Churn Prediction Analysis by Combining Machine Learning Algorithms and Best Features Exploration

Yasyn ELYUSUFI, M'hamed AIT KBIR
LIST Laboratory
STI Doctoral Studies Center
Abdelmalek Essaadi University, Morocco

***Abstract*—The market competition and the high cost of acquiring new customers have led financial organizations to focus more and more on effective customer retention strategies. Although the banking and financial sectors have low churn rates compared to other sectors, the impact on profitability related to losing a customer is comparatively high. Thereby, customer turnover management and analysis play an essential part for financial organizations in order to improve their long-term profitability. Recently, it appears that using machine learning to predict churning improves customer retention strategies. In this work, we discuss some specific machine learning models proposed in the literature that deal with this problem and compare them with some emerging models, based on Ensemble learning algorithms. As a result, we build a predictive churn approaches that look at the customer history data, check to see who is active after a certain time and then create models that identify stages where a customer can leave the concerned company service. Ensemble learning algorithms are also used to find relevant features in order to reduce their number which is of great importance when performing the training step with some classical models such us Multi-Layer Perception Neural networks. The proposed approaches can achieve up to 89% in accuracy when other research works, dealing with the same dataset, can achieve less than 86%.**

Keywords—Customer churn; prediction; machine learning

I. Introduction

Customer churn analysis deals with customer attrition rate in companies that offer some services. According to Forrester research statistics related to churn impact [1], it costs five times more to catch new customers compared to keeping the existing ones. As well, the Harvard Business School report states that on average, 5% increase in customer retention results in increase from 25% to 95% in profits. There are so many classic ways to keep customers from leaving [2]; by finding answers to the following issues:

- What are we, as a company, doing to cause customer turnover?
- What are our customers doing that is contributing to their leaving?
- How can we better manage our customer relationships to make sure it does not happen?

On the other hand, predictive analytics uses prediction models that approximate the risk of customer attrition and the degree of their dissatisfaction with regard to services offered by the organization. In addition, new technologies have increased bank access to customer data, which has made customer attrition analysis increasingly easy and accessible. In fact, predicting customer attrition can help banks to plan suitable marketing campaigns to convince clients who are potentially candidates for leaving [3].

Overall, actually demand for customer attrition analysis is increasing, the study of the characteristics related to the customers profile and behavior, by consulting their transactions history, remains the most widely used approach in research works related to this domain, most of these are statistical learning methods. This takes up the following question, which learning model can best predict customer churn? By taking a look at methods used in the literature, we can find that popular methods to predict churn likelihood are logistic regression (LR) [4], K-Nearest Neighbor (KNN) [5], decision trees (DT) [4] and SVM [6].

In this work, we seek to use and evaluate performances of new methods that can reach best results, compared to those cited above, at predicting customer churn which can give suitable responses to the following questions:

- How successful are the cited Machine learning methods in predicting customer churn?, answering this question will allow us to know which methods give the most reliable predictions based on many metrics;
- How are the relevant features according to some machine learning algorithms ?, answering this question allows us to pick up the most relevant features so as to use the training samples with reduced size vectors, which allows time saving in both training and generalization steps. Relevant features can be combined later with high accuracy machine learning algorithms to enhance performances.

However, methods cited above have reached their limits and have practical difficulties resulting in the emergence of a new generation of algorithms called Ensemble learning algorithms: Random Forest (RF), eXtreme Gradient Boosting (XGBoot) and Light Gradient Boosted Machine (LightGBM) models [7]. Light GBM model has many of XGBoost's assets, such as sparse optimization, parallel processing, regularization and bagging [8]. These algorithms belong to two families of ensemble decision tree models, Bagging and Boosting. We are also interested in this study by artificial neural networks

(ANN) based models, especially Multi-Layers neural network [22].

In this study, we are also interested by improving performances of machine learning models, related to predicting customer churn, by exploring ensemble learning algorithms capabilities to find relevant features. Indeed, the use of relevant features with some classical machine learning models is important when carrying out the training step.

This paper is a part of a series of research carried out by a team of our laboratory, interested in users profiling and related predictive analyzes [9] [10] [11]. The work is presented as follows; in the first section we present literature review of churn prediction works using machine learning models. In the second section we present data pre-processing, modeling, and comparison with some cited works. In the third section, we present our approach of best feature exploration in order to improve the quality of prediction using Multi-Layer Perceptron model. The results interpretation and conclusion comes in the last section.

