A PROJECT REPORT

ON

ANALYSIS OF A DATASET THROUGH DATA MINING ALGORITHMS

SUBMITTED TO UNIVERSITY OF ROME - TOR VERGATA, ROME
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE AWARD OF THE 45 DAYS SUMMER TRAINING PROGRAM
(VISHWANIKETAN UG-FELLOWSHIP)

BY

AKHIL MUKESH SETHIA

UNDER THE GUIDANCE OF
Prof. MAURO DE SANCTIS

UNIVERSITY OF ROME-TOR VERGATA, ROME

VIA DEL POLITECNICO,

1, 00133 ROMA RM

2017-18

UNIVERSITY OF ROME-TOR VERGATA, ROME

VIA DEL POLITECNICO,

1, 00133 ROMA RM

Certificate

This is to certify that Dissertation report entitled,

ANALYSIS OF A DATASET THROUGH DATA MINING ALGORITHMS

Submitted By

AKHIL MUKESH SETHIA

is a bonafide work carried out by them under the supervision of **Prof. Mauro De Sanctis** and it is submitted towards the partial fulfilment of the requirement of **University of Rome-Tor Vergata, Rome** for the award of the 45 days Summer Training Program (Vishwaniketan UG-Fellowship).

Dr. Aparna Bhirangi	Prof. Ernestina Cianca
Indian Supervisor	**Vice CTIF, Italy Coordinator**

Place: Rome, Italy
Date: 25/07/2017

UNIVERSITY OF ROME-TOR VERGATA

VIA DEL POLITECNICO,

1, 00173 ROMA RM

Certificate by Guide

This is to certify that**, Mr. Akhil Mukesh Sethia** have completed the dissertation work under my guidance and supervision and that, I have verified the work for its originality in documentation, problem statement, implementation and results presented in the dissertation. Any reproduction of other necessary work is with the prior permission and has given due ownership and included in the references.

.....

Prof. Mauro De Sanctis

University of Rome-Tor Vergata

Acknowledgment

I am grateful to University of Rome-Tor Vergata and Vishwaniketan iMEET for presenting me with an opportunity to be able to come here and collaborate with my mentor **Prof. Mauro De Sanctis.**

Without his guidance and insight I would not have been able to complete my project. His advice and support have helped me shape my project and my perspective toward approaching this project. I would like to express deepest gratitude towards **Prof. Mauro De Sanctis** and sincerely thank him for his time and his inputs.

I sincerely thank **Dr. Simone Di Domenico** for his supervision and his guidance. His continuous suggestions regarding the scope of the project were invaluable. I would also like to thank **Dr. Aparna Bhirangi** whose support throughout the summer training program and guidance on documentation and structuring of the report help me collate my findings.

Without the help and support of the aforementioned individuals, this project would have not have reached a conclusion.

This has been an excellent opportunity for me to get exposure to the technology and quality of research at University of Rome- Tor Vergata. I have personally gained a lot from this summer training. This UG-fellowship has been instrumental for my development, both as an individual and a professional and I am grateful to Vishwaniketan iMEET for the same.

Date: 25/07/2017 **AKHIL MUKESH SETHIA**

Place: Rome, Italy

List of Tables

List of Figures

Abstract

Data science lays the foundation to future technologies like artificial intelligence, automation and IoT. To make these technologies scalable and cost feasible it is fundamental for data mining algorithms to have low running time. Through maximisation of computational speed, we can implement real time artificial intelligence and IoT solutions from data mining algorithms. This report analyses the performance of data mining algorithms with respect to their accuracy and computational speed. Classification and clustering algorithms namely, naive Baye's, k-nearest neighbours, decision tree classification, agglomerative hierarchical clustering and k-means clustering algorithm have been analysed. Eleven classifier and cluster models have been generated each trained and tested on different combination of attributes of the Iris dataset for each algorithm. A performance metric is formulated, where ratio between normalised accuracy and testing time has been taken. The time and accuracy have been normalised to eliminate the results being skewed by one parameter. Based on these results, we observe that better performance can be achieved if algorithms operate on a subset of all attributes instead of the whole Iris dataset. A reduction in dimensionality of the dataset to optimise the speed of processing has been hypothesised.

Contents

CHAPTER 01
INTRODUCTION

Data mining is the process of evaluating patterns within a large data set. Data mining techniques include creating regression models of the data, recognising anomalies, associations and clusters within the data, and summarising the entire data. It transforms raw data into a comprehensible form.

Data mining algorithms have a variety of applications. Data mining algorithms are used for detecting similarity in DNA sequences. Patient demographic data is data mined to yield insight into ailments and genetic mutations with respect to geography and inheritance.

Pattern assessment in interpreting market research consumer, can aid businesses identify current and predict future market demand and trends and accordingly develop marketing strategies.

Data science forms the basis of machine learning, which is crucial in automation and artificial intelligence. Classification and clustering algorithms are important to give machine decision making and learning capability.

Optimisation of data mining techniques would effectively reduce the computational time and enable us to make larger computations which require hardware of higher specification more feasible, thus making it more accessible to real time solutions.

In this project we focus on analysis of clustering and classification analysis namely, k-means, agglomerative hierarchical clustering, naive Baye's, k-nearest neighbours and decision tree classification.

This project has a different approach in the analysis, and it tries to compare both efficiency between algorithms and also the efficiency between classifiers of same algorithm but which operate on a segment of the dataset. Many algorithms used here, depend on the dimensionality of the data. If we can reduce the dimensionality of the data, we would reduce the effective functioning time of the data by a large factor. Hence we have conducted experiments on the dataset to show that this is a possible and a feasible optimisation solution which can be applied to a dataset to improve performance.

CHAPTER 02

BACKGROUND

To understand the functioning of each algorithm and its strengths and weaknesses, a literary survey was performed on the current level of research already made on the analysis of datasets. To do this, research papers were referenced which introduced and analysed of each algorithm. The conventional algorithms along with their optimised versions were studied. The optimised versions were studied to understand the scope of improvement in the conventional algorithms.
The following papers were used for algorithmic support and understanding the use and application of data mining algorithms.

'A Survey Of Decision Tree Methodology' by S. Rasoul Safavian and David Landgrebe : was studied to understand the concept of decision trees.

'Naive Bayes classification algorithm based on small sample set' by Yuguang Huang and Lei Li : was studied to understand the calculations of naive Baye's methods.

'An efficient k-means clustering algorithm: analysis and implementation' by T. Kanungo , D.M. Mount , N.S. Netanyahu C.D. Piatko , R. Silverman and A.Y. Wu : was studied to understand efficient optimisations which can be made to k-means algorithm, and also to understand the applications of a kd-tree data structure to implement this algorithm.

'Evaluation of hierarchical clustering algorithms for document datasets' by
Ying Zhao and George Karypis: used to understand types of hierarchical algorithm, and its extensions, and the discrepancy in performance. It was used to study how conventional algorithm can be optimised.
'A k-nearest neighbor based algorithm for multi-label classification' by Min-Ling Zhang and Zhi-Hua Zhou: studied the convention k-nn classification model and studied its optimisations using lazy learning approach.

CHAPTER 03

PROBLEM DEFINITION AND SCOPE

The primary objective of this project is the analysis of various classification and clustering algorithms used to mine data. Parameters of performance comparison is the time computed while testing the algorithm, and the accuracy of a particular algorithm. Based on the observation results, further optimisation possibilities have been discussed in the project.

The report analyses classification and clustering algorithms. The classification algorithms analysed are naive Baye's, k-nearest neighbours and decision tree classification. Clustering algorithms analysed are k-means clustering and agglomerative hierarchical clustering. These algorithms can be classified into either supervised or unsupervised learning.

Supervised learning is a branch of data science where the classifier or the algorithms is trained initially, by data similar to that on which it is tested. Classification is a branch of supervised learning. Performance of classification algorithms depends on the amount of training data fed.

Classification algorithms make predictions about data based on its attributes values.

Unsupervised learning is when algorithms are directly run on the data and do not require training. These algorithms directly operate on the data, and produce expected results. Clustering algorithms fall under the unsupervised category, and are used to detect clouts and gropings within the data.

All classifiers have been tested and trained on the Iris flower dataset, where 50% of the dataset has been used for training, whereas the other 50% is used for testing the classifier. The results from the tested data is then converted into a confusion matrix, and the accuracy of classification is calculated from the confusion matrix. A ratio between accuracy and time is computed then to rate the best classifier. The most optimal classifier has then been chosen amongst all classifiers.

In clustering, the algorithms have directly run on the entire Iris flower dataset. The silhouette value and the purity of cluster model has been calculated. Then a performance metric with respect to silhouette value, purity and time has been computed and the most optimal cluster model has been chosen to predict the most optimal model.

CHAPTER 04

PROJECT PLAN

The Iris flower dataset had 150 tuples, with 4 attributes each. Tuples have individual value for these 4 attributes, namely 'Sepal Length', 'Sepal Width', 'Petal Length' and 'Petal Width'. Each of these tuples are divided into 3 respective classes, 'Setosa', 'Versicolor' and 'Virginica', depending on the value of their attributes.

For classification models, the data is randomly divided into 2 equal halves, of which one is used for training the classifier whereas the other is used for testing. In the testing set used for classification, the number of tuples for class 'Setosa' is 23, for 'Versicolor' is 25 and for 'Virginica' is 27 (For more detail, refer Appendix I).

For clustering models, the entire dataset has been used to test the algorithm, the set contains 50 tuples of each class which are then clustered by the model without prior knowledge about the data.(For detailed information on datasets, refer Appendix I)

For each classification and clustering algorithm 11 different classifiers have been designed. The difference between each classifier, is the training methodology used and the testing data which is fed to them. For each classifier the number of attributes and the combination of attributed used to test and train the classifier is different. The classifier then predicts the test dataset on the combination of attributes for which training has been provided.

Clustering models similar to the classification models, cluster the dataset only on specific attributes. They produce results based on the of specific combination of attributes.

Models consider each combination of attribute as an entire dataset and then accordingly cluster and classify on the test dataset.

The combination of attributes used on classification and clustering models are 'Sepal Length, Sepal Width, Petal Length And Petal Width', 'Sepal Length, Sepal Width And Petal Length', 'Sepal Width, Petal Length And Petal Width', 'Sepal Length, Sepal Width And Petal Width', 'Sepal Length, Petal Length And Petal Width', 'Sepal Width And Sepal Length', 'Petal Length And Sepal Length', 'Petal Width And Sepal Length', 'Petal Width And Sepal Length', 'Petal

Width And Sepal Width' and 'Petal Width And Petal Length'. (For detailed information on results of clustering and classification model, refer Appendix I)

The testing time for each classifier to classify all data points was recorded. The test time has been averaged over 10 iterations to get consistent results. From these classification results a confusion matrix has been calculated, which shows how many points have been classified in which class. It also shows the number of points which have been miscalculated and shows both the actual and the predicted class of the data point. This matrix has been used to calculate the average accuracy of each classifier. This has been tabulated in the results. A performance metric has been designed to normalise and then calculate the ratio between accuracy and time consumed to find out which is the most optimal classifier for an algorithm. An average performance metric and an average accuracy and testing time has also be computed so that we compare an algorithmic performance.

The clustering model was fed the entire dataset and it separated the data into clusters. In both k-means and agglomerative hierarchical clustering the number of clusters were preset to 3. In k-means the algorithm repeatedly iterates 4 times to pick the minimum starting cluster points, as this affects the performance of the cluster. The testing time for each classifier to classify all data points was recorded. The test time has been averaged over 10 iterations to get consistent data. To evaluate the accuracy of the clusters the purity within each cluster has been calculated. The silhouette value has been additionally calculated to show the density of each cluster. A performance metric has been calculated considering all 3 performance parameters. Based on this the cluster models have been compared and the optimal cluster model for this dataset has been shown.

In calculation of time, the disk access cost, processing time and hardware costs have not been factored out and is included within the results. Although, due to the averaged iteration results the timing results obtained are stable and predictions can be made on them. The stability of each classifier was verified as with a given dataset the classifier produced the same set of results each time. (For detailed information on comparison and analysis of clustering and classification model, refer Appendix I)

(For complete source code, refer Appendix II)

CHAPTER 05

DETAILED DESIGN

Naive Bayes Algorithm:

Computational Complexity: $o(nd)$

Where 'd' is the number of dimensions in the data and 'n' is the number of tuples in the data.

Figure 5.1: Naive baye's Algorithm

NAIVEBAYES ($\mathbf{D} = \{(\mathbf{x}_j, y_j)\}_{j=1}^{n}$):

for $i = 1, \ldots, k$ **do**
  $\mathbf{D}_i \leftarrow \{\mathbf{x}_j \mid y_j = c_i, j = 1, \ldots, n\}$ // class-specific subsets
  $n_i \leftarrow |\mathbf{D}_i|$ // cardinality
  $\hat{P}(c_i) \leftarrow n_i / n$ // prior probability
  $\hat{\boldsymbol{\mu}}_i \leftarrow \frac{1}{n_i} \sum_{\mathbf{x}_j \in \mathbf{D}_i} \mathbf{x}_j$ // mean
  $\mathbf{Z}_i = \mathbf{D}_i - \mathbf{1} \cdot \hat{\boldsymbol{\mu}}_i^T$ // centered data for class c_i
  for $j = 1, .., d$ **do** // class-specific variance for X_j
    $\hat{\sigma}_{ij}^2 \leftarrow \frac{1}{n_i} \mathbf{Z}_{ij}^T \mathbf{Z}_{ij}$ // variance
  $\hat{\boldsymbol{\sigma}}_i = \left(\hat{\sigma}_{i1}^2, \ldots, \hat{\sigma}_{id}^2\right)^T$ // class-specific attribute variances
return $\hat{P}(c_i), \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\sigma}}_i$ for all $i = 1, \ldots, k$

TESTING (**x and** $\hat{P}(c_i)$, $\hat{\boldsymbol{\mu}}_i$, $\hat{\boldsymbol{\sigma}}_i$, **for all** $i \in [1, k]$):

$\hat{y} \leftarrow \arg\max_{c_i} \left\{ \hat{P}(c_i) \prod_{j=1}^{d} f(x_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2) \right\}$

return $\hat{y}$

*Note: This algorithm has been sourced from'DATA MINING AND ANALYSIS.'[1]

K-Nearest Neighbours Algorithm:

Computational Complexity: $o(ndk)$

Where 'd' is the number of dimensions in the data, 'k' is the number of neighbours and 'n' is the number of tuples in the data.

Figure 5.2: K-Nearest Neighbours Algorithm

k-Nearest Neighbor
Classify (**X**, **Y**, *x*) // **X**: training data, **Y**: class labels of **X**, *x*: unknown sample
for *i* = 1 **to** *m* **do**
  Compute distance $d(\mathbf{X}_i, x)$
end for
Compute set *I* containing indices for the *k* smallest distances $d(\mathbf{X}_i, x)$.
return majority label for $\{\mathbf{Y}_i$ where $i \in I\}$

*Note: This algorithm has been sourced from 'A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images.'[2]

Decision Tree Algorithm:

Computational complexity: $o(dn^2 logn)$

Where d = number of dimensions, and n is the number of tuples.

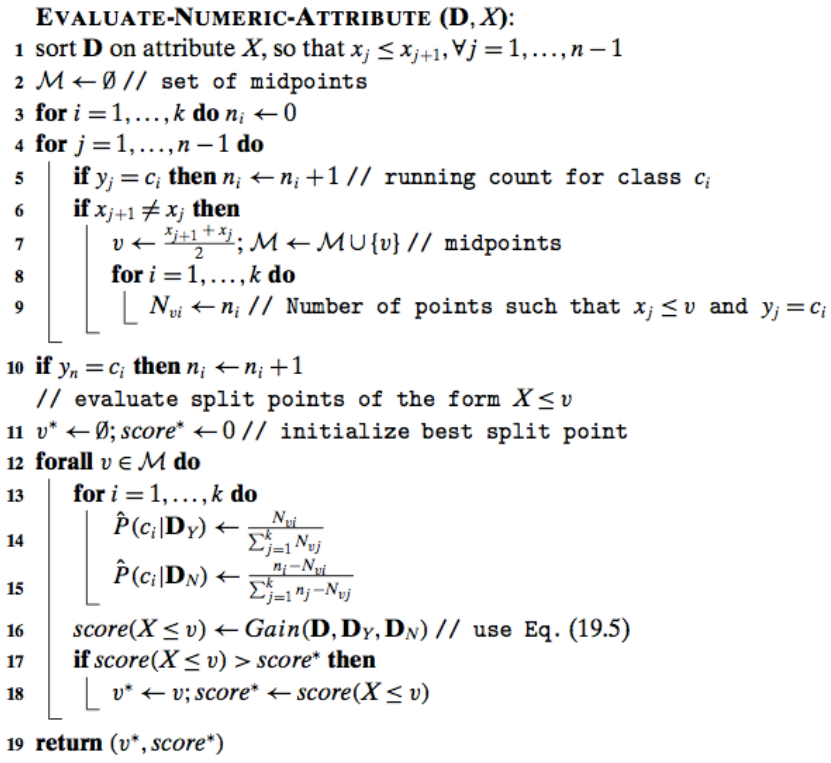
Figure 5.3: Decision Tree Algorithm

DECISIONTREE $(\mathbf{D}, \eta, \pi)$:

$n \leftarrow |\mathbf{D}|$ // partition size
$n_i \leftarrow |\{\mathbf{x}_j | \mathbf{x}_j \in \mathbf{D}, y_j = c_i\}|$ // size of class c_i
$purity(\mathbf{D}) \leftarrow \max_i \left\{\frac{n_i}{n}\right\}$
if $n \leq \eta$ *or* $purity(\mathbf{D}) \geq \pi$ **then** // stopping condition
    $c^* \leftarrow \arg\max_{c_i} \left\{\frac{n_i}{n}\right\}$ // majority class
    create leaf node, and label it with class c^*
    return
$(split\ point^*, score^*) \leftarrow (\emptyset, 0)$ // initialize best split point
foreach *(attribute* X_j*)* **do**
    if *(*X_j *is numeric)* **then**
        $(v, score) \leftarrow$ EVALUATE-NUMERIC-ATTRIBUTE$(\mathbf{D}, X_j)$
        if $score > score^*$ **then** $(split\ point^*, score^*) \leftarrow (X_j \leq v, score)$
    else if *(*X_j *is categorical)* **then**
        $(V, score) \leftarrow$ EVALUATE-CATEGORICAL-ATTRIBUTE$(\mathbf{D}, X_j)$
        if $score > score^*$ **then** $(split\ point^*, score^*) \leftarrow (X_j \in V, score)$
// partition $\mathbf{D}$ into $\mathbf{D}_Y$ and $\mathbf{D}_N$ using $split\ point^*$, and call recursively
$\mathbf{D}_Y \leftarrow \{\mathbf{x} \in \mathbf{D} \mid \mathbf{x}$ satisfies $split\ point^*\}$
$\mathbf{D}_N \leftarrow \{\mathbf{x} \in \mathbf{D} \mid \mathbf{x}$ does not satisfy $split\ point^*\}$
create internal node $split\ point^*$, with two child nodes, $\mathbf{D}_Y$ and $\mathbf{D}_N$
DECISIONTREE$(\mathbf{D}_Y)$; DECISIONTREE$(\mathbf{D}_N)$

*Note: This algorithm has been sourced from'DATA MINING AND ANALYSIS.'[1]

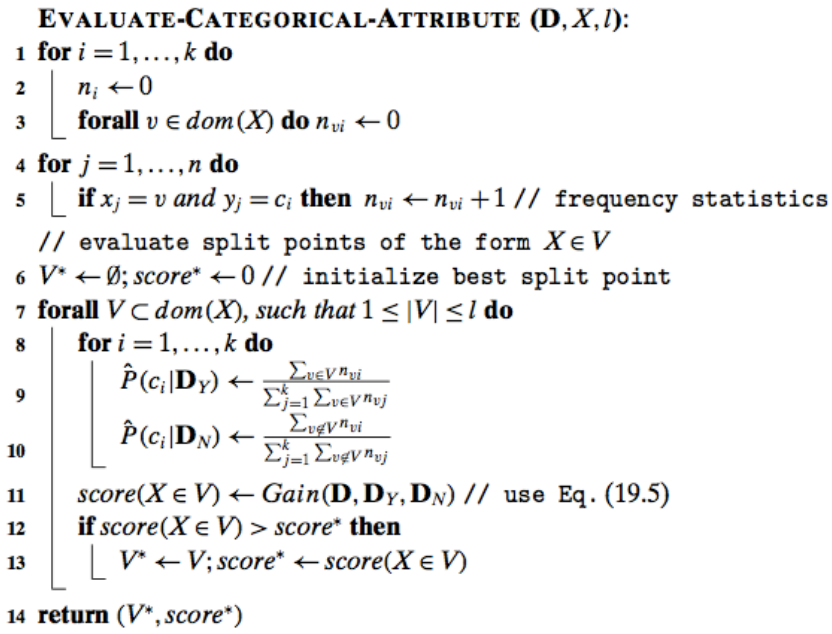
Figure 5.4: Gain of Numeric Attributes



```
EVALUATE-NUMERIC-ATTRIBUTE (D, X):
sort D on attribute X, so that x_j ≤ x_{j+1}, ∀j = 1, ..., n−1
M ← ∅ // set of midpoints
for i = 1, ..., k do n_i ← 0
for j = 1, ..., n−1 do
    if y_j = c_i then n_i ← n_i + 1 // running count for class c_i
    if x_{j+1} ≠ x_j then
        v ← (x_{j+1} + x_j)/2; M ← M ∪ {v} // midpoints
        for i = 1, ..., k do
            N_vi ← n_i // Number of points such that x_j ≤ v and y_j = c_i
if y_n = c_i then n_i ← n_i + 1
// evaluate split points of the form X ≤ v
v* ← ∅; score* ← 0 // initialize best split point
forall v ∈ M do
    for i = 1, ..., k do
        P̂(c_i|D_Y) ← N_vi / Σ_{j=1}^{k} N_vj
        P̂(c_i|D_N) ← (n_i − N_vi) / Σ_{j=1}^{k} (n_j − N_vj)
    score(X ≤ v) ← Gain(D, D_Y, D_N) // use Eq. (19.5)
    if score(X ≤ v) > score* then
        v* ← v; score* ← score(X ≤ v)
return (v*, score*)
```

*Note: This algorithm has been sourced from 'DATA MINING AND ANALYSIS.'[1]

Figure 5.5: Gain of Categorical Attributes



```
EVALUATE-CATEGORICAL-ATTRIBUTE (D, X, l):
for i = 1, ..., k do
    n_i ← 0
    forall v ∈ dom(X) do n_vi ← 0
for j = 1, ..., n do
    if x_j = v and y_j = c_i then n_vi ← n_vi + 1 // frequency statistics
// evaluate split points of the form X ∈ V
V* ← ∅; score* ← 0 // initialize best split point
forall V ⊂ dom(X), such that 1 ≤ |V| ≤ l do
    for i = 1, ..., k do
        P̂(c_i|D_Y) ← Σ_{v∈V} n_vi / Σ_{j=1}^{k} Σ_{v∈V} n_vj
        P̂(c_i|D_N) ← Σ_{v∉V} n_vi / Σ_{j=1}^{k} Σ_{v∉V} n_vj
    score(X ∈ V) ← Gain(D, D_Y, D_N) // use Eq. (19.5)
    if score(X ∈ V) > score* then
        V* ← V; score* ← score(X ∈ V)
return (V*, score*)
```

*Note: This algorithm has been sourced from 'DATA MINING AND ANALYSIS.'[1]

K-Means Clustering Algorithm:

Computational complexity: $o(tnkd)$

Where 't' is the number of iterations to convergence, and 'n' is the number of tuples.

