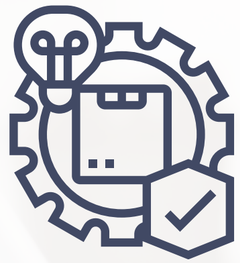
CHURN & RETENTION IN BANKING – A CASE STUDY

Big Data & MLOps
Final Project

Submitted By,
Shubham Gaddi

PROBLEM STATEMENTS



Customer Demographic Identification & Churn Modelling

Customer Segmentation for Churn Re-Entry Pilot Progrm

Ad Campaign Performance for Customer Retention Program

DEMOGRAPHIC Identification

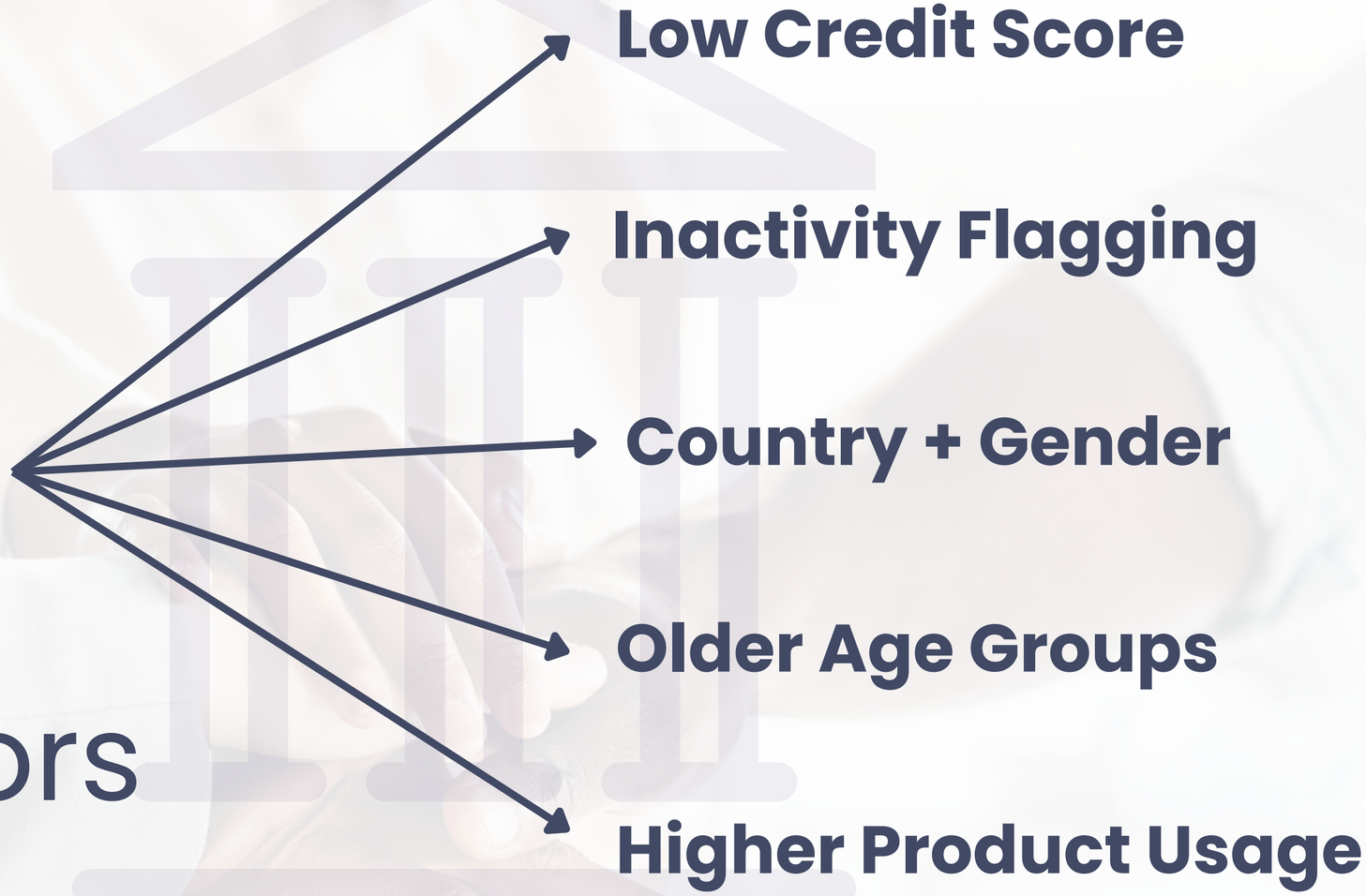
1. **Bank’s Target Customer Demographic**
2. **At-Risk Demographic - Churn Likelihood**
3. **Rule Based Screening Method for Churn**