II. Related Work

Today the industry is in competition with a limited number of potential customers because of the increasing saturation of the market [12]. Customers seek value-added relationships with their suppliers in order to stay loyal [13]. Companies are therefore looking for strategies to engage customers in their process in order to have a tangible idea of the required benefit. In order to improve this process, these companies are seeking to recognize customer behaviors that indicate a decline in his relationship with the company; this is established as customer relationship management (CRM). It's used by companies to efficiently assign their resources in order to maintain and improve customer relations [14], it's also used to lock onto customers churn rate. Churning specify the decrease in the consumption of a service provided by the company or the termination of that service, it can also be defined as an un-subscription. Several models have been proposed for customer behavior prediction in order to choose the auspicious time when the human resources department must give them more attention, after which incentive strategies should be followed in order to preserve this relationship and make it as profitable as possible. Many methods have been presented in the literature to determine the churn rate, they differ depending on the contractual and non-contractual nature of the work, the corresponding data structure availability and the amount of accessible transaction history. The following paragraph will be devoted to the presentation of research works related to our problematic and the promising machine learning models they use.

The results of some churn prediction models can vary on a wide spectrum; in this specific context a study of measuring the predictive accuracy of customer churn models was presented in [15]. In the same way, Breiman made the foundation for churn analysis based on classification and regression trees [16]. Upon this work, the cited author has also built additional methods by using bootstrap aggregators or bagging methods [8]. The bagging used proceeds by bootstrapping replicates of the training set, followed by creating the aggregation of predictors. In this use case Random forests introduced in [17] builds upon the previous model by introducing a new layer obtained by randomizing the bagging. In fact, a random subset of descriptors is used to expand trees, each one use a sample of the training set. It must be mentioned that RF method is known by its sensibility to data with unbalanced number of samples, which is typical to customer churn data sets, as the percentage of customers who churn is small or relatively unknown. In the same context, Xie proposes a new learning approach, called improved balanced random forests to produce better prediction results. In the present work we will assess the impact of using data re-sampling and scaling (robust-scaling method) before applying machine learning models. The following sections will be devoted to the preprocessing, modeling, evaluation, and comparison of machine learning model performances. These methods are used to help financial organizations to make decisions about the suitable strategy to prevent the customer from leaving the financial organization.

III. Our Work

A. Goal of the Work

The goal of this work is to explore whether bank's customers are about to leave the organization or not. In order to make an efficient model to predict customer churn we will use machine learning models. The event that defines the customer abandonment is the closing of the customer's bank account Fig.1

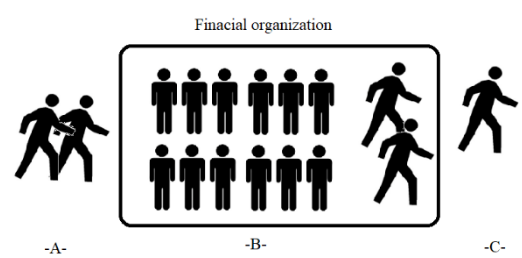


Fig. 1. Customer Churn Issue.

A-segment customers can need marketing campaigns to let them join the financial organization, when the B-segment customers can need to use machine learning models in order to predict who is about to leave and finally we can analyze C-segment customers so that more importance can be given to the suspicious profile thereby gaining more knowledge about how to make them more loyal.

B. Dataset

The dataset under study is called "Bank Customer" [18], the data was collected from an anonymous organization with customers in France, Spain and Germany. This public dataset consists of 10000 observations and 12 features, containing customer's information, where the "Exited" feature refers to customer abandonment status (target class). In order to find the churning client and to help making decision about the precaution that the financial organization should undertake we decided to study several machine-learning models, evaluate their performances and combine their decisions. The following table shows the used features, see Table I.

TABLE I. USED FEATURES

Feature	Signification
Surname	The surname of the customer
CreditScore	The credit score of the customer
Geography	The country of the customer(Germany/France/Spain)
Age	The age of the customer
Tenure	The customer's number of years in the bank
Balance	The customer's account balance
NumOfProducts	The number of bank products that the customer uses
HasCrCard	Does the customer has a card? (0=No,1=Yes)
IsActiveMember	Does the customer has an active mebership (0=No,1=Yes)
EstimatedSalary	The estimated salary of the customer
Exited	Churned or not? (0=No,1=Yes)

IV. DATA PRE-PROCESSING

Before the pre-processing, re-sampling and data scaling, we can analyze data relevance according to each feature and each machine learning algorithm. In fact, we decided to reveal the relevant feature by analyzing the columns one by one, looking for their dependency towards the churn. The following features are removed, because they do not have impact in the final decision:

- Row Number: It corresponds to the record row number
- Customer Id: It contains random values
- Surname: It represents the surname of a customer

Concerning the other features, we discuss their impact on customer churn by briefly analyzing the data.

