Table of Contents

1. Neo Telecom Customer Churn Analysis Report

1.1. Problem Statement

Neo Telecom is struggling with high customer churn rates, impacting revenue and growth. The goal is to predict which customers are likely to churn, enabling targeted retention efforts and reducing churn rates.

1.2. Problem Objective

- **Predict Customer Churn:** Develop a model to forecast which customers are at risk of leaving.
- **Deploy Models:** Implement the model in a production environment for real-time churn predictions.
- **Monitor and Improve:** Continuously track model performance and make updates as necessary.

1.3. Data Description

The dataset used includes:

- **Customer Demographics:** Age, gender, income, and tenure.
- **Usage Patterns:** Call minutes, data usage, and service usage.
- **Service History:** Subscription type, billing information, and customer service interactions.
- **Churn Labels:** Indication of whether the customer has churned (binary label).

1.4 Data Visualization

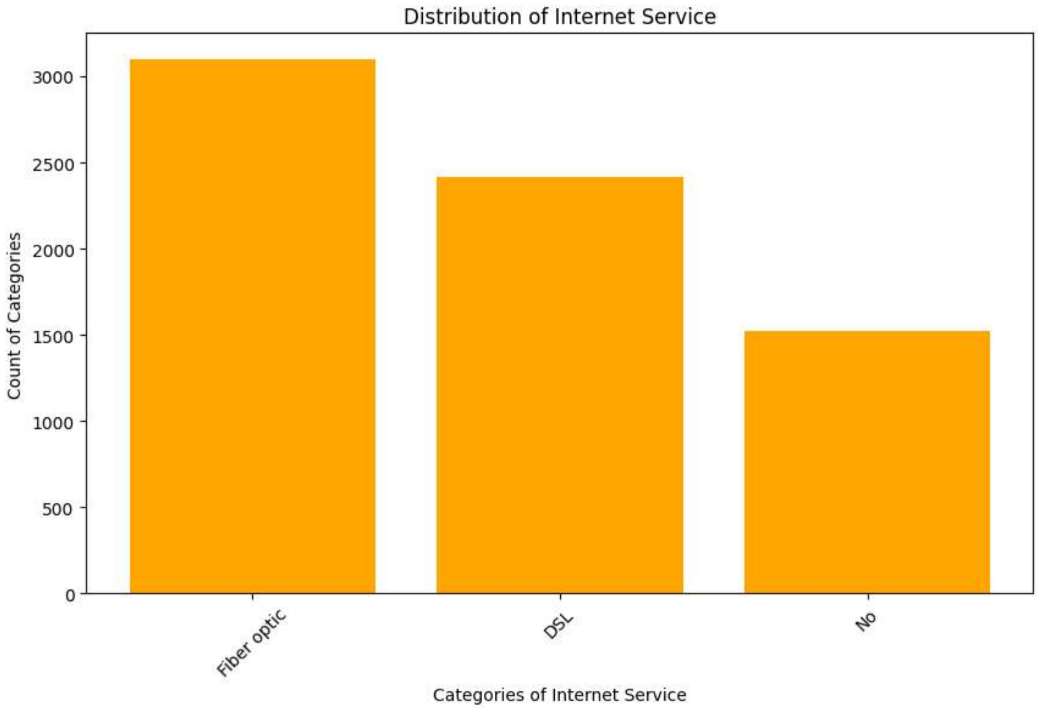

Figure 1: Distribution of Internet Service Customers.

