PROJECT REPORT

1.INTRODUCTION:

1.1 Project Overview:

In the telecommunications industry, customer churn represents a significant challenge as it leads to lost revenue and increased customer acquisition costs. Telecom companies face stiff competition, making customer retention as crucial as acquiring new customers. This project focuses on analyzing customer data to understand the factors driving churn and developing predictive models to identify at-risk customers. The ultimate goal is to implement strategies that reduce churn and enhance customer loyalty

The primary objective of this project is to analyze telecom customer data to identify patterns and factors contributing to churn. Using machine learning techniques, the project aims to predict which customers are likely to churn and provide actionable insights to help the company retain these customers, thereby reducing churn rates and improving overall customer satisfaction.

1.2 Purpose:

The main goal of this telecom customer churn analysis project is to develop a robust predictive model that accurately identifies customers at risk of churning. By analyzing customer data, we aim to uncover key factors contributing to churn and provide actionable insights that telecom companies can use to implement targeted retention strategies. Ultimately, the goal is to reduce the churn rate, increase customer retention, and enhance overall business profitability.

2. Problem Statement:

The telecommunications industry is highly competitive, with customers frequently switching providers in search of better deals, improved service quality, or other benefits. High churn rates result in lost revenue and increased costs for acquiring new customers. Therefore, it is crucial to identify the reasons behind customer churn and develop strategies to retain customers.

The project aims to answer the following things:

- Understanding the key factors that lead to customer churn.
- Predicting at-risk customers using machine learning models.
- Providing actionable recommendations to reduce churn rates based on model insights.

3. Proposed Solution:

3.1 Data Collection and Preprocessing:

- **Data Source:** The dataset is sourced from Kaggle and contains 7,043 rows and 21 columns. It includes customer demographics, account information, service usage, and churn status.
- **Data Cleaning:** The dataset was cleaned by addressing missing values in the TotalCharges column. Missing values were removed.

3.2 Feature Engineering:

- **Encoding Categorical Variables:** Categorical features like Contract, PaymentMethod, and InternetService were encoded using one-hot encoding to make them suitable for machine learning models.
- **Feature Scaling:** Numerical features like MonthlyCharges, tenure and TotalCharges were scaled using MinMaxScaler to ensure they are in the same range, improving the performance of the models.

3.3 Model Building:

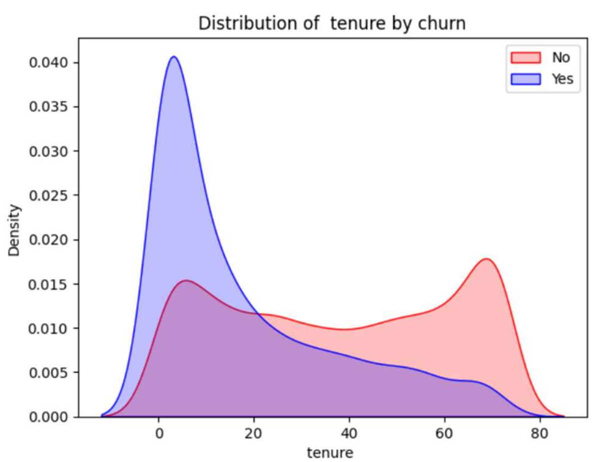
- **Training and Testing Split:** The data was split into training and testing sets in an 80:20 ratio to train the models on a substantial portion of the data and evaluate them on unseen data.
- **Model Selection:** Three machine learning models were selected:

