**List of Project’s for Learners**

|  |  |
| --- | --- |
| **Sr No** | **Project Name** |
| 1 | Predict the price of the house using Regression Algorithm |
| 2 | Credit Card Fraud Detection using Classification Algorithm |
| 3 | Loan Application Approval Prediction using Classification Algorithm |
| 4 | Big Mart Sales Data Prediction using Regression |
| 5 | Classification of Wheat Seed Using Classification Algorithm |
| 6 | E-mail Fraud detection using Classification Algorithm Project |
| 7 | Car Price Prediction Project using Linear Regression Algorithm |
| 8 | Classifying the fruits with respect to its weight, height and color score |
| 9 | Prediction of the salary of the employees using Regression Algorithm |
| 10 | Detection of Heart Disease using Classification Algorithm |

**Project Deliverables**

Given Below are important steps of each of the project

**[PROJECT 1&2]**

|  |  |
| --- | --- |
| **Predict the Prices of the Houses using Regression Algorithm** | **Credit Card Fraud Detection Project using Classification Algorithm** |
| **Plan the Model for prediction of the Prices and import the relevant Libraries to the project.**    **DATA PROCESSING STAGES**  **Data Acquisition: Import the required Dataset** | **Plan the Model for Credit Card Fraud detection and import the relevant libraries to the project**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required dataset** |
| **Data Processing: Load the data and concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data and Correlation between Attributes and target to choose that Attribute which has highest absolute correlation to perform Linear Regression.** | **Data Processing: Load the data and concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data and you can draw the pie Chart to get the visual representation of the proportion of Fraud and Genuine Transactions.**  **To simplify the Model, Use the Feature Selection method to select the attribute that has strongest relationship to the Target** |
| **Normalization of the Dataset: Prepare the data for Linear Regression Algorithm (all the values should lie on the common scale) to make the interpretation easy.**  **Check how well the model fit the dataset by Computing the Loss Function**  **Split the Dataset into training and Testing Dataset**  **Apply the Linear Regression**  **Model Training Visualization**  **Visualisation of the error Values**  **Train the Model with the different attributes and predict the prices.** | **Create the Gaussian Naive Bayes Classifier: Split The dataset and**  **Train the Model. Perform the Classification on the Testing dataset** |
| **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.** | **Evaluate the Model: Generate the Confusion Matrix to determine the accuracy of Classification by displaying the results for the testing dataset and visualize the performance of the Model** |
| **Dataset Link:** [**https://www.kaggle.com/vikrishnan/boston-house-prices**](https://www.kaggle.com/vikrishnan/boston-house-prices) | **Dataset Link :** [**https://www.kaggle.com/mlg-ulb/creditcardfraud**](https://www.kaggle.com/mlg-ulb/creditcardfraud) |

**[PROJECT 3&4]**

|  |  |
| --- | --- |
| **Loan Application Approval Prediction Project** | **Predict what is the sales of Big Mart in the upcoming months** |
| **Plan the Model and Import the relevant libraries**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required Dataset** | **Plan the Model and import the relevant libraries.**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required Dataset** |
| **Data Processing: Load the data and**  **Concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data.** | **Data Processing: Load the data and**  **Concatenate the data**  **Data Visualization: Description of the dataset to get the basic insight on the data.** |
| **Normalization of the Dataset: Prepare the data for any classification Algorithm (all the values should lie on the common scale) to make the interpretation easy**  **Split the Dataset into training and testing Dataset**  **Apply the Classification**  **Model Training**  **Visualisation of the error Values**  **Train the Model with the different attributes and predict the Approval.** | **Normalization of the Dataset: Prepare the data for Regression Algorithm (all the values should lie on the common scale) to make the interpretation easy.**  **Check how well the model fit the dataset by Computing the Loss Function.**  **Split the Dataset into training and testing Dataset**  **Apply The Relevant Algorithm to find out**   1. **Predict the average sales by product** 2. **Predict the sales of product for every outlet** |
| **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.**  **Apply Confusion Matrix and compute Precision and Recall** | **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.** |
| **Dataset Link :** [**https://www.kaggle.com/arbazkhan971/loan-approval-analysis**](https://www.kaggle.com/arbazkhan971/loan-approval-analysis) | **Dataset Link:** [**https://www.kaggle.com/mrmorj/big-mart-sales**](https://www.kaggle.com/mrmorj/big-mart-sales) |
|  |  |

