawn based on SL Benfica's data analysis conducted for the Field Lab.

Having this, there are three aspects that the club must consider prior to the application of ML: (1) quality and relevance of data, (2) capable human resources and (3) technology availability.

To have accurate results from the ML algorithms, it is essential to have (1) relevant data at its disposal. For instance, SL Benfica should collect information regarding its members' interests (such as favorite players and sports), which would be possible through a Facebook plug-in or, by other words, by allowing members and fans to login into *MyBenfica* platform with their Facebook account. Adding this semi-structured data to the ML models will provide more accurate results and predictions regarding future behavior both from the fans and members.

Furthermore, given the fact that ML algorithms are relatively complex and require high levels of programming, it is also important to have (2) capable human resources to deal with such techniques. And, although it is known that SL Benfica is hiring a data scientist to work on pilot projects in a period from 6 to 12 months, it is essential for the club to create a team dedicated to this field of business analytics.

Regarding the (3) technology factor, SL Benfica should leverage its partnership with Microsoft and integrate Microsoft Azure in the system, similarly to Portland Trail Blazers' case, previously mentioned in the Literature Review section. This solution would require three components from Microsoft Azure: *Azure SQL Data Warehouse*, *Azure Storage* and

Azure ML. The *Azure SQL Data Warehouse* would store several data from fans and members, while the *Azure Storage* would comply intermediary data (training and test datasets) that would be further used by *Azure ML* when running its algorithms.

Besides the general recommendations and through the members' data analysis conducted for the Field Lab, it was possible to generate more specific applications of ML algorithms and techniques, aiming for the optimization of the club's CRM practices and, ultimately, for the strengthening of its bond with the fans.

To begin with, it seems essential that SL Benfica becomes able to identify every fan that goes to the stadium to attend a match, which does not currently happen (unless the fans buy their tickets through *MyBenfica*'s platform). Once this information is gathered, the ML techniques can be applied to the computation of each fan's propensity to become a member, based on the frequency of matches previously attended.

Given the fact that the cost of acquiring a member is generally higher than the cost of retaining one, it is essential for SL Benfica to keep the churn rate as low as possible. As such, similarly to the applications that ML is having in conventional industries, its algorithms could be applied to members' data for the purpose of predicting which members are most likely to stop paying their fees. Once these results are obtained, it makes it possible to develop a more tailored loyalty program to retain these members.

On another hand, there is no denying that having a high number of season tickets sold is extremely important for sports clubs, since it decreases the volatility of stadium attendance throughout the sportive season and SL Benfica is not an exception. Having this, the club could exploit the ML algorithms to compute each fan's propensity to buy and renew the *RedPass* (name of SL Benfica's season ticket), taking into account his/her behavior during the past season.

In addition, since the club desires to increase the value of proposition for members and reinforce this relationship by being present in more important marks of their life, there is a pressing need to identify these life changing events. For instance, once SL Benfica is able to identify whether a member had become a parent, it could send a small gift to the newborn as well as a member proposal.

Until this point, all the previous recommendations apply supervised learning algorithms, since they consider historical data. However, unsupervised learning techniques can also be put in practice, such as Cluster Analysis (with K-Means or, more complex, Self-Organizing Maps), based on data from interactions like fees payment, merchandising and tickets purchases, among others. These algorithms can have an important role in the segmentation process of SL Benfica's fans and members.

It seems important to highlight that to achieve precise results with these recommendations and applications, SL Benfica needs to gather relevant and updated information about its fanbase, which might lead to the identification of triggers and, consequently, to the development of accurate predictions regarding future behavior.

Finally, the field of Machine Learning is limitless and is changing the way industries work and evolve. Although there is a lack of research regarding its application on the sports industry, this report will hopefully allow SL Benfica to optimize its CRM practices as well as contribute to further study in this area.

References

Adamson, G., & Tapp, A. (2005). From CRM to FRM: Applying CRM in the Football Industry. *Journal of Database Marketing & Customer Strategy Management*, *13*, 156–172.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. *Journal of Chemical Information and Modeling* (1st ed.). New York.

Cui, G., Wong, M., & Lui, H.-K. (2006). Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science*, *52*(4), 597–612.

