Capstone Project 2 - Advanced Data Science

Total Marks- 60

Problem Statement: Build machine learning models for prediction (regression/classification).

Dataset: Dataset used in Capstone Project 1

Important!!! Use the cleaned data which you have obtained at the end of Exploratory Data Analysis in Capstone Project-1.

Part 1: [25 Marks]

- Analyse the dataset and identify the task to be performed whether it is regression or classification.
- Perform the steps below to build any 5 machine learning models to perform regression or classification tasks considering the dataset.
    1. Data Splitting – Split the data into training and testing datasets.
    2. Build the model using training dataset
    3. Make predictions using the built model on test dataset.
    4. Evaluate model using appropriate evaluation metrics.
- Compare the accuracy of all 5 models and select the best one.
- Add Python comments to explain your code blocks.

Part 2: [30 Marks]

- Perform the following operations:
    1. Model building (any 5 ML models)
    2. Apply hyperparameter tuning using GridSearchCV for all the models.
    3. Create a dataframe to store the result of GridSearchCV for all the models which will includes model_name, best_score_ and best_params_

For example,

If regression task: (Any 5)

Model Name	Best Score	Best Parameters
Linear Regression		
Lasso Regression		
Ridge Regression		
SVR		
DecisionTreeRegressor		
KNeighborsRegressor		
RandomForestRegressor		

OR

If classification task:

Model Name	Best Score	Best Parameters
Logistic Regression		
SVC		
DecisionTreeClassifier		
KNeighboursClassifier		
RandomForestClassifier		

- Add Python comments to explain your code blocks.

Part 3: [5 Marks]

- Write conclusion to indicate your learning from part 1 and part 2 of the project.

Note –

1. The comments and conclusions/observations carry marks for each part.
2. Students should submit their work in the form of .ipynb code file to Ms. Samidha (samidhak@regenesys.net)
3. Please follow the filename format below: Name_of_student-Batch_name-Capstone_Project_2.ipynb.
4. Keep same naming convention as mentioned above for the subject of email.
5. The duration of the project work is set to one month. Last date of submission is 15 January 2024.