Figure 5.6: K-means Clustering Algorithm

K-MEANS $(\mathbf{D}, k, \epsilon)$:

$t = 0$
Randomly initialize k centroids: $\boldsymbol{\mu}_1^t, \boldsymbol{\mu}_2^t, \ldots, \boldsymbol{\mu}_k^t \in \mathbb{R}^d$
repeat
  $t \leftarrow t + 1$
  $C_j \leftarrow \emptyset$ for all $j = 1, \cdots, k$
  // Cluster Assignment Step
  foreach $\mathbf{x}_j \in \mathbf{D}$ **do**
    $j^* \leftarrow \arg\min_i \left\{ \left\| \mathbf{x}_j - \boldsymbol{\mu}_i^t \right\|^2 \right\}$ // Assign $\mathbf{x}_j$ to closest centroid
    $C_{j^*} \leftarrow C_{j^*} \cup \{\mathbf{x}_j\}$
  // Centroid Update Step
  foreach $i = 1$ *to* k **do**
    $\boldsymbol{\mu}_i^t \leftarrow \frac{1}{|C_i|} \sum_{\mathbf{x}_j \in C_i} \mathbf{x}_j$
until $\sum_{i=1}^k \left\| \boldsymbol{\mu}_i^t - \boldsymbol{\mu}_i^{t-1} \right\|^2 \leq \epsilon$

*Note:

This algorithm has been sourced from 'DATA MINING AND ANALYSIS.'[1]

Agglomerative Hierarchical Clustering Algorithm:

Computational complexity: $o(n^2 log n)$

Where 'd' is the number of dimensions and 'n' is the number of tuples.

Figure 5.7: Agglomerative Hierarchical Clustering Algorithm

AGGLOMERATIVECLUSTERING$(\mathbf{D}, k)$:

$\mathcal{C} \leftarrow \{C_i = \{\mathbf{x}_i\} \mid \mathbf{x}_i \in \mathbf{D}\}$ // Each point in separate cluster
$\boldsymbol{\Delta} \leftarrow \{\delta(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i, \mathbf{x}_j \in \mathbf{D}\}$ // Compute distance matrix
repeat
  Find the closest pair of clusters $C_i, C_j \in \mathcal{C}$
  $C_{ij} \leftarrow C_i \cup C_j$ // Merge the clusters
  $\mathcal{C} \leftarrow \left(\mathcal{C} \setminus \{C_i, C_j\}\right) \cup \{C_{ij}\}$ // Update the clustering
  Update distance matrix $\boldsymbol{\Delta}$ to reflect new clustering
until $|\mathcal{C}| = k$

*Note: This algorithm has been sourced from 'DATA MINING AND ANALYSIS.' [1]

CHAPTER 06

IMPLEMENTATION AND RESULT

Classification Algorithms:

The following tables show the result produced by each classifier. Each classifier is grouped according to its algorithm. The 'PARAMETERS' column states the attributes on which the classifier has been trained and tested upon. The 'AVERAGE' tuple shows the average performance of each classifier for a specific algorithm. The 'CONFUSION MATRIX' shows the number of misclassified points, and their erroneous classes.

The testing and training data for all classifiers across all algorithms are consistent so that comparable results can be formulated.

All the testings were performed on 1.3 GHz Intel Core i5 processor.

The measuring distance used for k-nearest neighbour is Euclidean distance.

Kd-tree data structure has been used to optimise the performance of the k-nearest neighbours algorithm.

Generally, points have been misclassified between class 'Versicolor' and the class 'Virginica'. This is because of a high degree of overlap between these classes.

Naive Baye's Classifier has an average accuracy of 91.64% and a mean testing time of 0.03065 seconds. K-nearest neighbours classifier has an average accuracy of 90.82% and a mean testing of 0.03601 seconds. Decision Tree Classifier has an average accuracy of 89.93% and a mean testing time of 0.023582 seconds. We can see that 96% is the highest accuracy we can achieve with a computational time of 0.021303 seconds by using naive Baye's algorithm trained on parameters 'Sepal Width, Petal Length And Petal Width' .

Naive Baye's classifier has the highest average accuracy of considering all attributional combinations and the Decision Tree classifier has the lowest computational testing speed of all algorithms. K-nearest neighbours has the highest computational testing time whereas decision tree classifier has the lowest accuracy.

(For detailed information on results of classification model, and comparison and analysis, refer Appendix I).

Table 6.1-Naive Baye's Classifier

PARAMETERS	ACCURACY	TESTING TIME	CONFUSION MATRIX	PERFORMANCE
Sepal Length, Sepal Width, Petal Length And Petal Width	94.6667%	0.1691s	23 0 0 0 22 3 0 1 26	3.8424
Sepal Length, Sepal Width And Petal Length	88%	0.020681s	23 0 0 0 21 4 0 5 22	4.3857
Sepal Width, Petal Length And Petal Width	96%	0.021303s	23 0 0 0 23 2 0 1 26	4.4681
Sepal Length, Sepal Width And Petal Width	94.6667%	0.023927s	23 0 0 0 22 3 0 1 26	4.4428
Sepal Length, Petal Length And Petal Width	94.6667%	0.018614s	23 0 0 0 22 3 0 1 26	4.4664
Sepal Width And Sepal Length	80%	0.010596s	22 1 0 0 19 6 0 8 19	4.3358
Petal Length And Sepal Length	89.3333%	0.016948s	23 0 0 0 20 5 0 3 24	4.4169
Petal Width And Sepal Length	90.6667%	0.013831s	23 0 0 0 22 3 0 4 23	4.4459
Petal Width And Sepal Length	90.6667%	0.016223s	23 0 0 0 22 3 0 4 23	4.4357
Petal Width And Sepal Width	94.6667%	0.013999s	23 0 0 0 23 2 0 2 25	4.4871
Petal Width And Petal Length	94.6667%	0.01191s	23 0 0 0 22 3 0 1 26	4.4964
AVERAGE VALUES	91.64%	0.03065s	23 0 0 0 22 3 0 3 24	4.3815

*Note: Self Generated data. This table shows the results attained by performing tests on classifiers running on Naive Baye's algorithm.

Table 6.2-K-Nearest Neighbours Classifier

ATTRIBUTES	ACCURACY	TESTING TIME	CONFUSION MATRIX	PERFORMANCE
Sepal Length, Sepal Width, Petal Length And Petal Width	96%	0.19601s	23 0 0 0 22 3 0 0 27	3.7519
Sepal Length, Sepal Width And Petal Length	94.6667%	0.026067s	23 0 0 0 22 3 0 1 26	4.4333
Sepal Width, Petal Length And Petal Width	93.3333%	0.027566s	23 0 0 0 21 4 0 1 26	4.4128
Sepal Length, Sepal Width And Petal Width	93.3333%	0.019711s	23 0 0 0 20 5 0 0 27	4.4476
Sepal Length, Petal Length And Petal Width	96%	0.024565s	23 0 0 0 22 3 0 0 27	4.4536
Sepal Width And Sepal Length	69.6667%	0.017959s	22 0 1 0 18 7 0 15 12	4.1682
Petal Length And Sepal Length	89.3333%	0.015347s	23 0 0 0 23 2 0 2 25	4.4234
Petal Width And Sepal Length	93.3333%	0.018308s	23 0 0 0 22 3 0 2 25	4.4539
Petal Width And Sepal Length	88%	0.018493s	23 0 0 0 18 7 0 2 25	4.3953
Petal Width And Sepal Width	92%	0.014737s	23 0 0 0 21 4 0 2 25	4.4556
Petal Width And Petal Length	93.3333%	0.01731s	23 0 0 0 20 5 0 0 27	4.4583
AVERAGE VALUES	90.82%	0.03601	23 0 0 0 21 4 0 2 25	4.3494

*Note: Self Generated data. This table shows the results attained by performing tests on classifiers running on K-Nearest Neighbours algorithm.

Table 6.3-Decision Tree Classifier

ATTRIBUTES	ACCURACY	TESTING TIME	CONFUSION MATRIX	PERFORMANCE
Sepal Length, Sepal Width, Petal Length And Petal Width	92%	0.129402s	23 0 0 0 19 6 0 0 27	3.9724
Sepal Length, Sepal Width And Petal Length	92%	0.007909s	23 0 0 0 19 6 0 0 27	4.4864
Sepal Width, Petal Length And Petal Width	93.3333%	0.008138s	23 0 0 0 20 5 0 0 27	4.4994
Sepal Length, Sepal Width And Petal Width	88%	0.007875s	23 0 0 0 17 8 0 1 26	4.4422
Sepal Length, Petal Length And Petal Width	92%	0.009617s	23 0 0 0 19 6 0 0 27	4.4785
Sepal Width And Sepal Length	69.3333%	0.0064s	22 1 0 2 16 7 0 13 14	4.4212
Petal Length And Sepal Length	92%	0.006715s	23 0 0 0 19 6 0 0 27	4.4915
Petal Width And Sepal Length	94.6667%	0.004689s	23 0 0 0 23 2 0 2 25	4.5291
Petal Width And Sepal Length	93.3333%	0.007387s	23 0 0 0 20 5 0 0 27	4.5028
Petal Width And Sepal Width	89.3333%	0.006183s	23 0 0 0 18 7 0 1 26	4.4647
Petal Width And Petal Length	93.3333%	0.005095s	23 0 0 0 20 5 0 0 27	4.5131
AVERAGE	89.93%	0.023582s	23 0 0 0 19 7 0 1 25	4.3942

*Note: Self Generated data. This table shows the results attained by performing tests on classifiers running on Decision Tree Classification**.**

Clustering Algorithms:

The following tables show the result produced by each cluster model. Each model is grouped according to its algorithm. The 'PARAMETERS' column states the attributes on which the cluster model has operated on. The 'AVERAGE' tuple shows the average performance of each classifier for a specific algorithm. The 'PURITY' and 'SILHOUETTE VALUE' shows the proportion of data points clustered together which belong to the same class in the test data and density of points within a cluster.

The complete Iris dataset has been used for testing both these algorithms.

Table 6.4-Agglomerative Hierarchical Clustering

PARAMETERS	PURITY	TESTING TIME	SILHOUETTE VALUE	PERFORMANCE
Sepal Length, Sepal Width, Petal Length And Petal Width	68%	0.009507s	0.6184	0.4165
Sepal Length, Sepal Width And Petal Length	68%	0.010582s	0.6433	0.4328
Sepal Width, Petal Length And Petal Width	68%	0.007566s	0.5985	0.4039
Sepal Length, Sepal Width And Petal Width	66.67%	0.004728s	0.5999	0.3981
Sepal Length, Petal Length And Petal Width	69.33%	0.005387s	0.5934	0.4092
Sepal Width And Sepal Length	35.33%	0.005569s	0.1028	0.0361
Petal Length And Sepal Length	67.33%	0.005797s	0.4068	0.2723
Petal Width And Sepal Length	34.67%	0.00942s	-0.3799	-0.1305
Petal Width And Sepal Length	66.67%	0.008267s	0.8019	0.5302
Petal Width And Sepal Width	66.67%	0.02518s	0.6777	0.4406
Petal Width And Petal Length	67.33%	0.008337s	0.6221	0.4154
AVERAGE VALUE	61.63%	0.009122s	0.4804	0.3295

*Note: Self Generated data. This table shows the results attained by performing tests on clustering models running on Agglomerative hierarchical clustering.

All the testings were performed on 1.3 GHz Intel Core i5 processor.

The measuring distance used for k-means and agglomerative hierarchical clustering is Euclidean distance.The number of clusters were preset in both models to 3 so that a comparison can be made between the result and the known data.

For k-means clustering the process has been replicated 4 times, so the most accurate clustering result has been chosen of all the iterations. The testing time given in the table is an average of all the 4 iterations. These iterations have been performed to negate the effect of a starting point in k-means algorithm. (For detailed information on results of clustering model, and comparison and analysis refer Appendix I).

Table 6.5-K-Means Clustering

ATTRIBUTES	PURITY	TESTING TIME	SILHOUETTE VALUE	PERFORMANCE
Sepal Length, Sepal Width, Petal Length And Petal Width	89.33%	0.0097835s	0.7357	0.6508
Sepal Length, Sepal Width And Petal Length	88%	0.0075335s	0.7330	0.6402
Sepal Width, Petal Length And Petal Width	95.33%	0.00829s	0.7659	0.7241
Sepal Length, Sepal Width And Petal Width	82.67%	0.0090135s	0.6654	0.5452
Sepal Length, Petal Length And Petal Width	89.33%	0.00589225s	0.7523	0.6681
Sepal Width And Sepal Length	82%	0.00891975s	0.6201	0.504
Petal Length And Sepal Length	88%	0.00694425s	0.7560	0.6607
Petal Width And Sepal Length	82.67%	0.0070455s	0.6813	0.5593
Petal Width And Sepal Length	92.67%	0.0097835s	0.7612	0.6985
Petal Width And Sepal Width	92.67%	0.0075335s	0.7067	0.65
Petal Width And Petal Length	96%	0.00829s	0.8055	0.7669
AVERAGE VALUE	88.97%	0.008093s	0.7257	0.6425

*Note: Self Generated data. This table shows the results attained by performing tests on clustering models running on Agglomerative hierarchical clustering.

CHAPTER 07

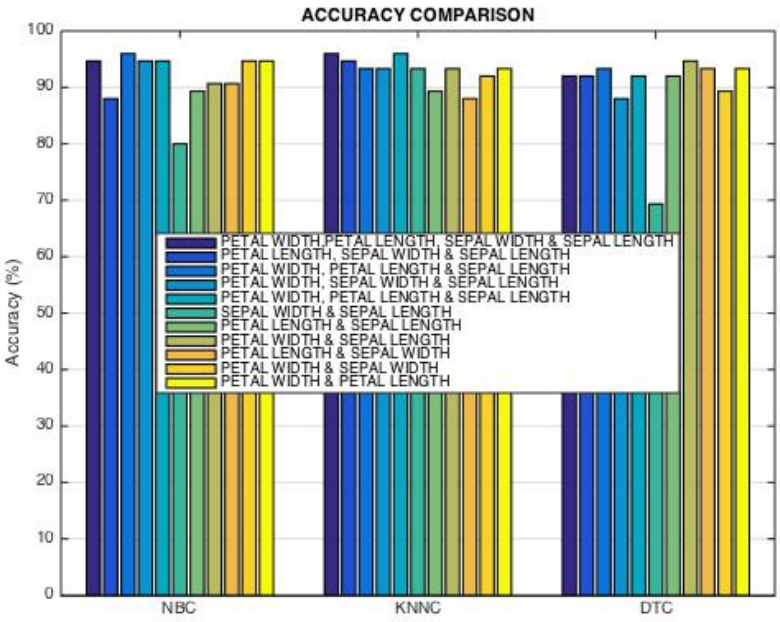
CONCLUSION AND FUTURE ENHANCEMENT

Classification Algorithms:

The performance of the algorithms can be compared by normalising both the accuracy and time and then calculating a ratio.The performance ratio can be calculated by finding a ratio between normalised accuracy and normalised time. $Performance = ln(a)/e^t$, where a=Accuracy and t=testing time. Decision Tree Classifier has an average performance ratio of 4.394, followed by naive Baye's 4.381 and then the K-Nearest Neighbours algorithm which has a ratio of 4.349. Hence, for general application we see that decision tree classification is the best classification algorithm. Although, for mid-sized datasets where accuracy has higher priority than testing time, naive Baye's algorithm is a preferred option.

The figure 7.1 shows the time and accuracy bar graphs of different classifiers.

Figure 7.1-Accuracy Comparison



*NOTE: All figures are self generated.

Figure 7.2(a)-Time Comparison

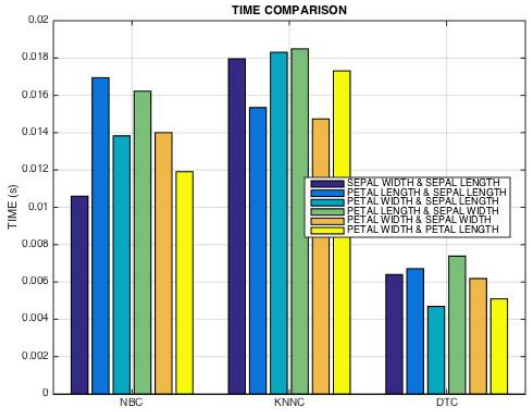
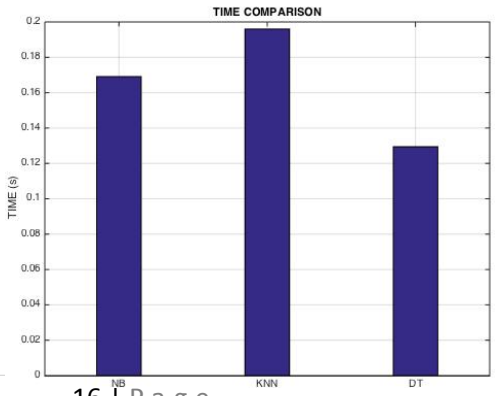


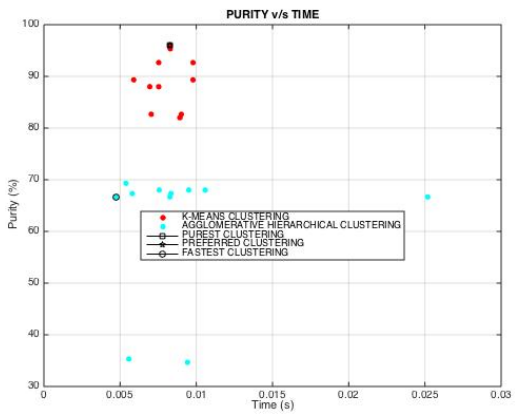
Figure 7.2(b)-Time Comparison



Clustering Algorithms:

The performance of clustering algorithms is measured by all 3 parameters involved in its measure. The performance can be defined as: $Performance = s * (p/e^t)$, where s=silhouette value, p=purity and t is the testing time of the graph. This measure would enable us to see the quality of the result our model produces. Using these values we see that k-means clustering algorithm is more optimal than hierarchical clustering and has an average performance metric of '0.6425' whereas hierarchical routing has a performance metric of '0.3925'. As shown by the graph 7.3 below, the k-means algorithm outclasses hierarchical clustering and has an average purity of 88.97% as compared to 61.63% of hierarchical routing. K-means clustering also has faster testing time of 0.008093 than 0.009122 of agglomerate hierarchical clustering.

Figure 7.3(a)(on the left)-Purity v/s Time and Figure 7.3(b)(on the right)-Purity v/s Silhouette Value



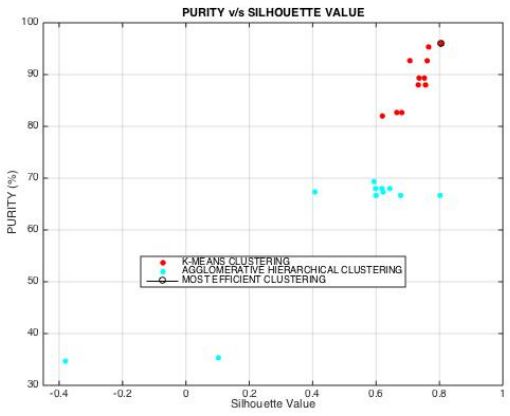
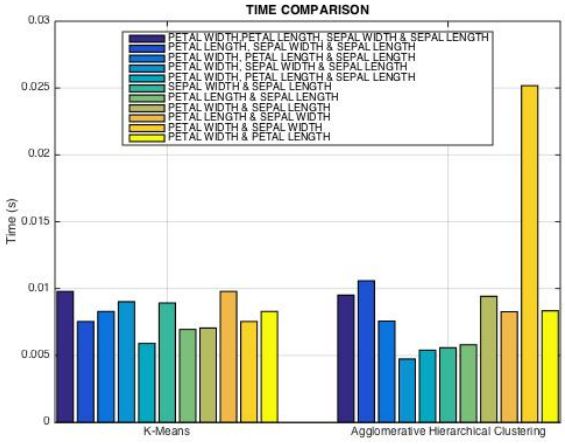
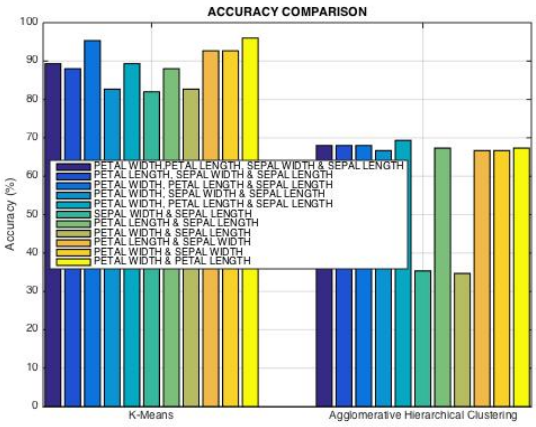


Figure 7.4(a)(on the left)-Time Comparison and Figure 7.4(b)(on the right) -Accuracy Comparison



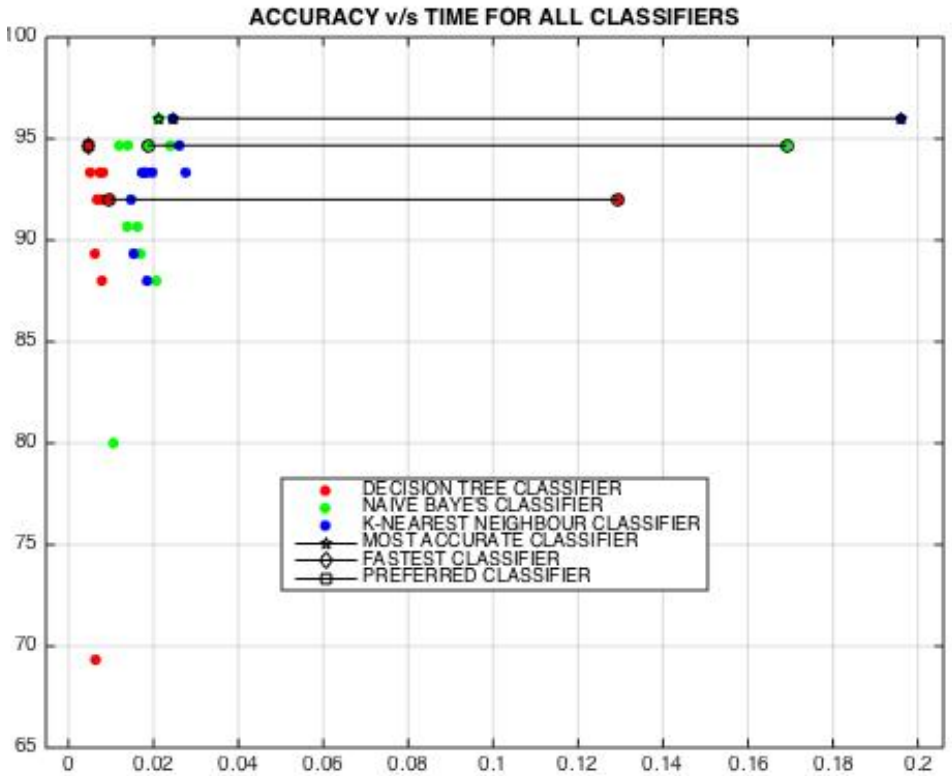


*Note: All figures are self generated.