CUSTOMER DEMOGRAPHICS

Segment/Demographic	Most Common Value (All Customers)	Highest Churn Proportion Segment
Credit Score	600 - 700	< 400
Country	France	Germany
Age	30 - 35	45+
Active Customer Status	Yes	No
No. of Products/Services	1	2+
Gender	Male	Female

RULE BASED CHURN MODELLING

Key Differentiators



MACHINE LEARNING FOR CHURN PREDICTION

Tree Based Models

Rationale

- No need for One Hot Encoding
- Easy to Ensemble and Aggregate
- Wide control over Hyperparameters
- Dynamic and Performant

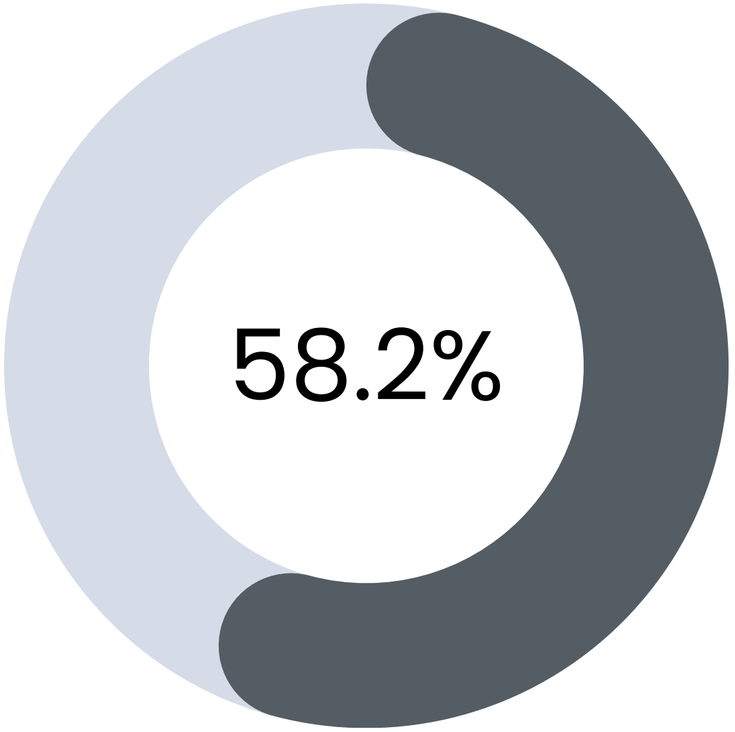
Models Trained

- Decision Tree (Singular Tree)
- Gradient Boosted Trees (Boosting)
- Random Forests (Bagging)

