

Assignment Report of SPEAKX



**Title: - Predicting Customer Churn
in Telecommunications Company**

Praveen Kumar Tiwari

12104582



L OVELY
P ROFESSIONAL
U NIVERSITY

LOVELY PROFESSIONAL UNIVERSITY(LPU)

Computer Science and Engineering

Date: - 09-06-2024

Customer Churn Prediction Report

Introduction

Customer churn is a significant concern for businesses, particularly in the telecommunications industry. This project aims to predict customer churn using machine learning techniques. The process involves data preprocessing, exploratory data analysis (EDA), feature engineering, model building, and evaluation. The dataset used in this project contains information about customers, their services, and whether they churned.

Task 1: Data Preprocessing

Steps:

1. Handling Missing Values:

- The TotalCharges column contained missing values, which were coerced to NaN and subsequently removed.

2. Encoding Categorical Variables:

- Binary categorical variables (gender, Partner, Dependents, PhoneService, PaperlessBilling, Churn) were encoded to numeric values.
- Other categorical variables with multiple categories were one-hot encoded.

3. Scaling Features:

- Continuous variables like tenure, MonthlyCharges, and TotalCharges were scaled using StandardScaler for better model performance.

Task 2: Exploratory Data Analysis (EDA)

EDA was performed to understand customer behavior and factors influencing churn. Key findings were visualized using graphs and charts:

1. Churn Distribution:

- Approximately 26.5% of customers churned, indicating class imbalance.

2. Tenure Analysis:

- Customers with shorter tenures (0-1 year) had higher churn rates.

3. Service Impact:

- Customers using fiber optic internet had higher churn rates compared to DSL.
- Additional services like online security, backup, and tech support reduced churn rates.

4. Contract Type:

- Month-to-month contracts had significantly higher churn rates compared to one-year and two-year contracts.

5. Payment Method:

- Customers using electronic checks had higher churn rates.

Task 3: Feature Engineering

Relevant features were created to improve model performance:

1. Tenure Grouping:

- Customers were grouped based on tenure: 0-1 Year, 1-2 Years, 2-4 Years, 4-5 Years, 5+ Years.

2. Total Services:

- A new feature representing the total number of services a customer subscribes to.

3. Monthly Charges Grouping:

- Monthly charges were categorized into Low, Medium, and High.

Task 4: Building the Churn Prediction Model

Several machine learning algorithms were implemented and compared:

1. Logistic Regression:

- A basic but interpretable model.
- Achieved an accuracy of 80.2%.

2. Random Forest:

- A robust and flexible model.
- Achieved an accuracy of 82.1%.

3. Gradient Boosting:

- A powerful model for handling complex relationships.
- Achieved an accuracy of 81.5%.

Task 5: Model Evaluation

The models were evaluated using accuracy, precision, recall, and F1-score:

1. Logistic Regression:

- Accuracy: 80.2%
- Precision: 70.3%
- Recall: 55.4%
- F1-score: 61.9%

2. Random Forest:

- Accuracy: 82.1%
- Precision: 72.8%
- Recall: 61.4%
- F1-score: 66.6%

3. Gradient Boosting:

- Accuracy: 81.5%
- Precision: 71.8%
- Recall: 60.4%
- F1-score: 65.6%

Challenges

1. Class Imbalance:

- The dataset had a significant class imbalance, which was mitigated using techniques like stratified sampling and class weighting.

2. Feature Selection:

- Identifying and engineering relevant features required careful analysis to avoid overfitting.

3. Model Selection:

- Balancing model complexity and interpretability was crucial in selecting the final models.

Conclusion

The project successfully built and evaluated models to predict customer churn. The Random Forest model performed the best with an accuracy of 82.1%. Insights from EDA and feature engineering helped improve model performance. Addressing challenges like class imbalance and feature selection was key to developing effective models. The results can be used by the telecommunications company to identify at-risk customers and take proactive measures to reduce churn.