

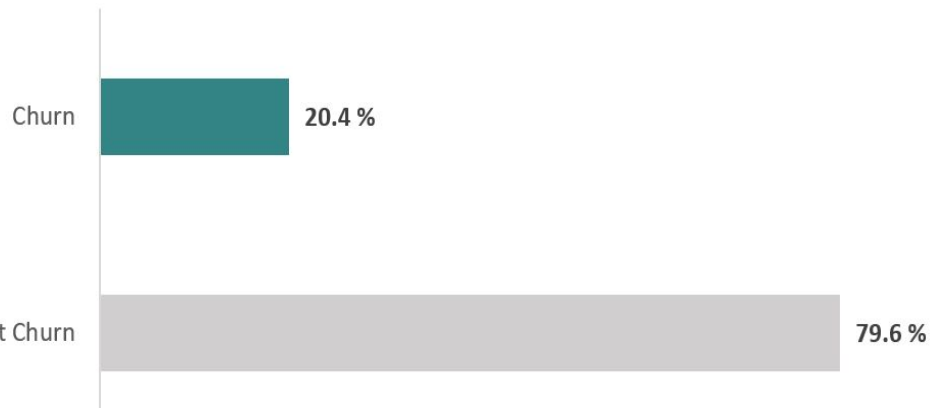
PREDICTING BANK CUSTOMER CHURN

CHETANA VYAS



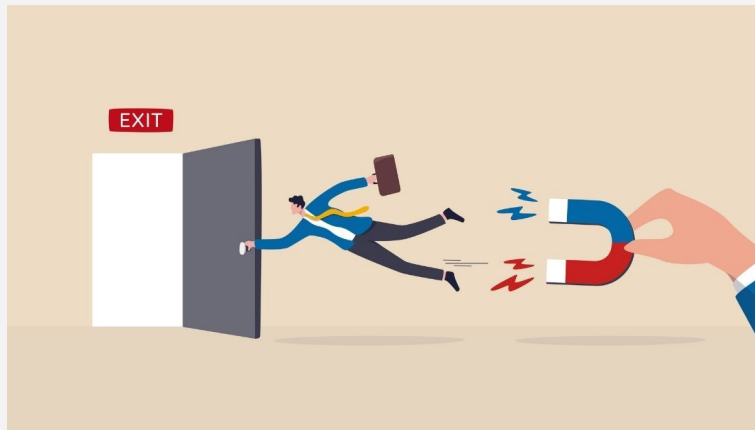
Why does it matter?

Customer Churn vs. Non-Churn



Involuntary Churn

- Closing the business
- Outdated Equipments



Avoidable Churn

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

INITIAL EDA



Customer Profiles that tend to Churn

- Age
- Number Of Products
- Germany
- Female

Baseline Model using Logistic Regression

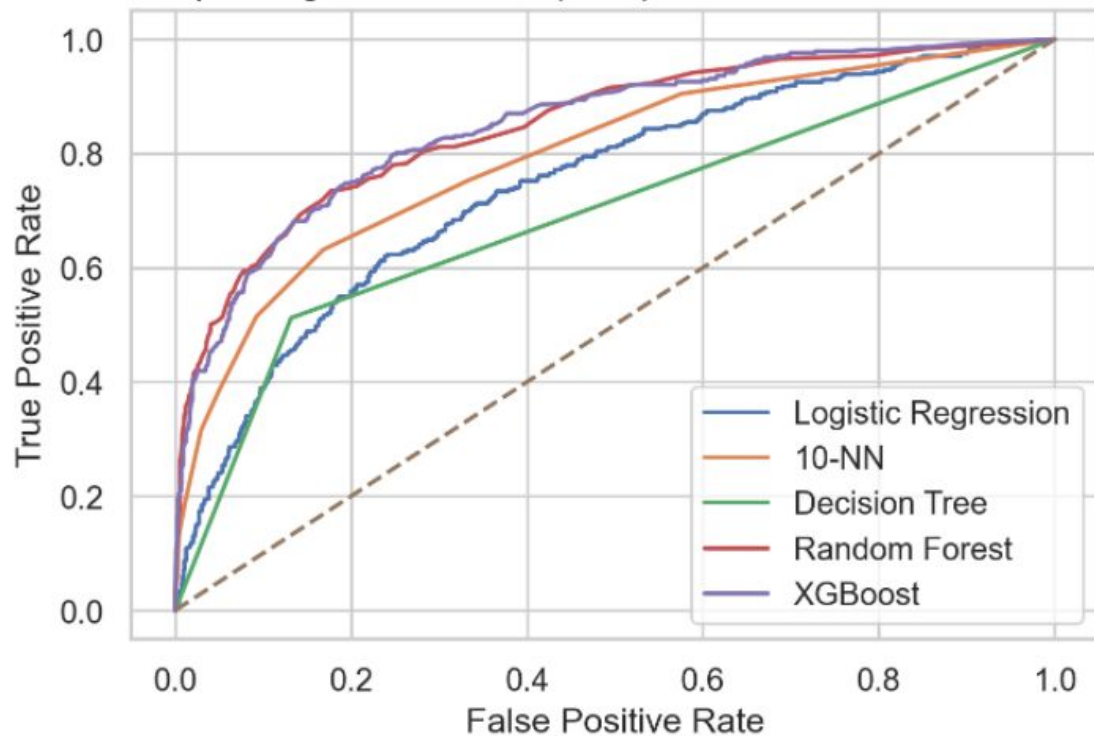
F_Beta Score=0.80
Beta = 2.5
(prefer recall)

