Graded Project

Machine Learning - Supervised Learning

Telecom Customer Churn Prediction

Learning outcome:

- Exploratory data analysis & preparing the data for model building.
- Training and making predictions using different classification models.
- Model evaluation.

Domain:

- Telecom

Dataset Description:

Each row represents a customer, each column contains customer's attributes described on the column Metadata. The data set includes information about:

- Customers who left within the last month – the column is called Churn (target)
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Project Objective:

Build a model that will help to identify the potential customers who have a higher probability to churn. This helps the company to understand the pinpoints and patterns of customer churn and will increase the focus on strategizing customer retention.

Steps to the project: [Total score: 50 points]

- Load and Explore the data (3 points)

1. Import the required libraries and read the dataset. (1 point)
2. Check the first few samples, shape, info of the data and try to familiarize yourself with different features. (2 points)

- Data Cleaning and Exploratory Data Analysis (16 points)

3. Check the percentage of missing values in each column of the data frame. Drop the missing values if there are any (2 points)
4. Check if there are any duplicate records in the dataset? If any drop them (2)
5. Drop the columns which you think redundant for the analysis (2 points)
6. Encode the categorical variables (2 points)
7. Perform a detailed univariate, bivariate, and multivariate analysis with appropriate plots and add detailed comments after each analysis. (8 points)

- Data Preparation for model building (6 points)

8. Store the target column (i.e. Churn) in the y variable and the rest of the columns in the X variable (2 points)
9. Split the dataset into two parts (i.e. 70% train and 30% test) (2 points)
10. Standardize the columns using z-score scaling approach (2 points)

- Model Building and Evaluation (25 points)

11. Write a function. (8 points)

    i) Which can take the model and data as inputs.

    ii) Fit the model with the train data.

    iii) Make predictions on the test set.

    iv) Returns the Accuracy Score.

12. Use the function and train a Logistic regression, KNN, and Naive Bayes, Decision tree, Random forest, Boosting and Stacking Classifier models and make predictions on test data. (8 points)

13. Display and compare all the models designed with their test accuracies.(4)

14. Select the final best trained model along with your detailed comments for selecting this model. (3 points)

15. Write your conclusion on the results.(2 points)

Submission:

- Please submit the solution file in .html and .ipynb format on Olympus
- Add necessary comments wherever required.