A New Model based on global hybridation of machine learning techniques for “Customer Churn Prediction In Banking Sector”

F. I. Khamlichi

DEVOTEAM

fahd.idrissikhamlichi@devoteam.com

H. Hachimi

USMS of Beni Mellal

hanaa.hachimi@usms.ma

***Abstract*— We present a new model based on a global hybridation of the most popular machine learning methods applied to the challenging problem of customer churning prediction in Banking Sector. In the first phase of our experiments, all models were applied and evaluated using cross-validation on a popular, public domain dataset. In the second phase, we describe our model and show the performance improvement. In order to determine the most efficient parameter combinations we performed a various simulation for each method and for a wide range of parameters. Our results demonstrate clear superiority of the proposed model against the popular existing ML models.**

Keywords*— *Churn prediction, machine learning techniques, boosting algorithm, banking sector.

I. INTRODUCTION

The regulatory framework within which financial institutions and insurance firms operate require their interaction with customers to be tracked, recorded, stored in Customer Relationship Management (CRM) databases, and then data mine the information in a way that increases customer relations, average revenue per unit (ARPU) and decrease churn rate.

According to scientific research, churn has an equal or greater impact on Customer Lifetime Value (CLTV) when compared to one of the most regarded Key Performance Indicator (KPI’s) such as Average Revenue Per User (ARPU).

As one of the biggest destructors of enterprise value, it has become one of the top issues for the banking industry.

The banking industry needs to intensify campaign to deliver a more efficient, customer focused and innovative offerings to reconnect with their customers. The problem of churn analysis is not peculiar to the banking industry.

In this paper we considered the churn detection problem in the Banking sector.

II. MOTIVATION & METHODOLOGY

A. Motivation

Churning is an important problem that has been studied across several areas of interest, such as mobile and telephony, insurance, and healthcare.

B. Methodology

Our paper aims to provide a new approach that allows a combination of existing models to yield even better results. The new approach is based on using a hybridization between a list of chosen models and build a model that chooses the most appropriate model for a new observation at inference time.

III. EVALUATION MEASURES

In order to evaluate classifiers performance in churn prediction for different schemes with their appropriate parameters, we use the measures of precision, recall, accuracy and F-measure, calculated from the contents of the confusion matrix.

True positive and false positive cases are denoted as TP and FP, respectively, while true negative and false negative cases are denoted as TN and FN.

These metrics are particularly important for the problem at hand (Churn Detection). They are the most significant metrics especially from the end-user point of view (Bank marketing teams and managers). A Bank usually assigns costs and benefits to clients that are predicted to stay and allows budgets to keep customers that are expected to churn.

The balance between the cost of a false positive and a false negative is thereby end-user dependent. The ‘best’ model is also by the same reason case dependent.

However, for benchmarking purposes, we have chosen the accuracy as the main metric to choose the best model and also to compare our overall results with the previous works.

IV. RELATED WORKS

Our work is inspired by the strategy of F. I. Khamlichi and R. Aboulaich in « global face recognition strategy based on multiple recognition systems combination » published in the international journal of imaging and robotics (issn 2231–525x), volume 6, number a11, autumn (2011). They extend multiple recognition systems for global face recognition. The idea behind this strategy is to combine the models based on the facial complexion as an input and basic criterion to specify which model to apply in order to obtain the best results in facial recognition.

To our knowledge, the other paper that also learn about the combined systems is this of F. I. Khamlichi, R. Aboulaich and A. E. El-Mrhari in global decision strategy based on combining multiple regression methods published in the international journal of statistics & statistics & economics ((ijse) (issn: 0975-556x), volume 8, number s12, 2012).

V. OUR ALGORITHM

As mentioned in the introduction, our approach is based on choosing the best combination of known models for the specific problem and apply the best model at the instance level.

The intuition behind this approach is better illustrated by an example. A model with a task of classification of cats and dogs into their sub-races performs less than two models, one trained on cats and another trained on dogs (The difference of performances has been proven for face recognition in one of our previous works [8] « Global face Recognition Strategy based on multiple recognition systems combination »).

Our approach applied to this problem would provide an overall model that first identifies if the input picture is of a cat or a dog. The second stage will be to apply the best model for the identified 'class'.

