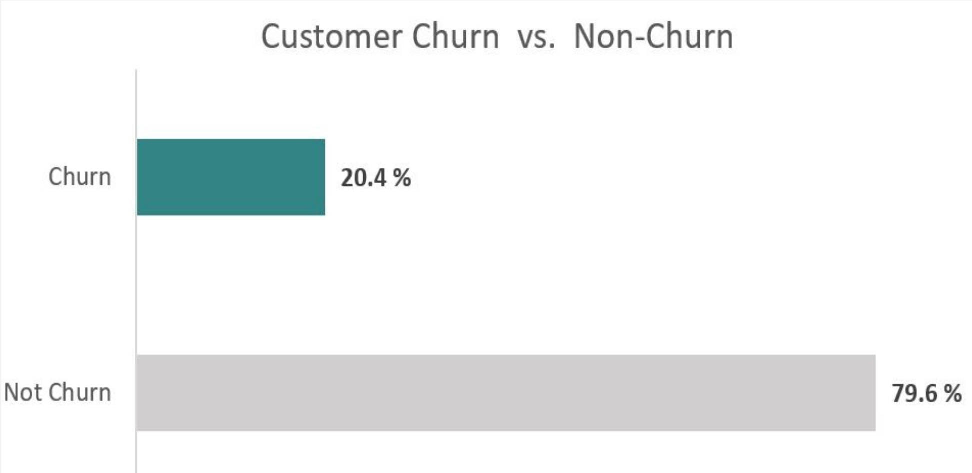
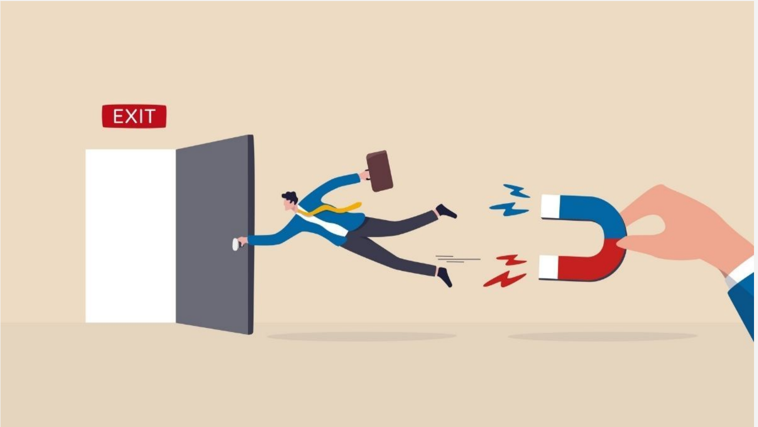
PREDICTING BANK CUSTOMER CHURN

CHETANA VYAS

Why does it matter?





Avoidable Churn

- Poor Customer Service
- Rigid Pricing
- Security Threats
- Complicated Interface

