MIT | Academy of Engineering

(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

CI Practical Project Demonstration

Bank Churning Prediction

Guided By Prof Smita Kulkarni

Group Members

Alvin Abraham	202201070132
Nishant Rajat	202201070131
Chinmay Parite	202201070134
Ritesh Rodge	202201070121

Contents

Introduction

- Bank churning refers to the process where customers frequently open and close bank accounts or switch between financial institutions.

- This often happens due to dissatisfaction with services, better offers from competing banks, or changes in financial needs.

- Churning can also describe a bank's efforts to replace departing customers with new ones, reflecting customer turnover rates.

Importance of Bank Churn Prediction for Banks

- Early Identification of At-Risk Customers:
- Improved Customer Retention
- Enhanced Customer Experience
- Personalized Offers and Services
- Optimized Marketing Costs
- Stronger Customer Relationships
- Better Financial Planning
- Competitive Advantage
- Insights into Customer Behavior

1. Data Preprocessing

- **Dropped irrelevant columns:** Removed columns like RowNumber, CustomerId, and Surname.
- **Checked for missing values:** data_cleaned.isnull().sum() to confirm the absence of null values.
- **Encoded categorical features:**
- Used label encoding for Gender (Male = 1, Female = 0).
- Applied one-hot encoding for Geography with pd.get_dummies().

2. Exploratory Data Analysis (EDA)

- **Distribution visualizations:**
- Used **countplot** for the churn distribution.
- Displayed gender distribution with a **pie chart.**
- **Scatter plot** for Age vs. Balance by churn.
- **Pairplot** to explore relationships between numerical features.
- **Correlation Heatmap:** Visualized feature correlations using a heatmap.

3. Feature Scaling

Standardized numerical features like

- CreditScore
- Age
- Balance
- EstimatedSalary using StandardScaler.

ML Models USED

Logistic Regression

Predicts the probability of customer churn to enable targeted retention strategies.

Decison Tree

Identifies clear rules like income or tenure thresholds that lead to churn for actionable insights.

Naive Bias

Quickly classifies churn likelihood for customers with independent feature relationships like geography and gender.

Random Forest

Handles diverse customer data to pinpoint key churn drivers and improve prediction reliability.

KNN

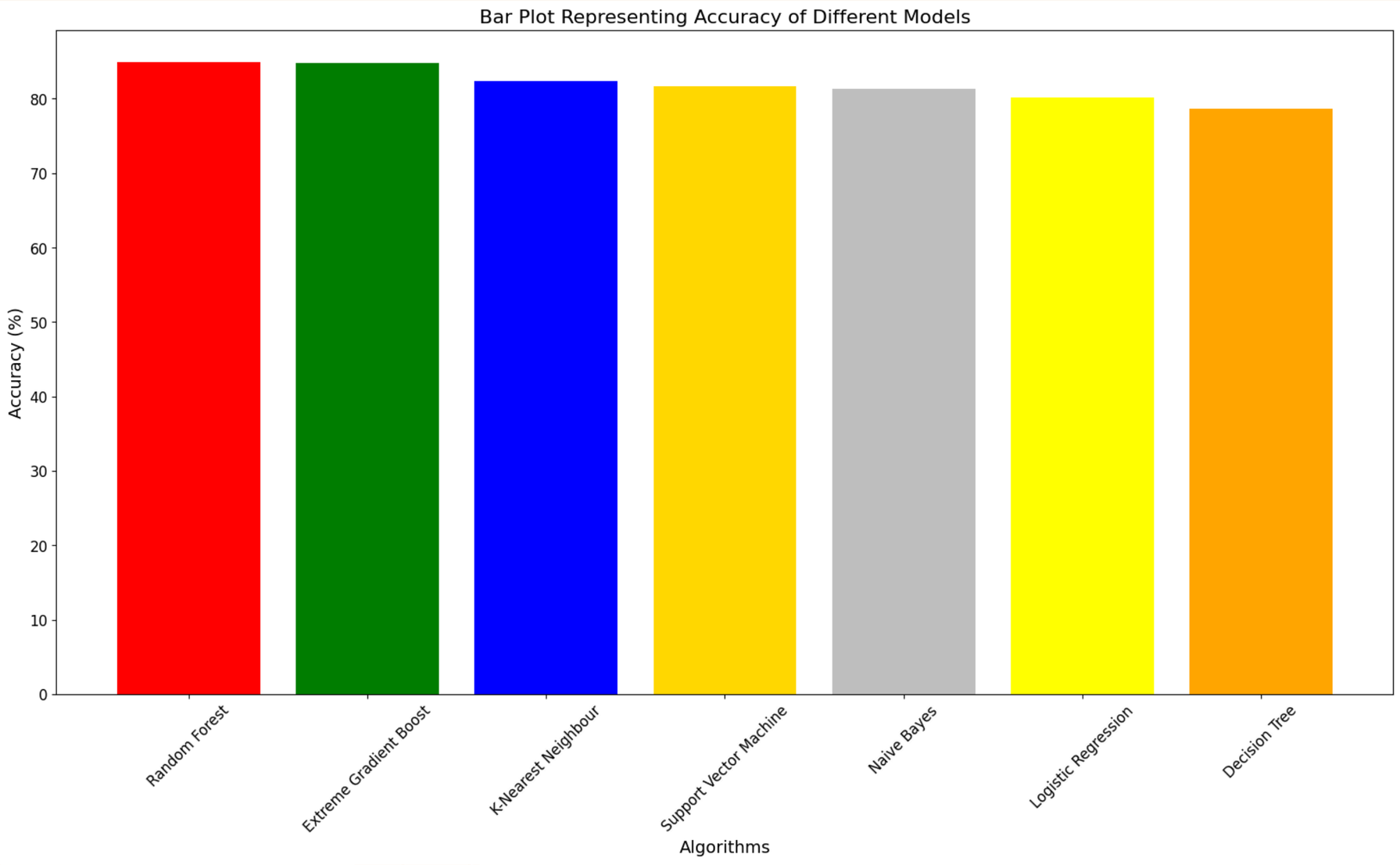
Identifies churn based on similarities between customers with shared attributes.