Dullaghan, C., & Rozaki, E. (2017). Integration of Machine Learning Techniques to Evaluate Dynamic Customer Segmentation Analysis for Mobile Customers. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *7*(1), 13–24.

Farquad, M. A. H., Ravi, V., & Bapi Raju, S. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing Journal*, *19*(1), 31–40.

Florez-Lopez, R., & Ramon-Jeronimo, J. M. (2009). Marketing Segmentation Through Machine Learning Models An Approach Based on Customer Relationship Management and Customer Profitability Accounting. *Social Science Computer Review*, *27*(1), 96–117.

Gazit, L. (2017). Trading Cancer for Data: Machine Learning for Cancer Diagnosis. *Workforce Solutions Review*, *1*(1), 11–15.

Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, *31*(3), 471–481.

Microsoft Ignite. (2017). *How the Portland Trail Blazers use Personalization and Acxiom Data to Target Customers*. Retrieved from https://www.youtube.com/watch?v=xtTUlSUG8KM

Patel, S. (2017). Integrating Machine Learning Techniques for Big Data Analytics. *International Journal of Advanced Research in Computer Science*, *8*(5), 2760–2763.

Rohm, A. J., Milne, G. R., & McDonald, M. A. (2006). A Mixed-Method Approach for Developing Market Segmentation Typologies in the Sports Industry A Mixed-Method Approach for Developing Market Segmentation Typologies in the Sports Industry. *Sport Marketing Quarterly*, *15*(1), 29–39.

Spann, M., Molitor, D., & Daurer, S. (2016). Tell Me Where You Are and I'll Tell You What You Want: Using Location Data to Improve Marketing Decisions. *GfK Marketing Intelligence Review*, *8*(2), 30–37.

Tarwani, P. K. M., Saudagar, P. S. S., & Misalkar, P. H. D. (2015). Machine Learning in Big Data Analytics : An Overview. *International Journal of Advanced Research in Computer Science and Software Engineering*, *5*(4), 270–274.

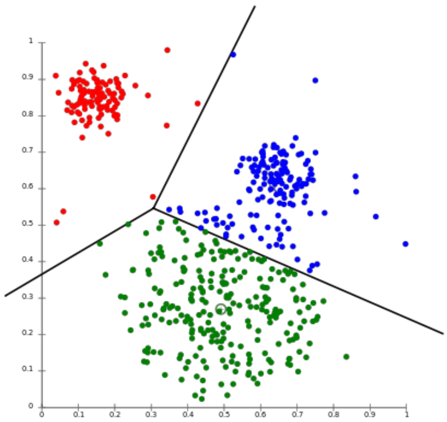
Tsiotsou, K., & Alexandris, R. (2012). Segmenting Soccer Spectators by Attachment Levels : A Psychographic Profile Based on Team Self- Expression and Involvement. *European Sport Management Quarterly*, *12*(1), 65–81.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*(1), 1–9.

Wellers, D., Elliott, T., & Noga, M. (2017). 8 Ways Machine Learning Is Improving Companies' Work Processes. *Harvard Business Review*, *1*(1), 2–6.

Appendices

Exhibit 1 – K-Means (on the left) and Self-Organizing Maps models (on the right)
Source: (Bishop, 2006)



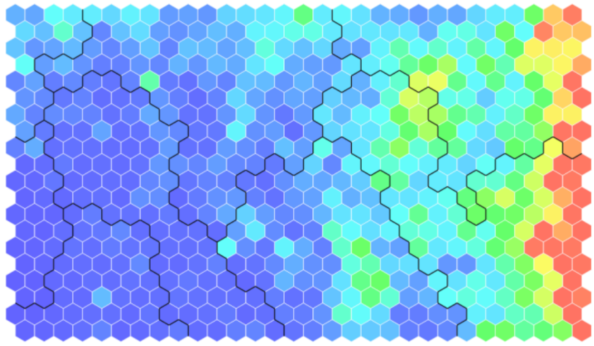


Exhibit 2 – Portland Trail Blazers workflow on Microsoft Azure platform.
Source: (Microsoft Ignite, 2017)