Involuntary Churn

- Closing the business
- Outdated Equipments

INITIAL EDA

Customer Profiles that tend to Churn

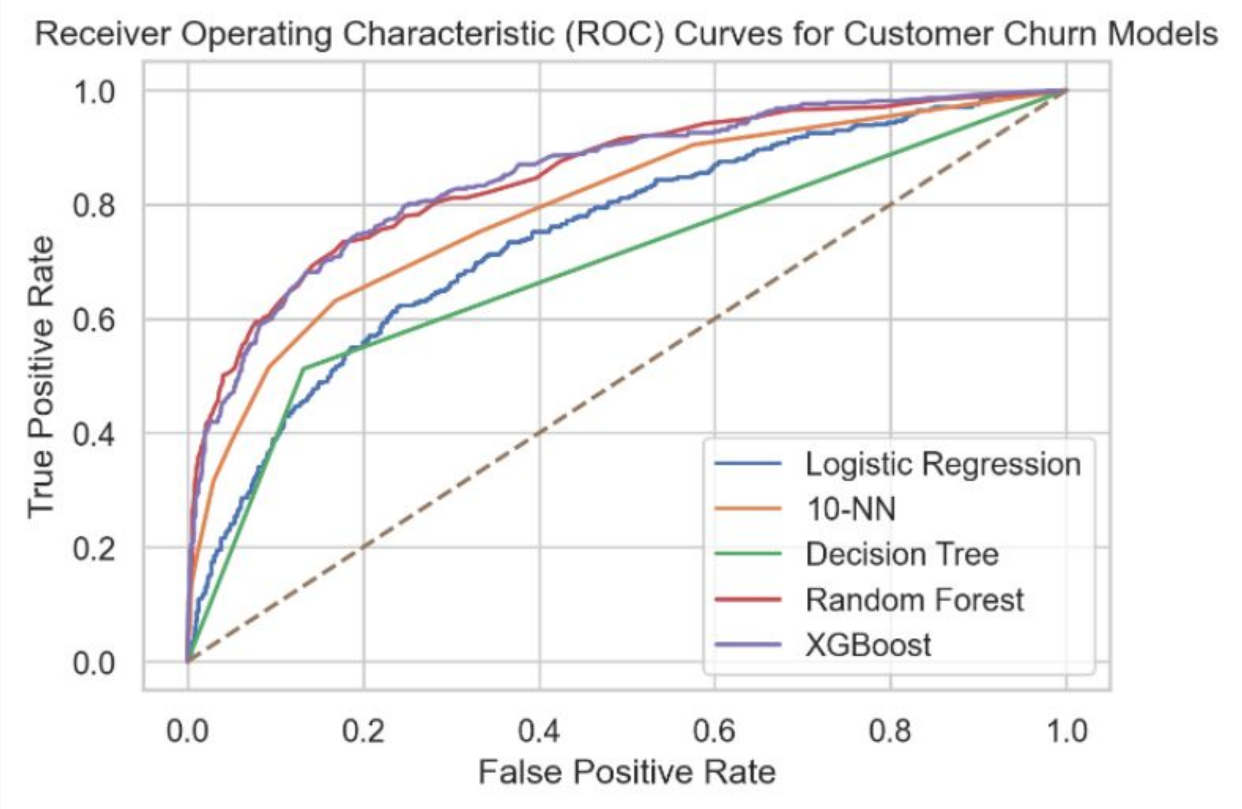
- Age
- Number Of Products
- Germany
- Female

Baseline Model using Logistic Regression

F_Beta Score=0.80
Beta = 2.5
(prefer recall)

Recall = 0.21

Classification Model Comparisons (ROC AUC)



- ➢ Random Forest
- ➢ XGBoost (Extreme Gradient Boosting)

Classification Model Comparisons (F Beta Score)

- ➢ Performance Metric : F_beta with beta = 2.5

- ➢ XGBoost (Extreme Gradient Boosting)
- ➢ Random Forest

Extreme Gradient Boosting

F_Beta Score=0.89
(Beta = 2.5)