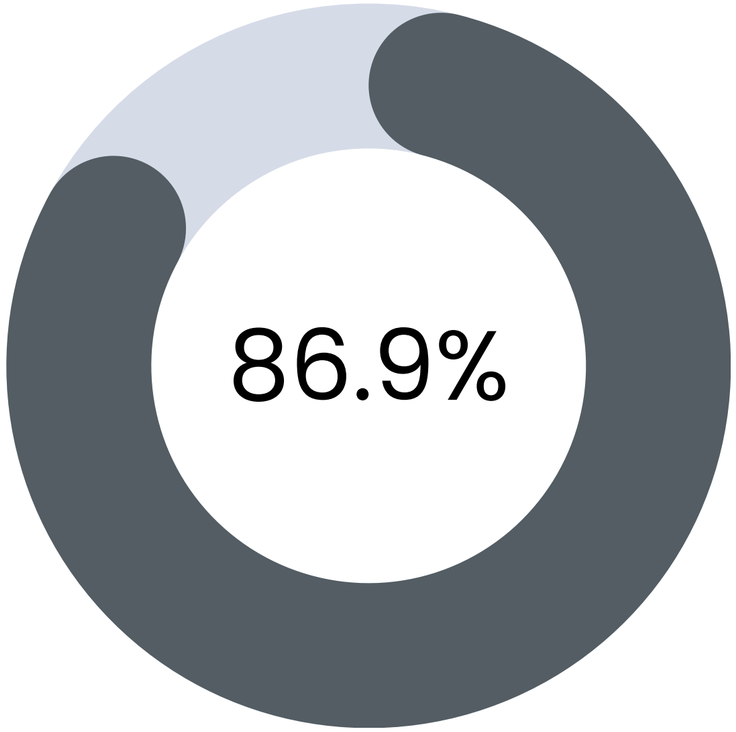
Evaluation

- Area Under ROC Curve
- Higher Score = Better Performance

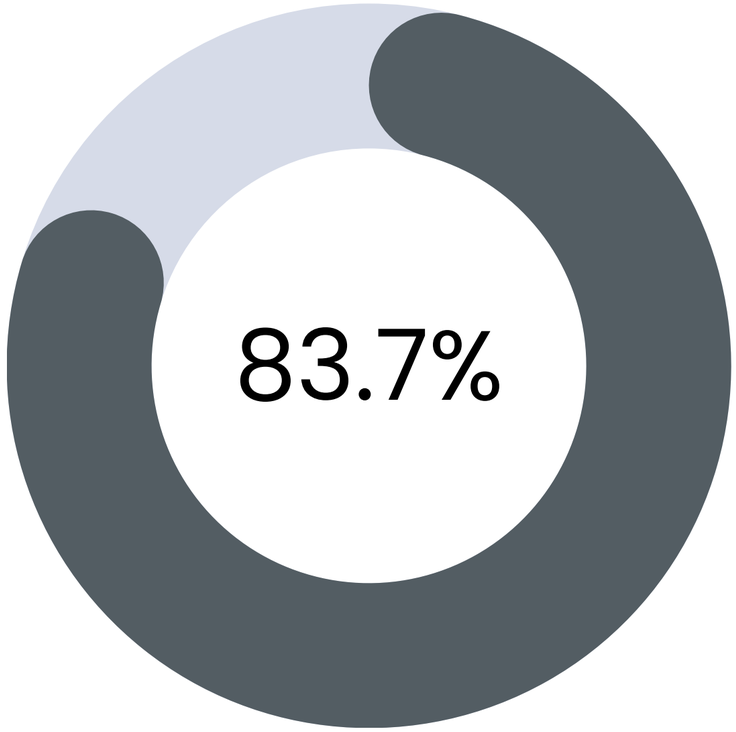
COMPARING MODEL PERFORMANCE



Decision Tree



Gradient Boosted Tree

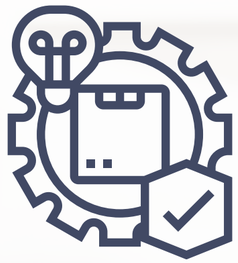


Random Forest

CUSTOMER Segmentation

- **Churned Customer Data - Income & Spending**
- **Identify Customer Segments for Re-Entry Pilot Program**
- **Suggest High Value Segments as Targets**