FUTURE ENHANCEMENT:

Figure 7.5 shows 2 classifiers, which are joined by a single line produce the same accuracy, but with different computational times. During experiment it has been noticed that the classifier classifying on the basis of all 4 attributes, and the classifier classifying on the basis of sepal length, petal length and petal width produce similar results for all 3 classifiers. Hence, we can observe that classification based on a combination of attributes in the entire dataset, can be used to model the classification done on an entire attribute set. Hence we can optimise performance of classification, by using this specific combination of attributes. We can thus reduce the dimensionality of a dataset to be classified, and greatly increase speed. In the Iris dataset we see that the classification based on sepal length, petal length and petal width for Naive Baye's algorithm runs 9.085 times faster, for K-Nearest Neighbours algorithm runs 8.32% faster and for Decision Tree runs 13.456 times faster. Hence, an enhancement to this project would be the optimisation of performance of classification algorithms by reducing the dimensionality of data.

Figure 7.5-Accuracy v/s Time



*NOTE: All figures are self generated.

Looking at the below table 7.1 showing that the performance score of both clustering and classification models based on 3 attributes outperforms that based on 4 attributes, we can hypothesise that a large dataset can be modelled by smaller representation of the same dataset. We can also see that despite lower accuracy in modelling with 2 attributes, they have a high performance score as they compute results significantly faster. Because we were functioning on a dataset with only 4 dimensions, the difference is not notable, but with detests regarding a higher number of dimensions, and especially categorical dimensions, this reduction in dimensionality will result in a significant difference in time between the reduced and the complete dataset.

In reference to the table and the calculated performance parameters the decision tree classifiers working on 2 attributes has the most optimum classifiers. It has an average accuracy of 87% and an average computational time of 0.006078 seconds. The best classifier is the decision tree classifier trained on 'Petal Width And Sepal Length'.

The k-means clustering algorithm, significantly outperforms the agglomerative hierarchical algorithm. The best performer was the k-means cluster model on all 4 attributes with an accuracy of 89.33% and testing time of 0.0097835.

(For detailed graphical analysis of both clustering and classification models, refer Appendix I).

Table 7.1-Performance Comparison

ALGORITHM	4 ATTRIBUTES	3 ATTRIBUTES	2 ATTRIBUTES
NAIVE BAYE'S	3.8424	4.4407	4.4284
K-NEAREST NEIGHBOURS	3.7519	4.4368	4.4392
DECISION TREE	3.9724	4.4766	4.4871
K-MEANS	0.6508	0.6444	0.6399
AGGLOMERATIVE HIERARCHICAL	0.4165	0.4111	0.22974

*Note: Self Generated Data. This table shows a performance comparison analysis with classifiers trained on different number of attributes.

Bibliography

1. Mohammed J. Zaki and Wagner Meira, 'DATA MINING AND ANALYSIS '.
2. Bunheang Tay, Jung Keun Hyun and Sejong Oh, 'A Machine Learning Approach for Specification of Spinal Cord Injuries Using Fractional Anisotropy Values Obtained from Diffusion Tensor Images'.
3. S. Rasoul Safavian and David Landgrebe, 'A Survey Of Decision Tree Methodology'.
4. Yuguang Huang and Lei Li, 'Naive Bayes classification algorithm based on small sample set'.
5. T. Kanungo , D.M. Mount , N.S. Netanyahu C.D. Piatko , R. Silverman and A.Y. Wu , 'An efficient k-means clustering algorithm: analysis and implementation'.
6. Ying Zhao and George Karypis, 'Evaluation of hierarchical clustering algorithms for document datasets'.
7. Min-Ling Zhang and Zhi-Hua Zhou, 'A k-nearest neighbor based algorithm for multi-label classification'.
8. Ahmad Ashari, Iman Paryudi and A Min Tjoa, 'Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool'.
9. Michael Steinbach, George Karypis and Vipin Kumar, 'A Comparison of Document Clustering Techniques'.
10. Rich Caruana and Alexandru Niculescu-Mizil, 'An Empirical Comparison of Supervised Learning Algorithms'.

Appendix I

Datasets

The training dataset used, is fundamental to the accuracy of the classifier. Training dataset, should be chosen such that that it should perfectly mimic the testing dataset. In the current report, the Iris dataset was divided into 2 equal halves by a random self-designed algorithm. The algorithm partitions the dataset into 2 equal halves and the class-wise bifurcation of the training data is as follows: 'Setosa' -27, 'Versicolor'- 25 and 'Virginica' is 23. Given below is both the training and testing dataset used.

For clustering method, the entire Iris dataset was used at once for operation.

*Note: All figures and tables under this section are self generated.

Table I.1-TRAINING DATASET

Index Number	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa'
3	4.7	3.2	1.3	0.2	setosa'
4	5.4	3.9	1.7	0.4	setosa'
5	4.4	2.9	1.4	0.2	setosa'
6	4.9	3.1	1.5	0.1	setosa'
7	4.3	3	1.1	0.1	setosa'
8	5.7	4.4	1.5	0.4	setosa'
9	5.4	3.9	1.3	0.4	setosa'
10	4.6	3.6	1	0.2	setosa'
11	5.1	3.3	1.7	0.5	setosa'
12	4.7	3.2	1.6	0.2	setosa'
13	4.8	3.1	1.6	0.2	setosa'
14	4.9	3.1	1.5	0.2	setosa'
15	5	3.2	1.2	0.2	setosa'
16	5.5	3.5	1.3	0.2	setosa'
17	4.4	3	1.3	0.2	setosa'

Index Number	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
18	5.1	3.4	1.5	0.2	setosa'
19	4.5	2.3	1.3	0.3	setosa'
20	5	3.5	1.6	0.6	setosa'
21	5.1	3.8	1.9	0.4	setosa'
22	4.6	3.2	1.4	0.2	setosa'
23	5.3	3.7	1.5	0.2	setosa'
24	7	3.2	4.7	1.4	versicolor'
25	6.9	3.1	4.9	1.5	versicolor'
26	5.5	2.3	4	1.3	versicolor'
27	6.5	2.8	4.6	1.5	versicolor'
28	4.9	2.4	3.3	1	versicolor'
29	5.2	2.7	3.9	1.4	versicolor'
30	5.9	3	4.2	1.5	versicolor'
31	5.6	2.9	3.6	1.3	versicolor'
32	6.7	3.1	4.4	1.4	versicolor'
33	6.2	2.2	4.5	1.5	versicolor'
34	5.9	3.2	4.8	1.8	versicolor'
35	6.3	2.5	4.9	1.5	versicolor'
36	6.4	2.9	4.3	1.3	versicolor'
37	6.8	2.8	4.8	1.4	versicolor'
38	5.7	2.6	3.5	1	versicolor'
39	5.5	2.4	3.8	1.1	versicolor'
40	5.5	2.4	3.7	1	versicolor'
41	5.8	2.7	3.9	1.2	versicolor'
42	6	2.7	5.1	1.6	versicolor'
43	5.4	3	4.5	1.5	versicolor'
44	6	3.4	4.5	1.6	versicolor'
45	6.3	2.3	4.4	1.3	versicolor'

Index Number	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
46	5.5	2.6	4.4	1.2	versicolor'
47	5	2.3	3.3	1	versicolor'
48	5.7	2.9	4.2	1.3	versicolor'
49	6.3	3.3	6	2.5	virginica'
50	5.8	2.7	5.1	1.9	virginica'
51	7.1	3	5.9	2.1	virginica'
52	7.6	3	6.6	2.1	virginica'
53	7.3	2.9	6.3	1.8	virginica'
54	6.7	2.5	5.8	1.8	virginica'
55	6.5	3.2	5.1	2	virginica'
56	6.4	2.7	5.3	1.9	virginica'
57	6.8	3	5.5	2.1	virginica'
58	5.7	2.5	5	2	virginica'
59	6.5	3	5.5	1.8	virginica'
60	7.7	3.8	6.7	2.2	virginica'
61	7.7	2.6	6.9	2.3	virginica'
62	6	2.2	5	1.5	virginica'
63	6.9	3.2	5.7	2.3	virginica'
64	5.6	2.8	4.9	2	virginica'
65	6.3	2.7	4.9	1.8	virginica'
66	6.7	3.3	5.7	2.1	virginica'
67	7.2	3	5.8	1.6	virginica'
68	7.4	2.8	6.1	1.9	virginica'
69	6.3	3.4	5.6	2.4	virginica'
70	6	3	4.8	1.8	virginica'
71	6.9	3.1	5.4	2.1	virginica'
72	6.9	3.1	5.1	2.3	virginica'
73	6.8	3.2	5.9	2.3	virginica'

Index Number	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
74	6.7	3.3	5.7	2.5	virginica'
75	5.9	3	5.1	1.8	virginica'

*Note: All TABLES are self-generated.

Figure I.2-TESTING DATASET

INDEX NUMBER	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
1	4.6	3.1	1.5	0.2	setosa'
2	5	3.6	1.4	0.2	setosa'
3	4.6	3.4	1.4	0.3	setosa'
4	5	3.4	1.5	0.2	setosa'
5	5.4	3.7	1.5	0.2	setosa'
6	4.8	3.4	1.6	0.2	setosa''
7	4.8	3	1.4	0.1	setosa'
8	5.8	4	1.2	0.2	setosa'
9	5.1	3.5	1.4	0.3	setosa'
10	5.7	3.8	1.7	0.3	setosa'
11	5.1	3.8	1.5	0.3	setosa'
12	5.4	3.4	1.7	0.2	setosa'
13	5.1	3.7	1.5	0.4	setosa'
14	4.8	3.4	1.9	0.2	setosa'
15	5	3	1.6	0.2	setosa'
16	5	3.4	1.6	0.4	setosa'
17	5.2	3.5	1.5	0.2	setosa'
18	5.2	3.4	1.4	0.2	setosa'
19	5.4	3.4	1.5	0.4	setosa'
20	5.2	4.1	1.5	0.1	setosa'
21	5.5	4.2	1.4	0.2	setosa'
22	4.9	3.6	1.4	0.1	setosa'

INDEX NUMBER	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
23	5	3.5	1.3	0.3	setosa'
24	4.4	3.2	1.3	0.2	setosa'
25	4.8	3	1.4	0.3	setosa'
26	5.1	3.8	1.6	0.2	setosa'
27	5	3.3	1.4	0.2	setosa'
28	6.4	3.2	4.5	1.5	versicolor'
29	5.7	2.8	4.5	1.3	versicolor'
30	6.3	3.3	4.7	1.6	versicolor'
31	6.6	2.9	4.6	1.3	versicolor'
32	5	2	3.5	1	versicolor'
33	6	2.2	4	1	versicolor'
34	6.1	2.9	4.7	1.4	versicolor'
35	5.6	3	4.5	1.5	versicolor'
36	5.8	2.7	4.1	1	versicolor'
37	5.6	2.5	3.9	1.1	versicolor'
38	6.1	2.8	4	1.3	versicolor'
39	6.1	2.8	4.7	1.2	versicolor'
40	6.6	3	4.4	1.4	versicolor'
41	6.7	3	5	1.7	versicolor'
42	6	2.9	4.5	1.5	versicolor'
43	6.7	3.1	4.7	1.5	versicolor'
44	5.6	3	4.1	1.3	versicolor'
45	5.5	2.5	4	1.3	versicolor'
46	6.1	3	4.6	1.4	versicolor'
47	5.8	2.6	4	1.2	versicolor'
48	5.6	2.7	4.2	1.3	versicolor'
49	5.7	3	4.2	1.2	versicolor'
50	6.2	2.9	4.3	1.3	versicolor'

INDEX NUMBER	SEPAL LENGTH	SEPAL WIDTH	PETAL LENGTH	PETAL WIDTH	CLASS
51	5.1	2.5	3	1.1	versicolor'
52	5.7	2.8	4.1	1.3	versicolor'
53	6.3	2.9	5.6	1.8	virginica'
54	6.5	3	5.8	2.2	virginica'
55	4.9	2.5	4.5	1.7	virginica'
56	7.2	3.6	6.1	2.5	virginica'
57	5.8	2.8	5.1	2.4	virginica'
58	6.4	3.2	5.3	2.3	virginica'
59	7.7	2.8	6.7	2	virginica'
60	7.2	3.2	6	1.8	virginica'
61	6.2	2.8	4.8	1.8	virginica'
62	6.1	3	4.9	1.8	virginica'
63	6.4	2.8	5.6	2.1	virginica'
64	7.9	3.8	6.4	2	virginica'
65	6.4	2.8	5.6	2.2	virginica'
66	6.3	2.8	5.1	1.5	virginica'
67	6.1	2.6	5.6	1.4	virginica'
68	7.7	3	6.1	2.3	virginica'
69	6.4	3.1	5.5	1.8	virginica'
70	6.7	3.1	5.6	2.4	virginica'
71	5.8	2.7	5.1	1.9	virginica'
72	6.7	3	5.2	2.3	virginica'
73	6.3	2.5	5	1.9	virginica'
74	6.5	3	5.2	2	virginica'
75	6.2	3.4	5.4	2.3	virginica'

Figure I.3-Bifurcation of data in classes

*Note: All tables are self generated.

Classification and Clustering Results:

In this section, we will study the prediction results made by all 11 cluster and classifier models of each algorithm. For each classifier the left had figure, shows the data points grouped by their actual class of the datapoint, whereas the right hand side figure shows the points grouped as their predicted class. The circled points are the one's which have been wrongly classified. This plot is used to study the specific strength and weakness of the classifier and spot critical zones, where the classifier generally misclassifies the point. Every 3-D plot has 4 different plots and every 2-D plot has 6 different plots, which cover all combination of axes possible. Despite a 4 dimensional plot is not possible, this approach can give us a complete visualisation of the data between any 2 possible attributes.

For every algorithm, we will first look at the classifier which classifies based on the whole dataset, then we look at the algorithm which classifies based on 3 attributes and then finally the classifier which classifies on only 2 attributes. For classifiers classifying on 4 attributes, both 3-dimensional and 2-dimensional plots have been given.

For decision tree classifiers, the decision tree has also been given, so we can understand the basic functioning of decision tree and how it classifies points.

The number of clusters which can be used have been preset to 3 for both algorithms. Both algorithms will generate a silhouette curve and will be assed on purity, silhouette value and testing time as seen in the report. Here the graph of clustering, the specific regions, and centroids achieved while clustering will be studied in detail.

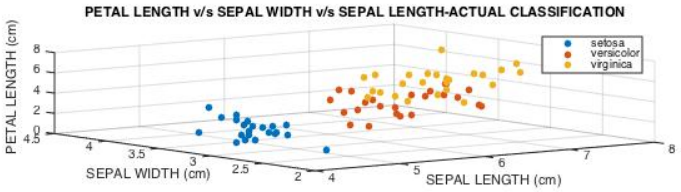
Here because of high time complexity of hierarchical routing we cannot divide the plot into cluster like that in clustering. Additionally, dendrograms are given for all hierarchical clustering models, to visualise how hierarchical routing takes place. Along with each figure, its purity, silhouette value and clustering time is mentioned.

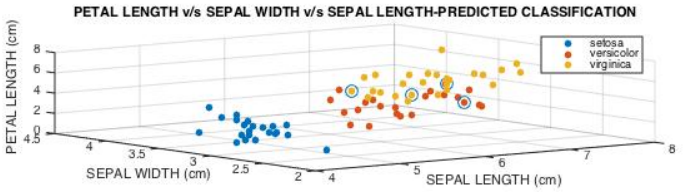
*Note: All figures and tables under this section are self generated.

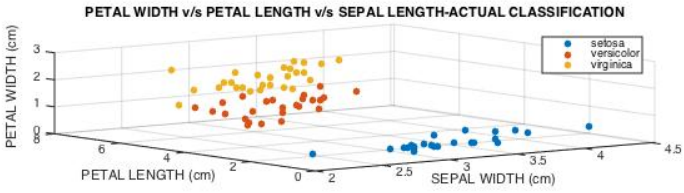
Naive Baye's Classifiers:

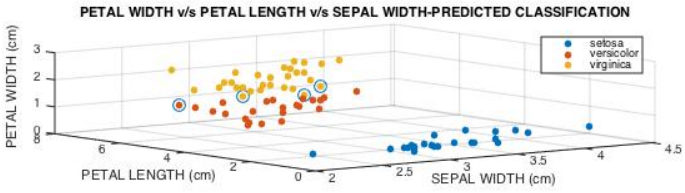
SEPAL LENGTH v/s SEPAL WIDTH v/s PETAL LENGTH v/s PETAL WIDTH:

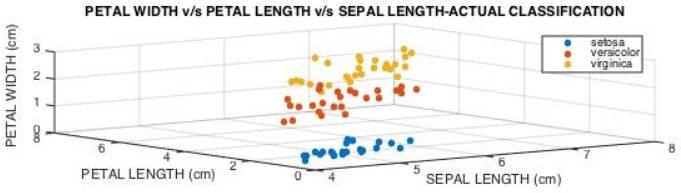
Figure I.1-Accuracy 94.67%, Testing Time 0.1691 s

