LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry*

Marcos Roberto Machado[1], Salma Karray[1]

***Abstract*— xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx.**

I. INTRODUCTION

In order to maximize profits, companies have invest their efforts in trying to retain customers, boosting customers satisfaction levels and as a consequence escalating their loyalty. The intensity of customers satisfaction is a key metric for a company's CRM (Customer Relationship Management) and as a consequence, loyalty can also measure and analyzed as well. Although customer satisfaction and loyalty are key mediators of profit, they cannot be taken as simple predictors of it. From a business standpoint, it is more important to identify and nurture relationships specifically with profitable customers (Kumar and Reinartz, 2012 [1]). In this context artificial intelligence can be use through the application of Machine Learning (ML) and Deep Learning (DL) models given the extraordinary amount of customer information to businesses.

ML and DL models can be apply within customer loyalty context from two different point of views. Firstly, it can be use to measure loyalty given customers information, and, secondly, it can be apply to improve the customer satisfaction by providing the right product or service in the right channel/place at the correct time. In both cases, ML/DL models would learn from previous customer interactions. So important it is be focus on customer satisfaction and loyalty that, according to Gerry Brown from IDC, 2017 [2], 65% of the marketing executives surveyed pointed out that "real-time personalized advertising insertions" and "optimized message targeting" will convey a compelling value by 2020. Also, a research published by MIT and Google said that 50% of businesses intend to apply ML for customers insights, and 48% expect it to earn a competitive advantage.

Some customer loyalty researches had applied ML techniques (Serkan, 2016 [3]; Ajay et al. 2019 [4] and Davies et al. 2018 [5]). More general, other business metrics such as customer lifetime value (CLV), customer retention or customer churn were also explored using ML models and different applications were found at the literature (Jamalian and Foukerdi, 2018 [6], Jing and Xing-hua, 2008 [7], Amin et al. 2017 [8] and Amin et al. 2018 [8]). Moreover, none of these applications studied customer loyalty for credit card customers using LightGBM model.

The aim of this study is to explore the application of LightGBM as a predictor for a customer loyalty score. The problem was motivated based in an online competition (Kaggle, [9]), where a credit card company is looking for a ML model that would make possible to predict a customer loyalty score for each single Card ID (for each single card issued). Company is also requesting a RMSE (Root Mean Square Error) to be presented along to the loyalty score predicted. These type of problem is getting more and more common in the financial industry, where, with more customer data it is more important to anticipate customers needs in order to retain them, increasing their loyalty. LightGBM as applied as the main ML method, mainly based on its capacity to uses GPU and uses less memory, the possibility to get high speed and handle large size of data. Also, because it present better accuracy than other decision tree gradient boosting models such as XBoosting (Ke et al., 2017 [10]) and by reason of its application was not found at the CRM/Loyalty literature.

II. LITERATURE REVIEW

A. Decision Tree Models

Falar um pouco de como decision tree models funcionam. Fazer uma revisao de literatura rasa.

B. Gradient Boosting Models

Fazer uma revisao de literatura rasa de modelos de grad boosting.

C. DTGB Models

Fazer uma revisao de literatura mais completa de como esses modelos funcionam, em especial, DT and GB.

1) LightGBM:

2) XBoosting:

III. METHODOLOGY

A. Datasets

Elo is one of the largest payment brands in South America. Elo results from a partnership of three of the largest banks in Brazil: Banco do Brasil, Bradesco and CAIXA. Elo offers credit, debit, and prepaid cards. At the end of 2018, Elo had issued more than 50 million cards. Also, Elo credit card supports installments payments, which allows a customer to arrange payments for a period of time.

*Submitted to Prof. Dra. Salma Karray as a report/article suggestion. This text was elaborated based on a ML grad course presentation performed at The Fields Institute in April 2019.

[1]University of Ontario Institute of Technology

Elo has built partnerships with merchants in order to offer promotions or discounts to cardholders. They have built machine learning models to understand the most important aspects and preferences in their customers lifecycle, from food to shopping. They also had built machine learning models to attribute a loyalty score for cluster of customers (per segment), however so far none of them is specifically tailored for an individual or profile (Kaggle, 2018 [9]). This is where the Kaggle competition and data made available comes in, the main objective is to develop algorithms to identify a customer loyalty score for each card identified at the datasets.