- Credit Score: People with a credit score between 680 and 689, about 342 customers, are more loyal than others and less likely to leave the organization.
- Age: As shown in Fig. 2; it's certainly clear that the age feature is relevant, since customers with age between 35 and 55 are more likely to leave their organization than customers with other ages.

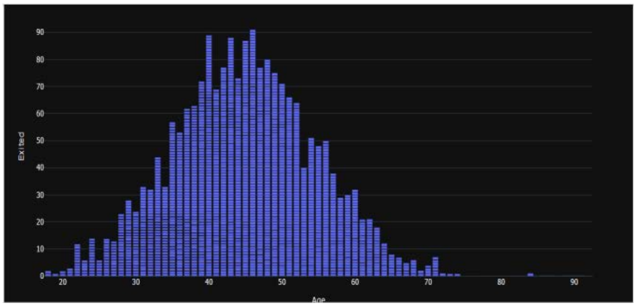


Fig. 2. Age Distribution.

- Tenure: It mentions the number of years spent by the customer with the organization. Normally, people with Tenure=7 are more loyal, about 851 customers.
- Balance: It's an important indicator, as people with a balance between 122k and 127k are more loyal to the organization.
- Number of Products: it indicates the number of products that a customer has purchased through the organization, people with Number of Products >=2 are more loyal to the organization, about 4242 customer.
- Has Credit Card: It indicates the fact that a customer has a credit card or not. People possessing a credit card are less predisposed to leave the organization, 30.09% of customers that churned have not a card, about 613 customer.
- Is Active Member: It indicates the presence of transactions over the customer account, active customers are more predisposed to stay with bank and 63.9% of customers that churned are not active.
- Estimated Salary: It indicates the estimated salary of the customer. The costumers with salary between 175k and 185k are more likely to leave the organization, about 122 costumers, while the more loyal costumer with salary between 77.5k and 82.5k, about 222 costumers.
- Gender: As shown in Fig. 3, we can infer that gender feature plays an essential role in churn prediction.
- Geography: It indicates customer's location. Being close to the bank can affect customer decision, especially with home changing.

In order to prepare data, we will apply the one hot encoding process for Geography feature to allow a more expressive categorical data, as shown in Table II.

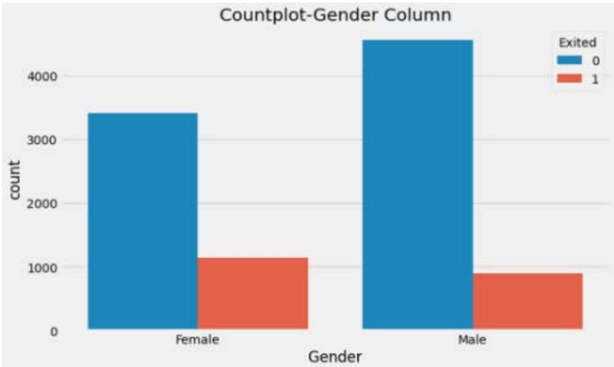


Fig. 3. Gender Distribution.

TABLE II. GEOGRAPHY FEATURE ONE HOTE ENCODING

Geography_France	Geography_Germany	Geography_Spain
1	0	0
0	0	1
1	0	0
1	0	0
0	0	1

According to Geography feature analysis, the histogram shown in Fig. 4 can infer that the total number of customers who exited is highest from Germany, which means that the bank needs to focus more on those customers followed by France customers and finally Spain customers.

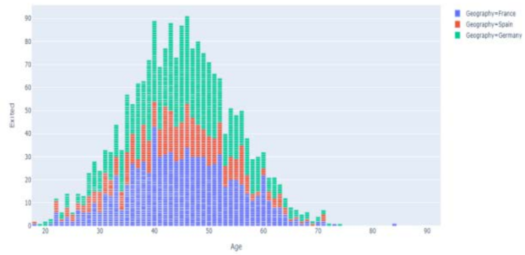


Fig. 4. Age Histogram Depending on Customer Location.

