Custom Algorithm Engine

This is a flask web/ML project where a dataset is uploaded in a csv format, the next page opens, which creates a dropdown menu of columns to be selected dynamically based on the uploaded dataset, where preferences are selected which include the algorithm to be used, test size ratio, columns to be fit to the model, outcome column. Then the selected algorithm is run on the dataset on the backend and results are shown on a webpage. The Engine works for classification and regression problems. It can handle categorical and continuous features and teh user needs to specify it on the preferences page. Preferences page has few suggestions and the few columns of the dataset to help the user make his choice properly.

Requirements for the application:
-python or anaconda should be installed
-install these packages :
Flask
numpy
pandas
scikit-learn
time

-Set the environment variable of python if using python without anaconda
-dataset should be in CSV format
-dataset should have the TARGET column as the last column
-dataset should not have NULL values
-User needs to select categorical and numerical values separately
-User gets a few intrinsically categorical columns selected on the preferences page to help in making a choice.

Algorithms to be chosen :

For Classifictaion :

Logistic Regression

Linear SVM

KNN

Decision Trees

MLP Classifier

For Regression :

Linear Regression

SVM

Decision Tree Regressor

MLP Regressor

Steps to run:

-open command prompt or conda prompt, change the directory to Cust_Alg_Engine the working folder of the application

-type python app.py and open "http://localhost:5000/" in your browser

-upload the dataset

-choose your preferences and click on Analyze

-See the results on the next page which include the Algorithm used, selected columns which were used to train the model, time taken to train and predict , confusion matrix, classifictaion report, accuracy and five rows from test data split.

Working of the application:

The chosen prefernces are used to run one of the supported algorithms on the platform which was chosen by the user on the backend to train and predict the class and it shows the Analysis results.

Application :

To provide a platform to use Machine Learning for everyone without the specific knowledge of the domain.

To give the user the freedom to choose the algorithm he wants with other preferences

To compare different types of classification algorithm

It can be used to predict the class after training on a dataset

Future Enhancements:

Multiple other algorithms can be added to the list

Multiple other dataset formats can be added

Data vsiualization can be added to give user a pictorial/graphical view of data

Feature to handle null values can be added

Features to manipulate data can be added

Feature to balance the dataset can be added for an imbalanced dataset