Churn Prediction in the Credit Card Industry

@justgrossi
https://github.com/justgrossi/Portfolio.git

***Abstract*—A classifier predicting credit card churns was built comparing the performances of *Decision Trees* and *Gradient Boosting Machine*. The latter was the best performing: Accuracy: 96.6%; Cohen's Kappa: 87.5%; F-Measure: 97.9% and the five most important features when making such predictions were: total number of transactions and amount transacted; revolving balance; difference in number of transactions and amount transacted compared to the previous twelve months.**

***Index Terms*—Classification, Decision Trees, Gradient Boosting Machine, Credit Card, Churns.**

I. Introduction

Multiple industries and business models benefit from early churn-detection systems to avoid customer and revenue losses. A popular streaming service once dominating the home-entertainment industry for instance made the news after losing a million customers in four months [1]. Being able to detect the early signs of potential defections hence becomes paramount. More traditional competitive landscapes such as the banking industry make no exception. As it seems demonstrated that acquiring new customers is far more expensive than retaining existing ones [2]; understanding the root cause of defections, and being able to anticipate them, might constitute a competitive advantage. Here is where machine learning and data mining techniques come into play creating tremendous impact on the business.

The analysed data set was retrieved from the *UCI Machine Learning Repository* [3].

II. Related Work

In the financial and investment sectors, churn is commonly defined as customers divesting portfolios below specific thresholds or closing their positions altogether [4] [5]. Research in the field however seems to display the tendency to overlook the critical issue of class imbalance. Most data mining techniques to output reliable predictions require a balanced representation of the instances pertaining to the target variable [6]. Insufficient examples of the minority class hinder algorithmic learning, and may result in skewed predictions favoring the predominant class [7] [8] [9]. *Decision trees* were employed in [10] [11] and [12] but it seems these studies failed to address class imbalance, compromising the reliability of their results. Despite being the best-performing method, the accuracy of the model may be misleading. An ensemble type of approach in churn detection is proposed in [13], since even though more mature methodologies exist, it still remains challenging to identify a consistently superior algorithm [14]. Research in the Chinese banking ecosystem showcased *Support Vector Machine* as the most accurate technique even though its performance was just above a random classifier with a 50% chance of making a correct prediction [15]. Contrasting results were recorded in [16] where, after addressing class imbalance and hyperparameter tuning, the model exhibited a 99% accuracy. In the Brazilian financial industry [17] *Random Forest* outperformed various models, including *Decision Trees*, *K-Nearest Neighbors*, *Elastic Net*, *Logistic Regression*, and *Support Vector Machine* in predicting churns.

TABLE I
ALGORITHMS COMPARISON - ACCURACY

Reference	Technique	Accuracy [%]
[10]	*Decision Tree*	*85.2*
[11]	*K-Nearest Neighbour*	*88.5*
[12]	*Support Vector Machine*	*97*
[13]	*Gradient Boosting*	*79.7*
[15]	*Support Vector Machine*	*60*
[16]	*Support Vector Machine*	*99*
[17]	*Random Forest*	*90*

III. Methodology

The Cross Industry Standard Process for Data Mining (i.e. *CRISP-DM*) [18] was the adopted methodology. Research questions were approached and facilitated by framework' six stages.

A. Business understanding

Multiple industries and business models can benefit from early churn-detection systems and financial institutions are no exception. High predictive capabilities would mean the ability to support business' long term revenue goals through effective customer relationship management strategies and tactics. Insights regarding the most important variables signalling potentially departing customers would be invaluable to build customer loyalty and increase their lifetime value.

B. Data understanding

The data base had 10,127 entries and 19 independent variables 5 of which were categorical and 14 numerical. The binomial target label *'Churned'* identified credit card cancellations. *'Income level'* was the only positively skewed variable with 35% of customers earning less than 40,000 and 7% more than 120,000 USD (Fig.1). There were no missing values yet the data set was imbalanced with only 16% of customers recorded as churning. Churns were equally distributed across genders with the average client being 46 of age, possessing an entry-level *'Blue'* type of credit card for an average tenure of 36 months. Those possessing a *'Blue'* card also churned in greater numbers (15%), recorded both the average highest credit card utilization ratio (i.e. 0.29) and number of customer service contacts (Fig.2). They also had the tendency to buy an average higher number of ancillary products (i.e. 3.85) than any other cohort (Fig.3).

