Bank Churn Modelling

Classification Project

ISSUE / PROBLEM

An European commercial bank seeks to improve client retention and answer the following question: **What's likely to make the client churn?**

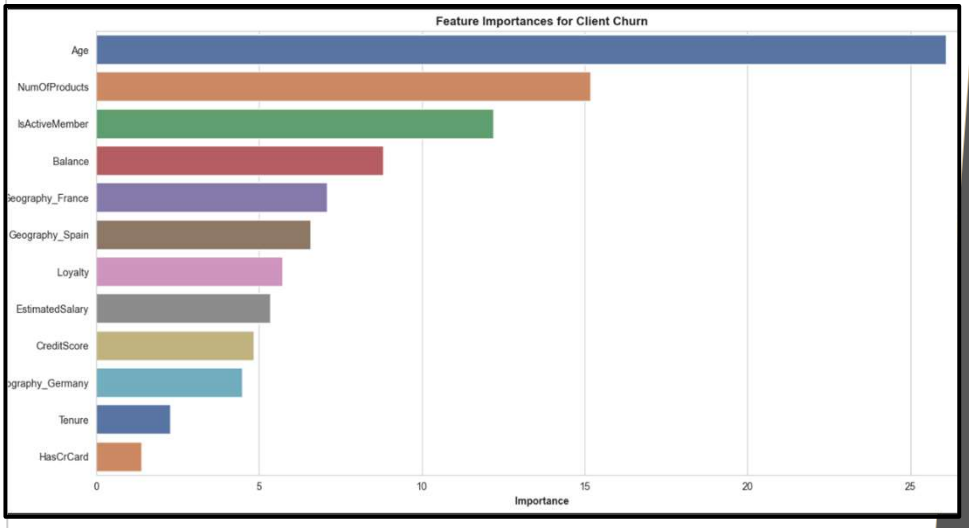
RESPONSE

Since the variable we are seeking to predict is categorical whether the client churns or not, we can build a classification model either a logistic regression or a naïve bayes model or a tree-based machine learning model. After modelling 5 different algorithms, we found that the best performing model is random forest model.

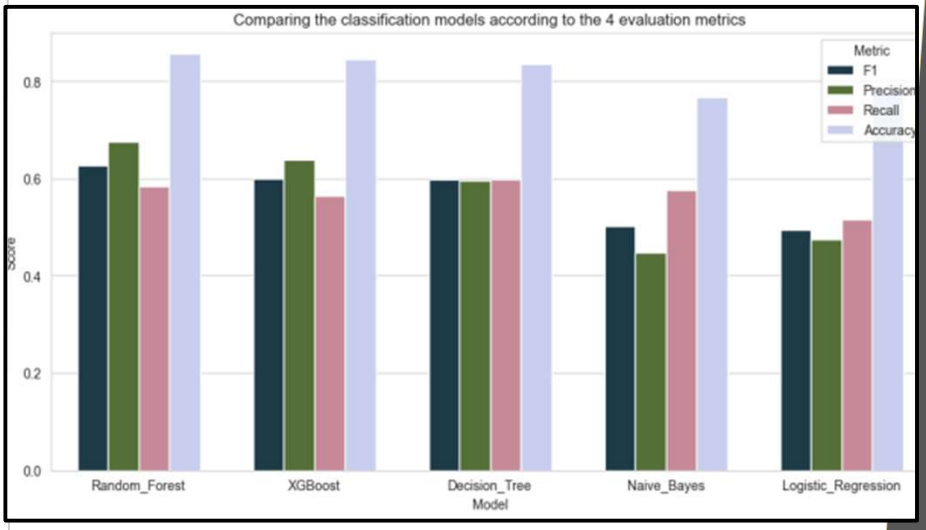
IMPACT

This model helps predict whether a client will churn and identify which factors are most influential.

These insights can help the customer service team make decisions to reduce client churn.



The plot above shows that in this random forest model, age, number of products, member activity and client's balance have the highest importance, in that order. These variables are most helpful in predicting the outcome variable, Exited, and they are the same as the ones referred to by the decision tree model.



We could observe that the Random Forest Model performs the best based on the F1 Score.

KEY INSIGHTS

- Engage the client in more than one product, as this will result in more client retention.
- Consider promoting the old clients or conduct further investigation about why older clients are more likely to churn.
- Encourage the clients who are classified as idle or inactive to be an active clients, as churned clients are more often to be inactive bank clients.
- If clients aren't satisfied with the bank's interest rate policies, inform them about this, since churned clients have more balance in their accounts by around 18000 dollars than retained clients.
- For all the constructed models, there aren't sufficient evaluation and testing scores, so the case study needs more predictive features to predict the target variable more precisely.