

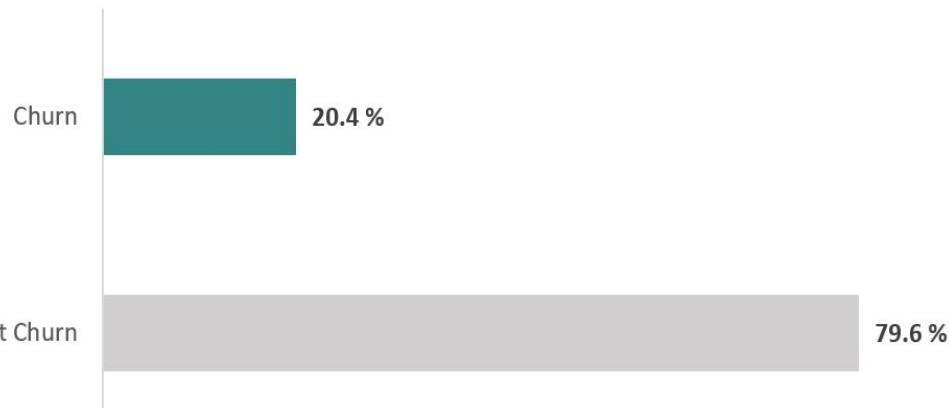
PREDICTING BANK CUSTOMER CHURN

CHETANA VYAS



Why does it matter?

Customer Churn vs. Non-Churn



Involuntary Churn

- Closing the business
- Outdated Equipments



Avoidable Churn

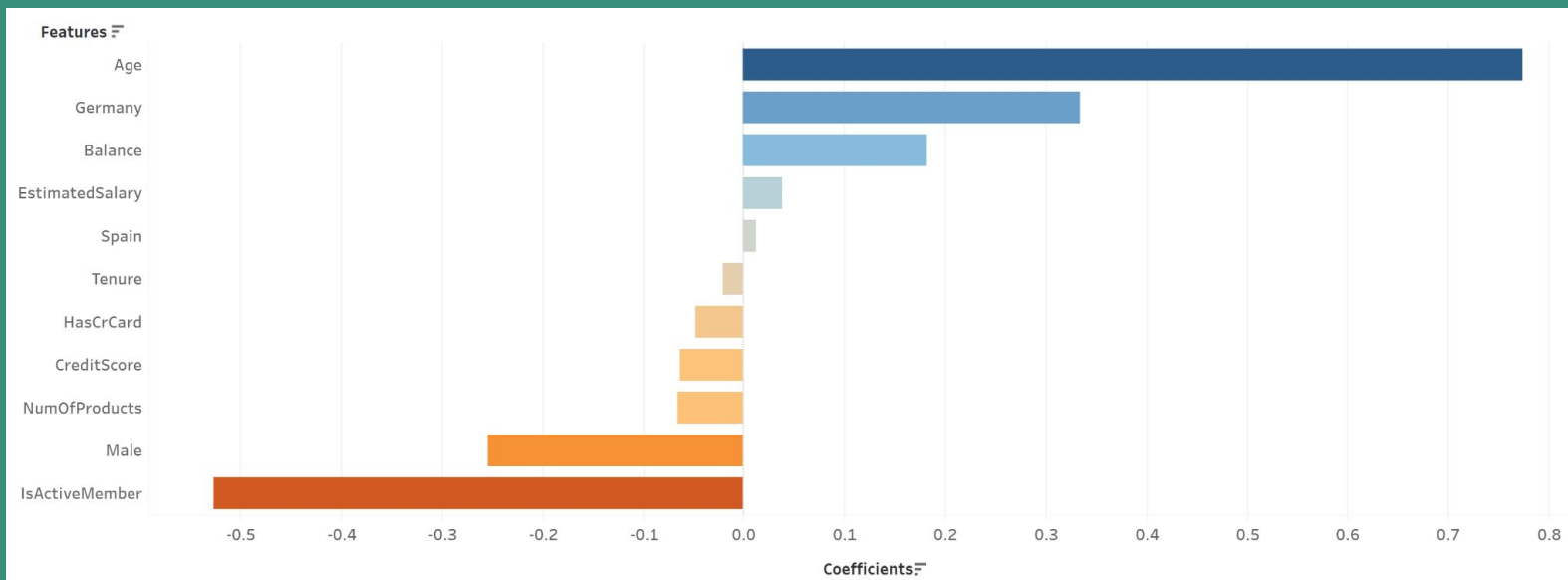
- Poor Customer Service
- Rigid Pricing
- Security Threats
- Complicated Interface

Baseline Model - Logistic Regression

F_Beta Score=0.80
Beta = 2.5
(prefer recall)