C. Algorithms Selection

Decision Trees and *Gradient Boosting Machine* were selected as they consistently recorded strong performances across the reviewed related works and tolerate well mixed data sets. They also are easy to interpret, the decision making process can be visualized, and features can be ranked based on their relative importance. *Gradient Boosting Machine* in particular allows predictive features to be ranked according to the *Gini*'s coefficient mean decrease: the higher the score linked to a variable, the higher its importance.

D. Data preparation

Class imbalance was resolved applying the synthetic minority oversampling technique (i.e. *SMOTE*) [19]. Since it requires numerical data as inputs, the categorical features were one-hot encoded prior to being fed to the model. In the resulting balanced data base 48% of records belonged to the minority class and 52% to the majority respectively. An 80/20 split was finally performed to create training and test sets. The balance of the target label within the two was preserved.

E. Modelling

A 10-folder cross-validation procedure was used to tune models' hyperparameters and assess in-sample performance with accuracy as the metric of reference. *Decision trees*' complexity parameter *'C'* optimal value was 0.01 (Fig.4). The tree was pruned accordingly (Fig.5). Following are *Gradient Boosting Machine*'s parameters that yielded the highest in-sample accuracy of 96.6% (Kappa: 87.6%; N.I.R.: 83.9%, p-value: .2e-16) (Fig.6):

- ***'interaction.depth'***: maximum depth of the individual trees in the ensemble: 7,
- ***'ntree'***: maximum number of trees to grown within each bootstrap sample: 300,
- ***'shrinkage'***: rate at which the boosting algorithm adapts or learns from each additional tree that's added to the ensemble: 0.3,
- ***'n.minobsinnode'***: minimum number of observations required in a terminal (leaf) node of the decision tree during the tree-building process: 15.

IV. Evaluation

Both models' performance exceeded the reviewed related works and resolving class imbalance surely contributed to such result. *Decision Trees*' accuracy (90.4%) was higher than the no information rate (i.e.*NIR* - 83.9%, p-value: 2.2e-16) hence the model was significantly better than a predictor always predicting the majority class. The adjusted version of the accuracy *Kappa*, accounting for correct predictions due to chance however, resulted to be lower (69.2%). With a 97.8% precision the model recorded a fairly high quality of the positive predictions. Similar and well balanced accuracy, precision, and recall values meant an healthy confusion matrix hence a model equally good at predicting both classes. As a matter of fact the *F-Score* was 94%. The *R.O.C.* curve closely resembled the ideal spot on the upper-left corner (Fig.7) and the *A.U.C.* was 90.0%. *Decision Trees* ranked as the top 5 features (Fig.8): credit card limit, total transacted amount, revolving balance, and difference in both the number of transactions and transacted amount compared with the previous twelve months. *Gradient Boosting Machine*'s accuracy was higher (96.6%) and much better than a model always predicting the majority class (*N.I.R.* 83.9%, p-value: .2e-16). *Kappa* was 87.6%. Even in this case, with a 98.3% precision, the confusion matrix showed the model was equally good at predicting both classes. The *F-Score* was 98.0%. The *R.O.C.* curve was extremely close to the ideal spot (Fig.9) and the *A.U.C.* was 94.5%. *Gradient Boosting Machine* considered the top 5 features to be (Fig.10): total number of transactions and total amount transacted, revolving balance on the credit card, and difference in both the number of transactions and transacted amount compared with the previous twelve months.

TABLE II
CONFUSION MATRIX

	Decision Tree		*GBM*	
	References			
Predictions	*No*	*Yes*	*No*	*Yes*
No	1540	34	1659	28
Yes	160	291	41	297

TABLE III
MODEL EVALUATION

Model	*Accuracy*	*Sensitivity*	*Specificity*	*Kappa*	*FScore*
DT	90.4%	90.6%	89.5%	69.2%	94%
GBM	96.6%	97.6%	91.4%	87.6%	98%

V. Conclusions and Future Work

A classifier was built to predict customers at risk of churning their credit card and the model based on *Gradient Boosting Machine* recorded the best performance with an accuracy of 96.6%, a Cohen's Kappa of 87.6% and an F-Measure of 98%. The total number of transactions and amount transacted, revolving balance on the credit card, and difference in both the number of transactions and transacted amount compared with the previous twelve months emerged as the most important features determining such predictions. Future research should consider expanding the data sets including both more records and features. Additional balancing strategies other than the minority oversampling techniques could be explored too, to resolve class imbalance.