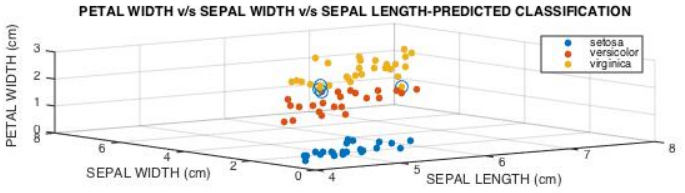


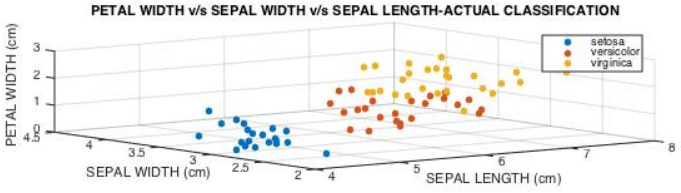


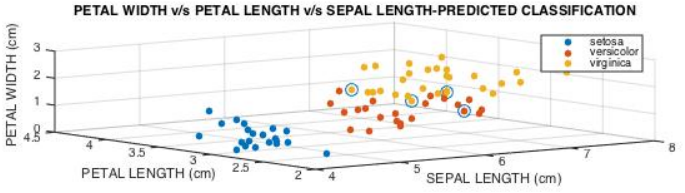


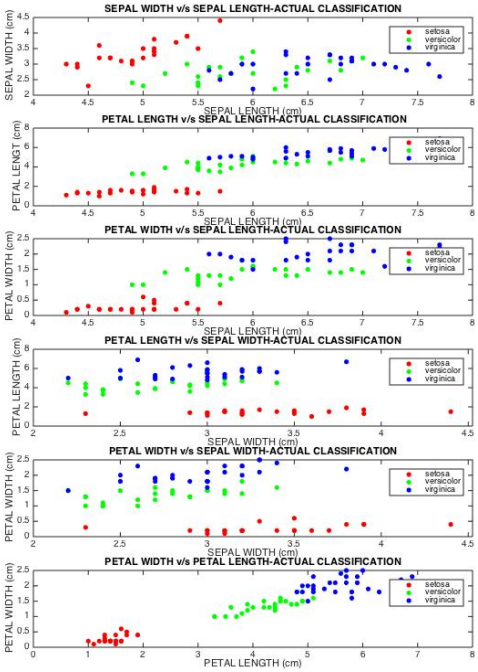


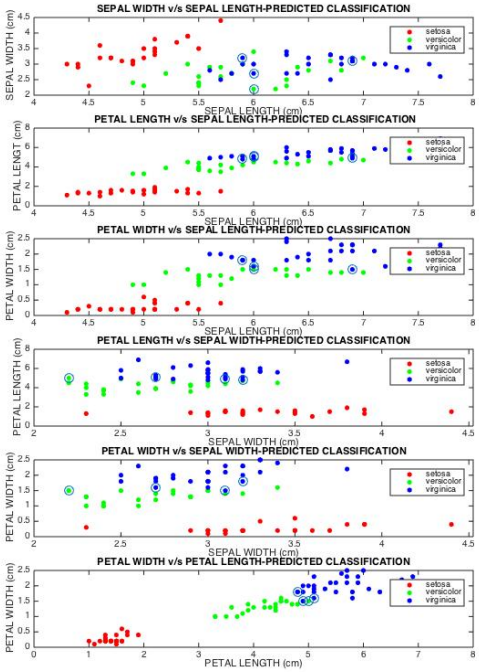






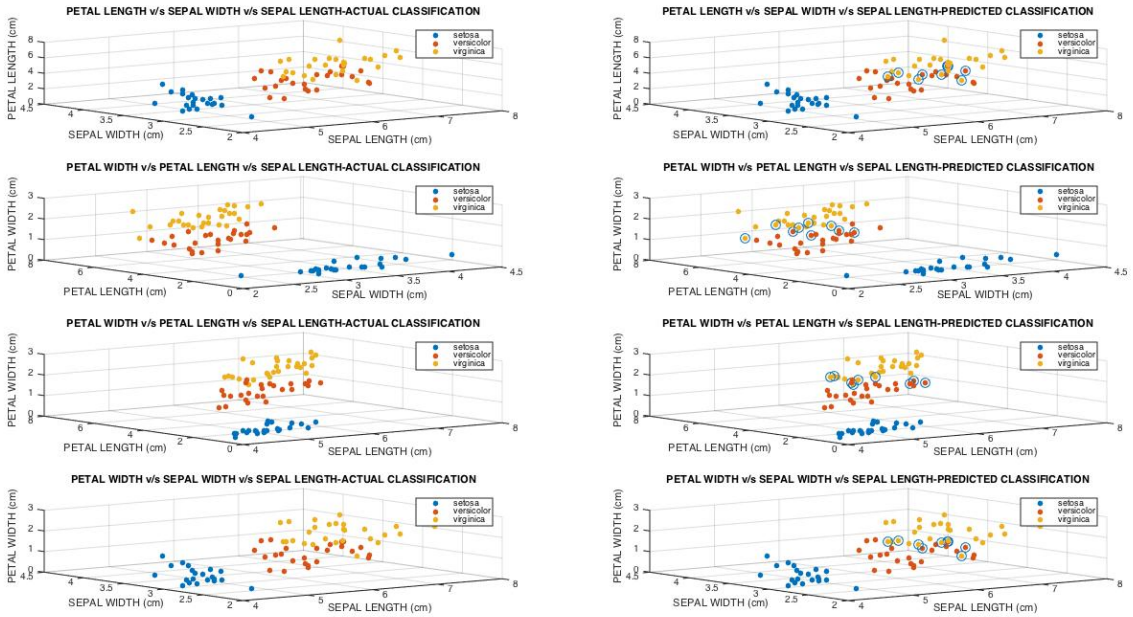






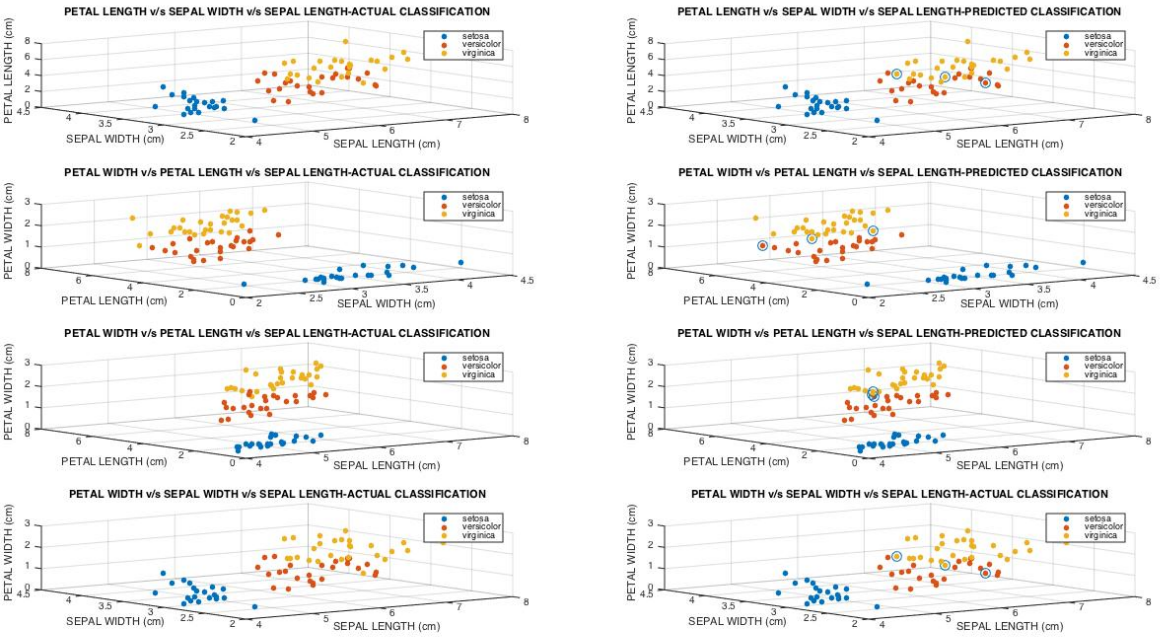
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL LENGTH:

Figure I.2-Accuracy 88%, Time 0.020681s



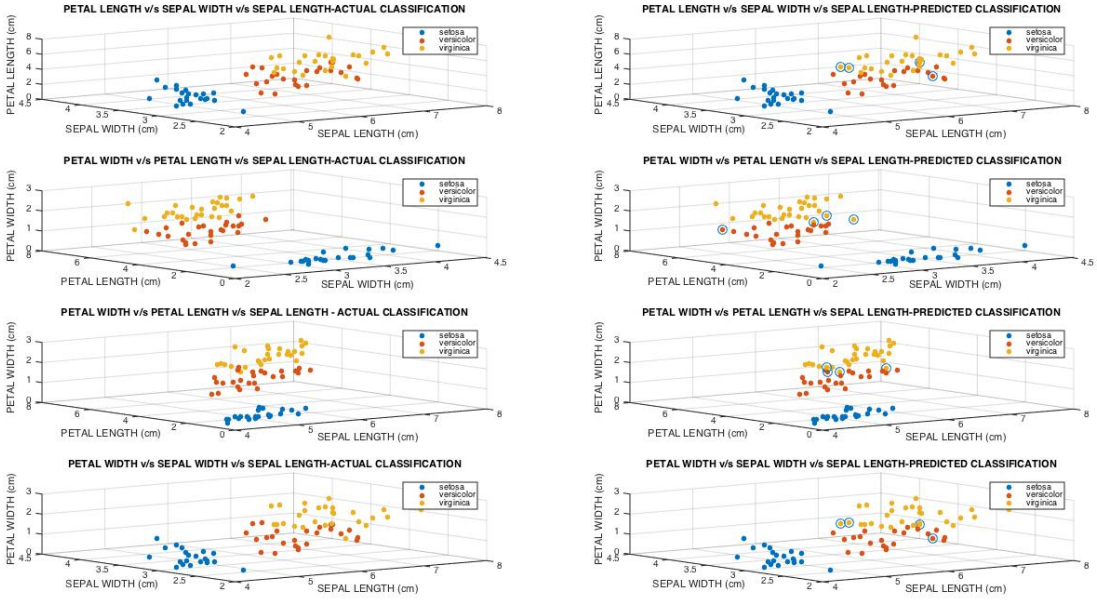
SEPAL WIDTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.3-Accuracy 96%, Time-0.021303s



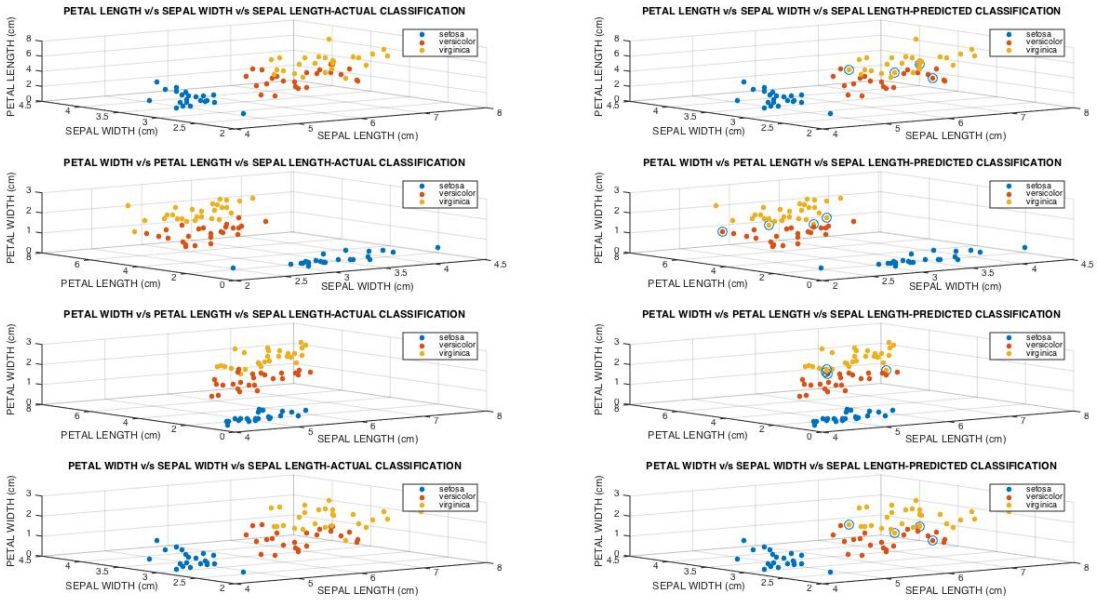
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL WIDTH:

Figure I.4-Accuracy 94.67%, Time 0.023927s



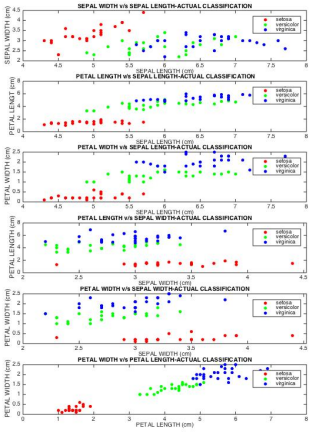
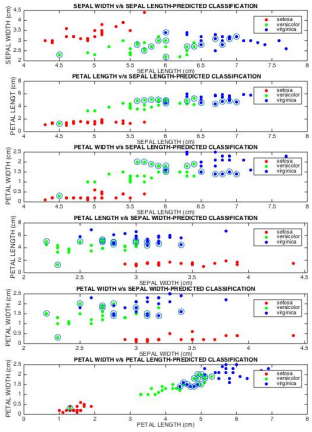
SEPAL LENGTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.5-Accuracy 94.67%, Time 0.018614s



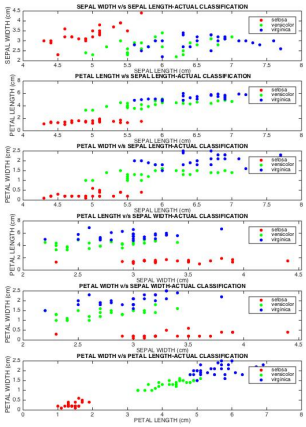
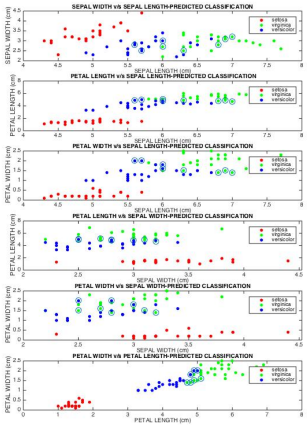
SEPAL LENGTH V/S SEPAL WIDTH:

Figure I.6-Accuracy 80%, Time 0.010596s

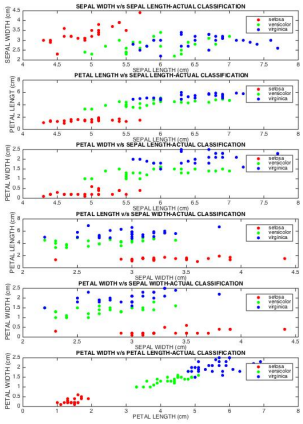
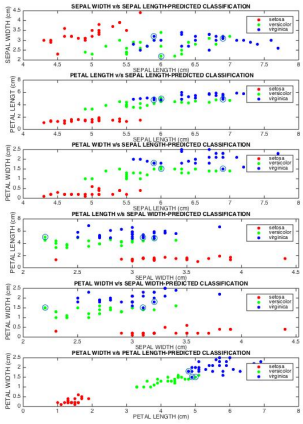
SEPAL LENGTH V/S PETAL LENGTH:

Figure I.7-Accuracy 89.33%, Time 0.016948s

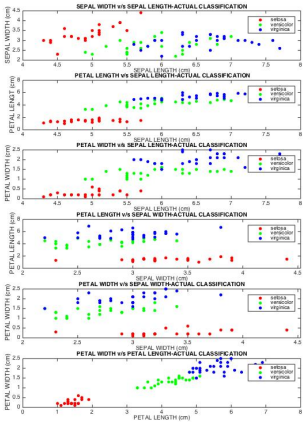
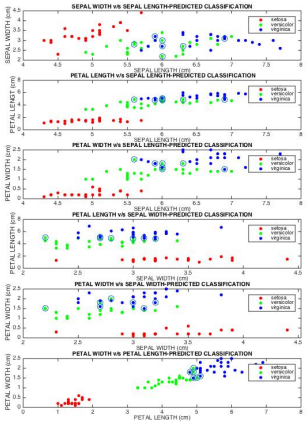
PETAL WIDTH V/S SEPAL LENGTH:

Figure I.8-Accuracy 90.67%, Time 0.013831s

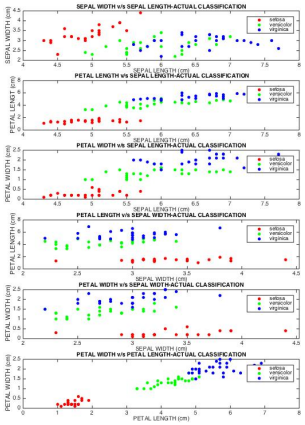
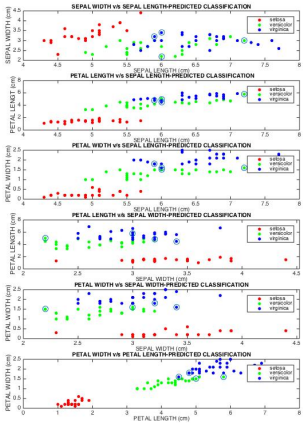
PETAL WIDTH V/S SEPAL LENGTH

Figure I.9-Accuracy 90.67%, Time 0.016223s

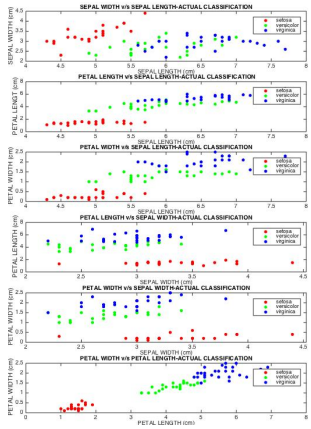
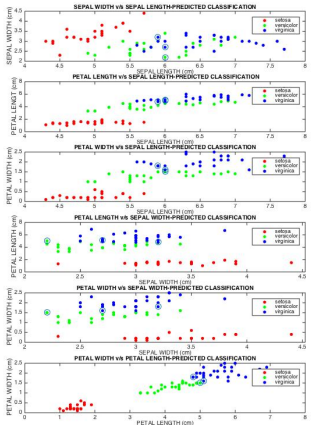
SEPAL WIDTH V/S PETAL WIDTH:

Figure I.10-Accuracy 94.67%, Time 0.013999s

PETAL LENGTH V/S PETAL WIDTH:

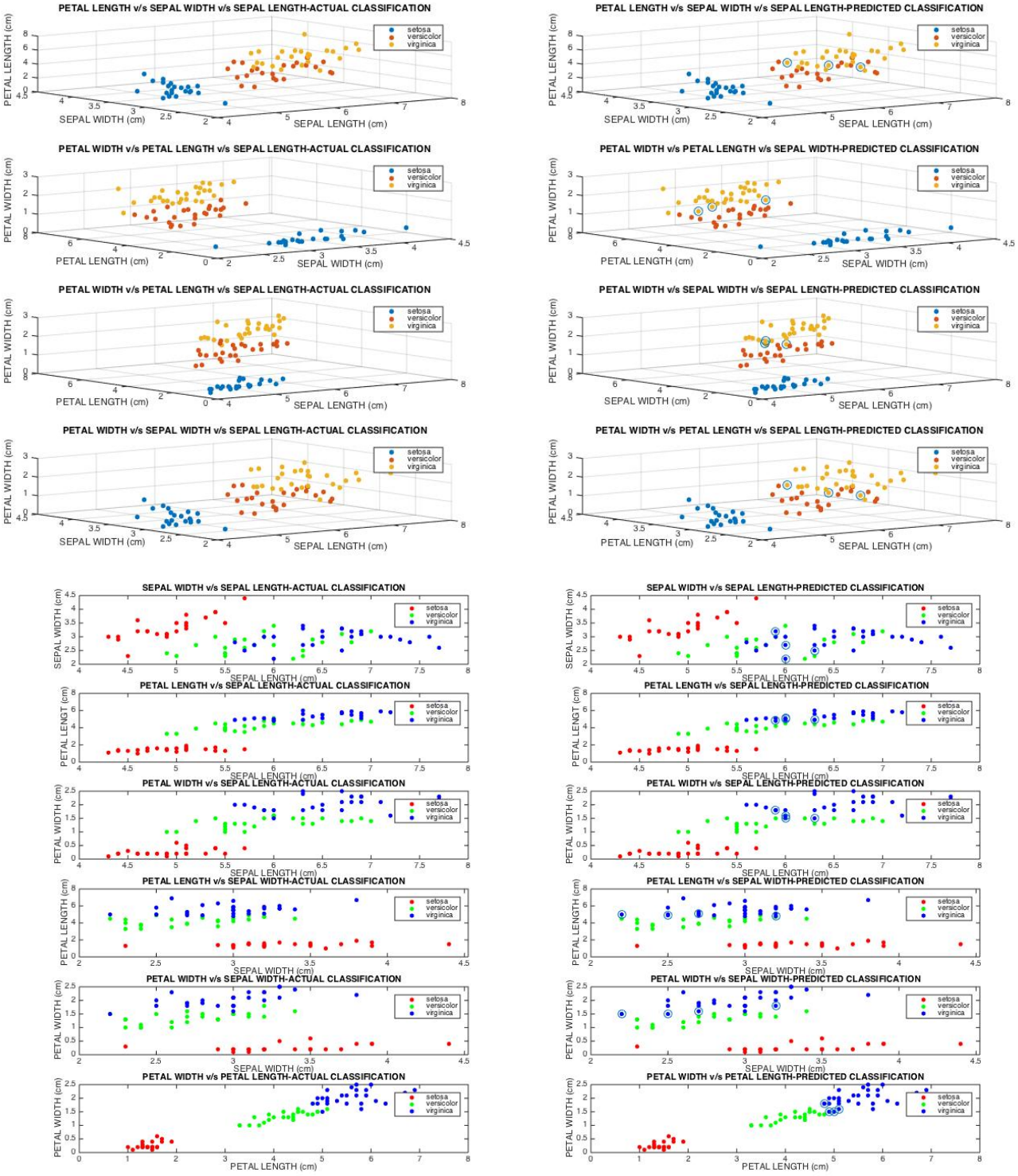
Figure I.11-Accuracy 94.67%, Time 0.01191s

K-Nearest Neighbours Classifiers:

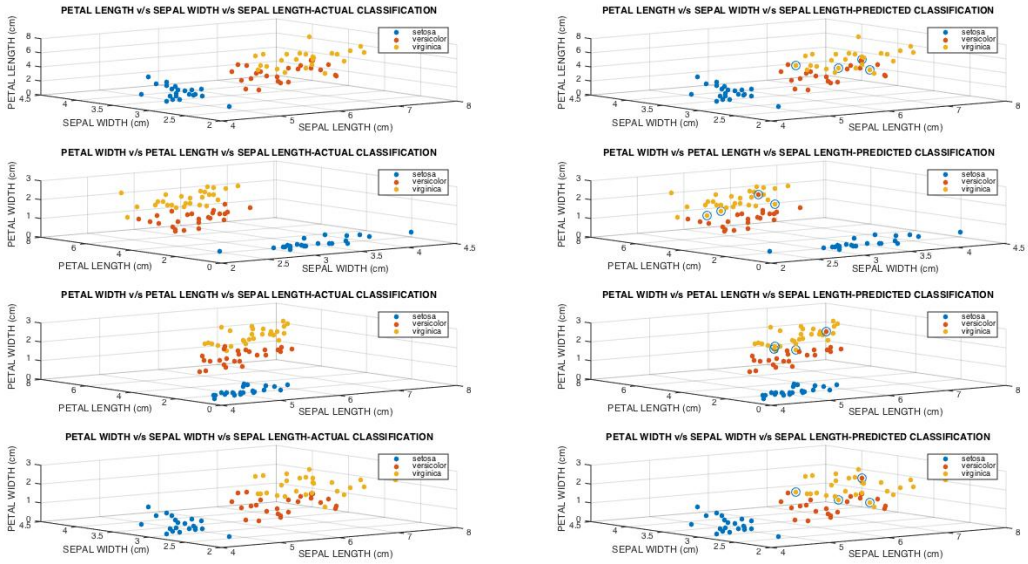
SEPAL LENGTH v/s SEPAL WIDTH v/s PETAL LENGTH v/s PETAL WIDTH:

Figure I.12-Accuracy 96%, Time 0.19601s



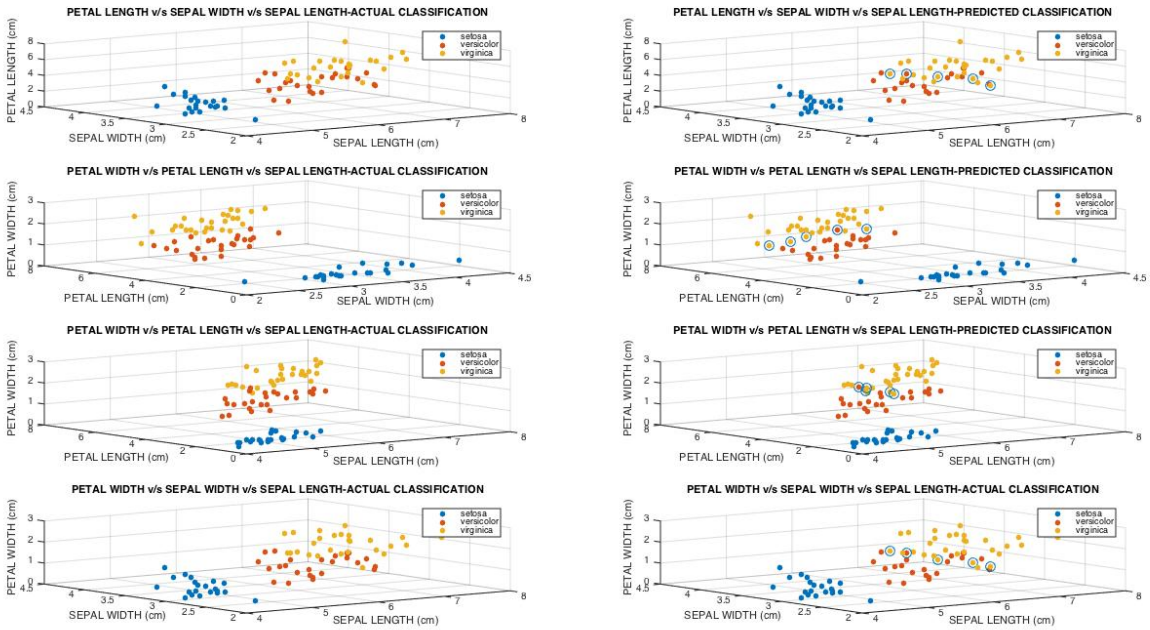
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL LENGTH:

Figure I.13-Accuracy 94.67%, Time 0.026067s



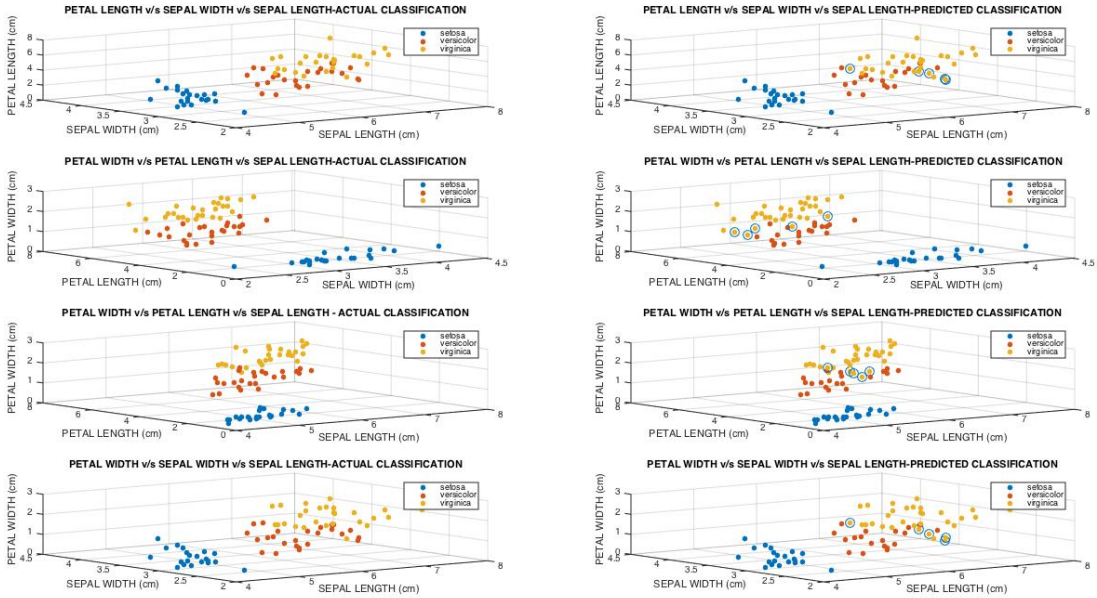
SEPAL WIDTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.14-Accuracy 93.33%, Time 0.027566s



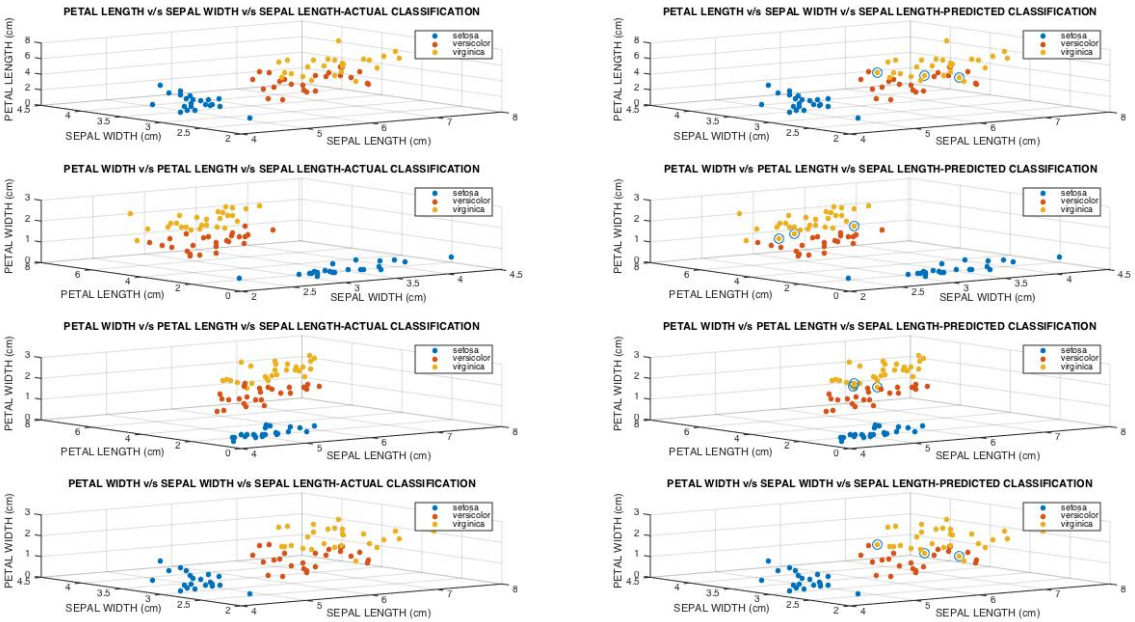
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL WIDTH:

Figure I.15-Accuracy 93.33%, Time 0.019711s



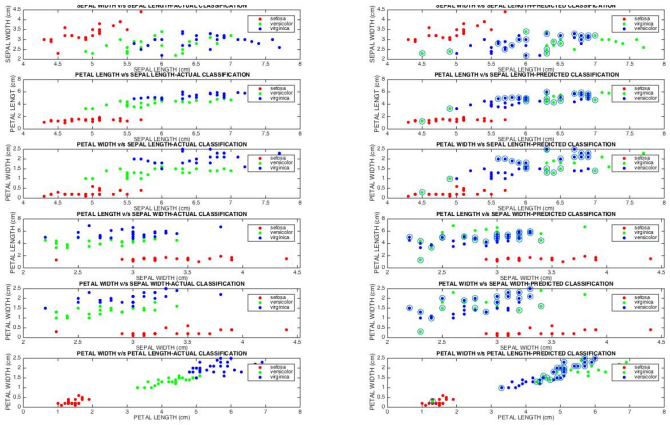
SEPAL LENGTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.16-Accuracy 96%, Time 0.024565s



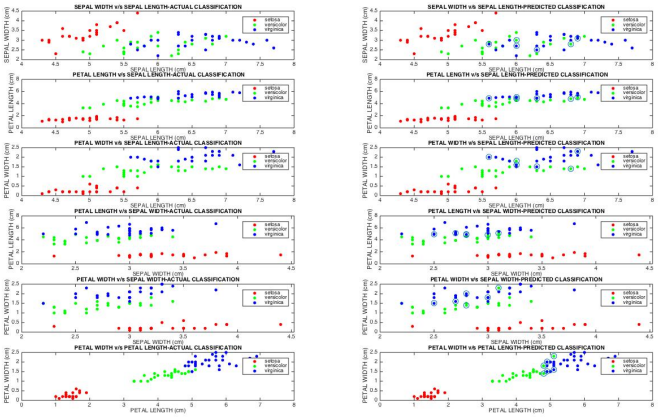
SEPAL LENGTH V/S SEPAL WIDTH:

Figure I.17-Accuracy 69.67%, Time 0.017959s



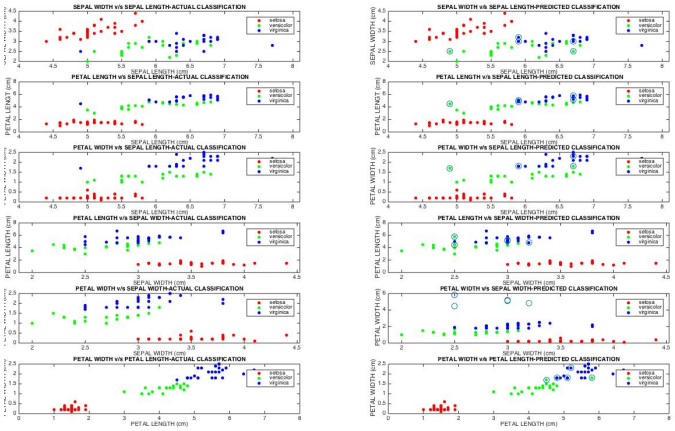
SEPAL LENGTH V/S PETAL LENGTH:

Figure I.18-Accuracy 89.33%, Time 0.015347s



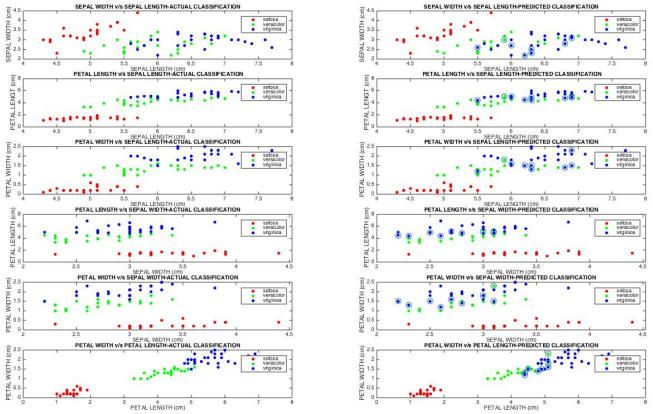
SEPAL LENGTH V/S PETAL WIDTH:

Figure I.19-Accuracy 93.33%, Time 0.018308s



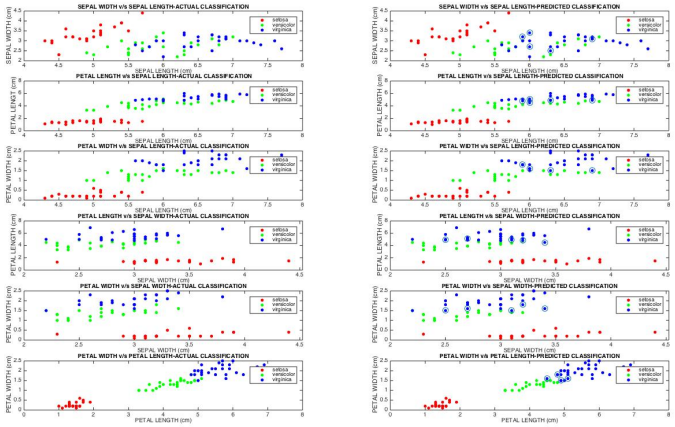
SEPAL WIDTH V/S PETAL LENGTH:

Figure I.20-Accuracy 88%, Time 0.018493s



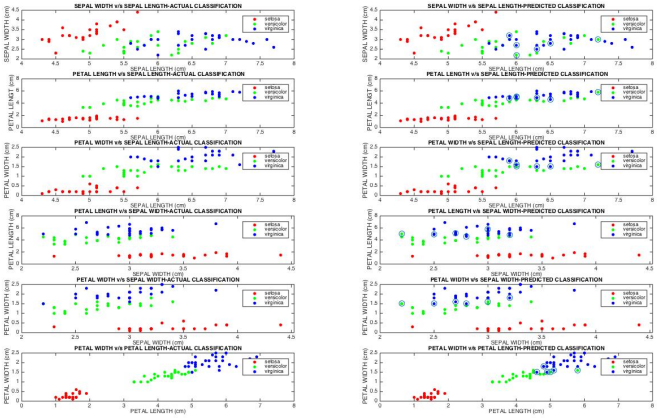
SEPAL WIDTH V/S PETAL WIDTH:

Figure I.21-Accuracy 92%, Time 0.014737s



PETAL LENGTH V/S PETAL WIDTH:

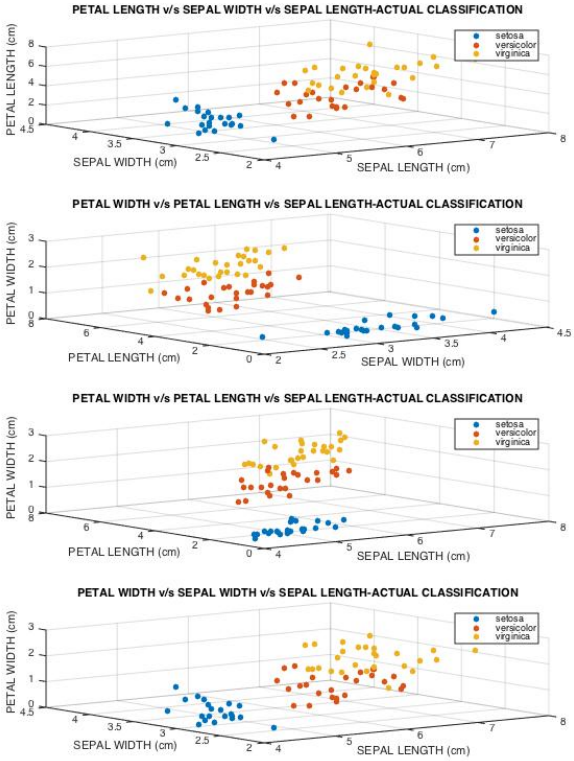
Figure I.22-Accuracy 93.33%, Time 0.01731s

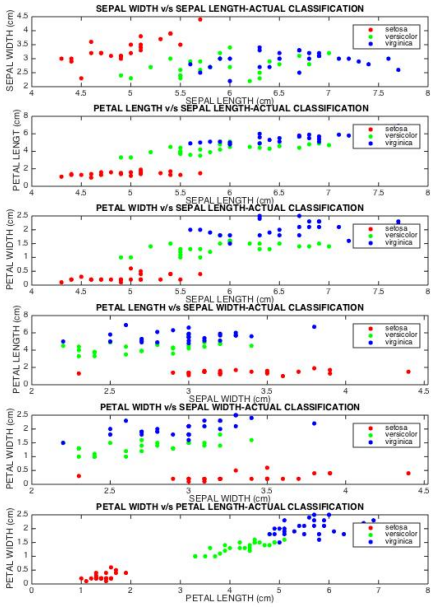


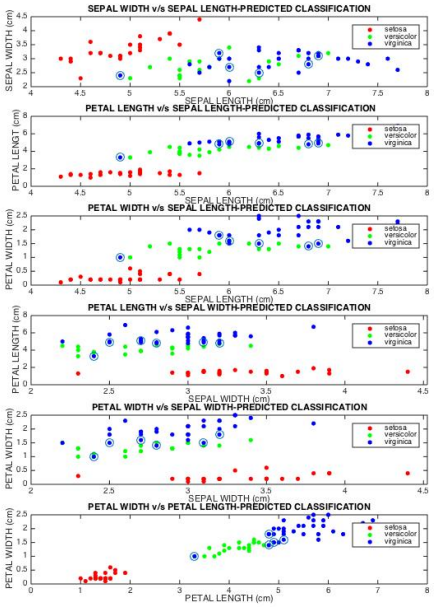
Decision Tree Classifier:

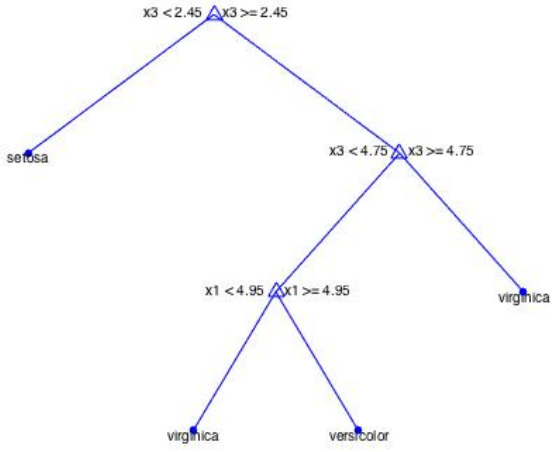
SEPAL LENGTH v/s SEPAL WIDTH v/s PETAL LENGTH v/s PETAL WIDTH:

Figure I.23-Accuracy 92%, Time 0.129402s



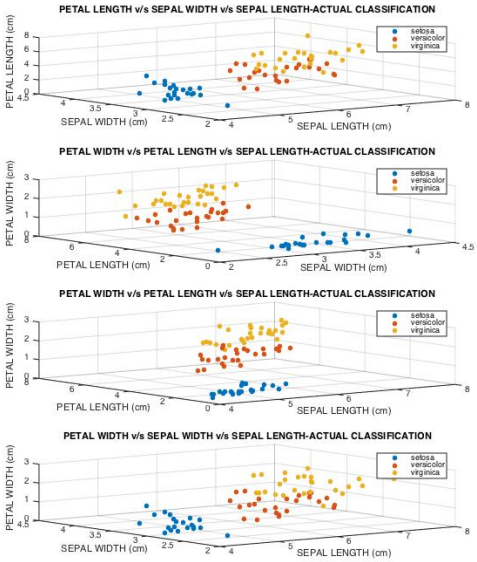


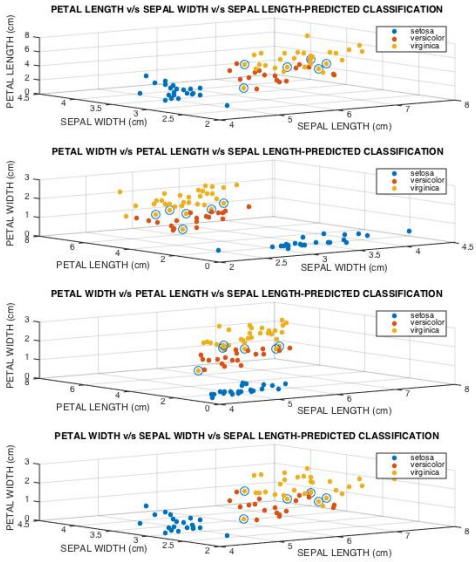


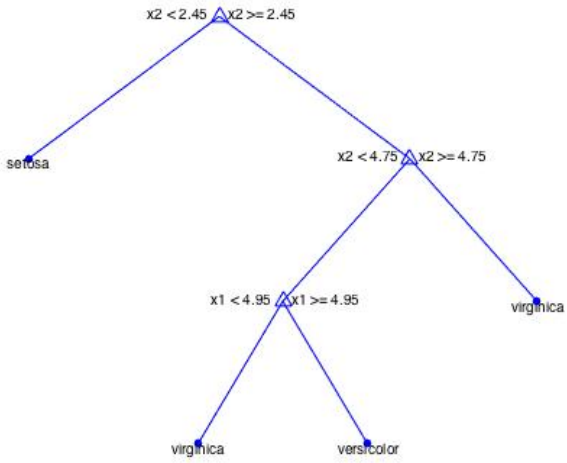


SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL LENGTH:

Figure I.24-Accuracy 92%, Time 0.007909s

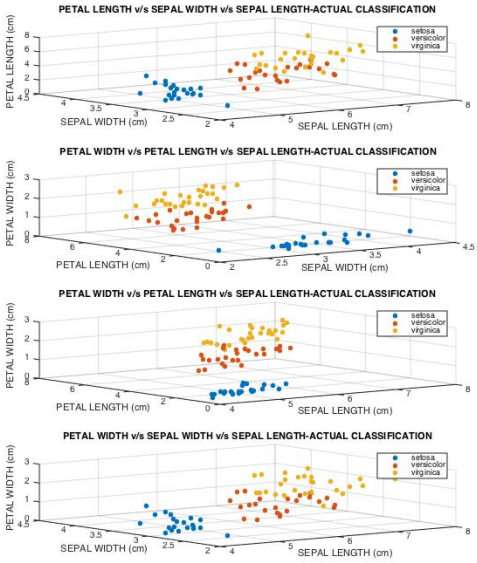


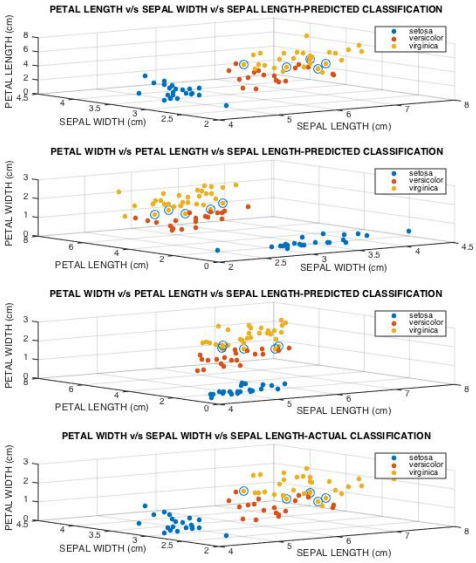


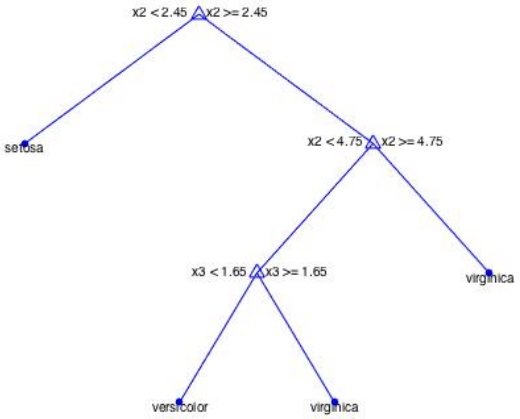


SEPAL WIDTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.25-Accuracy 93.33%, Time 0.008138s