Different datasets were made available (Figure 1). The historical dataset has transaction information, features such as purchase date and amount, number of installments, city and state, also a merchant identification and a sub sector identification of the merchant where card was used. The merchant dataset has additional information with respect to the places where cards were used. Features such as merchant id group, average sales, average purchase, active month lags, city and state of the merchant location were available. New merchant has same type of information that historical data frame has, however this set of data contains material info of merchants visited for the first time. Train dataset contains the first month of activation, the target (loyalty score) and other features. All five data sets also have anonymous information/features.

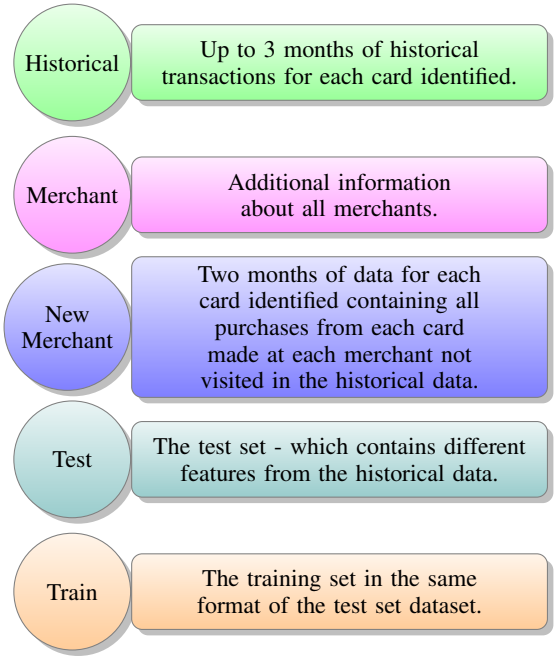

Fig. 1. Five different datasets were made available from ELO, a credit card company. Dataset are briefly described, each dataframe has different set of infromations. Data was published at kaggle.com in 2018 [9]

B. Data Preparation

Almost all ML algorithms cant not perform well with missing values. This was the very first thing treat at the Data Preprocessing Treatment (DPT), which is part of what we are calling Data Cleaning. At this stage, for all datasets missing values were replaced based on the following rule: if the number of missing observations were less than 5% of the total of instances for that feature, those observations were drop, in the other case, the missing value was replaced by the mode of that feature. In addition, an analysis over possible outliers was performed. It is important to highlight that, mainly for the target, 1.1% (2,207) observations were possible outliers, and given the relevance of the feature in the data, those values were marked, considering the inclusion of a dummy feature in order to identify those.

In a second stage of a DPT, categorical features were converted to numerical values in order to facilitate things later on at the pipeline process. Codes were attribute to different nominal observation and this standard was applied for all datasets. Then, feature engineering was performed. Firstly, all the datasets were unified/appended by the card identification, then more features were create based in the ones that the company made available. Feature transformers were applied in order to provide more data information to the models learn better and predict information with higher accuracy. At the fourth stage, a correlation analysis over the features was implemented. When correlation was higher than 90% features (new features created or previous ones) were drooped from the datasets. Finally, at the end of these workflow, a pipeline was implemented in order to select possible models and techniques to be use to train the data and do predictions. A workflow process is presented at Figure 2, where the bottom to top Data Preprocessing Treatment (DPT) performed in this study is presented.

C. Model Training

In order to perform predictions at the test set, the main parameters used at LightGBM training process were:

- Number of leaves: 31;
- Minimum amount of data in a leaf: 30;
- Objective: Regression;
- Number of threads: 4;
- Boosting: GBDT

Another important thing to mention with respect to the training process is that stratified random sampling method was used within this step. This is a method of sampling that was used to divide the data set that went through the DPT into smaller groups known as strata. In a stratified random sampling method, the strata are formed based on similar attributes or characteristics. Each new sample has the same size, in LightGBM it is necessary to define the number of strata, which in this study was define as 5 for all training performed. By using stratified random sampling method, it is possible to have a higher number of subsets to train the

