

## **Acknowledgement**

An industrial attachment cannot be completed without significant help from others. First we gratefully acknowledge the help and support from our parents, teachers, employers, friends and others, whose support has been invaluable for me. I would like to thank the following people for their contribution in this industrial attachment

## **Abstract**

This project aims to develop a Telecom Customer Churn Analysis System employing advanced data analytics and machine learning techniques. The primary goal is to build an effective model for predicting customer churn in the telecom industry, enabling proactive customer retention strategies and improved service quality.

The dataset utilized in this project comprises diverse customer-related attributes such as usage patterns, account information, service features, and customer demographics. Exploratory data analysis (EDA) is conducted to understand the relationships between these factors and customer churn. Machine learning algorithms are employed to predict customer churn.

Various models including logistic regression, decision trees, random forest, gradient boosting, and neural networks are trained on historical customer data, considering factors that influence customer behavior and retention.

The project aims to provide a user-friendly interface allowing telecom providers to input customer attributes and receive churn predictions. This system aids in customer relationship management, resource allocation, and service enhancements, ultimately reducing customer turnover and improving business sustainability.

## **Internship Experience & Job Description**

**Name and short Description of the project**

**Telecom Customer Churn Analysis Using Machine Learning:**

**Aims to predict whether a customer is likely to leave (churn) a telecom service provider based on historical data. By analyzing customer behavior, usage patterns, and demographics, machine learning models like decision trees, random forests, or logistic regression can be used to identify key factors driving churn. This helps telecom companies develop targeted retention strategies and improve customer satisfaction.**

**Involvement in Software/Hardware development process**

**Technology:**

**Front-End: Not Applicable**

**Back End: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn**

## Involvement in Planning

**Customer Data Collection:** Gathering relevant customer data including usage patterns, account details, and demographic information.

**Data Preprocessing:** Handling missing values, encoding categorical variables, and scaling numerical features for model compatibility.

**Exploratory Data Analysis (EDA):** Analyzing data distributions, correlations, and patterns related to customer churn.

**Feature Engineering:** Creating new features such as customer tenure, average usage, and churn-related indicators.

**Machine Learning Models:** Utilizing classifiers like Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Neural Networks for churn prediction.

**Model Evaluation:** Assessing model performance using metrics like accuracy, precision, recall, F1-score, and area under the ROC curve.

**Hyperparameter Tuning:** Optimizing model parameters using techniques like grid search or random search to improve performance.

**Cross-Validation:** Implementing k-fold cross-validation to ensure model generalization and reliability.

**Ensemble Methods:** Employing techniques like model ensembling to combine predictions from multiple models for enhanced accuracy.

**Feature Importance Analysis:** Determining key features influencing churn using techniques such as permutation importance or SHAP values

## **Involvement in Implementation, testing and documenting**

**Data Preprocessing:** Clean and prepare the telecom customer data, handling missing values, and outliers.

**Model Selection:** Choose appropriate machine learning models such as logistic regression, decision trees, random forests, or gradient boosting algorithms

**Model Validation:** Split the data into training and test sets (e.g., 80% training, 20% testing) . Perform cross-validation to prevent overfitting and ensure the model generalizes well.

**Process Documentation:** Document the entire process, from data gathering and cleaning to feature engineering and model selection.

