Abstract

Course: CSC 9025 Grand Challenges of Computing

Topic: Data Analytics Portal

The main idea of this project is to provide a platform for users to perform data analytics and evaluate different machine learning models on their data sets without any knowledge on scripting languages that are needed to perform such operations. The portal allows the user to upload their dataset in the csv format. The first feature will be the user providing their csv format dataset in the portal and select one of the various options for performing Exploratory Data Analysis (Histogram plot, Box plot, Scatter plot etc.). Once the user selects the options a python script that will perform the exploratory data analysis will be called from the backend and the plots are displayed to the user.

The other feature will be, apply one of the Decision tree classification or regression algorithms on the dataset and view the accuracy scores of the model. This will provide a common platform that does not require to write the scripts for each dataset, instead uses the same scripts by modifying the parameters as required by the scenario. This will also help the Data Analyst select the model that best fits the dataset and work on tuning performance using that model. The application will be in .NET and Python