**[PROJECT 5&6]**

|  |  |
| --- | --- |
| **Classification of Wheat Seed Using Classification Algorithm** | **E-mail Fraud detection using Classification** |
| **Plan the Model and import the relevant libraries.**  **DATA PROCESSING STAGES** **Data Acquisition: Import the required Dataset** | **Plan Model and import the relevant libraries.**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required Dataset** |
| **Data Processing: Load the data and**  **Concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data.**  **To simplify the Model, Use the Feature Selection method to select the attribute that has strongest relationship to the Target** | **Data Processing: Load the data and**  **Concatenate the data (features and targets)**  **Data Visualization**  **Description of the dataset to get the basic insight on the data.**  **To simplify the Model, Use the Feature Selection method to select the attribute that has strongest relationship to the Target**  **Apply Vectorization technique to convert Textual data into Numerical form** |
| **Apply The Relevant Classifier Algorithm**  **Split The dataset**  **Train the Model**  **Perform the Classification on the Testing dataset** | **Apply The Relevant Classifier Algorithm**  **Split The dataset**  **Train the Model**  **Perform the Classification on the Testing dataset** |
| **Evaluate the Model: Generate the Confusion Matrix to determine the accuracy of Classification by displaying the results for the testing dataset and visualize the performance of the Model.** | **Evaluate the Model: Generate the Confusion Matrix to determine the accuracy of Classification by displaying the results for the testing dataset and visualize the performance of the Model.** |
| **Dataset Link :** [**https://www.kaggle.com/jmcaro/wheat-seedsuci**](https://www.kaggle.com/jmcaro/wheat-seedsuci) | **Dataset Link:** [**https://www.kaggle.com/uciml/sms-spam-collection-dataset**](https://www.kaggle.com/uciml/sms-spam-collection-dataset) |

**[PROJECT 7&8]**

|  |  |
| --- | --- |
| **Car Price Prediction Project using Regression** | **Classifying the fruits with respect to its weight, height and color score** |
| **Plan the Model and import the relevant libraries.**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required Dataset** | **Plan the Model and Import the relevant Libraries.**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the required dataset.** |
| **Data Processing: Load the data and**  **Concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data.** | **Data Processing: Load the data and**  **Concatenate the data (features and targets)**  **Data Visualization: Description of the dataset to get the basic insight on the data.** |
| **Normalization of the Dataset: Prepare the data for Regression Algorithm (all the values should lie on the common scale) to make the interpretation easy and check how well the model fit the dataset by Computing the Loss Function.**  **Splitting the Dataset into training and testing Dataset.**  **Apply the Linear Regression:**  **Model Training Visualization**  **Visualisation of the error Values**  **Train the Model with the different attributes and predict the prices.** | **Apply The Relevant Classifier Algorithm**  **Split The dataset and Train the Model**  **Perform the Classification on the Testing dataset.** |
| **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.** | **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.** |
| **Dataset Link:** [**https://www.kaggle.com/goyalshalini93/car-data**](https://www.kaggle.com/goyalshalini93/car-data) | **Dataset Link:** <https://www.kaggle.com/mjamilmoughal/fruits-with-colors-dataset> |

**[PROJECT 9&10]**

|  |  |
| --- | --- |
| **Prediction of the salary of the employees using Regression Algorithm** | **Detection of Heart Disease using Classification Algorithm** |
| **Planning the Model and Import the relevant libraries.**    **DATA PROCESSING STAGES** **Data Acquisition: Import the Dataset** | **Planning the Model and import the relevant libraries.**  **DATA PROCESSING STAGES**  **Data Acquisition: Import the dataset** |
| **Data Processing: Load the data and**  **Concatenate the data (features and targets).**  **Data Visualization: Description of the dataset to get the basic insight on the data.** | **Data Processing: Load the data and**  **Concatenate the data (features and targets).**  **Data Visualization: Description of the dataset to get the basic insight on the data.**  **To simplify the Model, Use the Feature Selection method to select the attribute that has strongest relationship to the Target** |
| **Normalization of the Dataset: Prepare the data for Regression Algorithm (all the values should lie on the common scale) to make the interpretation easy.**  **Check how well the model fit the dataset by Computing the Loss Function.**  **Split the Dataset into training and testing Dataset**  **Apply the Linear Regression**  **Model Training Visualization**  **Visualisation of the error Values**  **Train the Model with the different attributes and predict the prices.** | **Apply The Relevant Classifier Algorithm**  **Split The dataset and Train the Model**  **Perform the Classification on the Testing dataset.** |
| **Evaluate the Model: Compute the accuracy between Actual values and Predicted Values.** | **Evaluate the Model: Generate the Confusion Matrix to determine the accuracy of Classification by displaying the results for the testing dataset and visualize the performance of the Model** |
| **Dataset Link :** [**https://www.kaggle.com/manishkc06/engineering-graduate-salary-prediction?select=Engineering\_graduate\_salary.csv**](https://www.kaggle.com/manishkc06/engineering-graduate-salary-prediction?select=Engineering_graduate_salary.csv) | **Dataset Link:** [**https://www.kaggle.com/ronitf/heart-disease-uci**](https://www.kaggle.com/ronitf/heart-disease-uci) |

Note: Participants are free to use other dataset based on similar theme as well