NAME :**Mannam Gnaneshwar**   Reg:12114807

Telecommunication Customer Churn Analysis Report

Abstract

This report explores the phenomenon of customer churn in a telecommunications company by leveraging various data science techniques. Customer churn, where customers discontinue their service, poses a significant challenge to the industry by impacting revenue and profitability. The primary objectives of this analysis are to identify the key factors driving customer churn, develop predictive models to foresee potential churn, and provide actionable strategies to enhance customer retention. Using a comprehensive dataset that includes customer demographics, service details, billing information, and churn status, we employed several machine learning models: Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Multi-Layer Perceptron (MLP), AdaBoost, and Gradient Boosting. The data preprocessing steps included handling missing values, encoding categorical variables, and normalizing numerical features.

Significant patterns were found using exploratory data analysis (EDA), including greater churn rates among senior individuals, independent consumers, and those with fibre optic internet access. Critical determinants were billing methods and contract types; month-to-month contracts and higher monthly prices were associated with higher turnover.

Accuracy, precision, recall, and F1-score measures were used to assess the prediction models; the Random Forest model offered important insights into the significance of the features. Significant churn predictors included tenure, monthly fees, contract type, online security, and technical support.

These results lead us to suggest various tactics, including contract incentives, enhanced service quality, focused retention marketing, streamlined billing, and ongoing customer feedback monitoring. By putting these tactics into practice, you may lower attrition rates, increase client retention, and make sure

Introduction

Customer churn, the phenomenon where customers stop using a company's services, is a critical issue for telecommunications companies. High churn rates can significantly impact a company's revenue and profitability, making it essential to understand and address the underlying causes. This project aims to leverage data science techniques to analyze customer churn data, identify the key drivers behind churn, and develop predictive models to help the company proactively manage and reduce churn.

Objectives

The primary objectives of this project are:

1. **Understand the factors influencing customer churn**: By analyzing various customer demographics, service usage patterns, and billing information, we aim to uncover the main reasons why customers leave.
2. **Build predictive models**: Develop machine learning models to accurately predict which customers are at risk of churning.
3. **Provide actionable recommendations**: Based on the insights gained, offer strategies for reducing churn and improving customer retention.

Data Description

The dataset used in this analysis includes customer demographics, service details, and churn status. The key variables in the dataset are:

- **CustomerID**: Unique identifier for each customer
- **Gender**: Male or Female
- **SeniorCitizen**: Binary (0: No, 1: Yes)
- **Partner**: Binary (Yes or No)
- **Dependents**: Binary (Yes or No)
- **Tenure**: Number of months the customer has stayed with the company
- **PhoneService**: Binary (Yes or No)
- **MultipleLines**: No phone service, One line, or Multiple lines
- **InternetService**: No, DSL, or Fiber optic
- **OnlineSecurity**: Binary (Yes or No)
- **OnlineBackup**: Binary (Yes or No)
- **DeviceProtection**: Binary (Yes or No)
- **TechSupport**: Binary (Yes or No)
- **StreamingTV**: Binary (Yes or No)
- **StreamingMovies**: Binary (Yes or No)
- **Contract**: Month-to-month, One year, or Two years
- **PaperlessBilling**: Binary (Yes or No)
- **PaymentMethod**: Electronic check, Mailed check, Bank transfer, or Credit card
- **MonthlyCharges**: Monthly amount charged to the customer
- **TotalCharges**: Total amount charged to the customer
- **Churn**: Binary (Yes or No) indicating if the customer churned

Methodology

Data Preprocessing

1. **Handling Missing Values**: Missing values were handled by appropriate imputation methods.
2. **Encoding Categorical Variables**: Categorical variables were converted into numerical values using one-hot encoding.
3. **Normalizing Numerical Features**: Numerical features were normalized to ensure all features contribute equally to the model performance.

Exploratory Data Analysis (EDA)

EDA was conducted to understand the distribution of data and identify any significant patterns:

- **Distribution Analysis**: Visualizations such as histograms and box plots were used to analyze the distribution of features.
- **Correlation Analysis**: Heatmaps and correlation matrices were used to identify relationships between features.
- **Churn Analysis**: The churn rate was analyzed across different features to identify significant predictors.