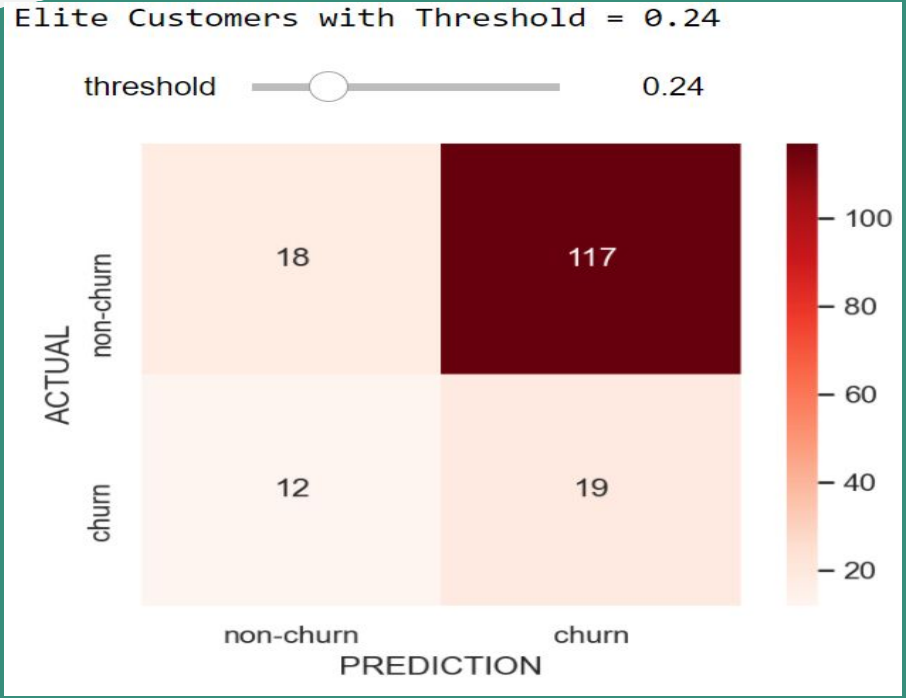
Recall = 0.54

EXAMPLES

ELITE CUSTOMERS

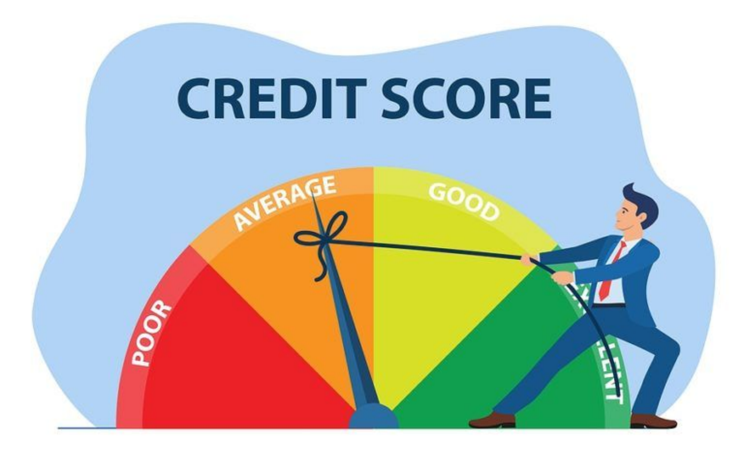
PRECISION = 14%

RECALL = 62%



Scoring the Random Forest Model on customers having

- ➢ **Bank balance > 100K**
- ➢ **Credit Score > 750**



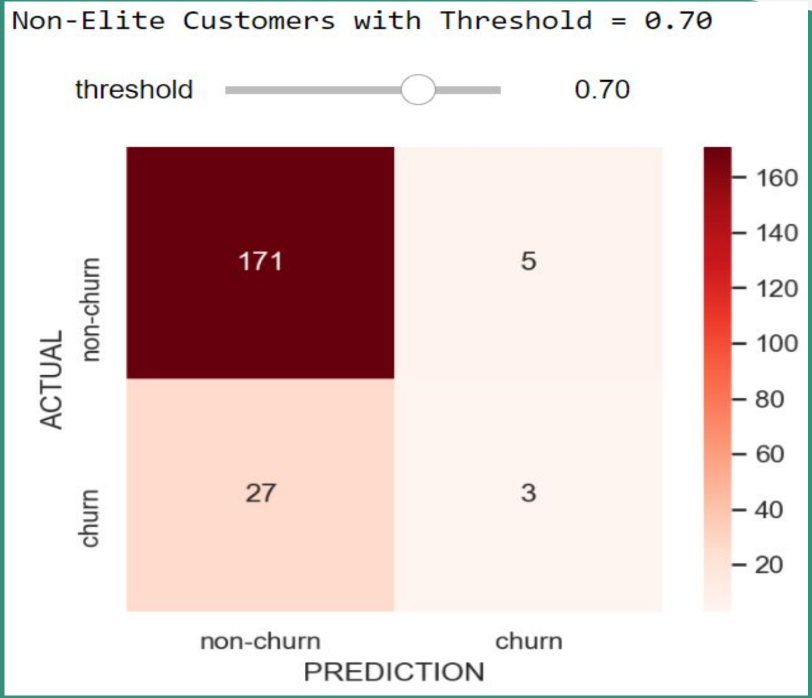
NON - ELITE CUSTOMERS

PRECISION = 38%

RECALL = 10%

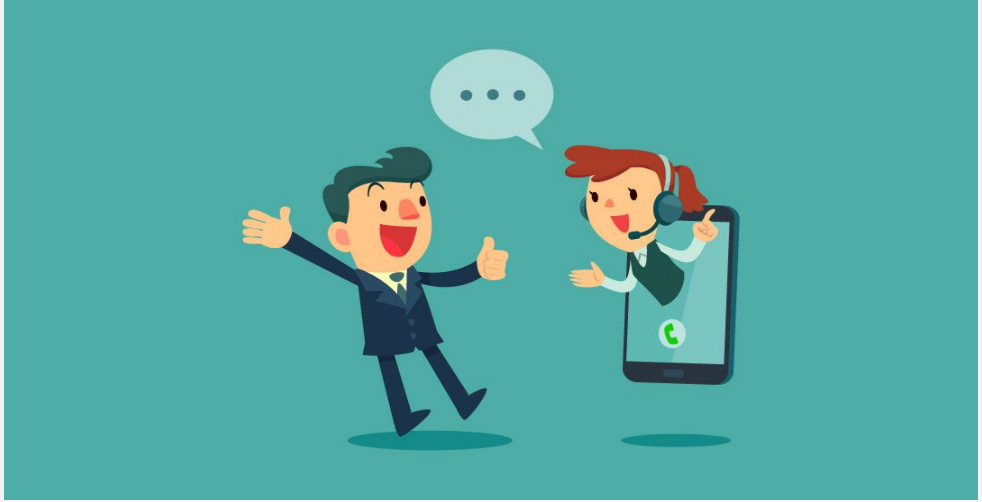
Scoring the Random Forest Model on customers having

- ➢ **Bank balance < 10K**
- ➢ **Credit Score < 600**



SO WHAT DO WE DO WITH OUR GREAT MODEL?