The Age feature is a numerical continuous number. According to Fig. 4, the age range between 35 and 55 are likely to leave. Furthermore, according to correlation study Fig. 5, the Age feature is correlated with 0.29 with the target and certainly will be a relevant variable for prediction.

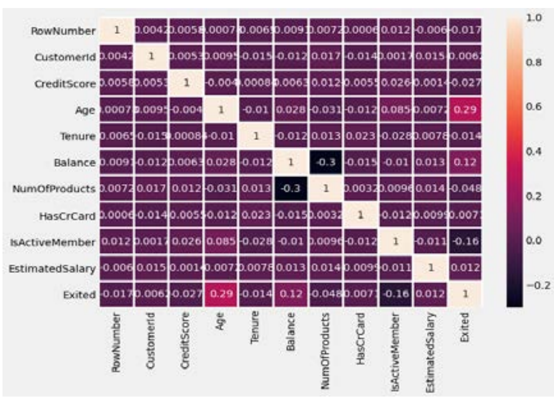
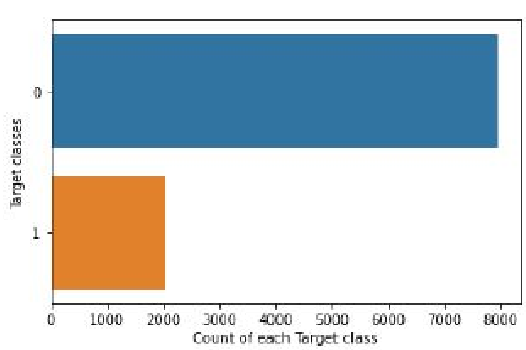


Fig. 5. Heatmap (Correlation with Target).

V. Re-Sampling And Modelling

A. Re-Sampling

In many works, the re-sampling technique is used to deal with unbalanced datasets [19] [20]. It is based on removing under-sampling respectively adding over-sampling samples from the majority in respective of the minority class. As shown in Fig. 6, we can check easily that the distribution of the target feature is unbalanced.



.

Fig. 6. Count Plots of Target Feature "Exited".

We decide to apply the over-sampling for increasing the number of samples as illustrated in Fig.7.

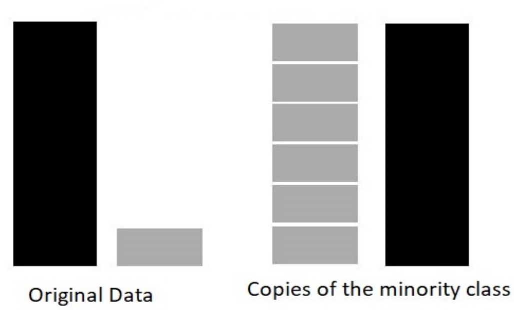


Fig. 7. Re-Sampling Method (Over-Sampling).

B. Evaluation Metrics

In order to assess machine-learning capabilities, the evaluation is based on many metrics. The most used ones are: accuracy, sensitivity, specificity, Matthew's Correlation Coefficient (MCC), that are based on confusion matrix information, presented by the following Table III, and Cohen's kappa and Matthew's correlation coefficient based on agreement and disagreement probabilities.

TABLE III. Confusion Matrix

		Actual Values	
		Success(1)	***Failure (0)***
Predicted Values	Success(1)	Truepositive	False negative
	Failure(0)	False positive	Truenegative

1) Accuracy: The accuracy calculates how many correct results your model managed to identify.

$$\text{Accuracy} = \frac{True\ positive + True\ negative}{Total\ number\ of\ samples} \quad (1)$$

2) Sensitivity: The sensitivity represents the fraction of relevant results that were retrieved; it uses also the information provided by the confusion matrix.

$$\text{Sensitivity} = \frac{number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives} \quad (2)$$

3) Specifity: The specificity also called true negative rate, measures how well a model can identify true negatives.

$$\text{Specificity} = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives} \quad (3)$$

4) Cohen's kappa: The Cohen's Kappa is a statistic measure, commonly referred to as inter-rater reliability. This metric reflects the reliability of two raters, who are rating the target result and the real one, and identifies how frequently the raters agree. Cohen's kappa, symbolized by letter $\boldsymbol{\kappa}$, this later range between -1 to 1, and it depends on the probability of agreement minus the probability of disagreement. When $\boldsymbol{\kappa}$=0 the amount of agreement has a random value, and κ=1 represents perfect agreement.