REFERENCES

[1] N. Sherman, J. Claiton, Netflix loses almost a million subscribers, 2022, BBC News, Available at: $https://www.bbc.com/news/business-$ 62226912

[2] J. Sterne, 2002, Web metrics: proven methods for measuring web site success, John Wiley and Sons, New York, p. 283.

[3] UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/ml/datasets.php

[4] J.F. Adolfsen, F. Kuik, E. M. Lis, T. Schuler, The impact of the war in Ukraine on euro area energy markets, European Central Bank Economic Bulletin, Issue 4/2022. Available at: www.ecb.europa.eu/pub/economic-bulletin/focus/2022/html/ecb.ebbox202204$_0$1 68ef3c3dc6.en.html

[5] N. Glady, B. Baesens, and C. Croux, Modeling churn using customer lifetime value, Eur. J. Oper. Res., vol. 197, no. 1, pp. 402–411, 2009, doi: 10.1016/j.ejor.2008.06.027.

[6] D. Van den Poel and B. Larivière, Customer attrition analysis for financial services using proportional hazard models, Eur. J. Oper. Res., vol. 157, no. 1, pp. 196–217, 2004, doi: 10.1016/S0377-2217(03)00069-9.

[7] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, On the Class Imbalance Problem, 2008, Fourth International Conference on Natural Computation, vol. 4, pp. 192–201. doi: 10.1109/ICNC.2008.871.

[8] P. Xenopoulos, Introducing DeepBalance: Random deep belief network ensembles to address class imbalance, 2017, IEEE International Conference on Big Data (Big Data), pp. 3684–3689. doi: 10.1109/BigData.2017.8258364.

[9] H. He and E. A. Garcia, Learning from Imbalanced Data, IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[10] I. Kaur and J. Kaur, Customer Churn Analysis and Prediction in Banking Industry using Machine Learning, 2020, Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 434–437. doi: 10.1109/PDGC50313.2020.9315761.

[11] A. Bhatnagar and S. Srivastava, A Robust Model for Churn Prediction using Supervised Machine Learning, 2019, IEEE 9th International Conference on Advanced Computing (IACC), pp. 45–49. doi: 10.1109/IACC48062.2019.8971494.

[12] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. Ch. Chatzisavvas, A comparison of machine learning techniques for customer churn prediction, 2015, Simul. Model. Pract. Theory, vol. 55, pp. 1–9, doi: 10.1016/j.simpat.2015.03.003.

[13] H. T. N. Thia, V. N. Thao, Building a proper churn prediction model for Vietnam's mobile banking service, 2022, Institute of Advanced Science Extension (IASE), doi: 10.21833/ijaas.2022.07.014

[14] G. Wang, L. Liu, Y. Peng, G. Nie, G. Kou, and Y. Shi, Predicting Credit Card Holder Churn in Banks of China Using Data Mining and MCDM, 2010, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, pp. 215–218. doi: 10.1109/WI-IAT.2010.237.

[15] 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, Oct. 2008, pp. 1–4. doi: 10.1109/WiCom.2008.2509.

[16] Benlan He, Yong Shi, Qian Wan, Xi Zhao, Prediction of Customer Attrition of Commercial Banks based on SVM Model, P$R.O.C.$edia Computer Science, Volume 31, 2014, Pages 423-430, ISSN 1877-0509, https://doi.org/10.1016/j.p$R.O.C.$s.2014.05.286.

[17] R. A. de Lima Lemos, T. C. Silva, B. M. Tabak, Propension to customer churn in a financial institution: a machine learning approach. Neural Comput and Applic 34, 11751–11768 (2022). doi.org/10.1007/s00521-022-07067-x

[18] CRISP-DM, Data Science Project Management. https://www.datascience-pm.com/crisp-dm-2/ (accessed Jun. 09, 2021).

[19] N.V. Chawla, K.W. Bowyer, L. O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, 2002, Journal of Artificial Intelligence, Research Volume, pp 321–357.