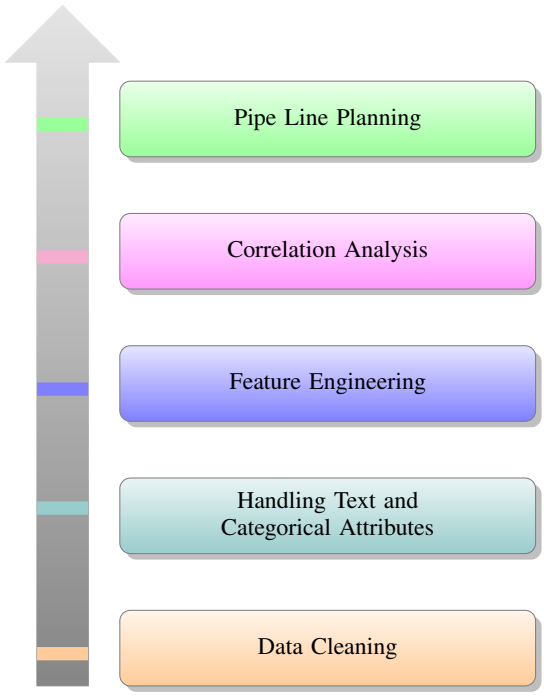

Fig. 2. Workflow from bottom to top of the Data Preprocessing Treatment (DPT) performed in the datasets in this study. All steps in this process was recommended by Geron, 2017 [11] and it is usually performed in an end-to-end ML modelling project.

model and by doing this give the model the chance to obtain a better performance, because in this case, each new subset is used once as a test set while the others are used as a training ones, until all of those be a test set once.

D. Evaluation Criteria

Falar sobre o RMSE e justificar o porque dele foi utilizado.

IV. RESULTS

This section aims to present how well LightGBM predict customer loyalty in the Financial sector, specifically for the case of Elo credit card company. RMSE is analyzed altering some of the parameters within the ML training algorithm such as learning rate and number of iterations, also a comparison between LightGBM and Xboosting is performed.

A. RMSE vs Learning Rate

Model was trained for different learning rates (0.001, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1) and the RMSE for the predicted loyalty score was calculated. Figure 3 presents these results. It is possible to verify that when learning rate is approaching zero RMSE is not one of the best results. That happens because given a poor learning rate, when the algorithm is looking for the best solution, a local minimum is found and it is understand to be the global minimum. On the other hand, when learning rate is increasing, after a certain point (around learning rate equals to 0.04), the RMSE is stable as a not optimal solution as well. This behavior is given by the reason of when the algorithm is looking for the best solution, its path is big enough to attribute big steps and passes by the global minimum, finding a local minimum as well. For this specific study, a learning rate of 0.01 is just right offering the lower RMSE and as well best loyalty score prediction.

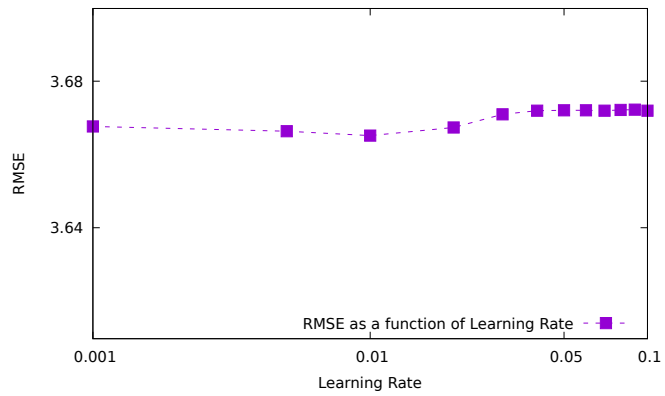

Fig. 3. LightGBM results: RMSE as a function of learning rate.

B. RMSE vs Number of Iterations

For a fixed learning rate of 0.01, the learning rate which provides a better RMSE as presented previously, different number of iterations was input during model training. Mainly because LightGBM has a Decision Tree model growing in its code, it is important to verify how the RMSE would behave for different number of iterations, mainly because in a decision tree model depending on the number of iterations the model will ended it up with a different tree and as a consequence a different prediction. In this setting, different number of iterations were considered and RMSE measured, results are presented in Figure 4.