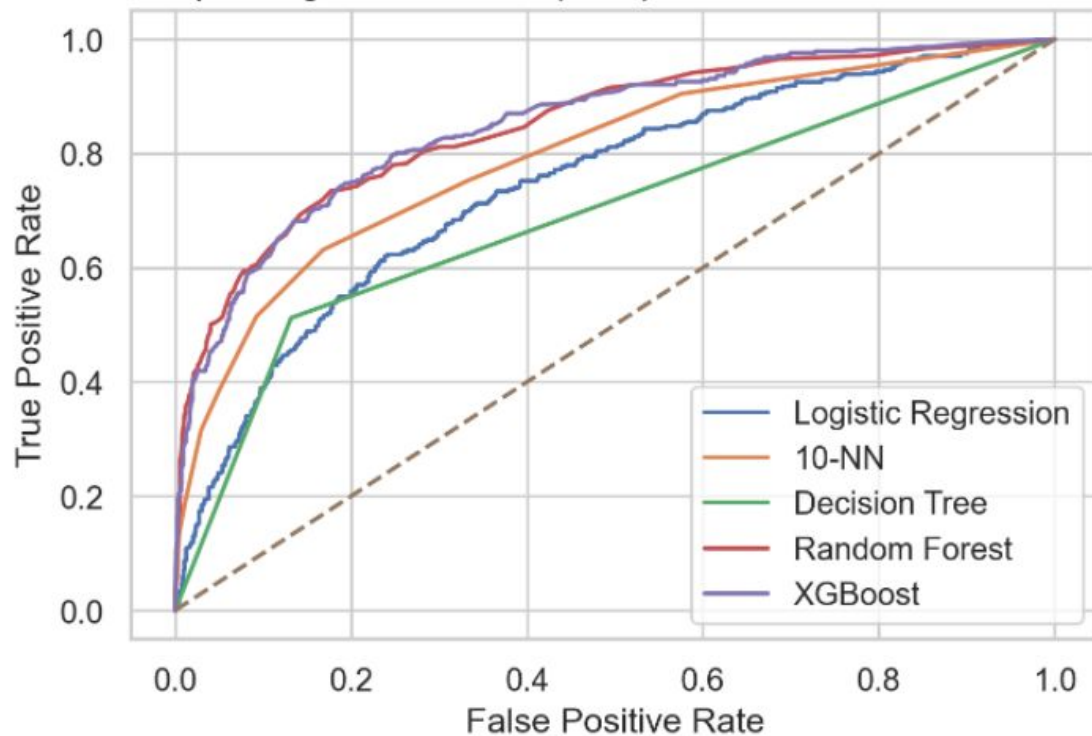
Recall = 0.21

FEATURE IMPORTANCE



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



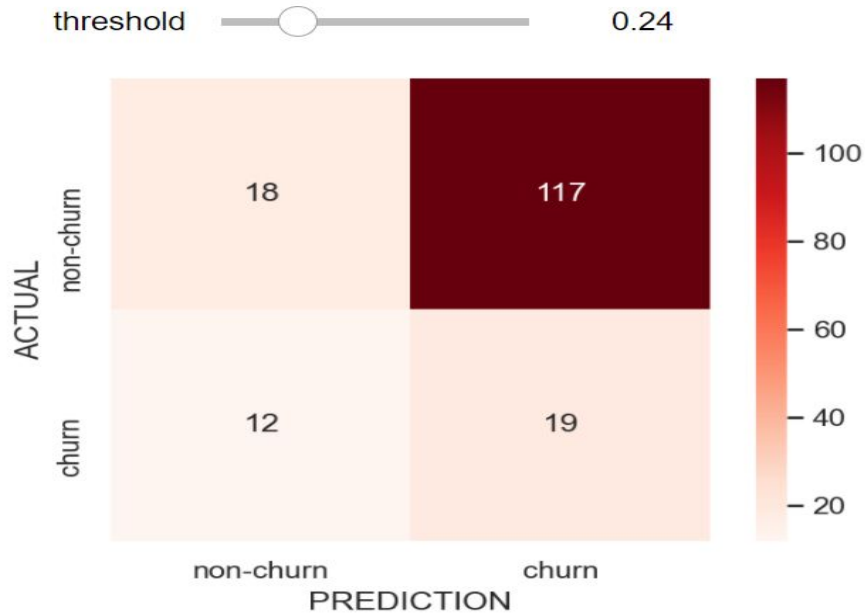
EXAMPLES

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%

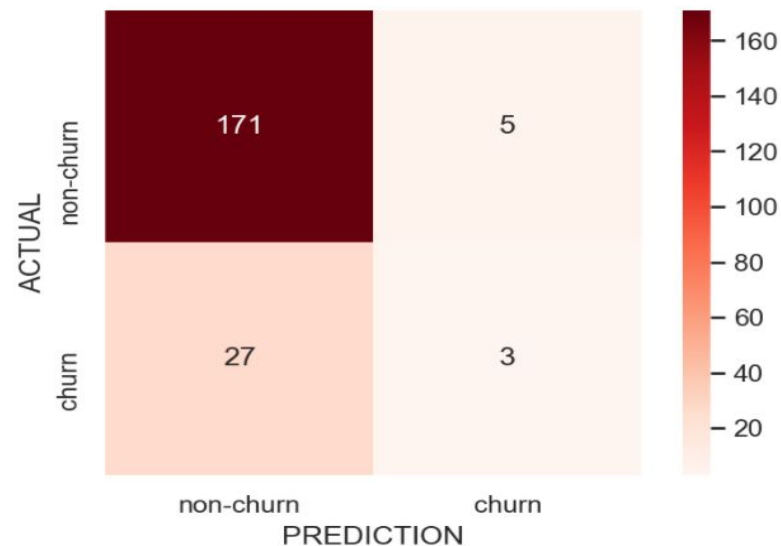
RECALL = 10%

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 BINARY
CLASSIFICATION MODEL?**

RECOMMENDATIONS

DAILY CHURN DETECTION

Build powerful
Machine Learning
Models to analyze
customer behaviour

CHURN PREVENTION

Loyalty & Retention
Programs

CONTINUOUS OPTIMIZATION

On-demand access
to predicted
customers at risk of
churning



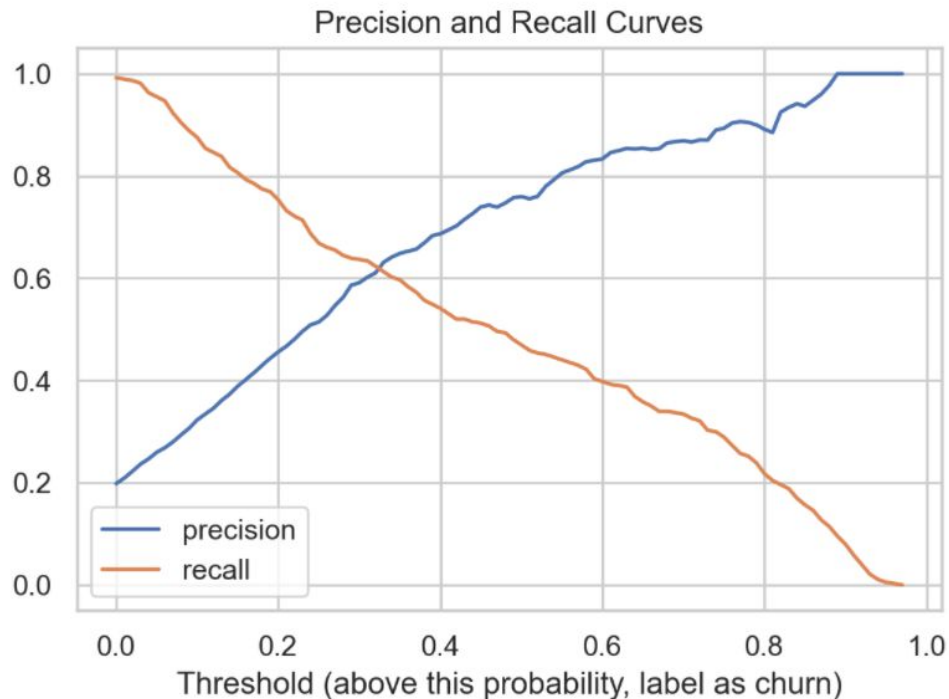


THANKS

- Chetana Vyas

APPENDIX - Precision & Recall Curve