$$\% \; of \; agreement = 100 \times \kappa^2 \quad (4)$$

5) Matthew's correlation coefficient (MCC): Matthew's correlation coefficient (MCC), is a metric that aims to evaluate the quality of binary classification. MCC is often used when dataset are unbalanced, and it's unimpressed by disproportions related to dependent variables. MCC metric takes values between -1 and 1. When MCC score is equal to -1, this reflects an unexceptionable misclassification, however when it's equal to 1, a perfect classification is detected, while zero value indicates that the model is no better than random values.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \quad (5)$$

C. Modeling

After the application of the process of over sampling using SMOTE library of python, we count 15926 samples while the original dataset count only 10000 observations. After, we need to standardize the data using a scaling stage. We used Robust-Scaling method [21], computed by subtracting the median from the feature value and dividing by the inter-quartile range (75% value - 25% value). This step is followed by dataset splitting, into testing and training sub-sets, the training subset count 90% of dataset. In our work, we will assess the impact of seven supervised machine-learning algorithms: Logistic Regression (LR) [4], K-nearest neighbors(KNN) [5], Random Forest (RF) [22], Decision Tree [4], Support Vector Machine (SVM) [6], XGB and Light GBM [7].

In order to choose the best model, for XGB and LightGBM, the following empirical hyper-parameter setting has been adopted see Table IV:

TABLE IV. HYPER PARAMETER USED FOR XGB AND LIGHTGBM MODELS

Learning Rate	0.05
Max Dept	5
Estimators	500

The Multi-Layer Perceptron (MLP) was configured so that the input layer has a number of neurons identical to the number of features. We proceed with two hidden layers; 15 neurons are used in the first hidden layer, 10 neurons in the second hidden layer. The stochastic gradient descent (SGD) was used as the solver function, with learning rate fixed to 0.05. Finally, the model will be trained for 500 epochs.

Parameters discussed in the precedent two paragraphs are used in all our experiments.

The evaluation process of each model, before and after over-sampling process, gives metric scores indicated respectively in Table IV and Table V.

After re-sampling and scaling, the accuracy score has been enhanced for all the models except for LR, this can be justified by the fact that this technique is susceptible to over-fitting and assumes linear relationship between independent attributes.

TABLE V. EVALUATION METRIC WITH ROBUST SCALING NORMALIZATION

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.803	0.213270	0.960710	0.227233	0.267877
KNN	0.830	0.402844	0.944233	0.404683	0.422475
CART	0.793	0.492891	0.873257	0.370652	0.370736
RF	0.849	0.436019	0.959442	0.465812	0.489577
SVM	0.847	0.345972	0.980989	0.415727	0.470908
XGB	0.851	0.478673	0.950570	0.489586	0.504743
LightGBM	0.857	0.507109	0.950570	0.515891	0.528854
MLP	0.847	0.454976	0.951838	0.469582	0.487264

TABLE VI. EVALUATION METRIC WITH ROBUST SCALING NORMALIZATION AND (OVER-SAMPLING)

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.824231	0.794155	0.853598	0.227233	0.649173
KNN	0.844319	0.841169	0.847395	0.404683	0.688586
CART	0.827997	0.858958	0.797767	0.370652	0.657615
RF	0.890772	0.885642	0.895782	0.465812	0.781516
SVM	0.861896	0.822109	0.900744	0.415727	0.725570
XGB	0.888889	0.872935	0.904467	0.489586	0.777994
LightGBM	0.889517	0.874206	0.904467	0.515891	0.779227
MLP	0.863779	0.834816	0.892060	0.469582	0.728408

These results show that ensemble decision tree models and Multi-Layer neural network produce higher accuracy when dealing with the "Bank Customer" dataset. However, finding best parameters for these methods is not a small task. In fact, it is time consuming with a high number of features. Consequently, using just relevant features to present dataset samples can be helpful when setting the appropriate model parameters and can affect the classification results.

The results in table VII were obtained by research works operating on the dataset described in this paper.

