Card Churn Modeling

Cindy Jeon

Contents

Introduction

Identify patterns and indicators
that precede churn
by leveraging the power of machine learning.

By doing so,
we aim to equip the business
with the tools to not only predict
potential customer turnover
but to take proactive countermeasures
to prevent it.

Business Question

What are the specific features associated with a high churn rate?

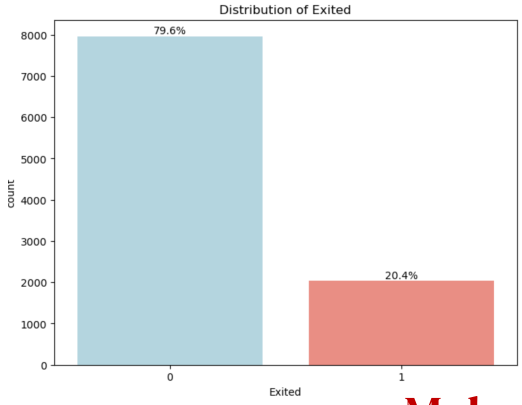
What type of classification models are the best to predict churn?

What do we need to improve the prediction quality?

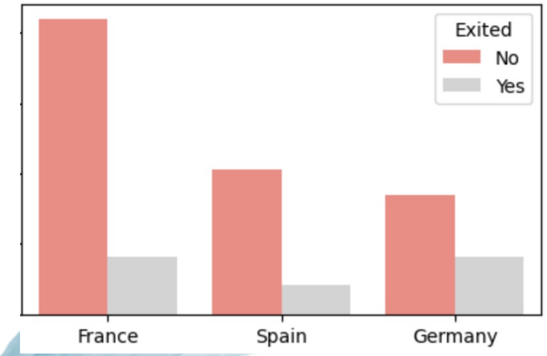
EDA

(Exploratory Data Analysis)

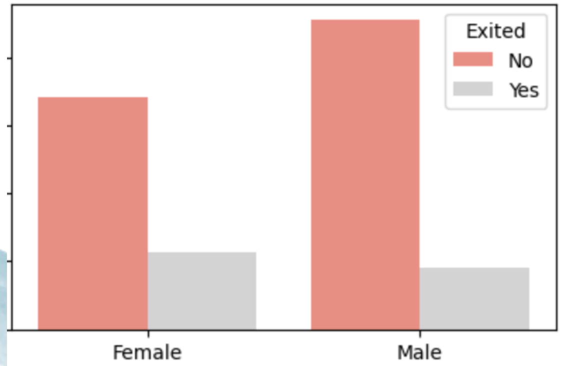
20% are churned



France has higher exited than other countries



Male tend to churn more than female



EDA

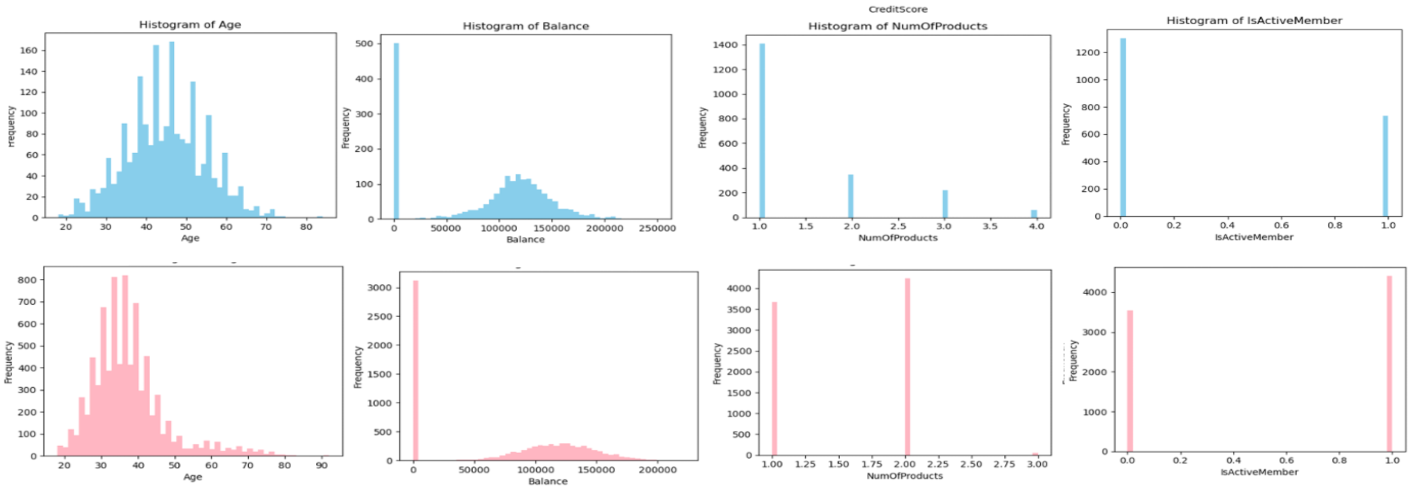
(Exploratory Data Analysis)