Graphic in Figure 4 shows that for a small number of iterations the model does not have time enough to learn and find best solution and as a consequence a higher RMSE is observed. As the number of iterations increase, the RMSE decrease, up to a certain point where it does not matter how much more the number of iterations increase the RMSE will not present a significant change, it will just increase the processing time. Therefore, it is plausible to choose between 900-1000 iterations in this study and guarantee a the RMSE of the series.

C. Training Time

Another important thing to decide, from the point of view of a prediction model like the one explore in this study is based on a trade-off: time versus accuracy. The number of iterations was fixed at 1000 and output measures such as time and RMSE was calculated (Figure 5). It is possible to observe that for small learning rate time is between the highest in the interval and the RMSE is not optimal, and as the learning rate increase the processing time decrease, however the RMSE does not present a significant absolute difference in its value.

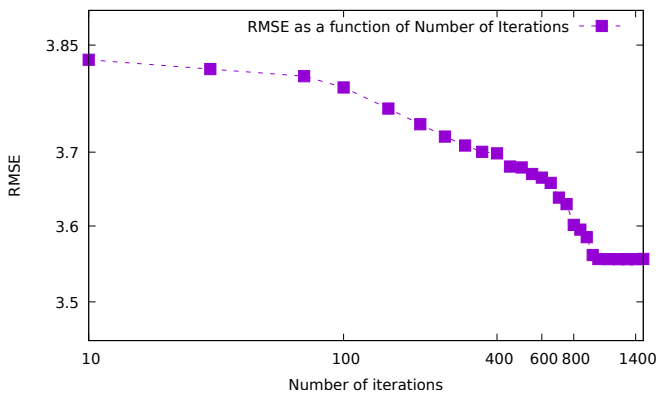

Fig. 4. LightGBM results: RMSE as a function of number of iterations.

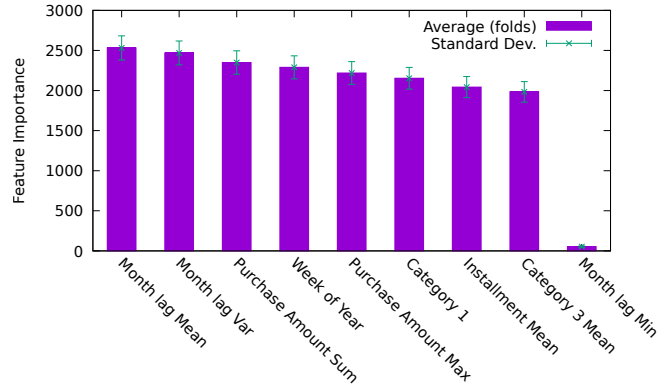

Fig. 6. LightGBM results: Feature Importance.

Therefore, depending on the application CRM will have to decide what is the primarily focus - time or accuracy.

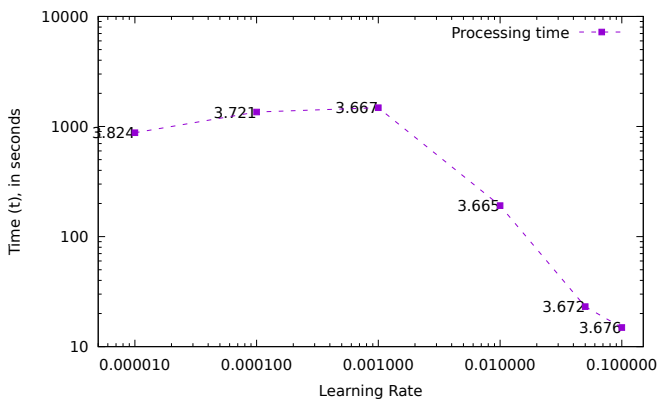

Fig. 5. LightGBM results: Time and RMSE as a function of learning rate. RMSE is presented as a label for each point at the dotted line.

D. Feature Importance

Before perform model training, we ended up with more than 230 features. Feature importance in predicting customer loyalty in this study is given through the LightGBM model. Importance of each feature is calculated based on how many times a feature was used to slip information (was used as a node) during the tree growing in the modelling process. Figure 6 presents the top 8 most important features and the least important to predict customer loyalty in this study, feature importance measure and standard deviation are presented. From approximately 200 features, only those are presented because they do not change when number of iterations or learning rate were altered. It is possible to observe that Month lag Mean, which is the average of months that a card identified was not used is the feature most used to split data in the decision tree process and the Month lag Minimum was the least one. Standard deviation slightly decrease as the feature importance decreases.