TABLE VII. RESULTS OF EVALUATION METRIC USING LR, RF, AND KNN MODELS PUBLISHED IN [5]

Name	Accuracy	Sensitivity	Specificity	Cohen's Kappa	MCC
LR	0.811	0.964	0.211	0.231	0.273
RF	0.866	0.969	0.465	0.513	0.539
KNN	0.836	0.962	0.342	0.376	0.409

Results presented in this table compared to those obtained in table VI show that the accuracy was improved for LR, RF and KNN models respectively by 1.3%, 2.4% and 1.4%. On the other hand, the accuracy given by the XGB and Light GMB models was respectively increased by 2.2%, 2.3% with respect to the highest accuracy result obtained in able VII. In the following section, we will proceed to feature importance analysis, using Scikit-Learn library [23], according to RF, XGB and Light GBM, models with the best obtained accuracy score. This analysis aims to enhance the quality of prediction and to make the training phase easier for some machine learning models such as Multi-Layer neural network.

VI. FEATURE RELEVANCE EXPLORATION

Feature relevance represents the reduction in node impurity weighted by the probability of reaching that node. The number of observations cumulated by the node divided by the total number of observations corresponds to the node probability. Higher values correspond to features that are more relevant. For each decision tree, nodes relevance is calculated, by taking only two child nodes (binary tree):

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} c_{right(j)} \quad (6)$$

ni_j: Importance of the node j

w_j: Proportion of observations that get at the node j

C_j: Impurity of node j

left (j) and right(j): Child node resulting from splitting the node j.

The relevance of each feature is then calculated by:

$$fi_i = \frac{\sum_{j: node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \epsilon\ all\ nodes} ni_k} \quad (7)$$

fi_i: The relevance of feature i

This term can be normalized by dividing by the sum of all feature relevance values:

$$\mathrm{normfi_j} = \frac{\mathrm{fi_j}}{\sum_{\mathrm{j} \in \mathrm{all\ features}} \mathrm{fi_j}} \quad (8)$$

The final feature relevance, at the Random Forest grade, is defined by the mean of all the trees. The sum of the feature's relevance value on each tree is calculated and divided by the total number of trees:

$$\mathrm{RFfi_i} = \frac{\sum_{\mathrm{j} \in \mathrm{all\ trees}} \mathrm{norm\ fi_{ij}}}{\mathrm{T}} \quad (9)$$

$\mathrm{RFfi_i}$: The relevance of feature is calculated from all trees in the RF model.

Norm ($\mathrm{RFfi_i}$): The normalized feature relevance for feature in the tree j.

T: Total number of trees.

The application of feature relevance calculation according to the three models with the best accuracy score > 88.8 % (RF, XGB and Light GBM) are shown in Fig. 8, Fig. 9, and Fig. 10.

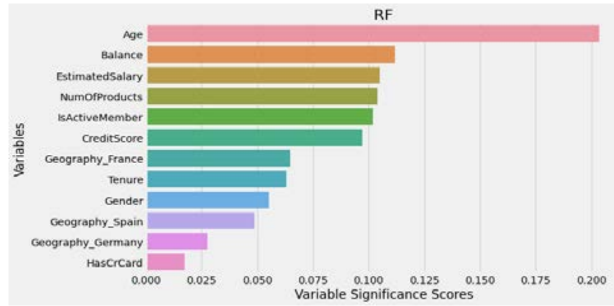


Fig. 8. RF Feature Relevance.

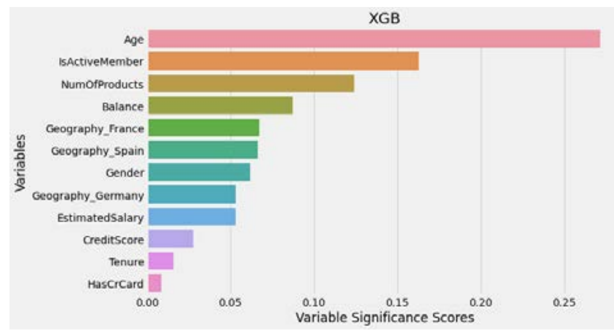


Fig. 9. XGB Feature Relevance.

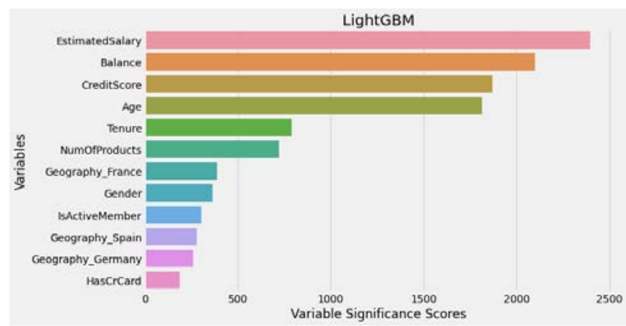


Fig. 10. Light-GBM Feature Relevance.