In the case of Churn Detection, we have to identify a way to split our dataset so we have a similar split between cats and dogs. The feature would ideally split our observations (customers) into few sub-populations that have different behaviours with regards to the chosen feature and also to our problem.

Such feature would show a variation of the target in a span basis (ie one class is predominant for observations having this feature within the same span). When the dataset is split based on this features' spans, the classification hopefully becomes a simpler problem.

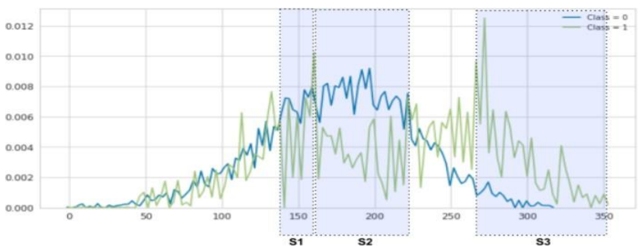

Fig. 1: Illustration: Examples of spans where one class is relatively predominant.

The number of possible features and their spans are very high for our dataset. This makes the identification via evaluation of every possibility computationally prohibiting. We have therefore decided to count on our understanding of the problem.

In our case, the problem of customers churn for Banks, we identified that the best assumption we can make is that the behavior of customers depends greatly on their age.

Visualizing the dataset shows that (3) spans is a good choice for the split. The next step is to split our data into subsets having the **Age** within three (3) defined spans. The resulting subsets are called clusters C1, C2 and C3.

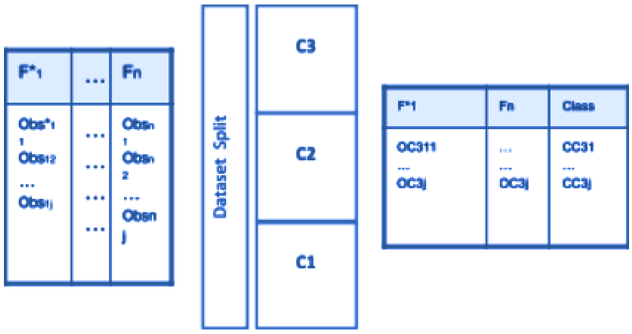

Fig. 2: Dataset split.

These clusters and their corresponding spans are as described in the following table.

Cluster	Span	Size
C1	0 - 30	3356
C2	30 - 45	3334
C3	45 - 100	3310

Table. 1: Dataset split results.

Once the data set is split, it is again split into training and validation sets (80% / 20%). The dataset is now ready for training and evaluation. We generally train n models M1, M2, ..., Mn, on each one of the subsets C1, C2 and C3.

Each model is then evaluated on the corresponding cluster and we compare the resulting performances: Gc1, Gc2, .. Gcn for each cluster c and identify the best model for the cluster c. The final results are the best identified models for each cluster: Mc1, Mc2 and Mc3 (Mci being the best model for cluster 'ci' where in general i ∈ [1, number of clusters]).

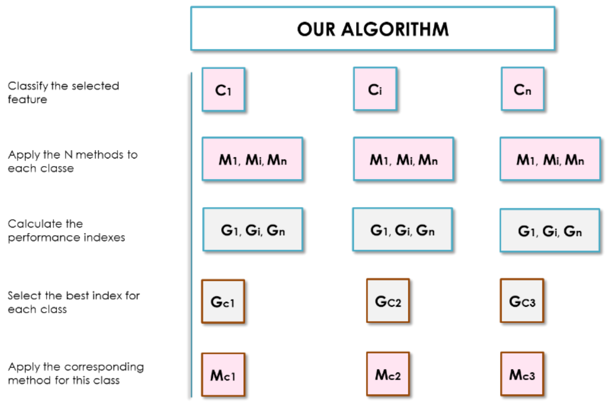

Fig. 3: The algorithm produces as many best models as the data split clusters.

At The inference time, each new observation is first placed into the corresponding cluster. We then use the best model for the given cluster to make the prediction.