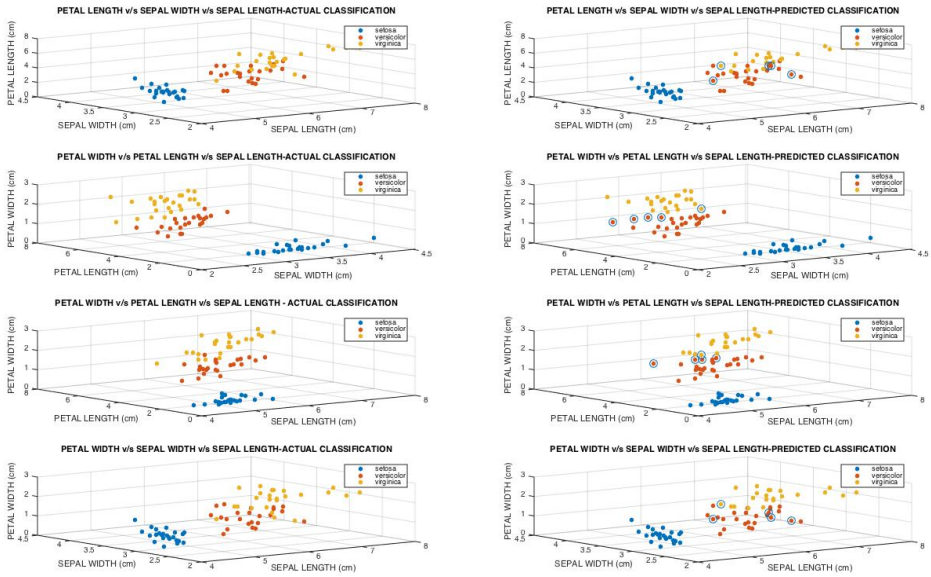


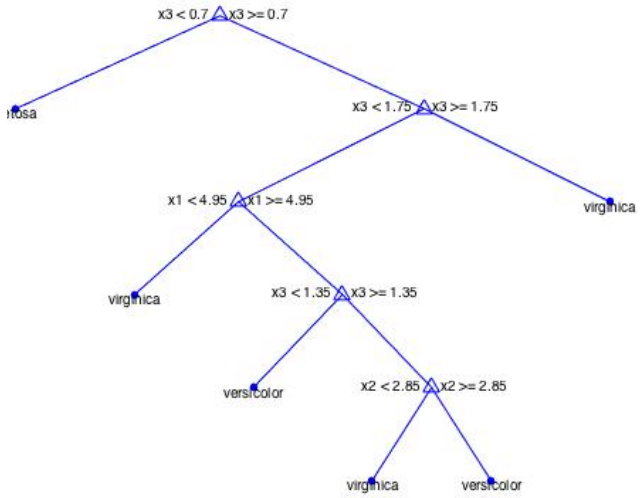




SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL WIDTH:

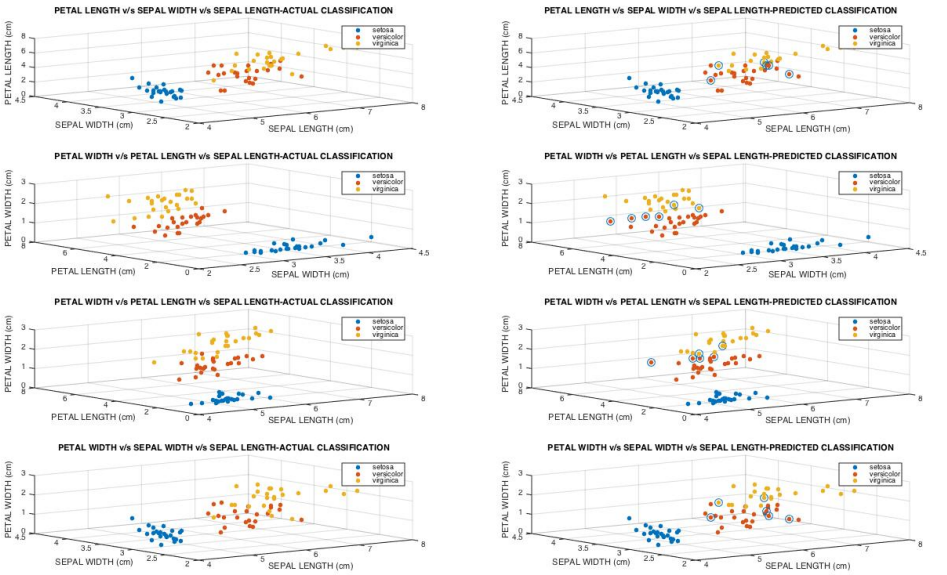
Figure I.26-Accuracy 88%, Time 0.007875s

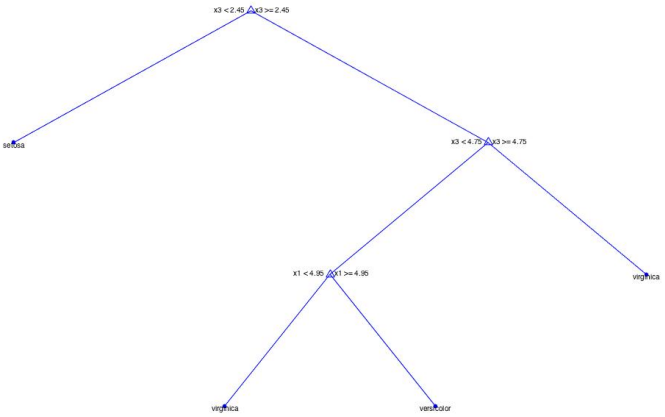




SEPAL LENGTH V/S PETAL LENGTH V/S PETAL WIDTH:

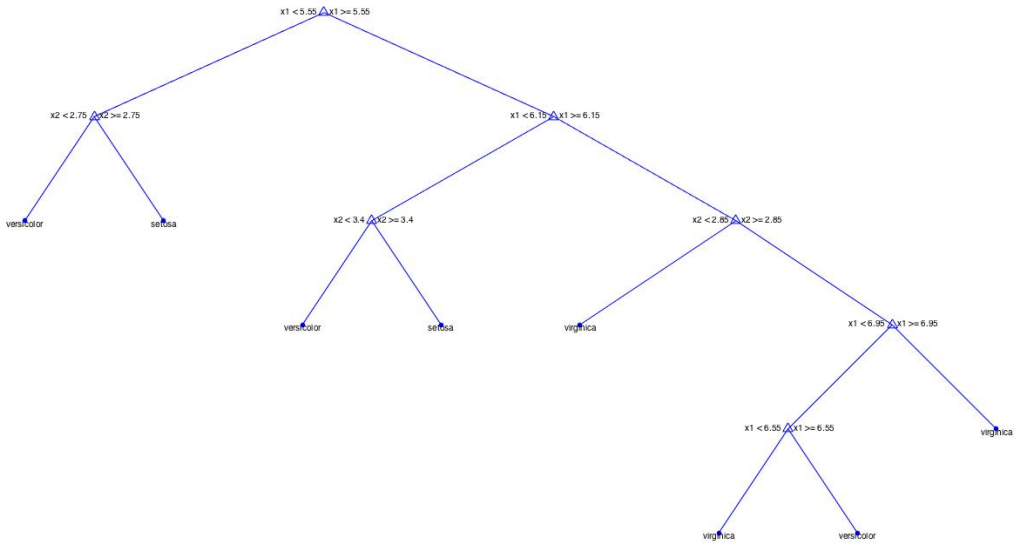
Figure I.27-Accuracy 92%, Time 0.009617s

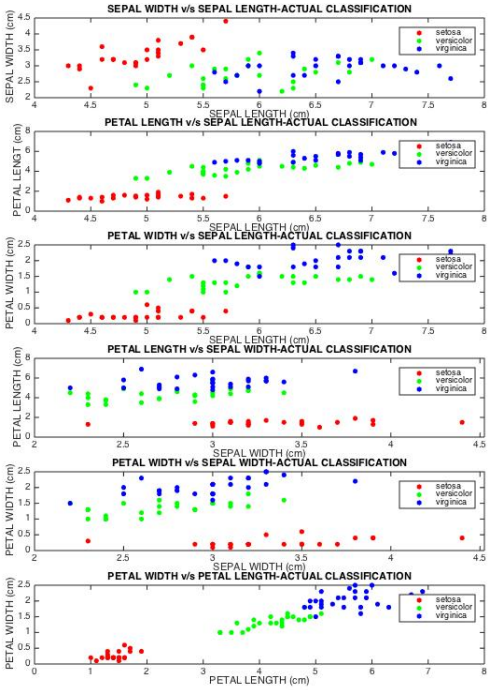


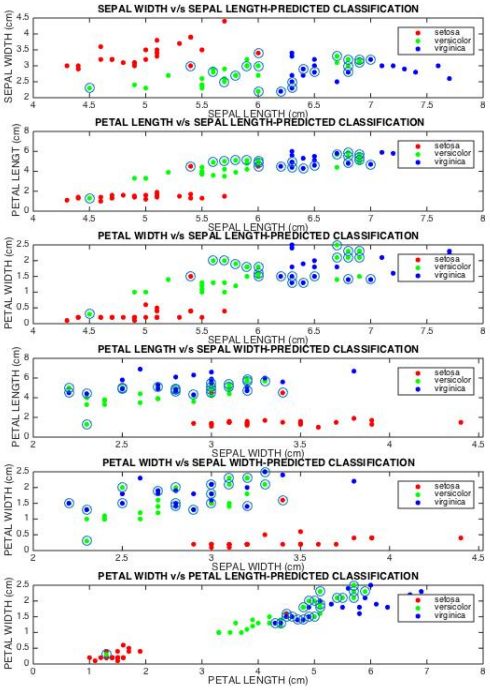


SEPAL LENGTH V/S SEPAL WIDTH:

Figure I.28-Accuracy 69.33%, Time 0.0064s

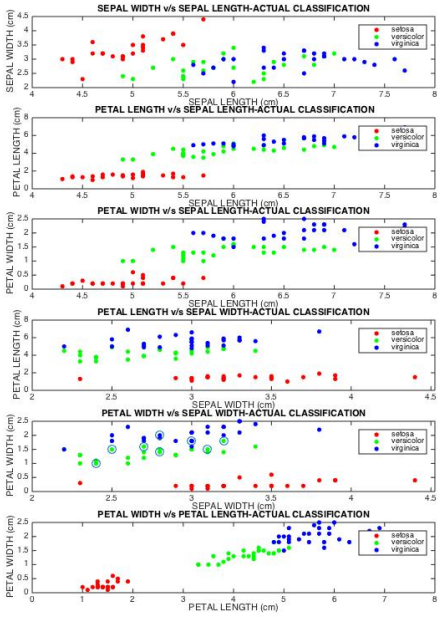


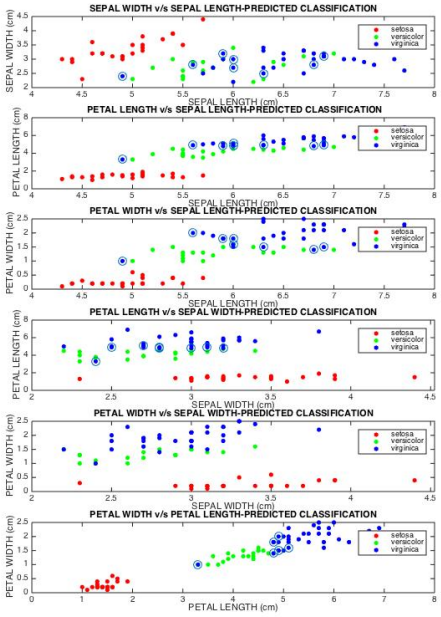


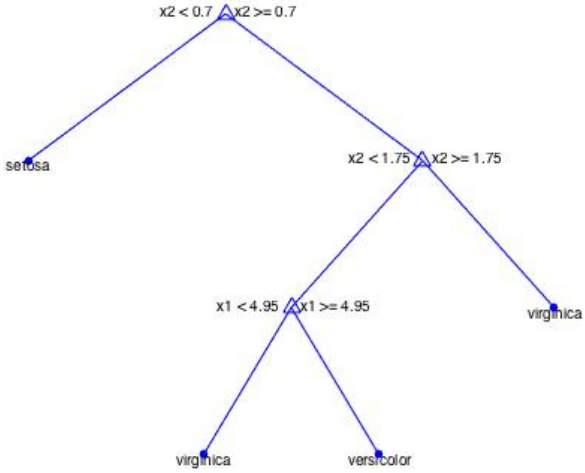


SEPAL LENGTH V/S PETAL LENGTH:

Figure I.29-Accuracy 92%, Time 0.006715s

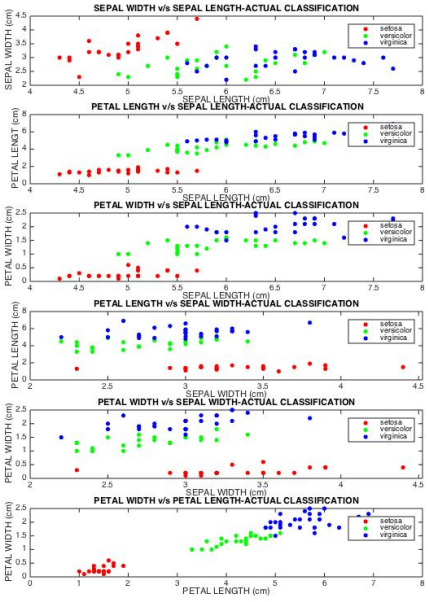


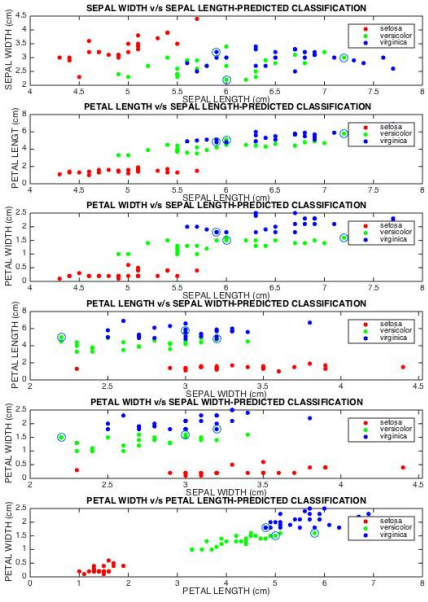


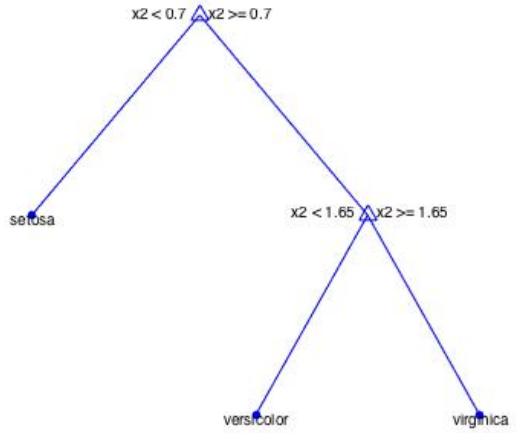


SEPAL LENGTH V/S PETAL WIDTH:

Figure I.30-Accuracy 94.67%, Time 0.004689s

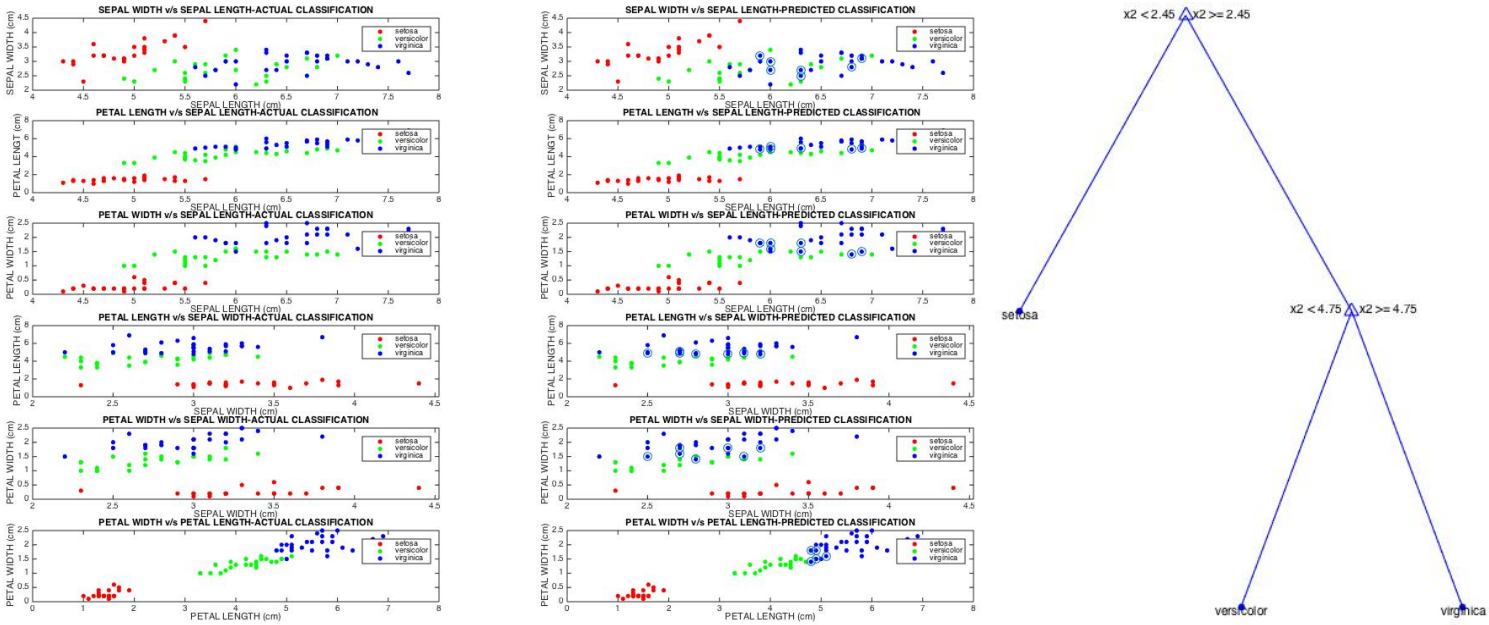






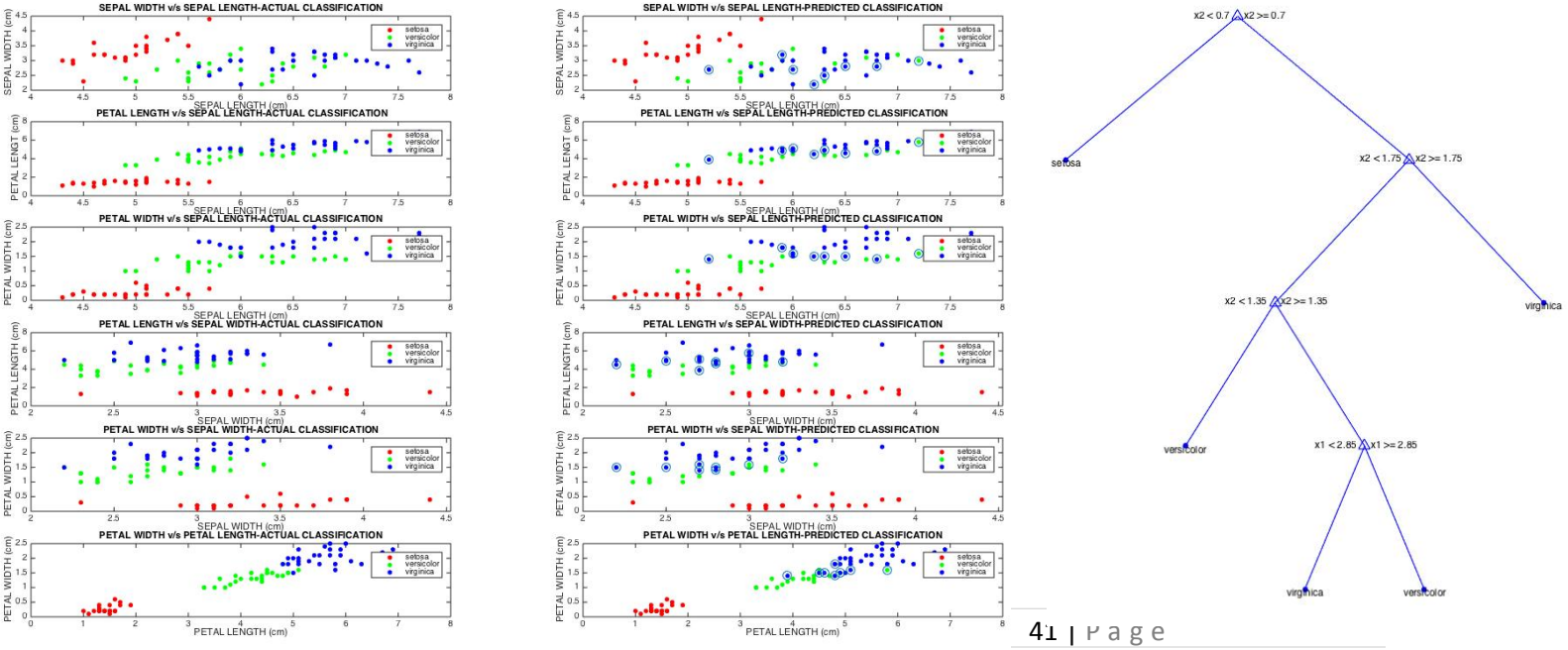
SEPAL WIDTH V/S PETAL LENGTH:

Figure I.31-Accuracy 93.33%, Time 0.007387s



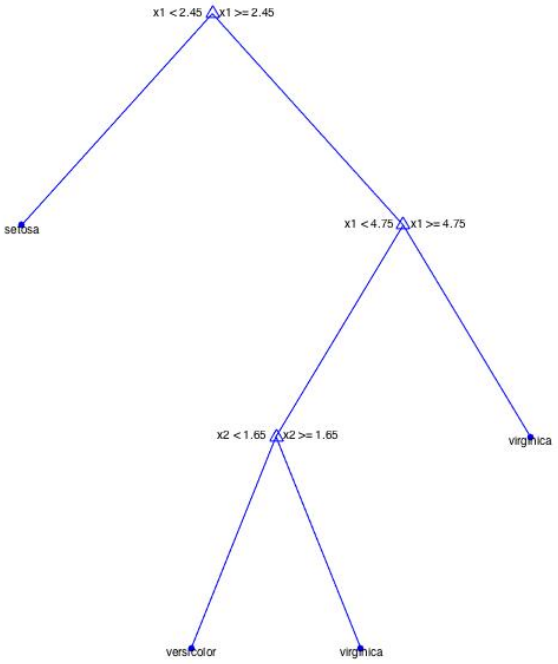
SEPAL WIDTH V/S PETAL WIDTH:

