BANK CHURN PREDICTION

Testing classification models

CODER HOUSE

March, 2022

MAIN OBJECTIVES

Being able to identify the factors that make a bank customer churn (or not),

Find the best model that predicts which customers are likely to churn

SOME INSIGHTS

- About 20% of the customers have churned
- There are clients who have churned, but still have money deposited in the account. That could probably mean that they are going to retire that amount of money (185 millions, that represents almost the 25% of all the money deposited in the bank).
- 75% of the clients are under 44 years old, and most of them are between 30 and 40 years old
- Older customers have a higher churn rate. In this sense, the bank should take action in this regard: that is, carry out customer retention campaigns according to age groups, focusing especially on those over 40 years of age.
- Regarding the years of permanence, the clients who churn, do so or more at the beginning, or rather after the 7 years passed. They should also carry out retention campaigns in order to keep their long-standing clients.
- Customers with larger accounts leave the bank the most. This is an important warning for the bank, since the most important clients are leaving
- While the largest number of clients come from France, the largest number of customers who churn are from Germany.
- Women churn in a greater proportion than men.

MODEL CHOSEN

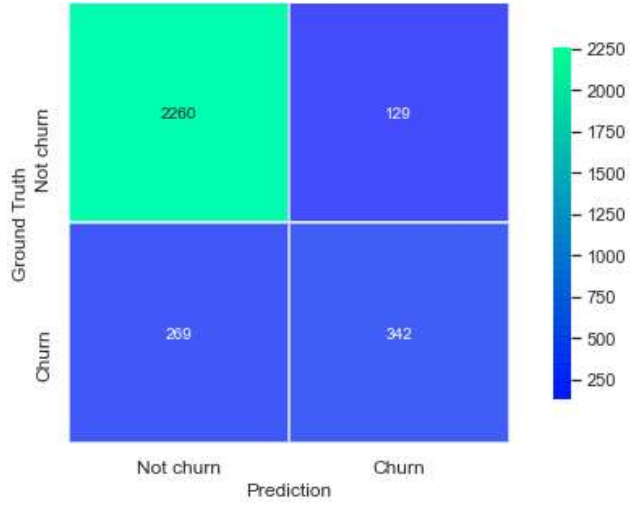
All the models tested where optimized, and a last model was created, by ensembling the two most performing models (Random Forest and Gradient Boosting). This lead into a model which had the highest ROC_AUC of all: this is the chosen model, the "ensemble – RFC and GBC"

	Model	Accuracy Test	Accuracy Train	Precision Test	Precision Train	Recall Test	Recall Train	F1 Score Test	F1 Score Train	AUC
8	Ensemble - RFC y GBC	0.8673	0.8814	0.7261	0.7709	0.5597	0.5947	0.6322	0.6714	0.8816
2	Gradient Boosting	0.8690	0.8761	0.7914	0.8239	0.4845	0.4986	0.6010	0.6212	0.8795
1	Random Forest	0.8517	0.8829	0.6305	0.7067	0.6563	0.7265	0.6431	0.7165	0.8773
5	SVM	0.8507	0.8741	0.6321	0.6868	0.6383	0.7027	0.6352	0.6946	0.8418
0	Decision Tree	0.8283	0.8643	0.5784	0.6705	0.5794	0.6564	0.5789	0.6634	0.8137
7	Gaussian Naive Bayes	0.8140	0.7953	0.5431	0.4975	0.5466	0.4881	0.5449	0.4927	0.8109
3	KNN	0.8423	0.8776	0.6708	0.7973	0.4435	0.5351	0.5340	0.6404	0.7956
4	Logistic Reggression	0.7567	0.7430	0.4338	0.4117	0.6383	0.6094	0.5166	0.4914	0.7921
6	Stochastic gradient descent	0.7377	0.7210	0.4109	0.3878	0.6645	0.6388	0.5078	0.4826	0.7785

MODELS RESULTS

This is the confusion matrix for the chosen model :"Ensemble – RFC and GBC"

The model predicts correctly that 2260 customers will not churn and that 342 will churn, and fails to predict that 269 clients will not churn, when in real life they will, and that 129 will churn, when they will actually not churn.



RECOMMENDATIONS

They should carry out retention campaigns in order to keep the clients with the higher churn rate

1. The country with the highest churn rate is Germany, almost duplicating the rate of Spain and France.

This is something to consider, it's important to know why this is happening. Maybe because of the kind of services offering in Germany, perhaps is not as good as in France

2. As women churn in a greater proportion than man, some actions must be taken in this area too, in order to let women engage with the bank as well as men.

3. There's also recommended to take action regarding the "older" customers, as the are more likely to churn.

4. One thing that's really important to take care, are the customers with largest accounts, as they are churning faster than the ones with less money in their account, and this are the most important clients of all.

THANK YOU FOR YOUR TIME

carolina.swoboda@gmail.com