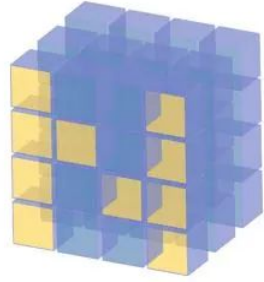
Random Forest Classifier



APPENDIX - Row data

Exited	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Germany	Spain	Male
1	619	42	2	0.00	1	1	1	101348.88	0	0	0
0	608	41	1	83807.86	1	0	1	112542.58	0	1	0
1	502	42	8	159660.80	3	1	0	113931.57	0	0	0
0	699	39	1	0.00	2	0	0	93826.63	0	0	0
0	850	43	2	125510.82	1	1	1	79084.10	0	1	0
...
0	771	39	5	0.00	2	1	0	96270.64	0	0	1
0	516	35	10	57369.61	1	1	1	101699.77	0	0	1
1	709	36	7	0.00	1	0	1	42085.58	0	0	0
1	772	42	3	75075.31	2	1	0	92888.52	1	0	1
0	792	28	4	130142.79	1	1	0	38190.78	0	0	0

TOOLS



NumPy

matplotlib



Pandas



Seaborn

CLASSIFICATION ALGORITHMS

01

Logistic Regression

02

K - Nearest Neighbours

03

Decision Trees

04

Random Forest

05

XG Boost