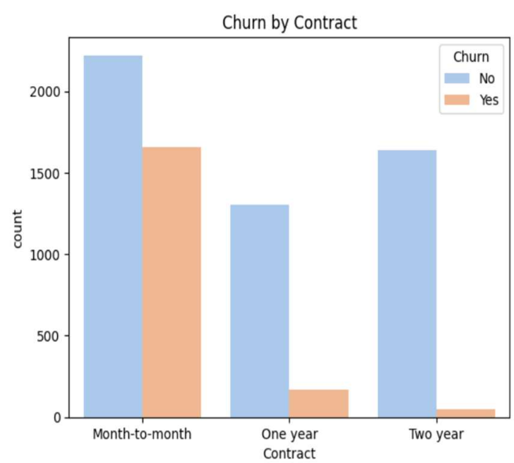
A. Artificial Neural Network (ANN): Implemented using Keras, focusing on capturing complex relationships between features.
B. Random Forest: A robust model implemented using scikit-learn, known for handling large datasets and reducing overfitting.
C. Logistic Regression: Used as a baseline model due to its simplicity and interpretability.

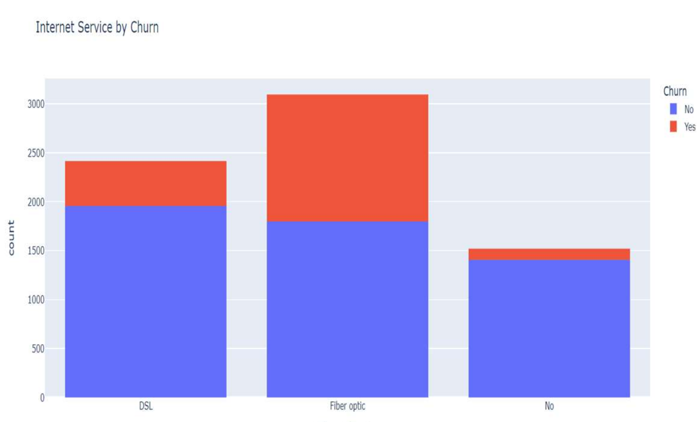
4. Data Analysis and Insights:

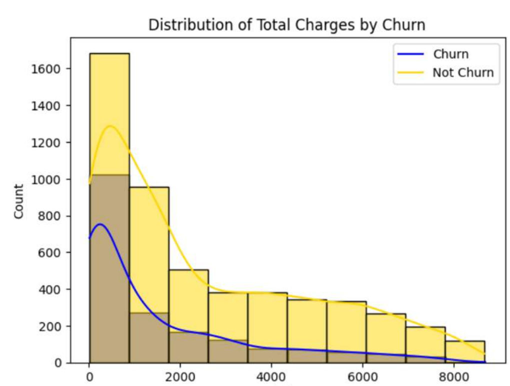
4.1 Data Exploration:

- **Customer Tenure:** Shorter tenures were associated with higher churn rates, indicating that customers who have not been with the company for long are more likely to leave.
- **Contract Type:** Customers with month-to-month contracts had the highest churn rates, suggesting that longer-term contracts might help in retaining customers.
- **Service Type:** Customers with Fiber Optic Internet Service were more likely to churn, pointing to possible issues with service quality or pricing.
- **Charges:** Higher monthly charges were associated with a higher likelihood of churn, indicating that pricing strategies might need to be revisited.









4.2 Model Performance:

- Logistic Regression: Provided slightly better performance in some areas, offering a good balance between accuracy and interpretability. It serves as a useful baseline model.
- Random Forest and ANN: Both models achieved similar results, demonstrating high accuracy in predicting customer churn. Random Forest provided a good balance between accuracy and interpretability, while the ANN model offered high predictive accuracy at the cost of greater computational complexity and reduced interpretability.

4.3 Insights:

- Short Tenure Churn: Customers with shorter tenures should be targeted with retention strategies early in their relationship with the company.
- Contract Promotion: Offering incentives for long-term contracts could help reduce churn among customers with month-to-month contracts.
- Service Improvement: Investigating and addressing issues with Fiber Optic Internet Service could reduce churn rates among these customers.
- Pricing Review: Adjusting pricing strategies, particularly for high-paying customers, could improve retention.