APPENDIX

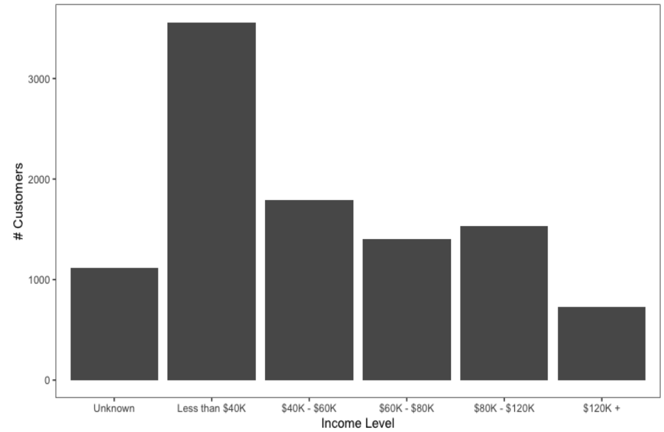

Fig. 1. Income Level Distribution

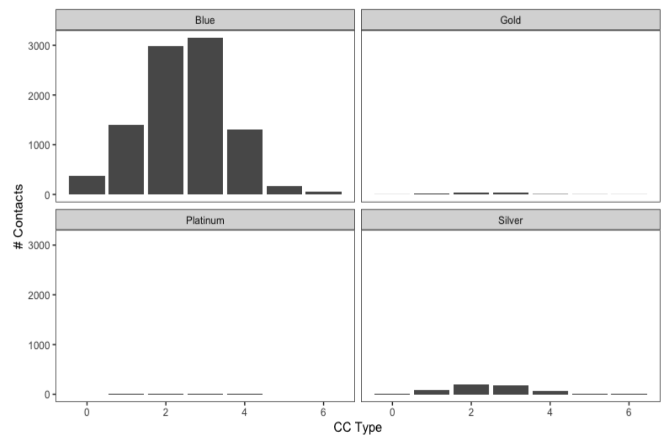

Fig. 2. Number of Contacts vs CC Type

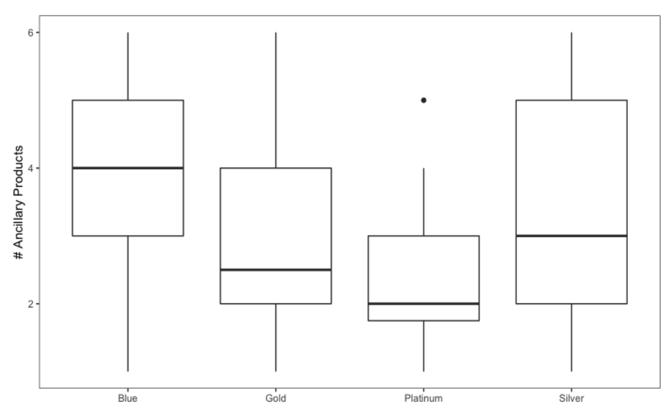

Fig. 3. Ancillary Products vs CC Type

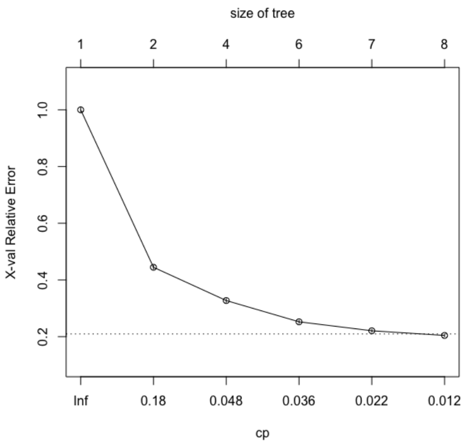

Fig. 4. Complexity Parameter *'C'*: Decision Trees in-sample accuracy

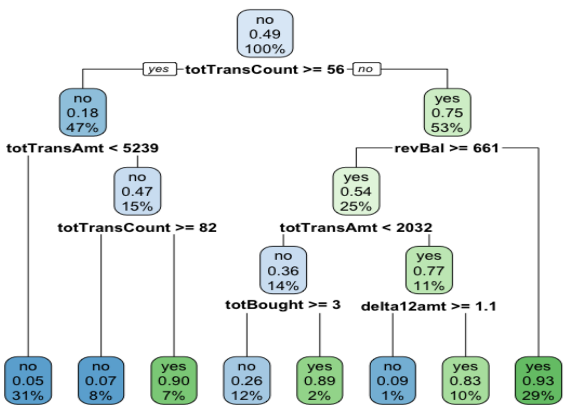

Fig. 5. Decision Tree - Pruned Tree