Recall = 0.21



Classification Model Comparisons (ROC AUC)

Receiver Operating Characteristic (ROC) Curves for Customer Churn Models



- 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

Decision Tree

0.8023

Logistic Regression

0.8065

KNN

0.8363

Random Forest

0.8675

XGBoost

0.8891



Extreme Gradient Boosting

F_Beta Score=0.89
(Beta = 2.5)

Recall = 0.54



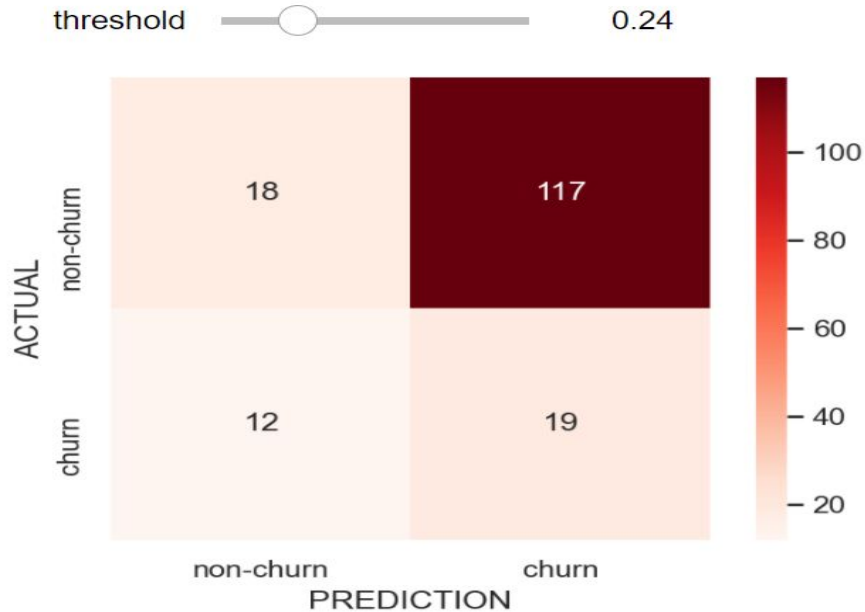
EXAMPLES

ELITE CUSTOMERS

PRECISION = 14%

RECALL = 62%

Elite Customers with Threshold = 0.24



Scoring the Random Forest Model on customers having

- Bank balance > 100K
- Credit Score > 750

CREDIT SCORE



NON - ELITE
CUSTOMERS

PRECISION = 38%

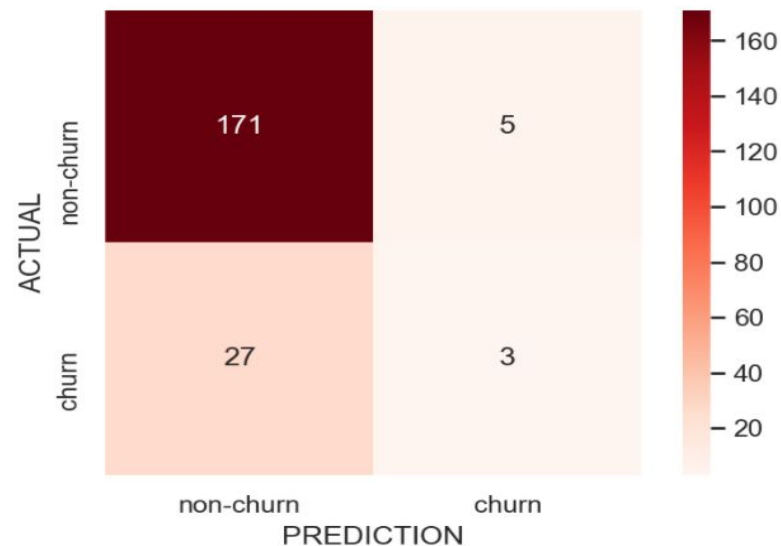
RECALL = 10%

Scoring the Random Forest Model on
customers having

- Bank balance < 10K
- Credit Score < 600

Non-Elite Customers with Threshold = 0.70

threshold 0.70



**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



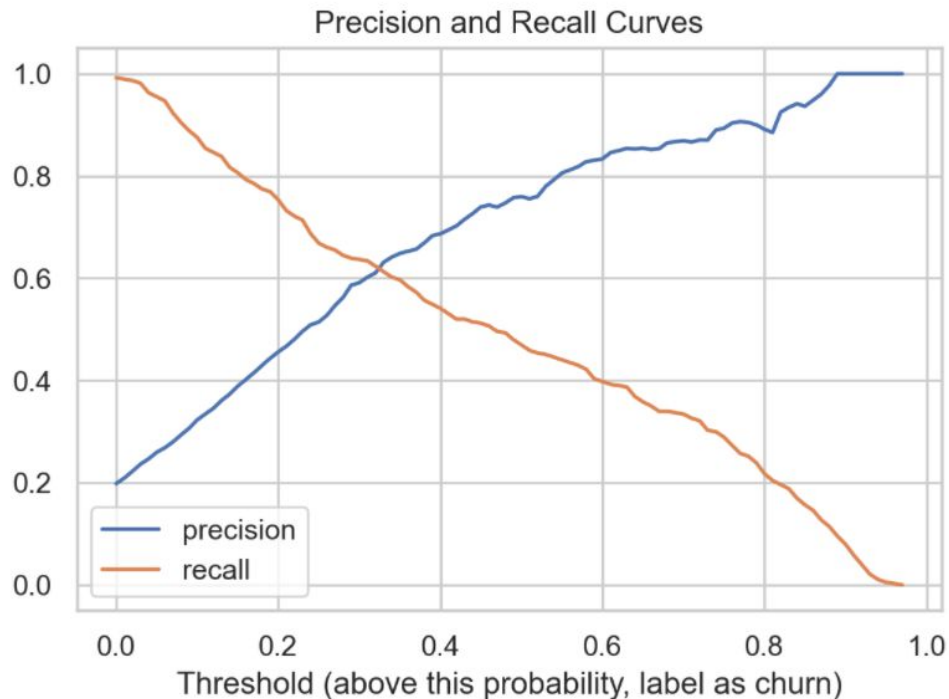
THANKS

Chetana Vyas



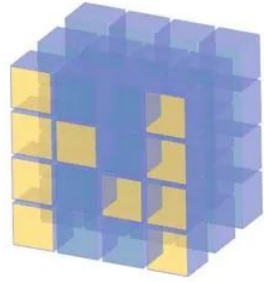
APPENDIX - Precision & Recall Curve

Random Forest Classifier



The figure displays a 10x10 grid of plots, where each row represents a different variable and each column represents the 'Survived' variable. The variables are: Survived, Age, Sex, Pclass, SibSp, Parch, Fare, Cabin, Embarked, and Name. The plots show the distribution and relationship of each variable with survival status. The variables are ordered as follows: Survived, Age, Sex, Pclass, SibSp, Parch, Fare, Cabin, Embarked, and Name. The plots include histograms, scatter plots, box plots, and density plots, illustrating the distribution and relationship of each variable with survival status.

TOOLS



NumPy

matplotlib



Pandas



Seaborn

CLASSIFICATION ALGORITHMS

01

Logistic Regression

02

K - Nearest Neighbours

03

Decision Trees

04

Random Forest

05

XG Boost