Binary Classification on ‘Customer_Churn’ using Keras

Tejaswini Anil Kale

Date

12/06/2024

Course title

Data Science and Artificial Intelligence offered by IITM Pravartak and Intellipaat

Table of Contents

1. Problem Statement
2. Project Objective
3. Data Description
4. Data Pre-processing Steps and Inspiration
5. Data Insights
6. Choosing the Algorithm for the Project
7. Motivation and Reasons for Choosing the Algorithm
8. Assumptions
9. Model Evaluation and Techniques
10. Inferences from the Same
11. Future Possibilities of the Project
12. Conclusion
13. Reference

1. Problem Statement:

Telecom company “Leo” is facing issues in retaining their customers. Customer churn, which refers to the loss of customers, significantly impacts revenue and growth. The goal is to analyse the data of company and find insights and stop customers from churning out to other telecom companies, predict customer churn based on historical data and customer attributes.

2. Project Objective:

- To identify key factors influencing customer churn.
- To develop a predictive model to forecast customer churn.
- To provide insights that can help in devising strategies for customer retention.

---

3. Data Description:

Dataset Overview:

- **Source**: Kaggle dataset (Telco Customer Churn)
- **Number of Rows**: 7043
- **Number of Columns**: 21
- **Columns and Data Types**:
    - customerID: Customer ID (object)
    - gender: Gender of the customer (object)
    - SeniorCitizen: Indicates if the customer is a senior citizen (int64)
    - Partner: Whether the customer has a partner (object)
    - Dependents: Whether the customer has dependents (object)
    - tenure: Number of months the customer has stayed with the company (int64)
    - PhoneService: Whether the customer has phone service (object)
    - MultipleLines: Whether the customer has multiple lines (object)
    - InternetService: Customer’s internet service provider (object)
    - OnlineSecurity: Whether the customer has online security (object)
    - OnlineBackup: Whether the customer has online backup (object)
    - DeviceProtection: Whether the customer has device protection (object)
    - TechSupport: Whether the customer has tech support (object)
    - StreamingTV: Whether the customer streams TV (object)

- StreamingMovies: Whether the customer streams movies (object)
- Contract: Type of contract (object)
- PaperlessBilling: Whether the customer uses paperless billing (object)
- PaymentMethod: Payment method (object)
- MonthlyCharges: Monthly charges (float64)
- TotalCharges: Total charges (float64)
- Churn: Whether the customer churned (object)

4. Data Pre-processing Steps and Inspiration:

- **Handling Missing Values**: Identified and removed rows with missing TotalCharges.
- **Data Type Conversion**: Converted TotalCharges from object to float. Convert Yes and No to 1 or 0.
- **Categorical Encoding**: Replaced categorical text values with numerical values for analysis.
- **Feature Engineering**: Extracted meaningful features from existing columns.
- **Outlier Handling**: Applied techniques to handle outliers.

---

5. Data Insights:

- The dataset consists of 7043 rows and 21 columns.
- Data types were appropriately converted, with particular attention to TotalCharges being converted to a numeric type.
- Unique values in categorical columns were inspected, and unnecessary categories like 'No internet service' and 'No phone service' were simplified to 'No'.
- Columns with 'Yes' and 'No' values were converted to binary (1 and 0).
- Key insights included:
    - Customers with month-to-month contracts are more likely to churn.
    - Higher monthly charges are associated with higher churn rates.
    - Longer-tenured customers are less likely to churn.

6. Choosing the Algorithm for the Project:

Given the classification nature of the problem, various algorithms were considered, including logistic regression, decision trees, random forests, and neural networks. A neural network using the Keras library was selected due to its ability to capture complex relationships in the data.

7. Motivation and Reasons for Choosing the Algorithm:

- **Neural Networks**: Capable of capturing complex patterns and interactions in the data.
- **Keras**: Offers a user-friendly and efficient framework for building deep learning models.
- **Performance**: Neural networks are expected to outperform simpler models in handling the intricate nature of churn prediction.

8. Assumptions:

- The data is assumed to be representative of the current customer base.
- The selected features are assumed to be relevant predictors of customer churn.
- The neural network model will generalize well to new, unseen data.

9. Model Evaluation and Techniques:

- **Train-Test Split**: The data was split into training and testing sets (80% training, 20% testing).
- **Metrics Used**: Accuracy, precision, recall, and F1-score were used to assess the model's performance.
- **Model Performance**: The neural network achieved an accuracy of 82%, with the following detailed metrics:

- **Precision**: 0.82

10. Inferences from the Same:

- The model successfully identified key factors contributing to customer churn, such as contract type, monthly charges, and tenure.
- Customers with month-to-month contracts and higher monthly charges are more likely to churn.
- Long-term customers and those with annual contracts are less likely to churn.
- The neural network model captured the underlying patterns effectively, providing a solid basis for churn prediction.

11. Future Possibilities of the Project:

- **Enhanced Feature Engineering**: Incorporate more features, such as customer demographics and usage patterns.
- **Advanced Models**: Explore ensemble methods and more advanced neural network architectures.
- **Real-Time Predictions**: Implement the model in a real-time system to provide immediate churn predictions and interventions.

12. Conclusion:

The project successfully developed a neural network model to predict customer churn. Data cleaning and pre-processing were crucial steps in improving model performance. The model provided valuable insights into factors influencing churn, which can guide business strategies for customer retention.

13. Reference:

- Dataset: Kaggle Telco Customer Churn Dataset
- Documentation: Keras Documentation
- Articles and papers referenced during the project.