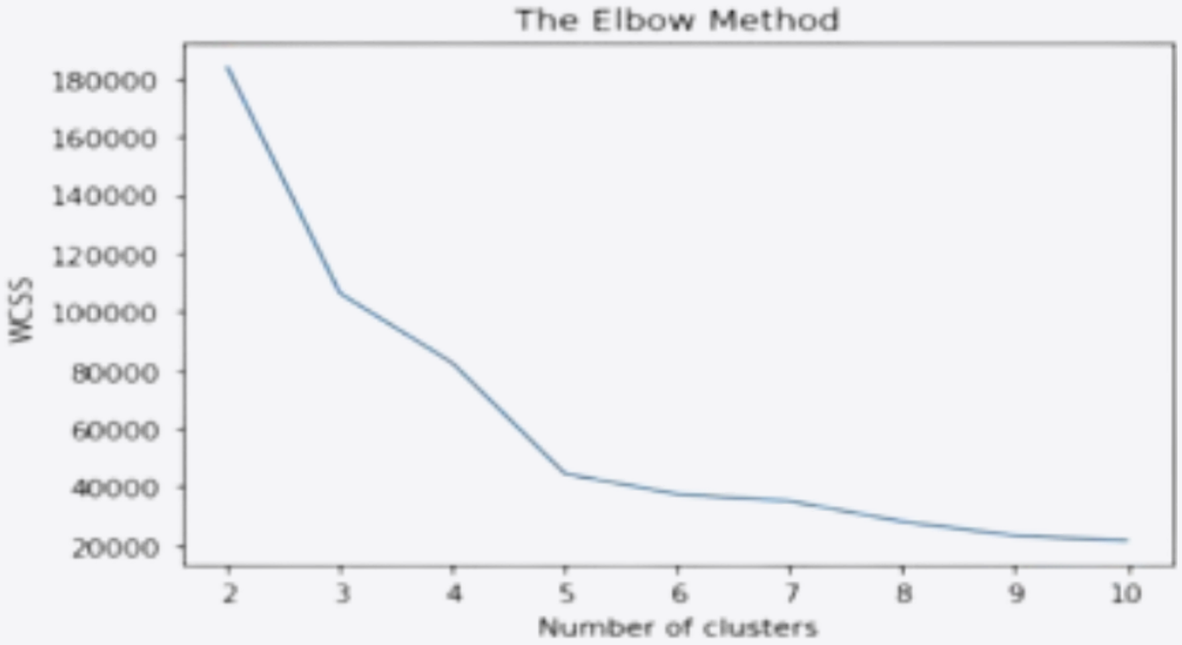
IDENTIFYING CUSTOMER SEGMENTS



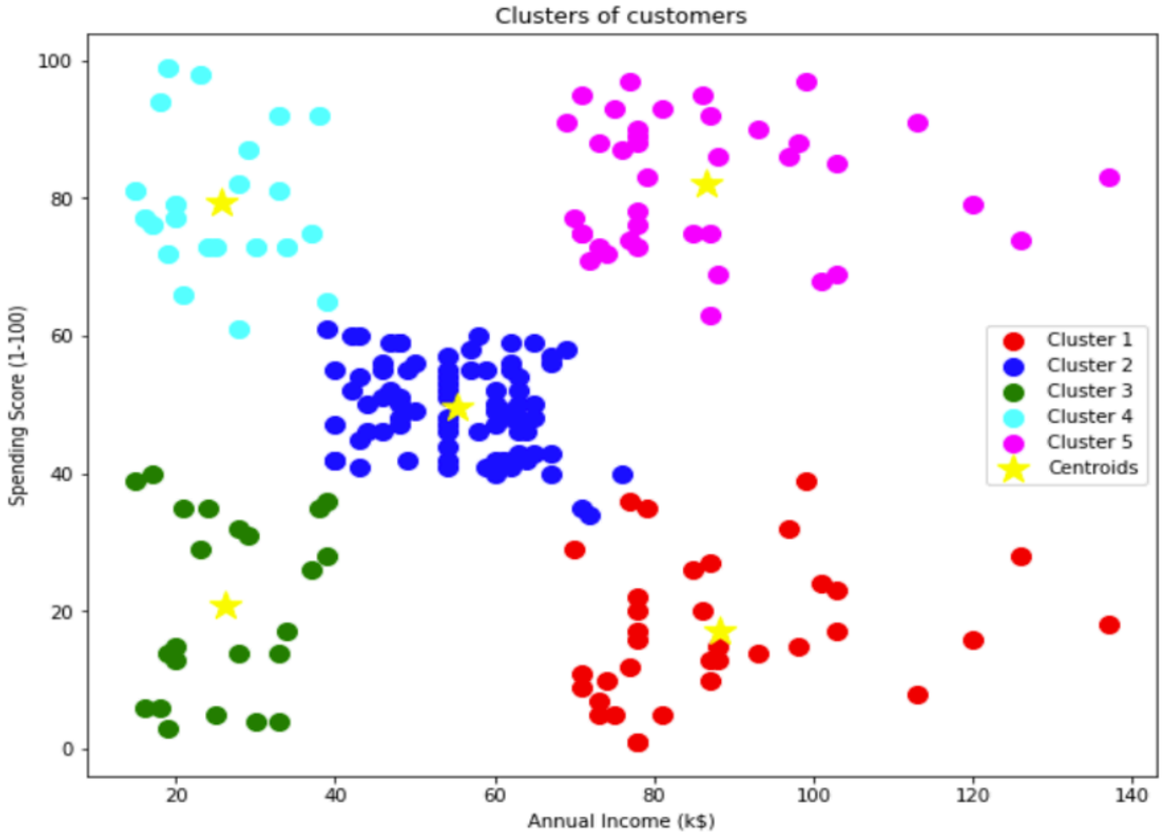
Identify Optimal Number of Segments - Elbow Method