The feature selection process was performed using the empirical hyper-parameters setting cited below.

A. Feature Selection usingXGB and LightGBM

After the calculation of feature relevance for each machine learning algorithm, the above results show that the most eight relevant features according to RF are:

- Age,
- Balance,
- Estimated Salary,
- Number Of Products,

- Is Active Member,
- Credit Score,
- Geography_France,
- Tenure.

The most eight relevant features according to XGB are:

- Age,
- Is Active Member,
- Number Of Products,
- Balance,
- Geography_France,
- Geography_Spain,
- Gender,
- Geography_Germany.

While the most eight relevant features according to Light-GBM model are:

- Estimated Salary,
- Balance,
- Credit Score,
- Age,
- Tenure,
- Number Of Products,
- Geography_France,
- Gender.

The goal of this part of study is to find relevant features according to each machine-learning algorithm and to try after to use just those features with other algorithms such as MLP neural network. This can make the training process easier and leads to time saving in both training and generalization steps.

According to the chosen hyper-parameters, the accuracy reaches 88.88% for the XGB model and 88.95% for Light GBM model.

B. *Using MLP Model with Relevant Features*

After implementing feature scaling for all inputs variables in order to have values with comparable ranges, we will study the effect of using relevant features with MLP model. The most relevant features are those that appear at least twice in the union of relevant features sets elaborated by RF, XGB and Light-GBM models. These features are:

- Age,
- Balance,
- Estimated Salary,
- Number Of Products,
- Is Active Member,
- Credit Score,
- Geography_France,
- Tenure,
- Gender.

We did the simulation with nine relevant features with MLP model.

After applying the feature selection with ensemble decision tree models and using these features to train MLP model, only the Cohen's Kappa metric has been improved, which reflect the perfect agreement between the observed and predicted outcomes, it has been improved by 19.40%. According to the results shown in Table VIII, we can deduce that even after removing three features, the MLP performances are not drastically affected.

TABLE VIII. OBTAINED METRICS 9 MOST RELEVANT FEATURES WITH MLP MODEL

Name	**Accuracy**	**Sensitivity**	**Specificity**	**Cohen's Kappa**	**MCC**
MLP	0.831764	0.844981	0.818859	0.663601	0.663902

VII. RESULTS INTERPRETATIONS

In this research, we have compared the results of applying eight machine-learning algorithms, LR, KNN, CART, RF, SVM, XGB, Light GBM and MLP with other studies using "Churn Customer" dataset [5]. Our study aims to predict accurately customers predisposed to stay with bank and to honor their commitments. For a given model, a high accuracy would indicate that the model is able to predict the decision that a customer can make (exit the bank/ stay with the bank). After the evaluation of the eight models, we notice that RF model got the best accuracy score with 89.07%. It must be mentioned that we have adopted three pre-processing steps, which consists of data re-sampling, scaling and hyper parameter setting. Light GBM model was the second more accurate model, with an accuracy of 88.95%. The third more accurate model was the XGB model with an accuracy score of 88.88%. While Artificial Neural Network (MLP) classifier accuracy score got 86.37%.

Concerning feature number reduction, the use of relevant features as input of the MLP neural network shows that the predicting capabilities of this model were not seriously affected. In fact, the accuracy is still around 83.17%, even if three features were not used. This result is important because the training phase is easier to engage when using a reduced number of features.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we proposed an approach for churn customer prediction, using a dataset with 10000 observations. After the comparison of our results to those obtained by other existing approaches [5], with respect to the used dataset, ensemble decision tree and Multi-layer neural network models have demonstrated good performances. In this work, we were also interested in improving the obtained results by exploring ensemble learning algorithms capabilities to find relevant features that can be fed to some classic machine learning models, as an example MLP neural network. The MLP model performances were not seriously affected by the reduction of features number, from 12 to 9, even some metrics were improved.

We can notice that this approach can be a motivation to make better decisions when dealing with personalized data and can also be generalized to deal with sophisticated concepts such as NLP (Natural language processing), in order to semantically analyze customer posts. We are also interested, as a perspective, in assisting MLP implementation by using Big Data tools to enhance its performances. In fact, it is possible to involve Spark libraries to enable running various machine-learning algorithms on distributed systems. Integrating these tools to this work will give the speed and the capacity to perform better results with a large amount of data.