RECOMMENDATIONS



DAILY CHURN DETECTION

Build powerful Machine Learning Models to analyze customer behaviour

CONTINUOUS OPTIMIZATION

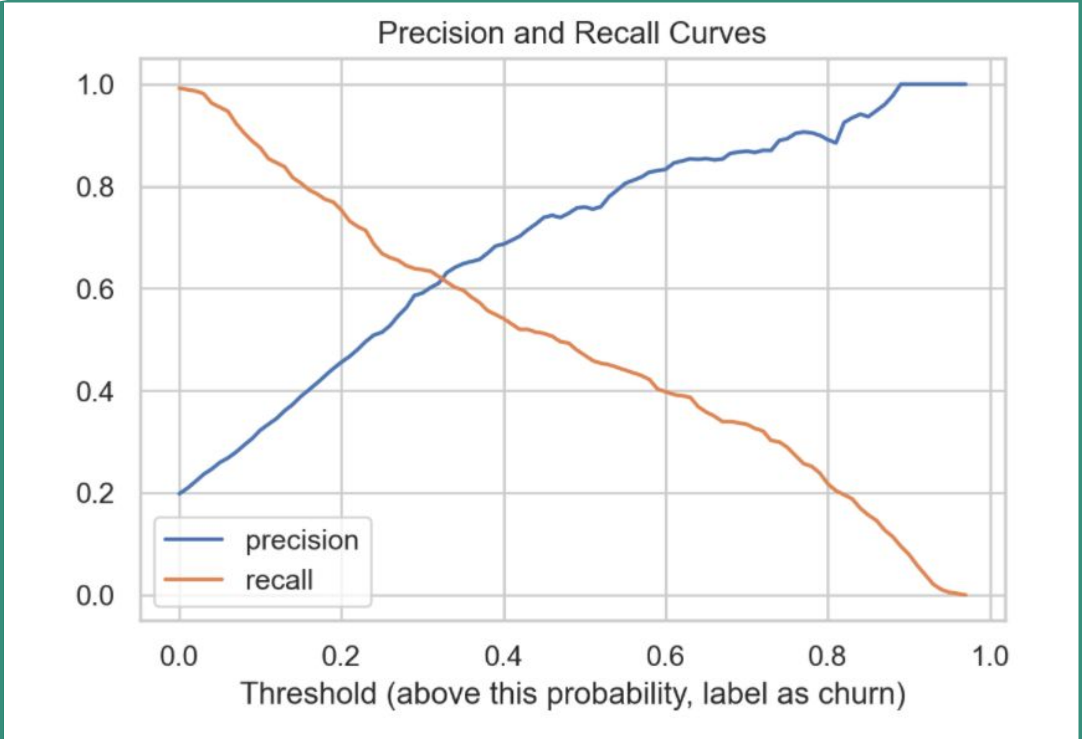
On-demand access to predicted customers at risk of churning

THANKS

Chetana Vyas

APPENDIX - Precision & Recall Curve

Random Forest Classifier



APPENDIX - Pairplot

TOOLS

NumPy

matplotlib

Pandas

scikit
learn

Seaborn

CLASSIFICATION ALGORITHMS

01
Logistic Regression

02
K - Nearest Neighbours

03
Decision Trees

04
Random Forest

05
XG Boost