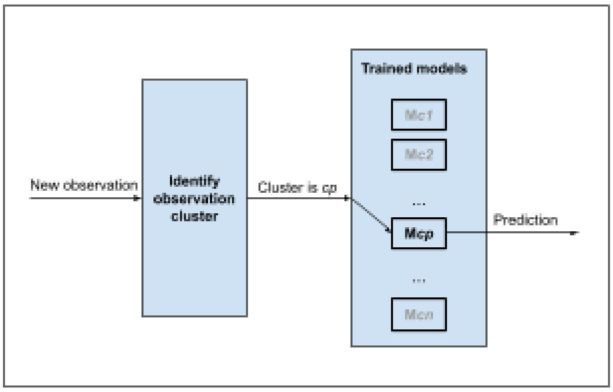

Fig. 4: Two stages inference.

VI. SIMULATION

A. Simulation Setup

Our main goal is to compare different models of classification on churn prediction and build the best one with a better prediction. in this sense, we implemented our simulation in three main steps using Python language following its power in the field of data science. First, we implemented different classification models throughout the dataset, the main models are first Logistic Regression, Support Vector Machine (SVM), Decision Tree Classification, Random Forest and XGBoost.

The next step is to select the influencing feature to divide it into classes and apply the model to each class in order to choose the perfect one, then merge the result into one to build a single model that groups the models previously builts (according to the number of classes devised).

B. Results

To evaluate the performance of the tested classifiers, we worked with a churn dataset from the UCI Machine Learning Repository, this version contains 10000 samples and 21 features plus the variable churn, presented in Table.2.

```
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   RowNumber        10000 non-null  int64
 1   CustomerId       10000 non-null  int64
 2   Surname          10000 non-null  object
 3   CreditScore      10000 non-null  int64
 4   Geography        10000 non-null  object
 5   Gender           10000 non-null  object
 6   Age              10000 non-null  int64
 7   Tenure           10000 non-null  int64
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64
 10  HasCrCard        10000 non-null  int64
 11  IsActiveMember   10000 non-null  int64
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64
dtypes: float64(2), int64(9), object(3)
```

C. Classifiers performance with hybrid method

Next, we divide the data set into three classes according to the feature that hopefully has the most impact on the target (in our case: Age) using the library named qcut and apply the models built previously in each class to choose those with the best results and merge them into one over the entire dataset, respecting the model used for each class.

Based on the results obtained in the Table.3 we choose to work the first class with SVM-POLY classifier with two parameters, C with 0.5, and gamma with 10.

For the second class we worked with the Logistic Regression LR classifier with penalty and C as the main parameters of this model, penalty =l1 and C=0.1.(Table.4)

In the case of the third class we used the Decision Tree Classifier using random_state, min_samples_split, min_samples_leaf and max_features having successively those values random_state=123, min_samples_split= 10, min_samples_leaf= 1, max_features=sqrt.(Table.5).

SVM (POLY)	Accuracy	False Positive	True Negative	False Negative	True Positive
C1	99%	1	138	1	28
C2	97%	2	137	3	24
C3	98%	1	148	2	16

Table.3: Metrics obtained using SVM(POLY) model on the different clusters.

Logistic Regression	Accuracy	False Positive	True Negative	False Negative	True Positive
C1	98%	1	140	2	25
C2	99%	0	140	1	25
C3	97%	3	141	1	22

Table.4: Metrics obtained using Logistic Regression model on the different clusters.

Decision Tree	Accuracy	False Positive	True Negative	False Negative	True Positive
C1	98%	1	141	3	25
C2	98%	2	150	1	14
C3	99%	1	143	0	23

Table.5: Accuracy, False Positive, True Negative, False Negative and True Positive obtained using Decision Tree model on the different clusters.

In this table (Table.6) we summarize the best results chosen for each cluster.

	Logistic Regression	SVM (POLY)	Decision Tree
C1	98%	**99%**	98%
C2	**99%**	97%	98%
C3	97%	98%	**99%**

Table.6: *Best Accuracy results for each cluster by model*

5.4. Performance comparison – Discussion

Using the mechanism described in "Fig. 4: Two stages inference", we evaluated the mechanism overall performance and compared it to the results from separate models. For the single models , the evaluation has been held on the whole dataset without separation based on any feature. Table.4 presents the resulting performance with a visible increase, especially the F-measure metric, with the new approach.