5. Performance Testing:

```
accuracy_score(y_pred1,y_test)
```

0.8038379530916845

```
recall_score(y_pred1,y_test)
```

0.6534653465346535

```
precision_score(y_pred1,y_test)
```

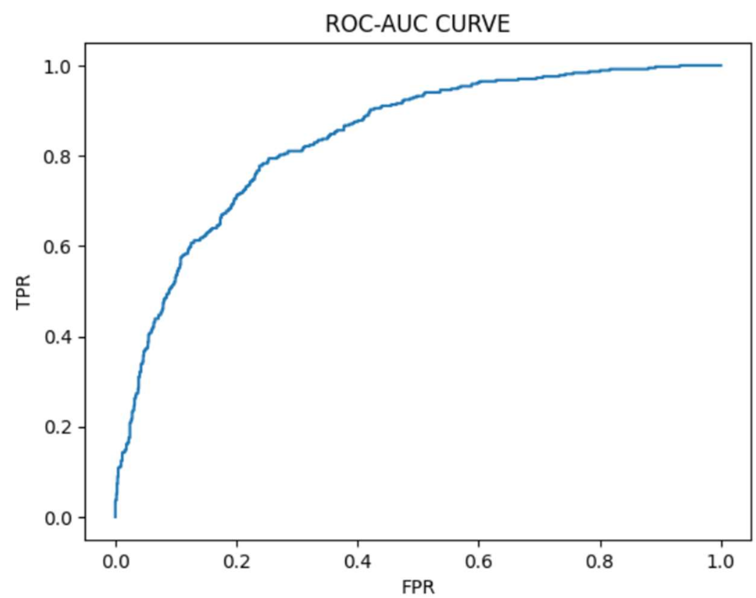
0.5365853658536586

```
f1_score(y_pred1,y_test)
```

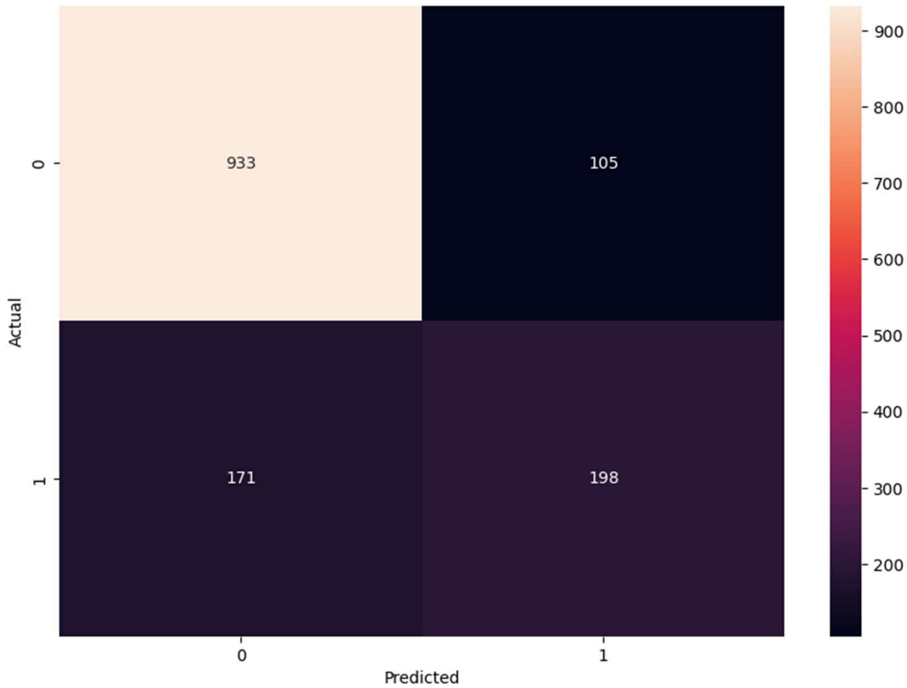
0.5892857142857143

```
roc_auc_score(y_pred1,y_test)
```