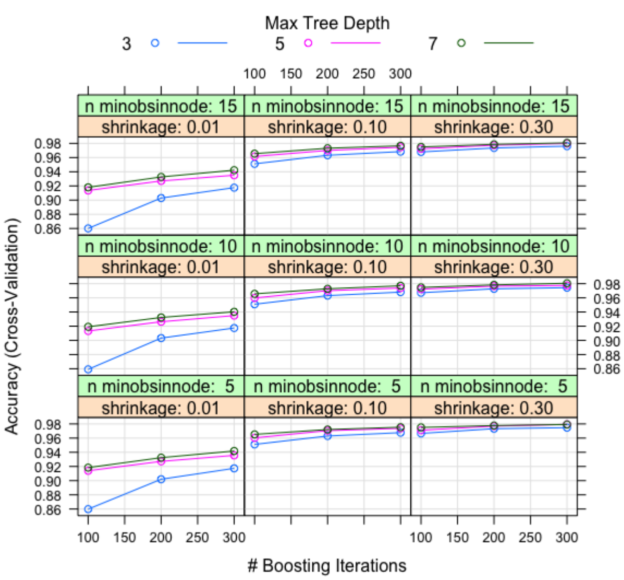

Fig. 6. Gradient Boosting Machine: Hyperparameters tuning in-sample accuracy

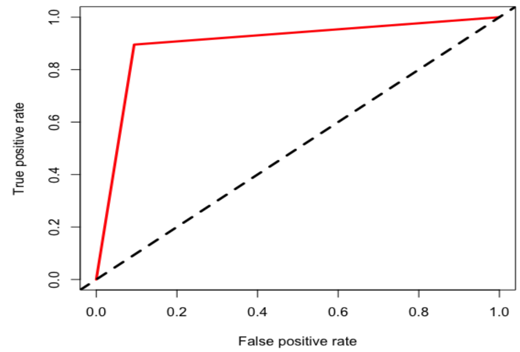

Fig. 7. Decision Tree - *R.O.C.* Curve

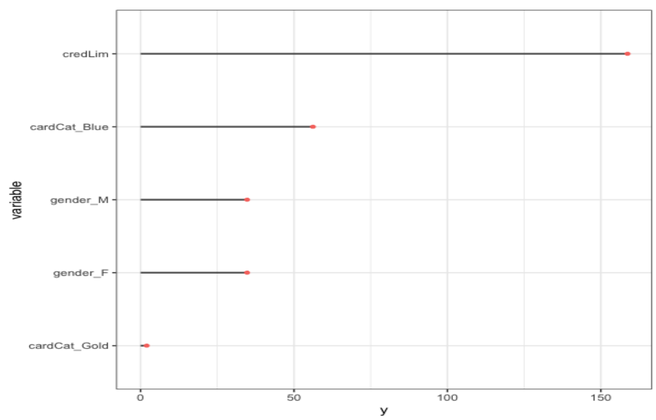

Fig. 8. Decision Tree - Features Importance

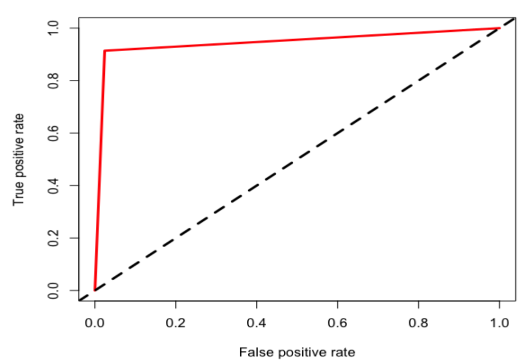

Fig. 9. Gradient Boosting Machine - *R.O.C.* Curve

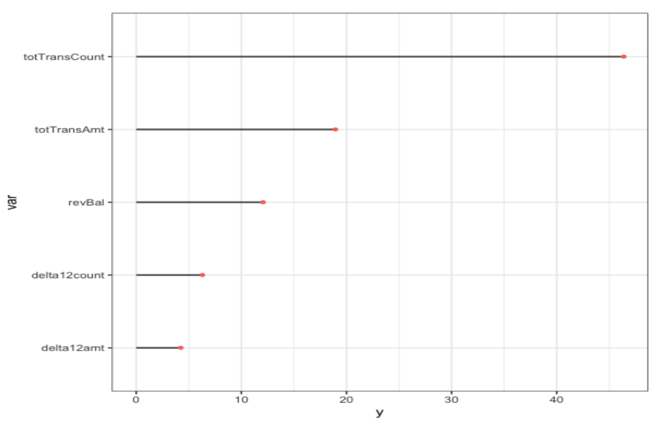

Fig. 10. Gradient Boosting Machine - Features Importance