## **Involvement in Deployment and Maintenance**

### **Deployment:**

Convert the model to a production environment

Model versioning

Cloud or On-premise Deployment

### **Maintenance:**

Performance Tracking

Data Drift Detection

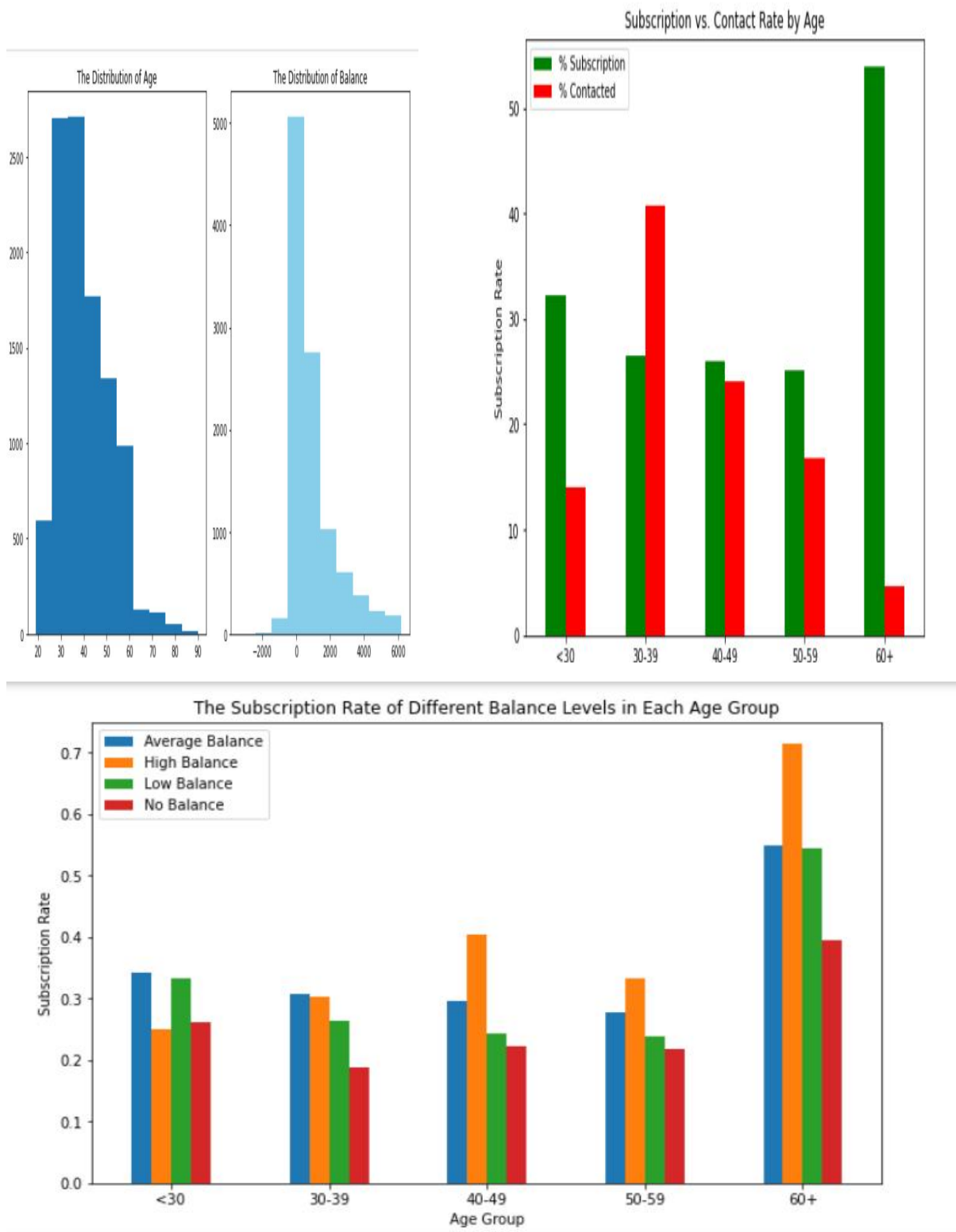
New Data Incorporation

Error Handling

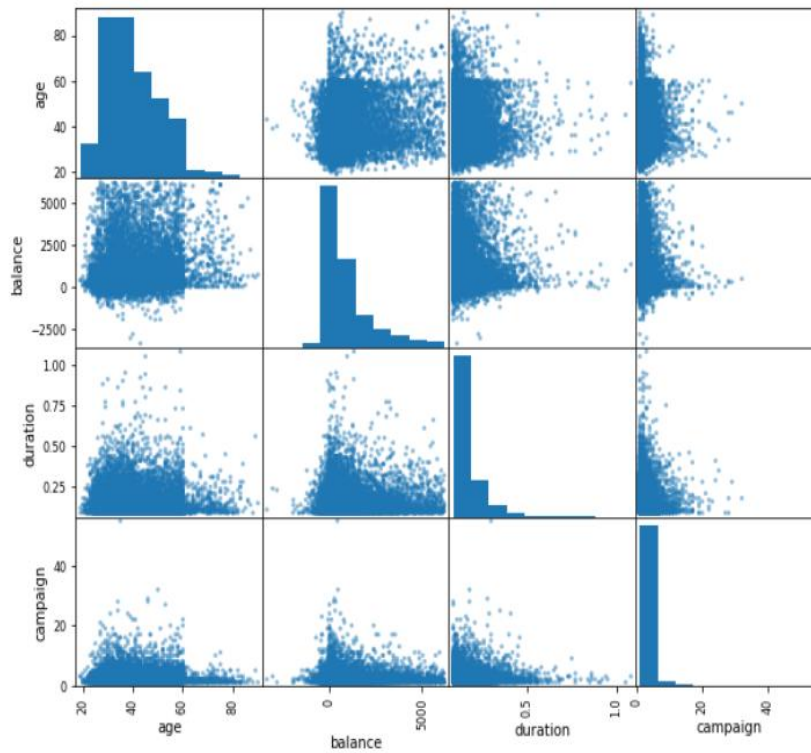
Logging and Debugging

User Feedback and Continuous Improvement

## Images/Screen shots



The Scatter Matrix of Age, Balance, Duration and Campaign



Accuracy of Logistic Regression: 73.3044733044733 %



## **Conclusion**

**Improve Targeting:** By predicting which customers are more likely to respond positively (e.g., subscribe to term deposits), the bank can focus its telemarketing efforts on the right customer segments.

**Optimize Marketing Spend:** Reducing marketing towards less responsive customers minimizes wasted effort and costs.

**Enhance Customer Satisfaction:** Personalized campaigns lead to better customer engagement and fewer unwanted calls or advertisements.