E. LightGBM vs XBoosting

LightGBM and XBoosting are based on Decision Tree and Gradient Boosting (DTGB) techniques. However, based on the Ke et al., 2017 [10] affirmation that statement that LightGBM performs better than Xboosting as a DTGB combined model, this section aims to test this hypothesis. Same set of parameters were set up for both scenarios, information such as learning rate and number of iterations were fixed and RMSE was measured for both cases and are presented in Figure 7. It is possible to verify that, for this study, LightGBM perform better than XBoosting for all interval of learning rate used. Therefore, it is plausible to affirm that LightGBM can also be used as a DTGB model in the financial sector to predict customer loyalty.

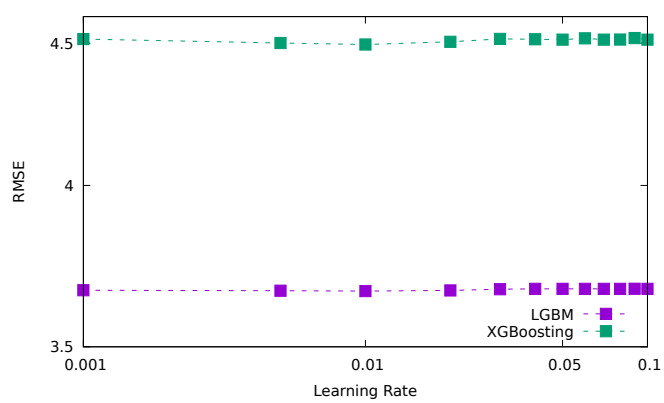

Fig. 7. LightGBM vs XBoosting: RMSE for customer loyalty prediction.

V. CONCLUSION

Olhar pra conclusao do artigo do hendrick pra ter uma ideia da ordem e do que escrever.

REFERENCES

[1] V. Kumar and W. Reinartz, *Customer Relationship Management: Concept, Strategy, and Tools*, 2nd ed. Berlim: Springer Berlin Heidelberg, 2012.

[2] b. G. B. IDC International Data Corporation, "Can machines be creative? how technology is transforming marketing personalization and relevance," 2019. [Online]. Available: https://www.criteo.com/digital-marketing-reports/can-machines-be-creative-how-technology-is-transforming-marketing-personalization-and-relevance/

[3] S. Varol, "Analyzing brand loyalty in automotive sector using the hidden markov model and support vector machine," Ph.D. dissertation, 2016, copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2017-09-13. [Online]. Available: http://search.proquest.com.uproxy.library.dc-uoit.ca/docview/1867750556?accountid=14694

[4] A. Aluri, B. S. Price, and N. H. McIntyre, "Using Machine Learning To Cocreate Value Through Dynamic Customer Engagement In A Brand Loyalty Program," *Journal of Hospitality & Tourism Research*, vol. 43, no. 1, pp. 78–100, 2019. [Online]. Available: https://doi.org/10.1177/1096348017753521

[5] A. Davies, M. A. Green, and A. D. Singleton, "Using machine learning to investigate self-medication purchasing in england via high street retailer loyalty card data," *PLoS ONE - Academic OneFile*, vol. 13, no. 11, p. p. e0207523, 2018. [Online]. Available: http://link.galegroup.com.uproxy

[6] E. Jamalian and R. Foukerdi, "A Hybrid Data Mining Method for Customer Churn Prediction," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2991–2997, 2018. [Online]. Available: www.etasr.com

[7] J. Zhao and X.-H. Dang, "Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example," *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, vol. 1, no. 3, pp. 1–4, 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4680698

[8] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, no. October 2017, pp. 1–12, 2018. [Online]. Available: https://doi.org/10.1016/j.jbusres.2018.03.003

[9] E. C. Cards, "Elo merchant category recommendation - help understand customer loyalty," 2019. [Online]. Available: https://www.kaggle.com/c/elo-merchant-category-recommendation

[10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3146–3154. [Online]. Available: http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf

[11] A. Géron, *Hands-On Machine Learning with Scikit-Learn*. O'Reilly, 2017.