SVM

Differentiates churners and non-churners in complex datasets with overlapping features.

5. Visualization of Model Performance

Bar plot displaying model accuracy, enabling a quick comparison of performance.



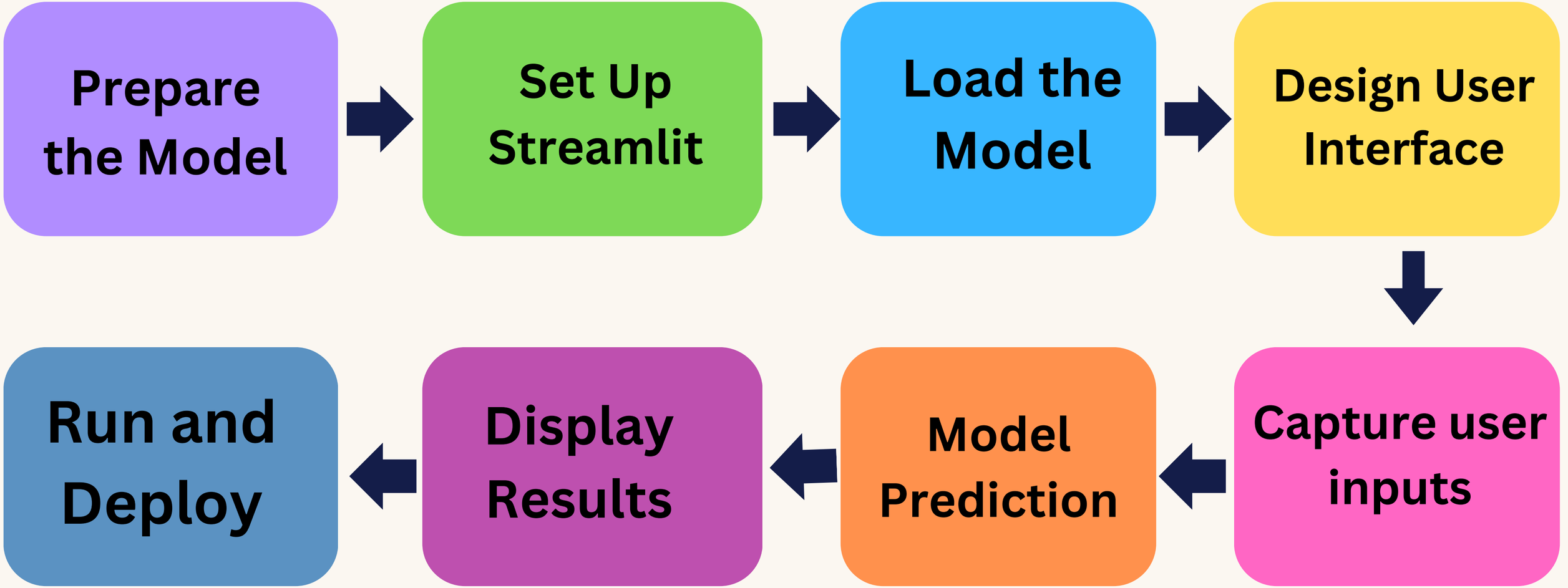
6. K-Fold Cross-Validation

- Performed cross-validation with k=10 for selected models.
- Reported mean and standard deviation of accuracy for each model.

7. Hyperparameter Tuning

- Applied GridSearchCV to optimize Random Forest hyperparameters.
- Reported the best parameters and evaluated the optimized model.

Model Deveployment Process(MLOps)



BANK CHURN PREDICTION

Age:

40

Credit Score:

400

Balance:

7000

Tenure (Years with Bank):

4

Number of Products:

1

Has Credit Card (1 = Yes, 0 = No):

1

Is Active Member (1 = Yes, 0 = No):

1

Predict

Prediction: Likely to Churn

BANK CHURN PREDICTION

Age:

40

Credit Score:

800

Balance:

70000

Tenure (Years with Bank):

10

Number of Products:

1

Has Credit Card (1 = Yes, 0 = No):

1

Is Active Member (1 = Yes, 0 = No):

1

Predict

Prediction: Unlikely to Churn

THANK YOU!