There are difference in

Age, Balance, Number of cards, active member **for churned and NOT churned**

[Churned]

[Not Churned]



For age, those who exited are predominantly 40-50 years old.

Balances are concentrated around $100K – 140K.

The number of products is mostly one, particularly among inactive members

Insights

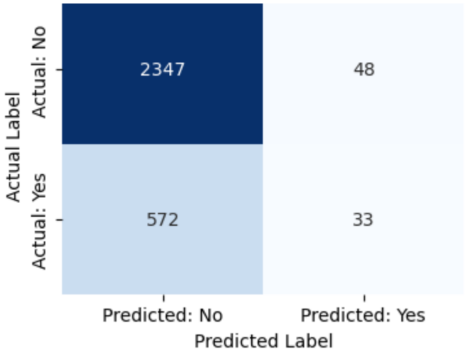
(Predictive Analysis)

1. Logistic Regression

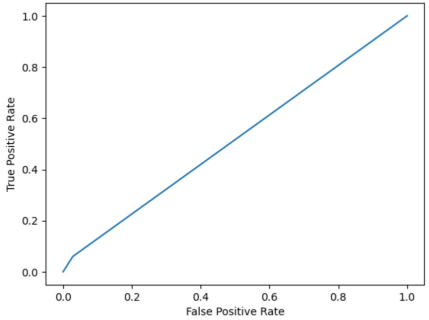
Recall/ Precision/ AUC

0.41 0.5 0.52

Confusion Matrix



ROC Curve



- Recall/ Precision/ AUC all around 0.5 which means that the model is not very accurate
- False Positives (FP): 48 instances were incorrectly predicted as "Yes" (Type 1 error)
- False Negatives (FN): 572 instances were incorrectly predicted as "No" (Type 2 error)

Insights

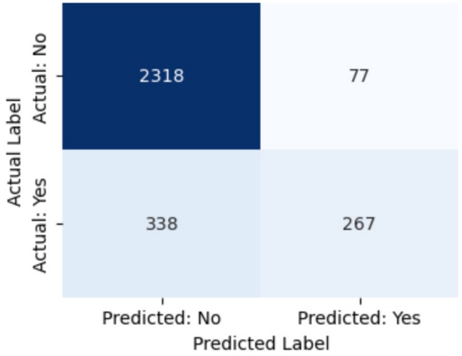
(Predictive Analysis)

2. Random Forest

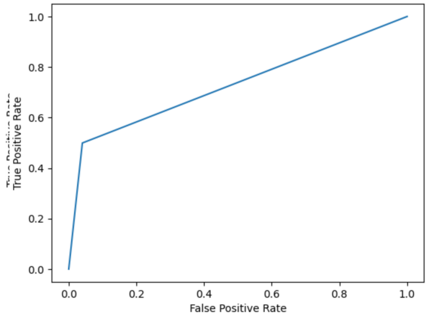
Recall/ Precision/ AUC

0.78   0.44   0.71

Confusion Matrix



ROC Curve



- Recall/ Precision/ AUC all around 0.7 indicating a fair level of distinction for classification
- False Positives (FP): 77 instances were incorrectly predicted as "Yes" (Type 1 error)
- False Negatives (FN): 338 instances were incorrectly predicted as "No" (Type 2 error)