Customer Clustering with Optimal Segments - K Means

OPTIMAL CLUSTER COUNT



CONSUMER CLUSTERING



HIGH VALUE CLUSTERS

High Income + High Spending

- Good candidates for short term lines of credit - Credit Card
- Good Candidates for long term lines of credit - Loans

Low Income + High Spending

- Good candidates for short term lines of credit - Credit Card
- Potential for missed payments - Interest on Credit Card payments

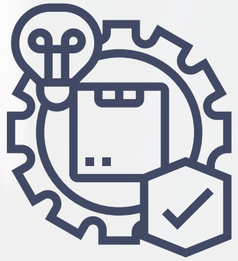
AD CAMPAIGN PERFORMANCE

- Data for 10 Ads displayed to 10,000 users
- Boolean Variables for Each User <> Ad Interaction
- Identify Best Performing Ad in Minimum Iterations

COMPLICATION
Explained

Running 10 Ad Campaigns Expensive

Simultaneous Exploration & Exploitation of Data Distributions



Multi Armed Bandit Problem - Reinforcement Learning

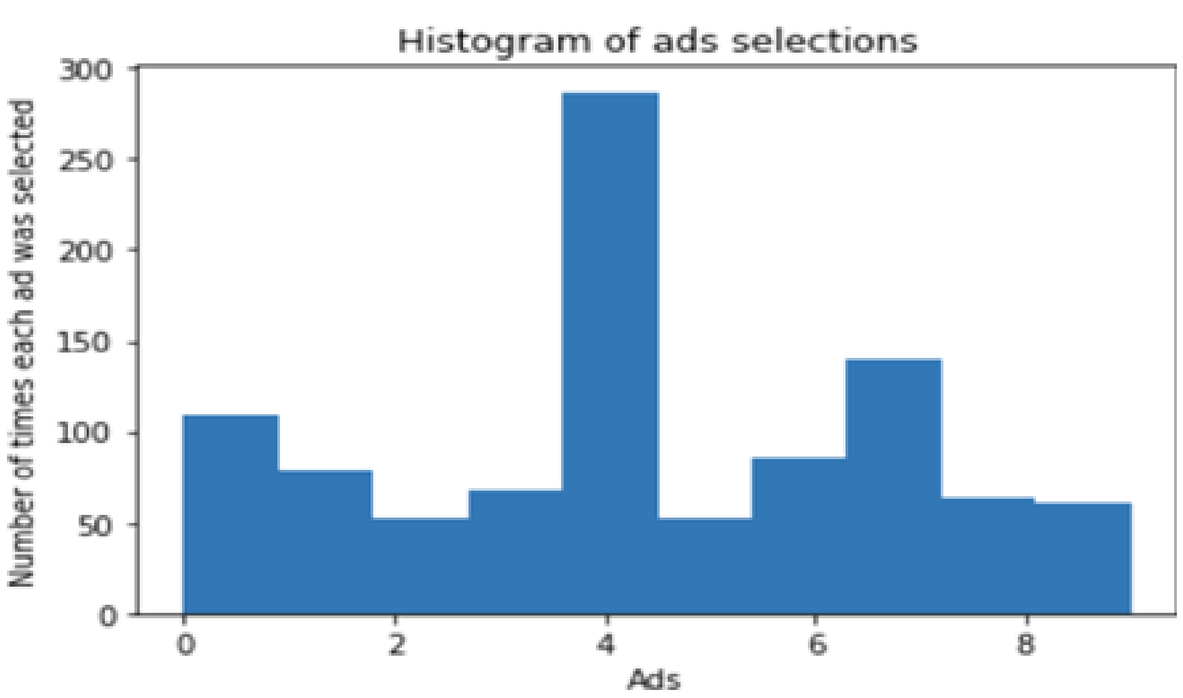
UPPER CONFIDENCE BOUND

- Assumes a starting point and expected value
- Created Upper Confidence Bound based on performance
- Highest UCB Chosen
- Probabilistic Model
- Min Iterations to Converge = 1000

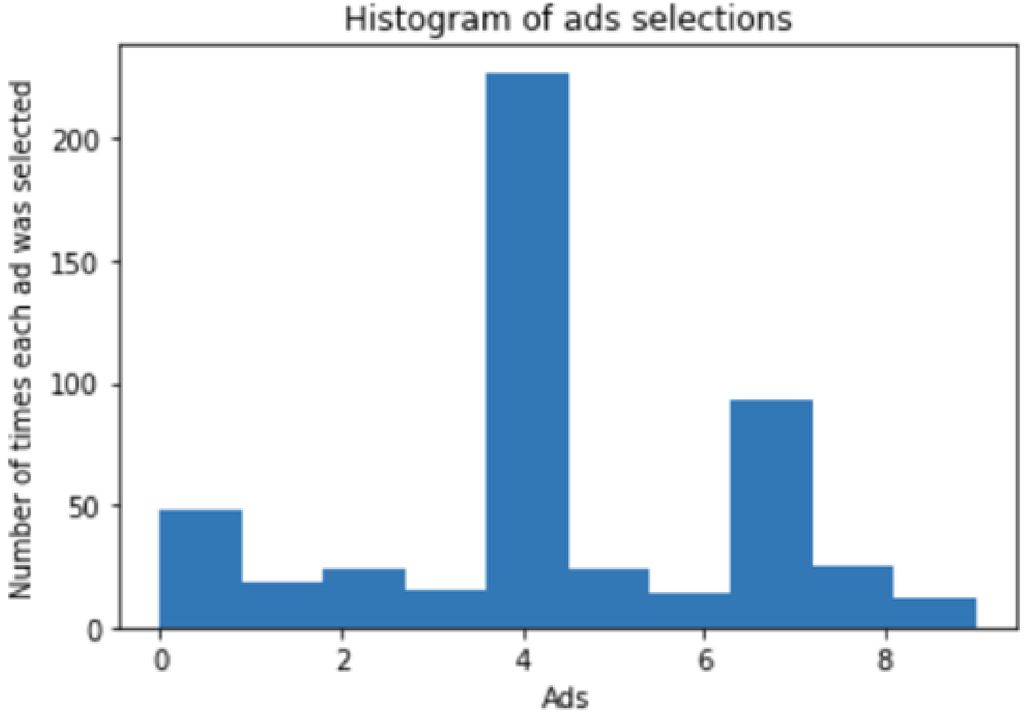
THOMPSON SAMPLING

- Based on the concept of Bayesian Inference
- Rewards best Ad at every iteration based on a Bayesian Distribution
- Picks the Ad with the Highest Reward
- Deterministic Model
- Min Iterations to Converge = 500

MEASURING CONVERGENCE SPEED



UCB Best Performance



TS Best Performance

SCOPE FOR IMPROVEMENT

- Hyperparameter Tuning for Churn Prediction Modelling
- Using KNN++ to Avoid Random Initialization Trap
- Fine Tune Reward Mechanism for Quicker Convergence