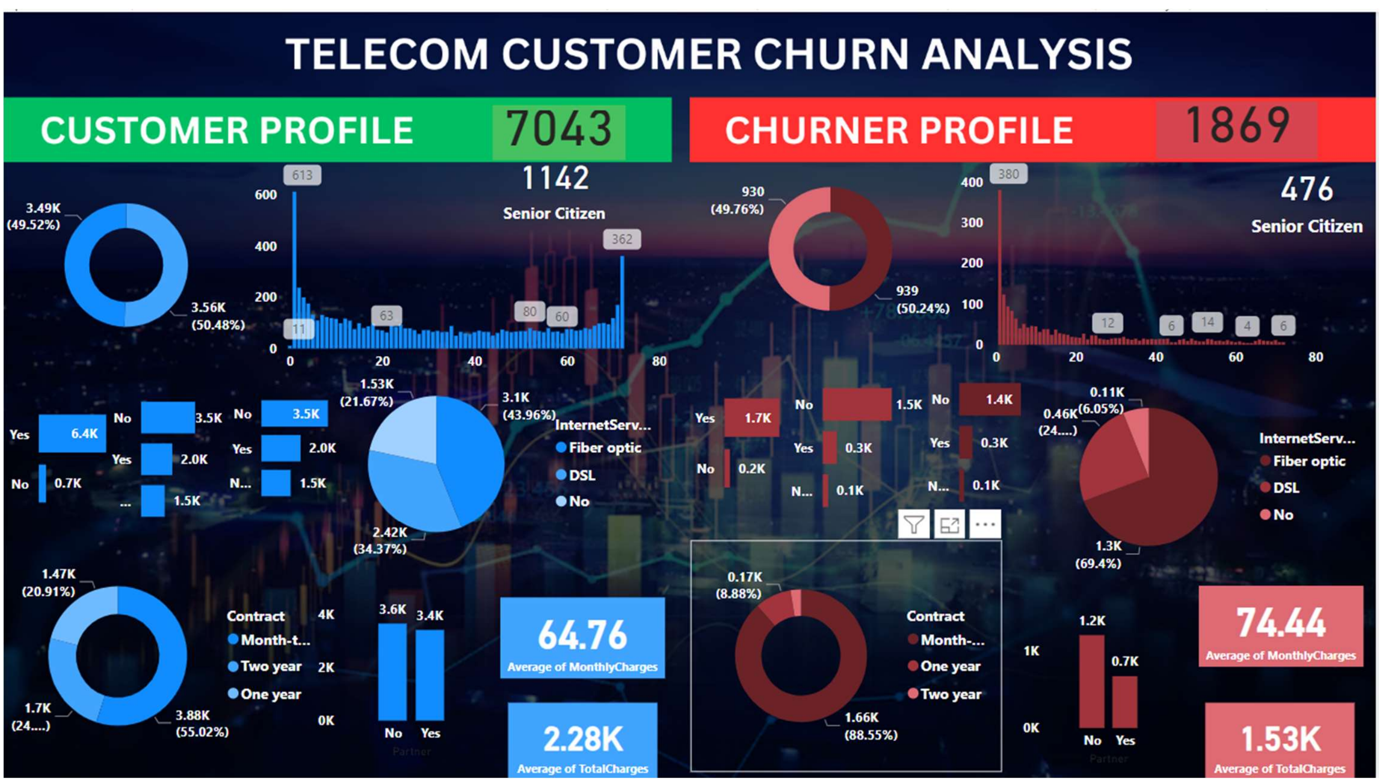
0.7492870210934138



Confusion Matrix:



POWER BI:



CONCLUSION:

The project successfully highlighted the key factors influencing customer churn and demonstrated the effectiveness of various machine learning models in predicting churn. By applying these insights and recommendations, telecom companies can enhance their strategies to reduce churn, improve customer retention, and boost overall profitability.