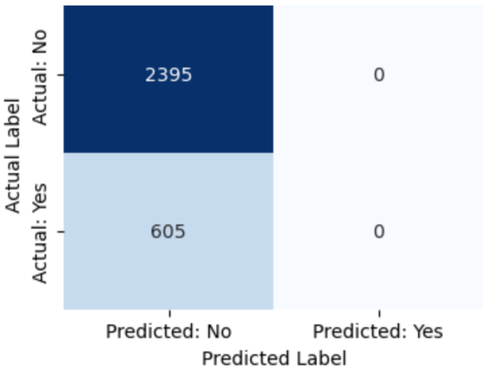
Insights
(Predictive Analysis)

3. Support Vector Machine

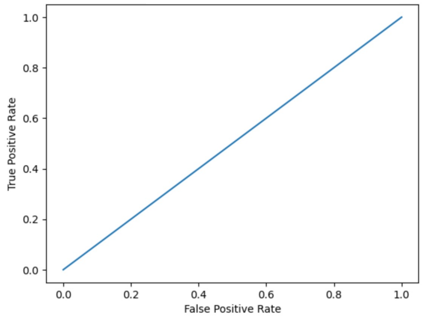
Recall/ Precision/ AUC

0.00 0.00 0.5

Confusion Matrix



ROC Curve



- Recall/ Precision are zero and AUC is 0.5 which means that the model has no ability to classify
- False Positives (FP): 0 instances were incorrectly predicted as "Yes" (Type 1 error)
- False Negatives (FN): 605 instances were incorrectly predicted as "No" (Type 2 error)

Insights

(Predictive Analysis)

- **Model Performance:** The Random Forest shows the best performance which can be used to predict churn in the future
- **Type 2 Error Analysis:** There are 338 case for the type 2 error, which means that the model might not be able to predict real churn as churn for 338 cases
- **Predictive Power and Business Impact**: The insights indicate that while the model can predict churn to a certain degree (evidenced by an AUC of 0.7 mentioned earlier), there is room for improvement to reduce the costs associated with false negatives, which in business terms means missed opportunities for customer retention.
- **Data Limitations and Potential:** The current limitation of having only 20% of data related to churn instances presents an insight that the model's predictive power could be significantly improved by increasing the volume of churn-related data.

Business Action

1) Focus on **specific features** associated with a high churn rate

(Geography(France), Gender(Male), Age group(40-50 years old), Balances($100K – 140K), Inactive members)

- Plan a retention program to keep the customers from the specific features
- In addition, make a targeted marketing plan for loyal customers

2) Utilize the **Random Forest model** to proactively predict and prevent customer churn

3) **Increase the data related to churn** to improve the model's accuracy (currently, only 20% of the data covers churn information)

Business Outcome

- **Improved Prediction and Retention:** By focusing on specific customer features associated with high churn rates, the business can target its retention efforts more effectively.

- **Mitigation of Type 2 Errors:** The presence of 338 Type 2 errors indicates that some actual churn cases might not be predicted. Actions to mitigate this could reduce the chance of missed churn predictions.

- **Proactive Churn Prevention:** Utilizing the Random Forest model for proactive churn prediction will enable the business to anticipate and address potential churn before it happens.

- **Data-Driven Marketing:** Creating a targeted marketing plan for loyal customers could increase customer lifetime value and reinforce brand loyalty, possibly turning them into advocates for the company.

- **Model Refinement with More Data:** This could lead to fewer false negatives and positives, making the model more reliable and saving resources by avoiding misdirected retention efforts.

- **Competitive Advantage:** Continuously improving the model and its predictions can provide a competitive advantage by retaining valuable customers, improving satisfaction, and maintaining a steady revenue flow.

Summary

- The Random Forest model outperformed others, suggesting it's well-suited for this dataset.

- Through EDA, highlighted key customer features that influence churn, discovered, such as geography, gender, age, account balances, and activity levels.

- A notable number of Type 2 errors (false negatives) were identified, indicating areas where the model could be missing true churn cases.

- Insights suggest the need for a more balanced dataset, as currently only 20% pertains to churn, which limits the model's learning potential.

- The findings can directly inform targeted retention strategies and proactive churn prevention measures, tailoring efforts to customer segments most at risk.

- The project underscores the importance of a proactive approach to customer retention through data-driven strategies, highlighting opportunities for further data enhancement and model refinement to better predict and prevent customer churn.

Thank You!