References

[1] S. Brinks, "Improving Customer Experience And Revenue Starts With The App Portfolio", A Forrester Consulting Thought Leadership Paper, March 2020.

[2] A. Gallo, "The value of Keeping the Right Customers", Harvard Business Review, October 29,2014.

[3] J. Ganesh,M.J Arnold & K.E Reynolds, “Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers”. July 2000, Journal of Marketing 64(3), pp. 65-87.

[4] A. De Caigny,K. Coussement, & K.W De Bock, 2018. "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees". European Journal of Operational Research, 269(2), pp. 760-772.

[5] I.Tandan, E.Goteman, "Bank Customer Churn Prediction, a comparison between classification and evaluation method", UPPSALA University, June 4, 2020.

[6] W. Verbeke et al. 2012. "New insights into churn prediction in the Telecommunication Sector: A profit driven data mining approach". European Journal of Operational Research, 218(1), pp. 211-229.

[7] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, "Light-GBM: A Highly Efficient Gradient Boosting Decision Tree", NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 2017. Pages 3149–3157.

[8] L. Breiman. "Bagging predictors", Kluwer Academic Publishers, Boston, Machine Learning, 24(2):123–140, Aug 1996.

[9] Y. Elyusufi, Z.Elyusufi, M. Ait Kbir "Social Networks Fake Profiles Detection Based on Acount Setting and Activity", SCA '19: Proceedings of the 4th International Conference on Smart City Applications, October 2019 Article No.: 37. pp. 1-5.

[10] Y. Elyusufi, Z., Elyusufi, M. Ait Kbir "Customer profiling using CEP architecture in a Big Data context", SCA '18: Proceedings of the 3rd International Conference on Smart City ApplicationsOctober 2018 Article No.: 64Pages 1–6.

[11] Y. Elyusufi, H.Seghiouer, M.A.Alimam, "Building profiles based on ontology for recommendation custom interfaces", 2014 International Conference on Multimedia Computing and Systems (ICMCS), 14-16 April 2014.

[12] W. J. Ferrier, G. K. Smith, and C. M. Grimm. "The role industry of competitive action in market share erosion and dethronement: A study of industry leaders and challengers". The Academy of Management Journal, 42(4):372–388, 1999.

[13] W. J. Reinartz and V. Kumar. "On the profitability of long-life customers in a non contractual setting: An empirical investigation and implications for marketing". Journal of Marketing, 64(4):17–35, 2000.

[14] K. Coussement, D. F. Benoit, and D. V. Poel. "Improved marketing decision making in a customer churn prediction context using generalized additive models". Expert Systems with Applications, 37(3):2132 – 2143,2010.

[15] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason. "Defection detection: Measuring and understanding the predictive accuracy of customer churn models". Journal of Marketing Research, 43(2):204–211,April 2006.

[16] L. Breiman. "Classification and Regression Trees". Routledge Taylor & Francis Group, Published January 1, 1984 by Chapman and Hall/CRC, ISBN 9780412048418.

[17] L. Breiman." Random forests". Machine Learning, 45(1):5–32, Oct 2001.

[18] A. Manus “Predicting Churn for Bank Customers”, (Kaggle : October 2018): https://www.kaggle.com/adammaus/predicting-churn-for-bank-customers.

[19] O. Arbelaitz, I. Gurrutxaga, J.Muguerza, and J. María Pérez: "Applying Resampling Methods for Imbalanced Datasets to Not So Imbalanced Datasets", Springer-Verlag Berlin Heidelberg 2013, : CAEPIA 2013, LNAI 8109, pp. 111–120, 2013.

[20] F. Alahmari, "A Comparison of Re-sampling Techniques for Medical Data Using Machine Learning", Journal of Information & Knowledge Management, Vol(19), No. 1(2020).

[21] J.Hale," Scale, Standardize, or Normalize with Scikit-Learn", Towards data Science Corporation, Mars 4, 2019.

[22] Y. Xie, X. Li, E.W.T. Ngai, and W. Ying. "Customer churn prediction using improved balanced random forests". Expert Systems with Applications,36(3, Part 1):5445 – 5449, 2009.

[23] S.Ronaghan, "The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark", Towards data Science Corporation , May 11, 2018.