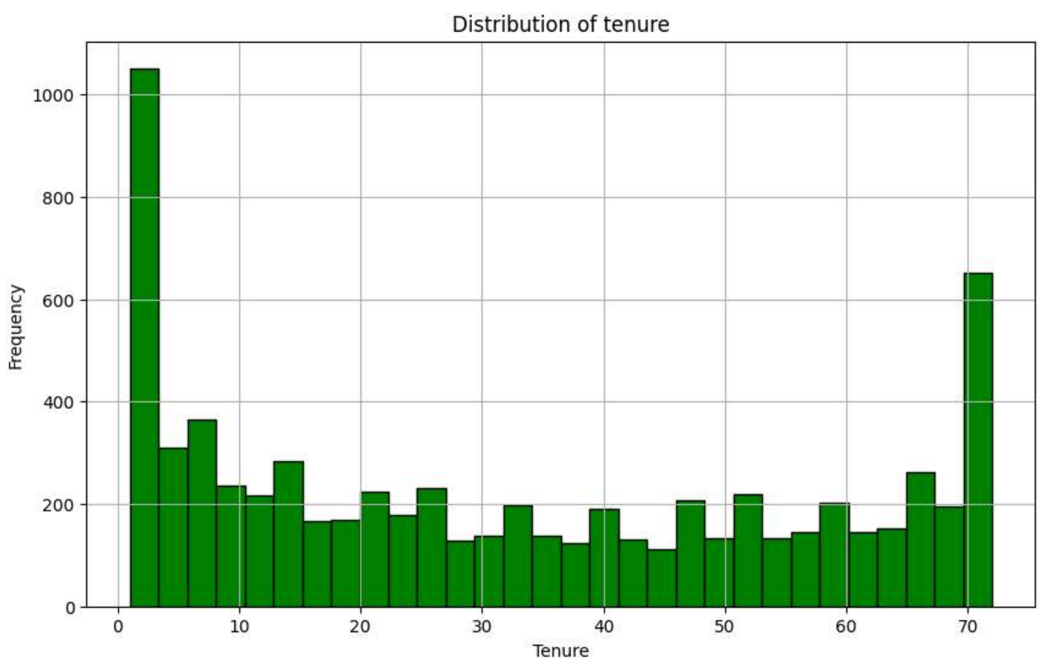

Figure 2: Distribution of Tenure

1.5. Preprocessing Steps

- **1.5.1. Handling Null Values**
  - Null values were handled using imputation techniques for numerical features and mode imputation for categorical features.
- **1.5.2. Feature Engineering**
  - Created new features such as customer lifetime value, average call duration, and interaction frequency.
  - Encoded categorical variables and scaled numerical features for model compatibility.
- **1.5.3. Data Normalization**
  - Applied normalization techniques to ensure feature values are on a comparable scale, improving model performance.

1.6. Choosing Algorithm

- **Logistic Regression:** Used for its simplicity and interpretability in binary classification.
- **Random Forest:** Chosen for its robustness and ability to capture complex relationships between features.

1.7. Model Deployment and Monitoring

- **Deployment:** The models were deployed using AWS SageMaker, enabling scalable and efficient prediction services.
  - **AWS SageMaker Features:** Automated model deployment, scaling, and management.
- **Monitoring:** Leveraged AWS CloudWatch for monitoring model performance metrics, such as accuracy and precision, as well as detecting model drift.
  - **Alerts:** Set up alerts for significant performance degradation or anomalies.

1.8. Model Evaluation and Techniques

- **1.8.1. Accuracy and Performance Metrics**
  - **Logistic Regression:** Achieved 80% accuracy with an AUC of 0.78.
  - **Random Forest:** Provided 83% accuracy with an AUC of 0.81.
  - Metrics used include precision, recall, F1-score, and ROC-AUC.
- **1.8.2. Model Interpretation**
  - **Feature Importance:** Random Forest identified key factors influencing churn, such as tenure, usage patterns, and service interactions.
  - **Coefficient Analysis:** Logistic Regression coefficients were analyzed to understand the impact of individual features on churn probability.
- **1.8.3. Real-time Predictions**
  - Integrated with Neo Telecom’s CRM system for immediate churn risk assessments and automated retention strategies.

1.9. Inferences from the Analysis

- **Key Drivers of Churn:** Identified factors such as high data usage, frequent customer service interactions, and service downgrades as significant predictors of churn.
- **Targeted Interventions:** Recommendations for targeted retention campaigns based on predicted churn risk, improving customer engagement and reducing churn rates.

1.10. Future Possibilities for the Project

- **Enhanced Modeling:** Explore advanced algorithms like Gradient Boosting or Deep Learning for potentially better performance.
- **Additional Features:** Incorporate external data sources, such as market trends or competitor analysis, for a more comprehensive churn model.
- **Personalized Strategies:** Develop customized retention strategies based on churn prediction insights, tailored to specific customer segments.

1.11. Conclusion

The churn prediction models have successfully identified key factors driving customer churn and enabled real-time predictions. The deployment on AWS SageMaker ensures scalable and efficient operations, while continuous monitoring allows for timely updates and improvements. These efforts contribute to more effective retention strategies and overall business growth.

1.12. References

- AWS SageMaker Documentation
- Scikit-learn Documentation
- Neo Telecom Customer Data (Internal Dataset)
- Industry reports on customer churn analysis and prediction