Classifiers	Accuracy	F-measure
	Single model	
Random Forest	**95%**	79%
XGBoost	**95%**	**80%**
Support Vector Machine	**93%**	73%
Decision Tree	**90%**	64%
Logistic Regression	**86%**	26%
K-Nearest Neighbors	**81%**	28%
Combined models approach		
Hybrid model	**99%**	**97%**

Table.7: *Performance comparison for models' classifiers with and without hybridization.*

VII. CONCLUSION AND SUMMARY

Our simulations were performed using six of the most popular classification methods for the churn prediction problem of Bank customers based on a publicly available dataset. First of all, all methods were tested without the use of boosting under different settings. The two top performing methods in our test were XGBoost classifier and Random Forest classifier, both methods achieved accuracy 95% and F-measure 80% approximately. The Support Vector Machine classifier (POLY kernel) obtained accuracy of 93% and 73% about F-measure. Decision Tree method achieved the accuracy of 90% and F-measure 64%. Logistic Regression and K-Nearest Neighbors methods fail short with accuracy 86% and 81% also an F-measure of about 26% and 28% respectively.

At a later stage, we considered to work with the entire dataset by dividing it into an impacting feature (classes) and implementing the algorithms on each slice then merging the results into a single algorithm. In our case we choose to work mainly with Support Vector Machine-POLY, Logistic Regression and Decision Tree. Overall, this method was the best way of classification with accuracy of almost 99% and F-measure over 97%.

This work has brought to light a new way of exploiting several algorithms in one according to the problem and the data we have, and still the best way to use the most popular algorithms and exploit them into one. Also, to use a larger and more detailed dataset can maximize the statistical significance of the result of our algorithm.

REFERENCES

[1] The Chartered Institute of Marketing, Cost of customer acquisition versus customer retention (2010).

[2] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, A. Rehman, Telecommunication subscribers' churn prediction model using machine learning, in: Digital Information Management (ICDIM), 2013 Eighth International Conference on, IEEE, 2013, pp. 131–136.

[3] K. Kim, C.-H. Jun, J. Lee, Improved churn prediction in telecommunication industry by analyzing a large network, Expert Systems with Applications.

[4] C. Kirui, L. Hong, W. Cheruiyot, H. Kirui, Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining, International Journal of Computer Science Issues (IJCSI) 10 (2).

[5] G. Kraljevi´c, S. Gotovac, Modeling data mining applications for prediction of prepaid churn in telecommunication services, AUTOMATIKA: ˇcasopis za automatiku, mjerenje, elektroniku, racunarstvo i komunikacije 51 (3) (2010) 275–283.

[6] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 321–357.

[7] Dudyala, A. K., & Ravi, V. (2008). Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies, 1(1), 4-28.

[8] Farquad, M. A., Ravi, V., & Raju, S. B. (2009). Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards. In Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (pp. 390-397). Springer Berlin Heidelberg,

[9] F. I. Khamlichi, « Analysis and identification of the feelings or the behavior of a person through web pages from social web sites, using Fuzzy Logic method », JICCLAP2018, Qatar 2018.

[10] M. Kaicer, S. Mhir and F. I. Khamlichi, « Vital Prognosis of Intoxicated Patients, Modeling by SVM », IProcedia Computer Science, Volume 127, 2018, Pages 154-160.

[11] F. I. Khamlichi, A. Boufrahi, « Mammography diagnostic using recognition system based on closed loop fuzzy logic controller », International Journal of Imaging and Robotics (ISSN 2231–525X), 2017.

[12] F. I. Khamlichi, R. Aboulaich and A. E. El Mrhari, « Quantitative-trading strategy based on fuzzy logic model », International Journal of Mathematics & Computation (IJMC) (ISSN: 0974-5718), Volume 19, Number 2, 2013.

[13] R. Aboulaich, F. I. Khamlichi and M. Kaicer, « Microcredit scoring using fuzzy logic model », International Journal of Statistics & Economics (IJSE) (ISSN: 0975-556X), Volume 10, Number 1, 2013.

[14] F. I. Khamlichi, R. Aboulaich and A. E. El Mrhari, « Global Decision Strategy based on combining multiple regression method