Department of Computer Science and Engineering
Faculty of Engineering
University of North Texas
Final Project CSCE5300 Spring 2024

Due on or before 30th of April 2024

Predicting Customer Churn in Telecom Industry using Random Forest

The telecommunications industry faces the challenge of customer churn, where customers switch from one service provider to another. In this project, you'll develop a predictive model using Random Forest to identify customers who are likely to churn based on their usage patterns and demographics. The goal is to help telecom companies proactively retain customers by targeting them with personalized retention strategies.

Project Tasks:

1. Data set and Exploration:
    - Use the dataset provided (Telecom_Churn.csv).
    - Explore the data to understand its distribution, identify missing values, outliers, and correlations between different features.
2. Data Preprocessing:
    - Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features as required.
    - Perform feature engineering to create new features or transform existing ones that may be more informative for predicting churn.
3. Model Development:
    - Implement a Random Forest classifier using libraries found in pyspark (use python as the base language).
    - Split the data into training and testing sets and train the Random Forest model on the training data.
    - Tune the hyperparameters of the Random Forest model using techniques like grid search or random search to optimize performance.
4. Model Evaluation:
    - Evaluate the performance of the trained Random Forest model using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score.

- Visualize the model's performance using techniques like ROC curves and confusion matrices to gain insights into its strengths and weaknesses.

5. Deployment and Interpretability:
    - Deploy the trained Random Forest model in a production environment where it can be used to predict customer churn in real-time.
    - Interpret the model's predictions and identify key features contributing to churn prediction using techniques like feature importance.
6. Business Insights and Recommendations:
    - Analyze the results of the predictive model to gain actionable insights into factors influencing customer churn.
    - Provide recommendations to the telecom company on targeted marketing strategies, service improvements, and customer retention initiatives based on the model's findings.
7. Documentation and Presentation:
    - Document the entire project, including data preprocessing steps, model development, evaluation results, and business insights.
    - Prepare a comprehensive report and presentation summarizing the project's objectives, methodologies, results, and recommendations for stakeholders.

Submission guidelines:

Submit your code as a notebook file. Include the documentation as a MS word file. Do not use screenshots. Include your code as text.