Model Building and Evaluation

Six machine learning models were built and evaluated:

1. **Logistic Regression**: A statistical model that predicts the probability of a binary outcome.
2. **K-Nearest Neighbors (KNN)**: A non-parametric method used for classification based on feature similarity.
3. **Random Forest**: An ensemble learning method that uses multiple decision trees to improve predictive accuracy.
4. **Multi-Layer Perceptron (MLP)**: A class of feedforward artificial neural network models.
5. **AdaBoost**: An ensemble method that combines weak classifiers to create a strong classifier.
6. **Gradient Boosting**: An ensemble technique that builds models sequentially to reduce errors.

The models were evaluated using the following metrics:

- **Accuracy**: Proportion of correctly classified instances.
- **Precision**: Proportion of true positive predictions among all positive predictions.
- **Recall**: Proportion of true positive predictions among all actual positives.
- **F1-Score**: Harmonic mean of precision and recall.

Results

Model Performance

The performance of each model is summarized below:

- **Logistic Regression**:
  - Accuracy: [accuracy_lr]
- **K-Nearest Neighbors (KNN)**:

- Accuracy: [accuracy_knn]
- **Random Forest**:
  - Accuracy: [accuracy_rf]
- **Multi-Layer Perceptron (MLP)**:
  - Accuracy: [accuracy_nn]
  - F1-Score: [mlp_f1]
- **AdaBoost**:
  - Accuracy: [accuracy_ada]
- **Gradient Boosting**:
  - Accuracy: [accuracy_gb]

Key Findings

1. **Demographic Insights**:
   - Senior citizens are more likely to churn.
   - Customers without dependents have a higher churn rate.
2. **Service Usage**:
   - Customers with fiber optic internet service have higher churn rates compared to DSL users.
   - Availability of online security and technical support services decreases churn.
3. **Billing and Contract**:
   - Customers with month-to-month contracts are more likely to churn.
   - Higher monthly charges are associated with higher churn rates.

Feature Importance

The Random Forest model was used to identify key features influencing churn:

- **Tenure**: Longer tenure significantly decreases the likelihood of churn.
- **MonthlyCharges**: Higher monthly charges increase the likelihood of churn.
- **Contract Type**: Month-to-month contracts increase the likelihood of churn.
- **OnlineSecurity**: Availability of online security decreases churn.
- **TechSupport**: Availability of technical support decreases churn.

Recommendations

Customer Retention Strategies

1. **Contract Incentives**:
   - Offer incentives such as discounts or additional services for customers who choose longer-term contracts.
2. **Service Improvement**:
   - Enhance the quality of fiber optic internet service to reduce churn.
   - Improve customer support services, including online security and technical support.
3. **Targeted Campaigns**:
   - Develop targeted retention campaigns for senior citizens and customers without dependents.
   - Focus on new customers in their initial months of tenure with special offers and personalized communication.

1. **Flexible Billing**:
   o Simplify billing procedures and offer flexible payment options to reduce the hassle for customers.
   o Encourage the use of automatic payment methods to ensure timely payments and reduce churn.
2. **Value-Added Services**:
   o Introduce value-added services or bundled packages that provide more value for the same or slightly higher cost, making it harder for customers to leave.

1. **Regular Monitoring**:
   o Continuously monitor churn rates and customer feedback to identify any emerging trends or issues.
   o Use predictive analytics to identify customers at high risk of churning and intervene proactively.
2. **Customer Feedback**:
   o Actively seek and incorporate customer feedback to improve services and address pain points.
   o Implement a robust customer relationship management (CRM) system to track interactions and preferences.

Conclusion

The analysis of customer churn for the telecommunications company revealed several key factors influencing churn, including demographic characteristics, service usage, billing methods, and contract types. By implementing targeted strategies based on these insights, the company can significantly reduce churn rates and enhance customer retention.

Continuous monitoring, coupled with regular feedback loops, will ensure that the company can adapt to changing customer needs and preferences, ultimately leading to increased customer loyalty and business success.

	LogisticReg	KNN	RandomForest	ADA Boost	GradientBoost
AccuracyScore	0.79	0.75	0.79	0.79	0.8