Projects

Instructions

- Please complete all the project before 23rd September, 2020, 06:00 PM
- Any submission after that will not be accepted.
- All projects and exercise should be submitted in a single file (No separate file for each project)
- File should be in .ipynb format only.
- Copy and Paste in any form is not recommended and will have negative impact on the evaluation of project.
- New idea and concept will be highly appreciated.
- Practice: Exercise 1 is for practice and is not Considered for Evaluation.

Project 1: Profit Predicting

From Data given it is required to develop the model to predict the profit of the company. Dataset file name: Companies-profit.csv

Task:

1) Load the dataset
2) Analyse the data visually and give justification for selecting particular algorithm for developing the model.
3) Split it into test and train.
4) Develop a model for predicting the profit of the company
5) Measure all the parameters of the model
6) Measure the performance of the model

Project 2: Profit Predicting

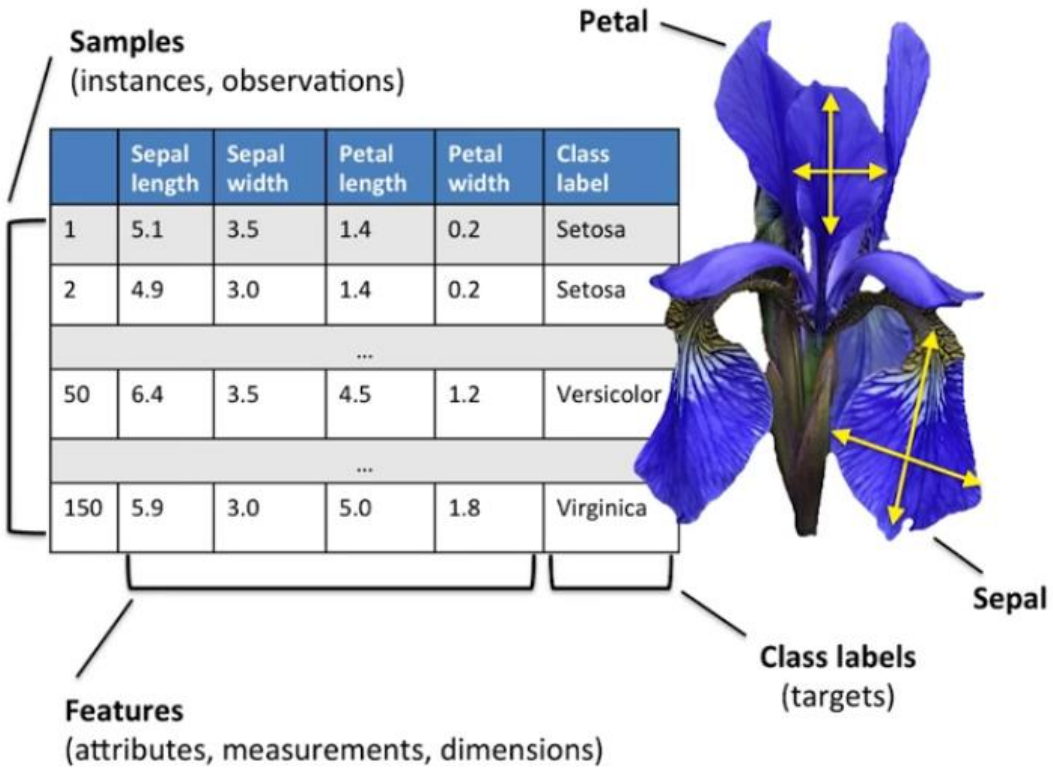
- Use wine dataset from sklearn
- Load the dataset and split it into test and train.
- Develop the model to Classify wines into different categories.
- Train the model using Gaussian and Multinominal classifier, please mention which model is better and why.
- Use the trained model to perform some predictions on test data.

Project 3: Operation successful or not

- Datasets for Kyphosis Disease operation is provided. Dataset file name: kyphosis.csv
  1. Load the dataset and split it into test and train.
  2. Analyse the Data Visually.
  3. Find limitation with data if any in developing the model
  4. Develop the model to identify whether operation successful or not using Decsion Tree algorithm
  5. Measure the performance of the model
  6. Try to improvise the model by some other technique.
  7. Measure the performance of the model

Project: 4

- Use iris flower dataset from sklearn library and try to form clusters of flowers using petal width and length features.
- Analyse the Data Visually.
- Drop other two features for simplicity.
- Try to Use K Mean Clustering
- Dataset from Sklearn dataset Library



	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Practice: Exercise 1

1. What is K-Nearest Neighbor(KNN) Algorithm for Machine Learning
2. How does K-NN work?
3. Implementation of the KNN algorithm- Take dataset: diabetes.csv
   - Data Pre-processing step-if any
   - Fitting the K-NN algorithm to the Training set
   - Predicting the test result
   - Test accuracy of the result