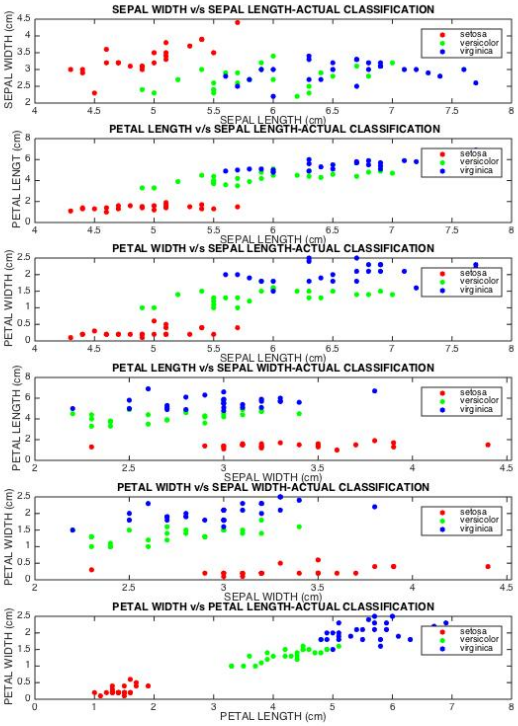
Figure I.32-Accuracy 89.33%, Time 0.006183s

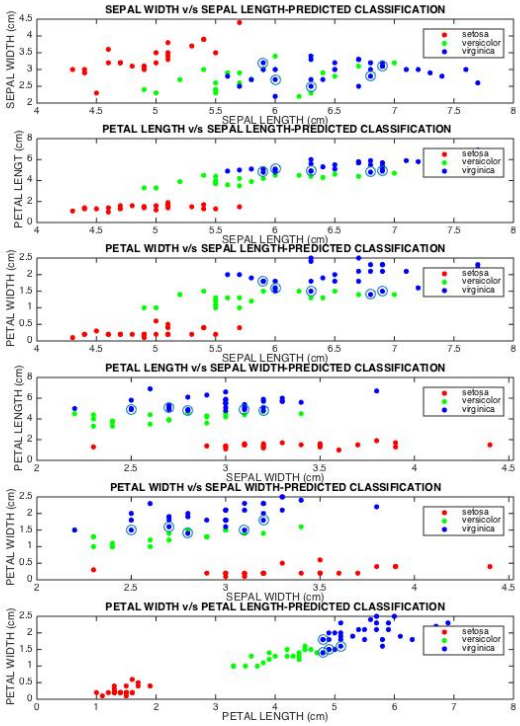


PETAL LENGTH V/S PETAL WIDTH:

Figure I.33-Accuracy 93.33%, Time 0.005095s



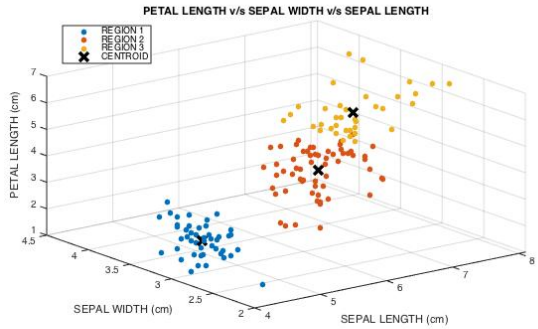


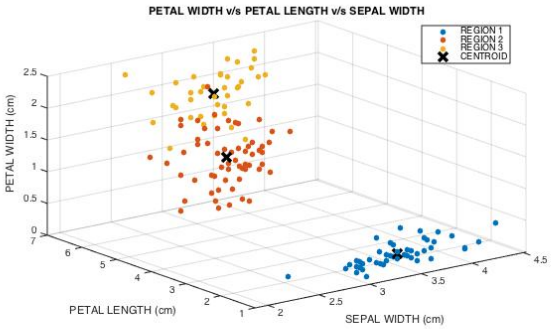


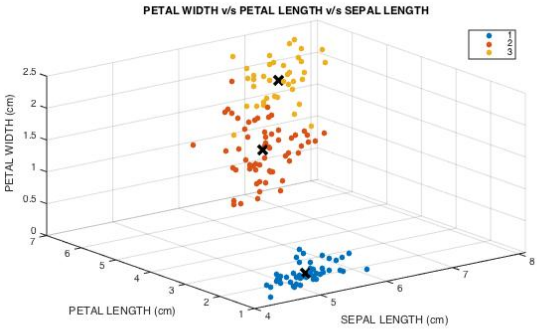
K-Mean Clustering:

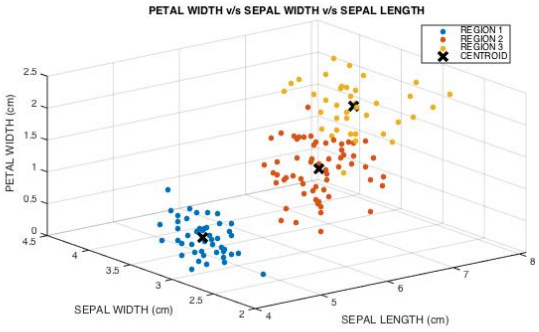
SEPAL LENGTH v/s SEPAL WIDTH v/s PETAL LENGTH v/s PETAL WIDTH:

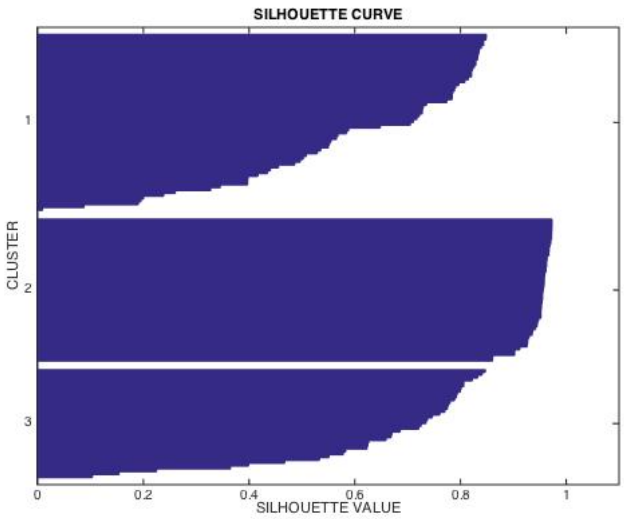
Figure I.34-Accuracy 89.33%, Silhouette Value 0.7357, Testing Time 0.0097835s

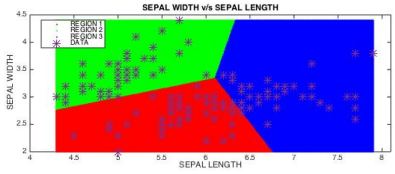


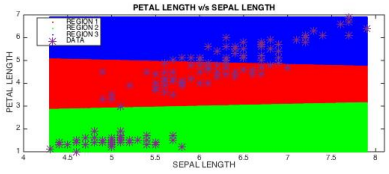


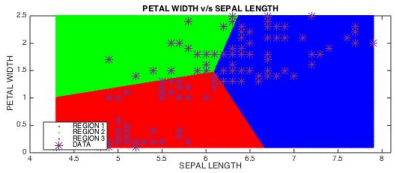


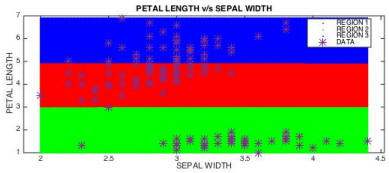


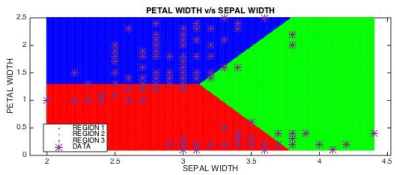


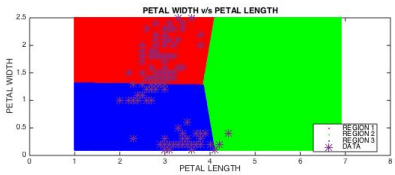








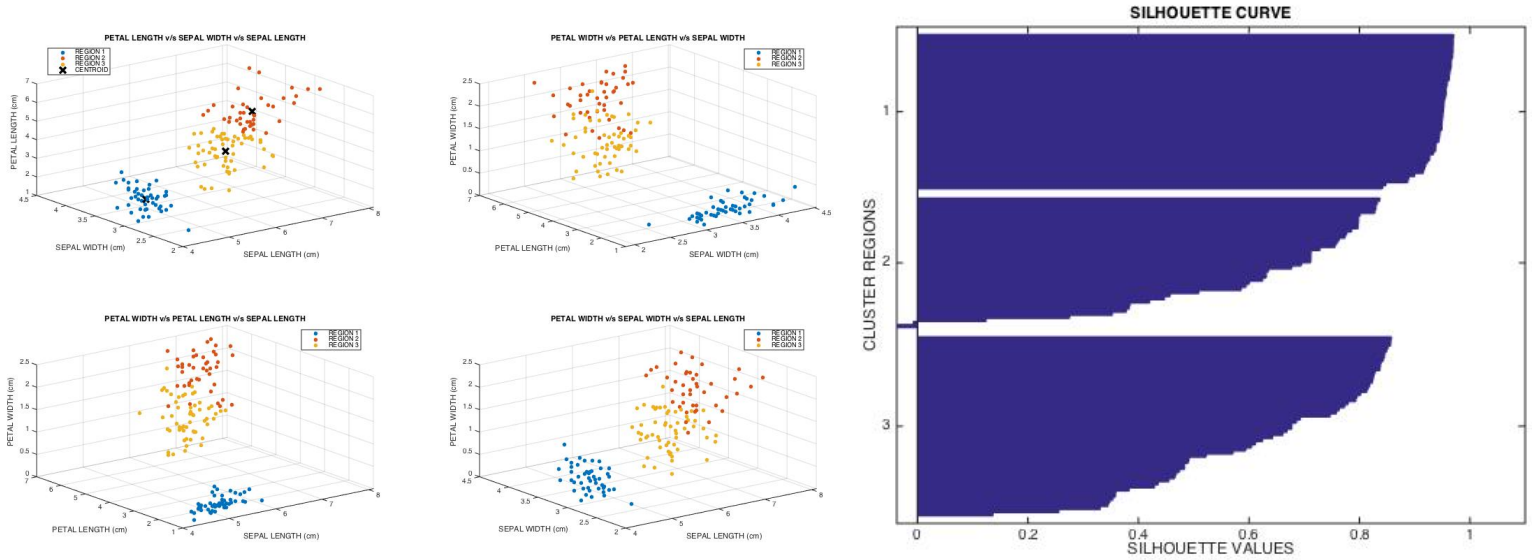




Value 0.7357, Testing Time 0.0097835s

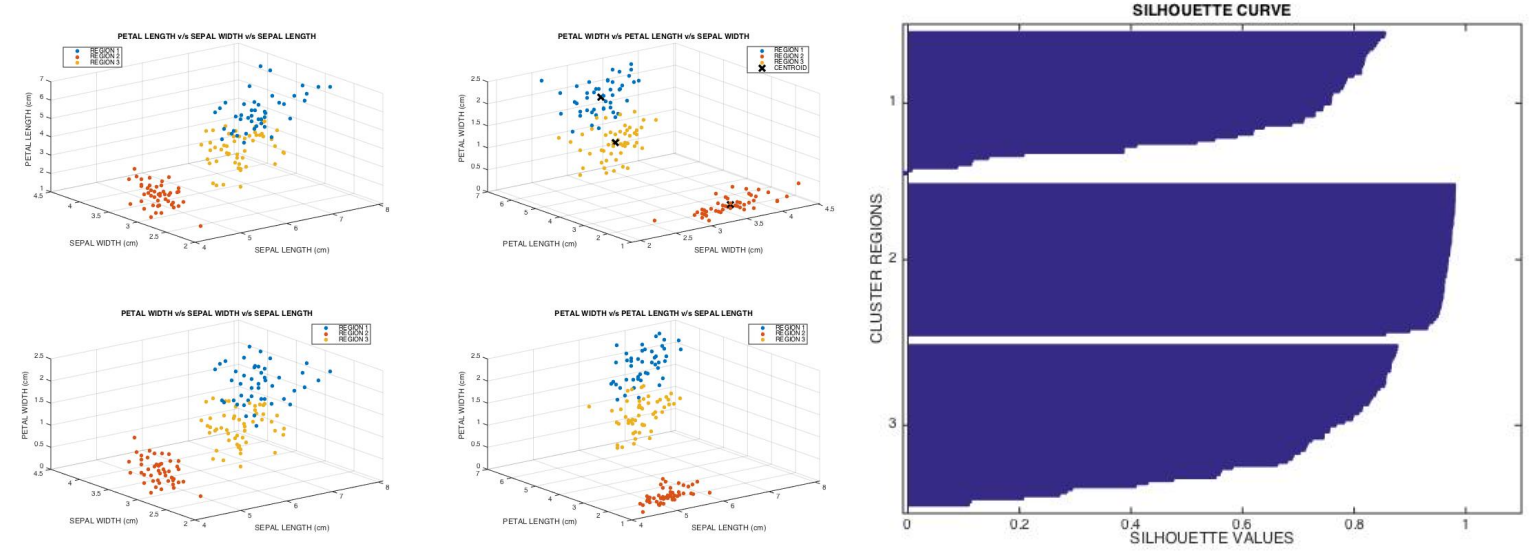
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL LENGTH:

Figure I.35-Accuracy 88%, Silhouette Value 0.7330, Testing Time 0.0075335s



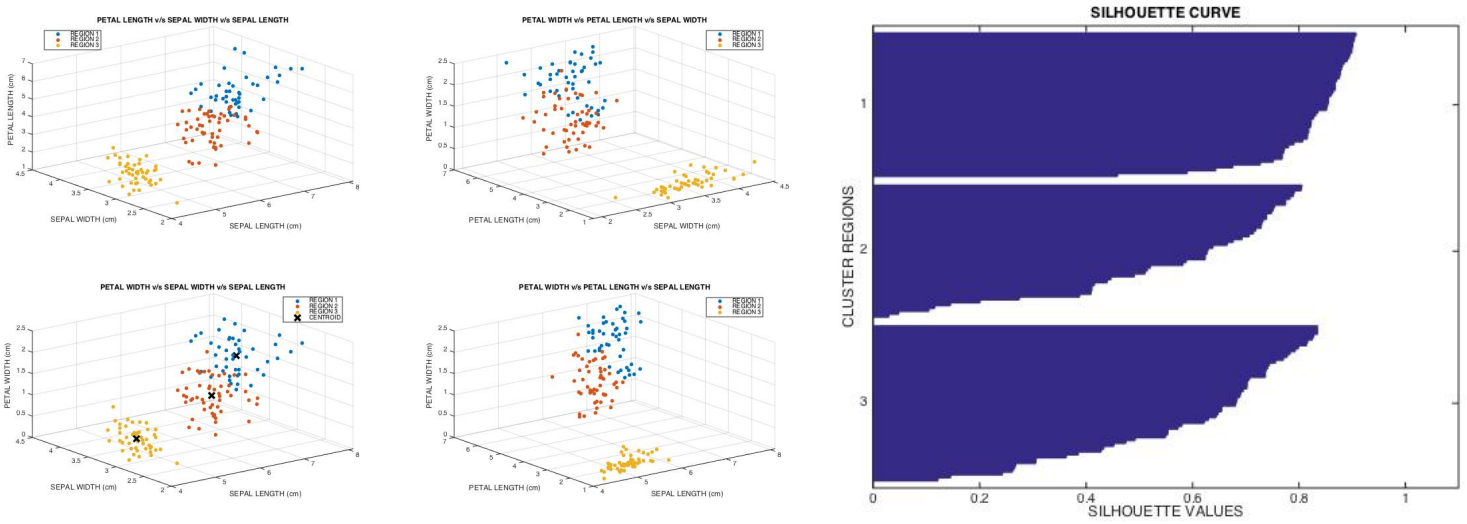
SEPAL WIDTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.36-Accuracy 95.33%, Silhouette Value 0.7659, Testing Time 0.00829s



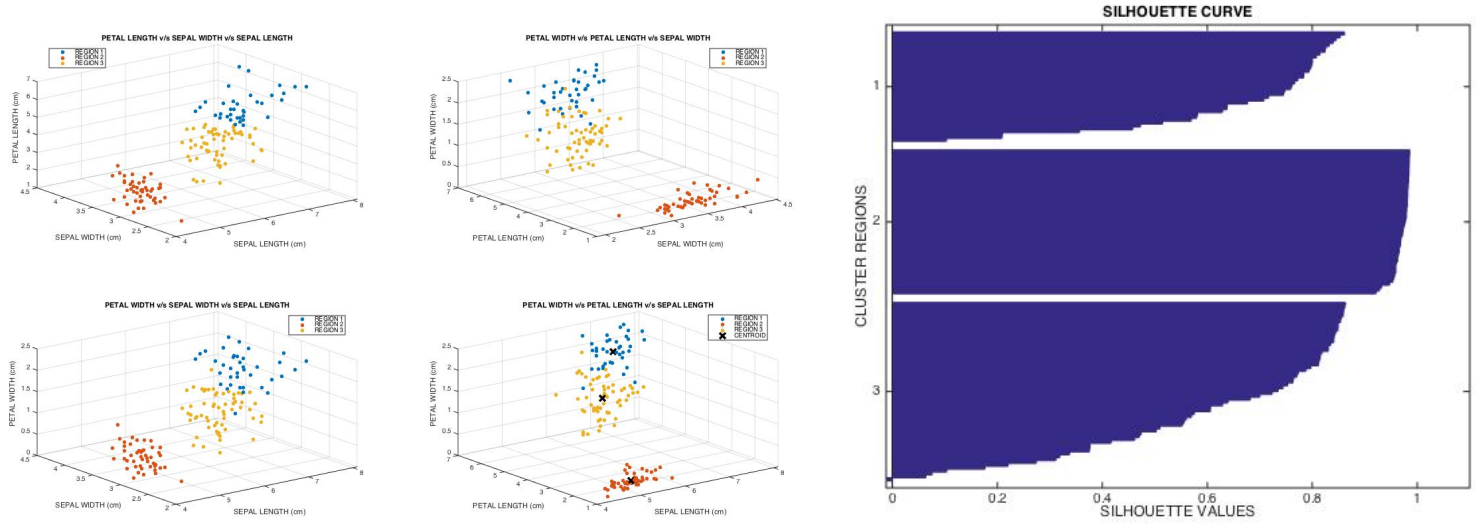
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL WIDTH:

Figure I.37-Accuracy 82.67%, Silhouette Value 0.6654, Testing Time 0.0090135s



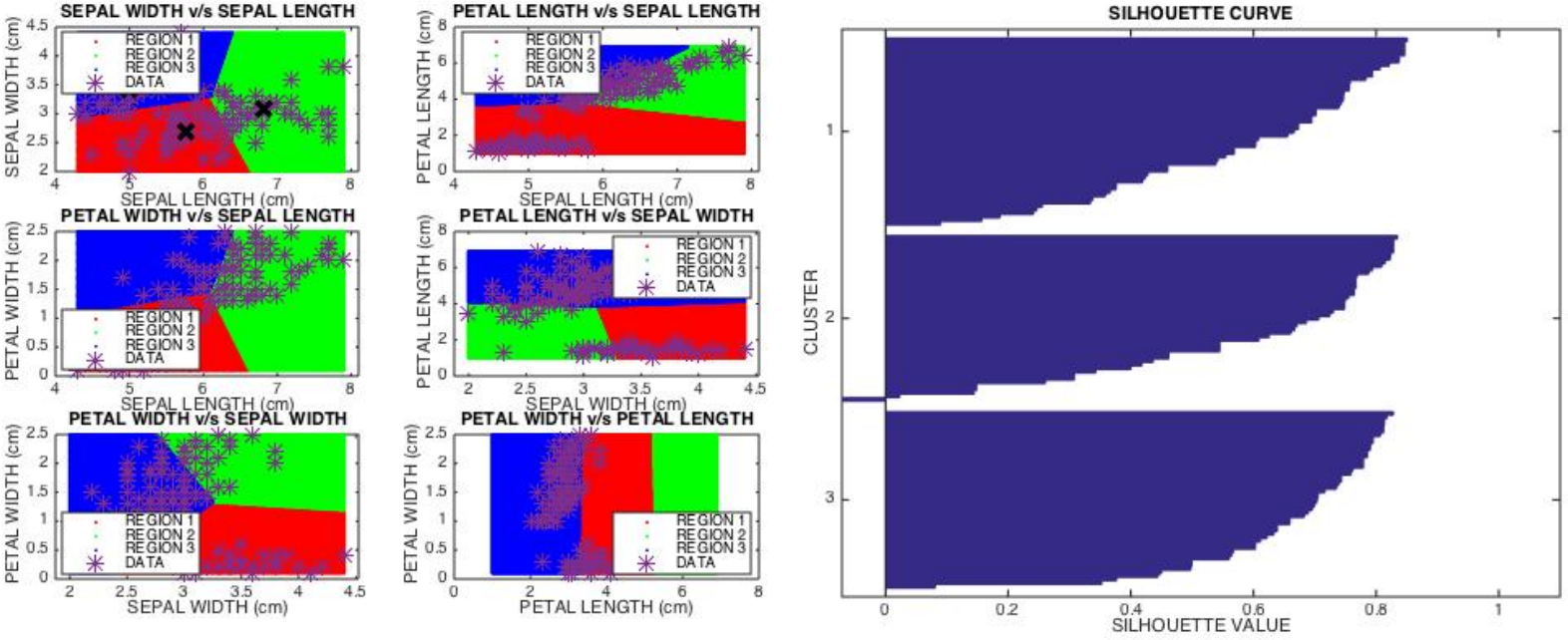
SEPAL LENGTH V/S PETAL LENGTH V/S PETAL WIDTH:

Figure I.38-Accuracy 89.33%, Silhouette Value 0.7523, Testing Time 0.00589225s



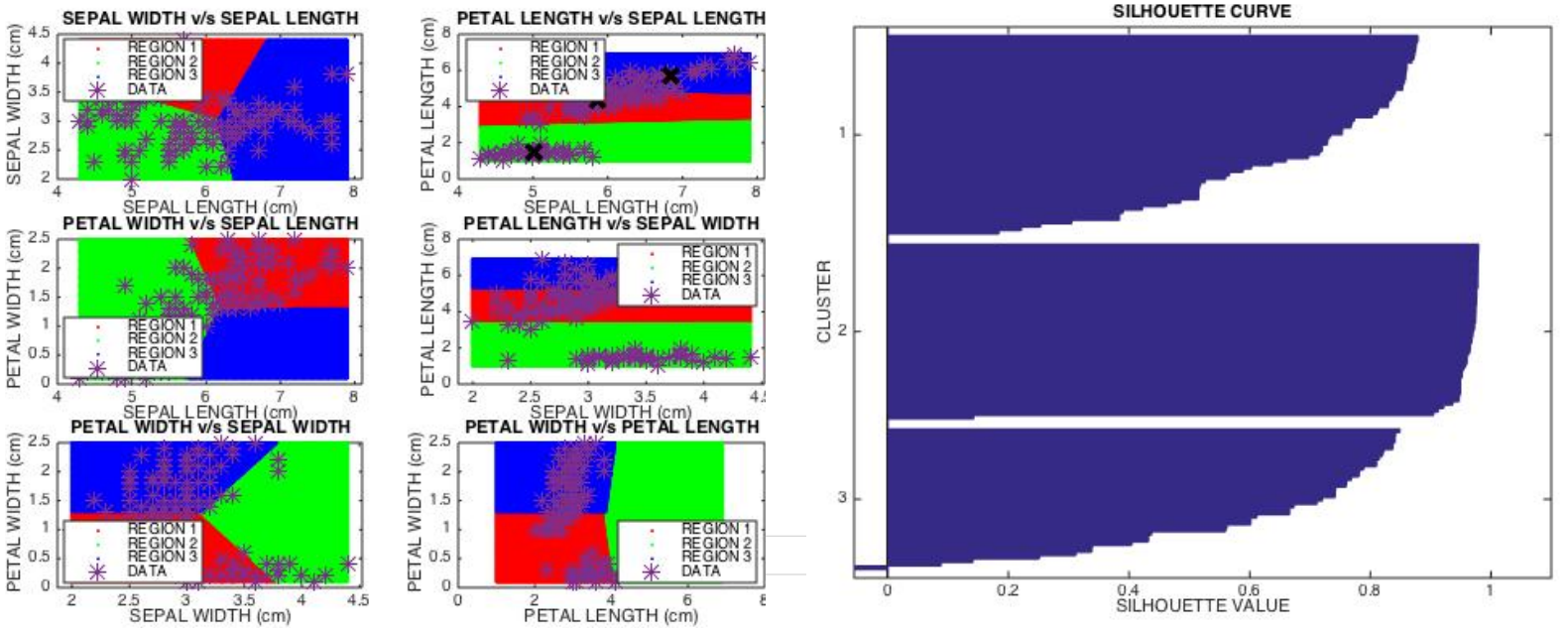
SEPAL LENGTH V/S SEPAL WIDTH:

Figure I.39-Accuracy 82%, Silhouette Value- 0.6201, Testing Time 0.00891975s



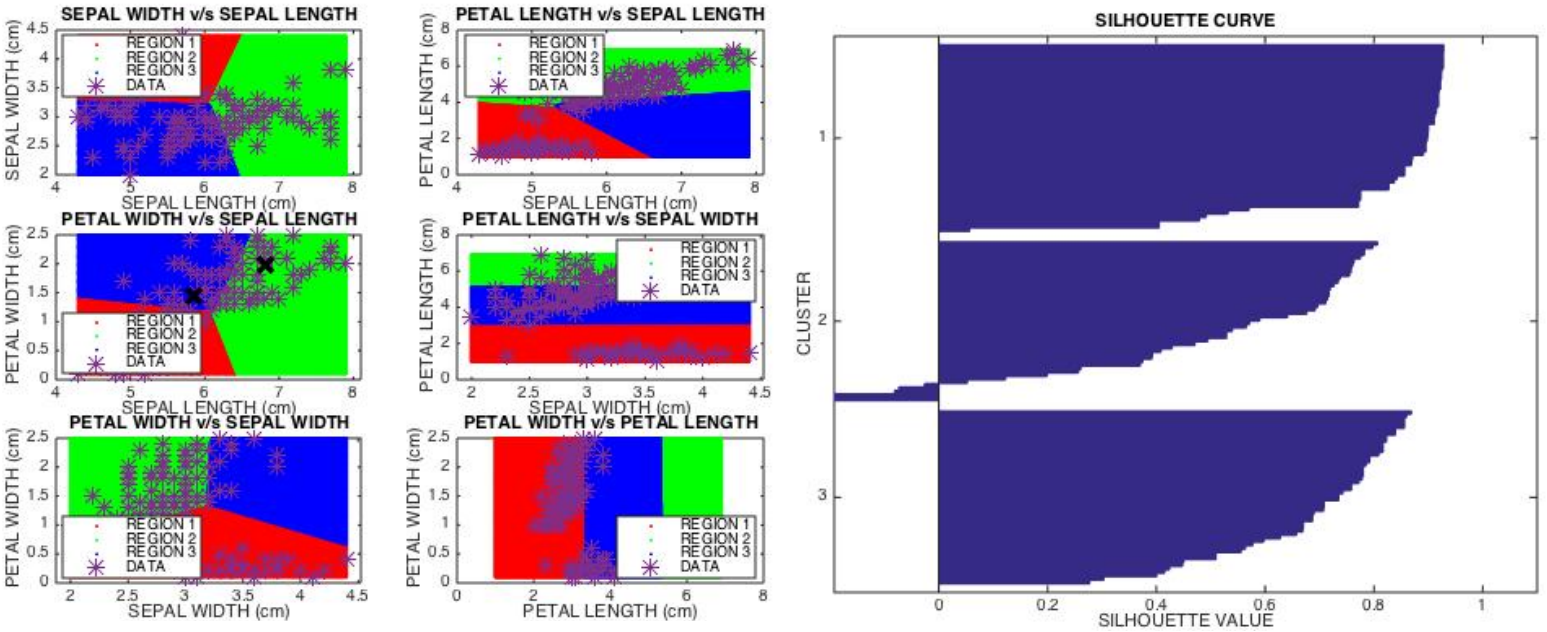
SEPAL LENGTH V/S PETAL LENGTH:

Figure I.40-Accuracy 88%, Silhouette Value 0.7560, Testing Time 0.00694425s



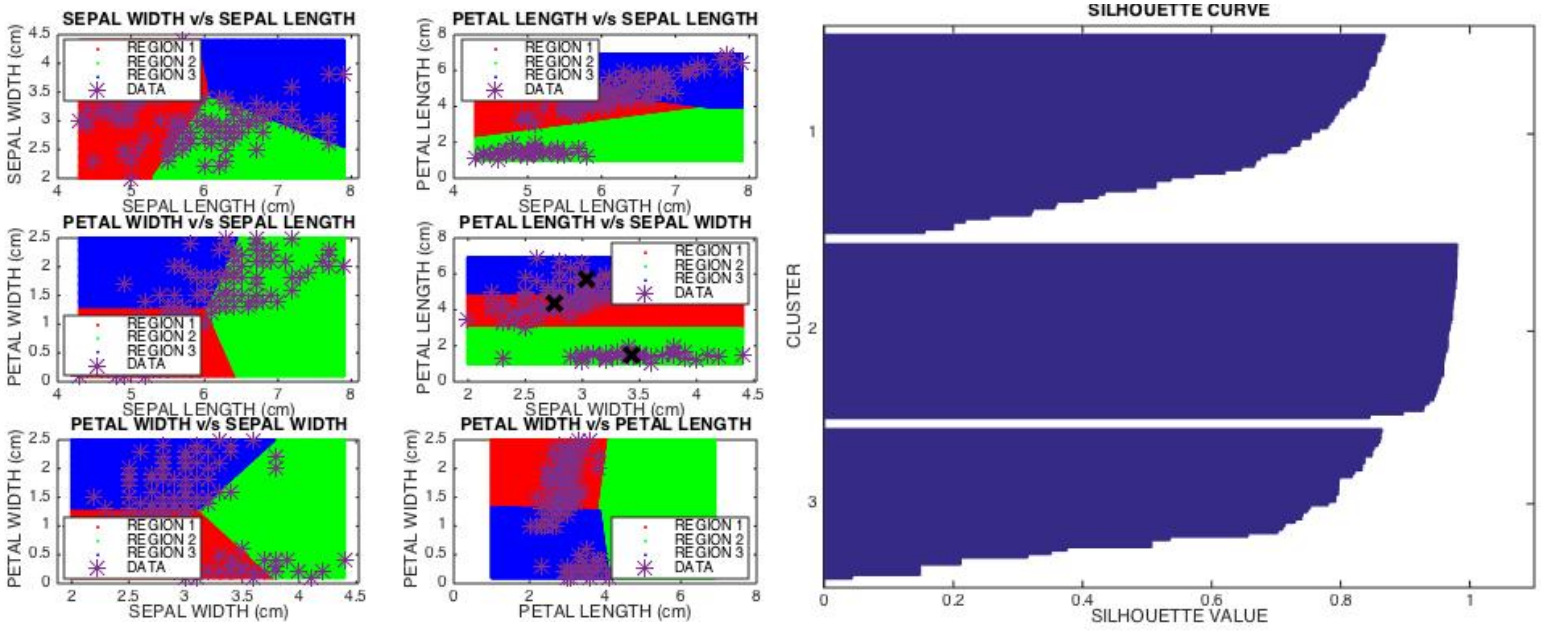
SEPAL LENGTH V/S PETAL WIDTH:

Figure I.41-Accuracy 82.67%, Silhouette Value 0.6813, Testing Time 0.0070455s



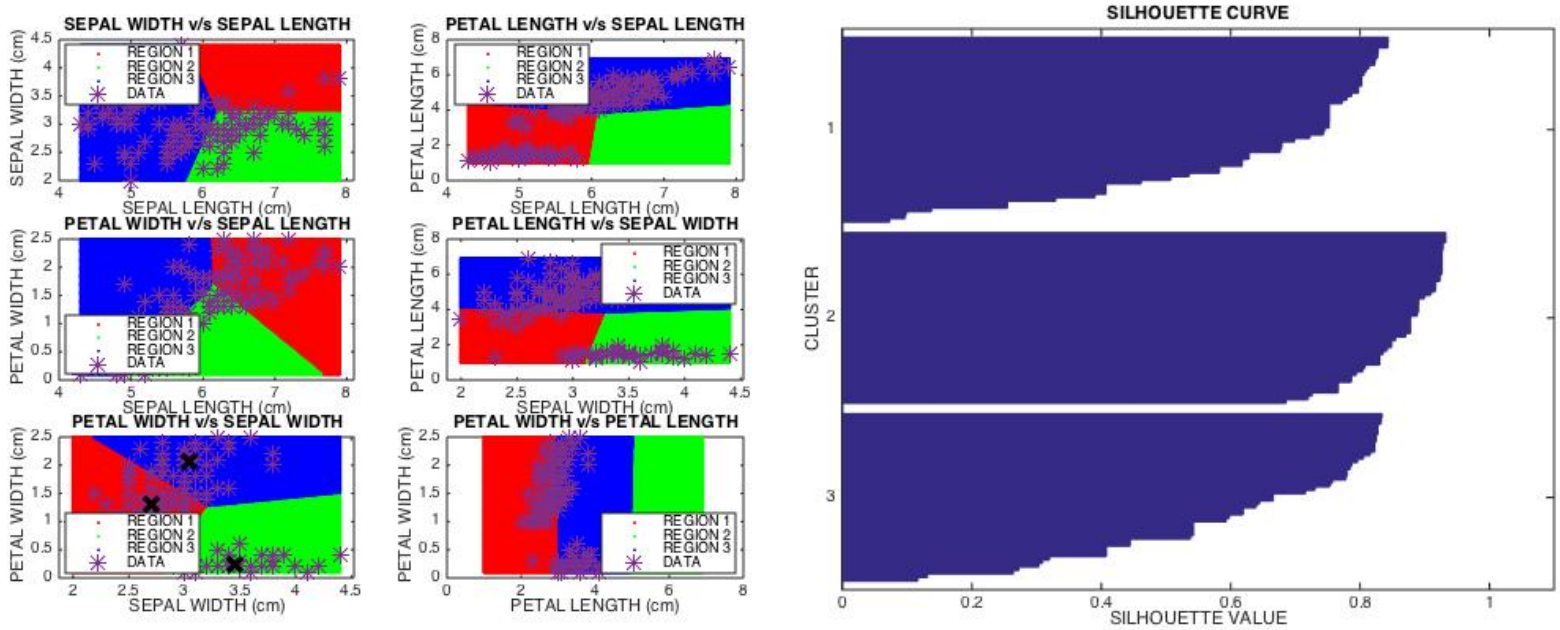
SEPAL WIDTH V/S PETAL LENGTH:

Figure I.42-Accuracy 92.67%, Silhouette Value 0.7612, Testing Time 0.0097835s



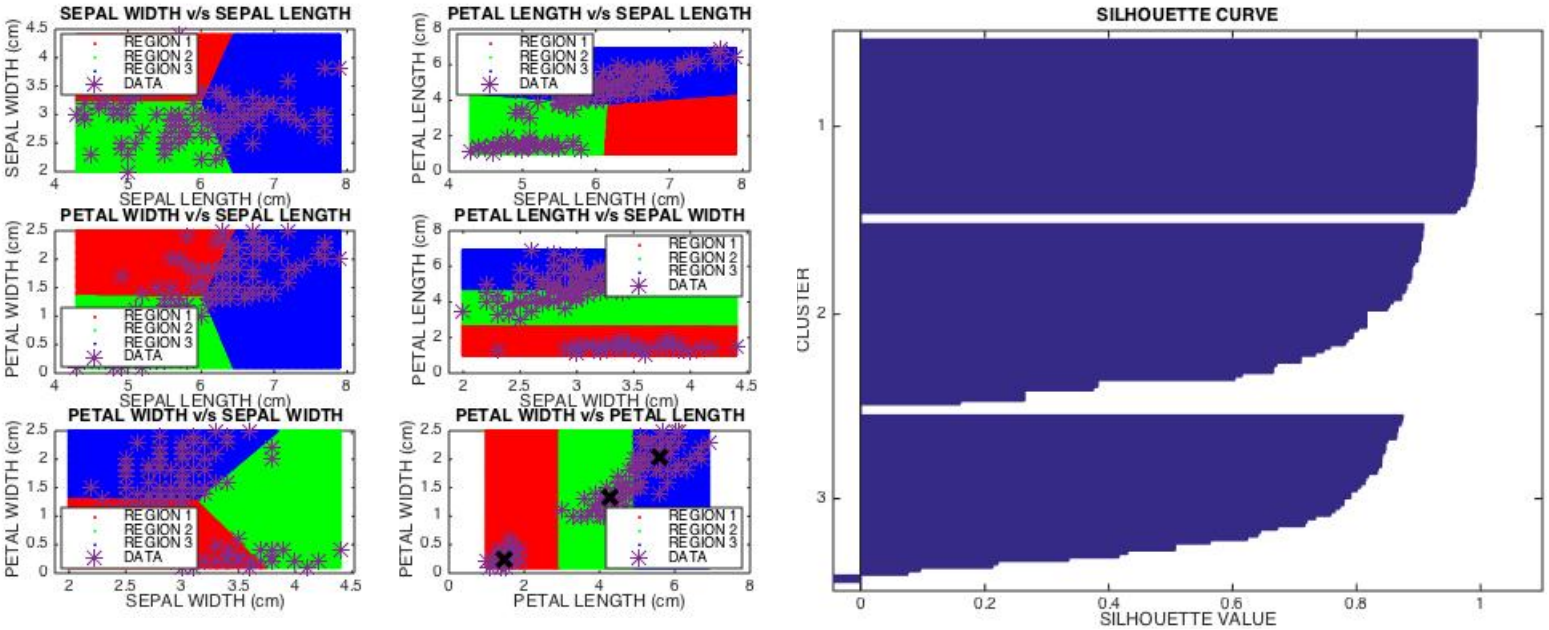
SEPAL WIDTH V/S PETAL WIDTH:

Figure I.43-Accuracy 92.67%, Silhouette Value 0.7067, Testing Time 0.0075335s



PETAL LENGTH V/S PETAL WIDTH:

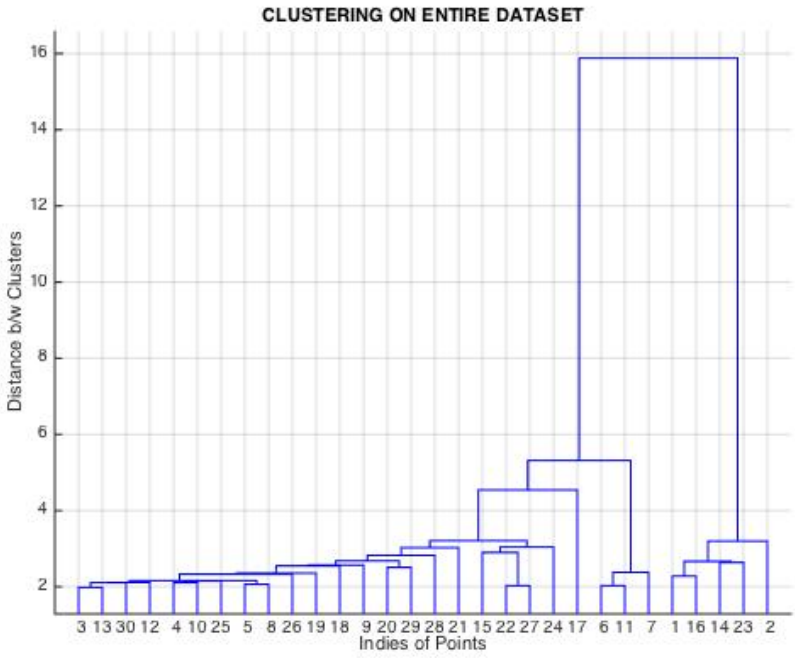
Figure I.44-Accuracy 96%, Silhouette Value 0.8055, Testing Time 0.00829s

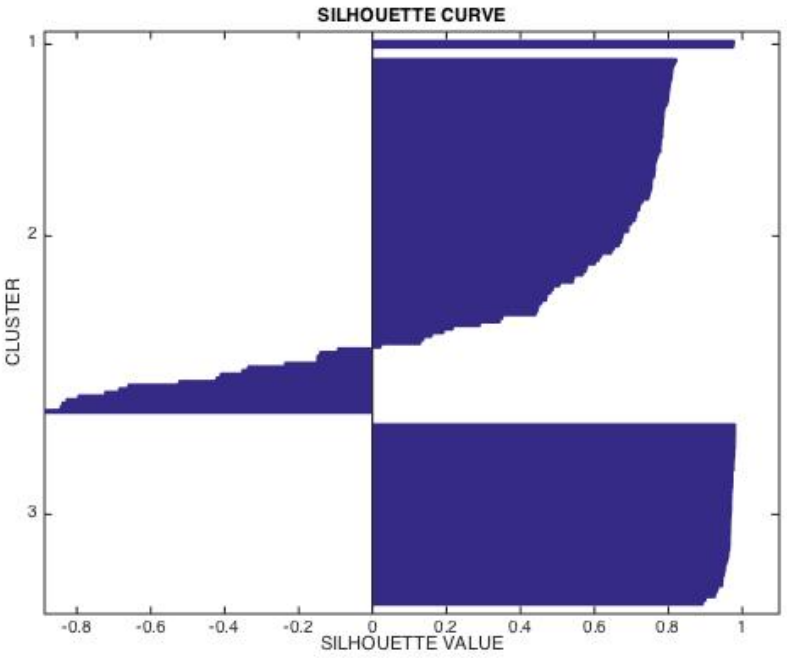


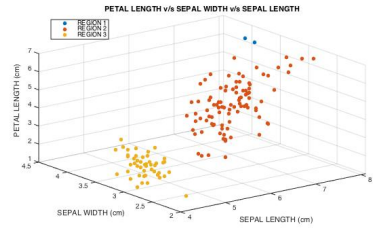
Agglomerative Hierarchical Clustering:

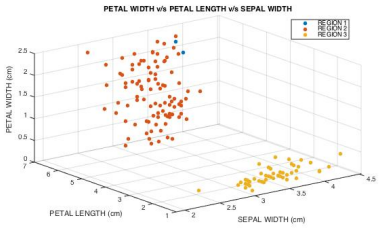
SEPAL LENGTH v/s SEPAL WIDTH v/s PETAL LENGTH v/s PETAL WIDTH:

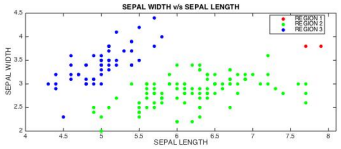
Figure I.45-Purity 68%, Silhouette Value 0.6184, Testing Time 0.009507s

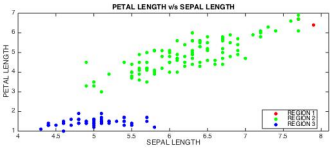


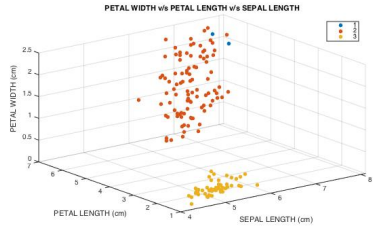


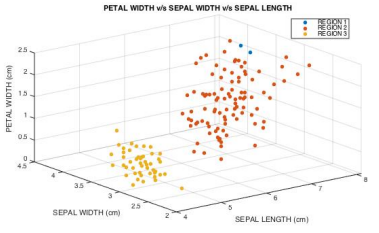


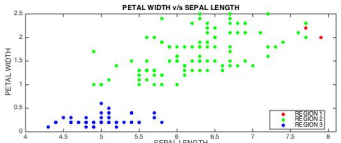


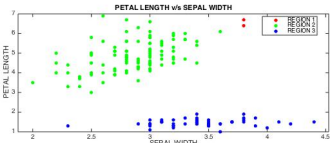


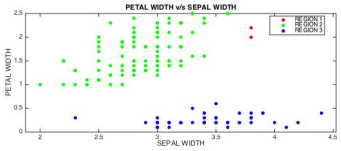


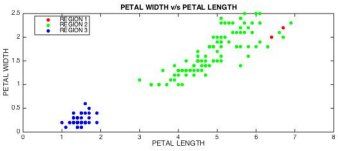






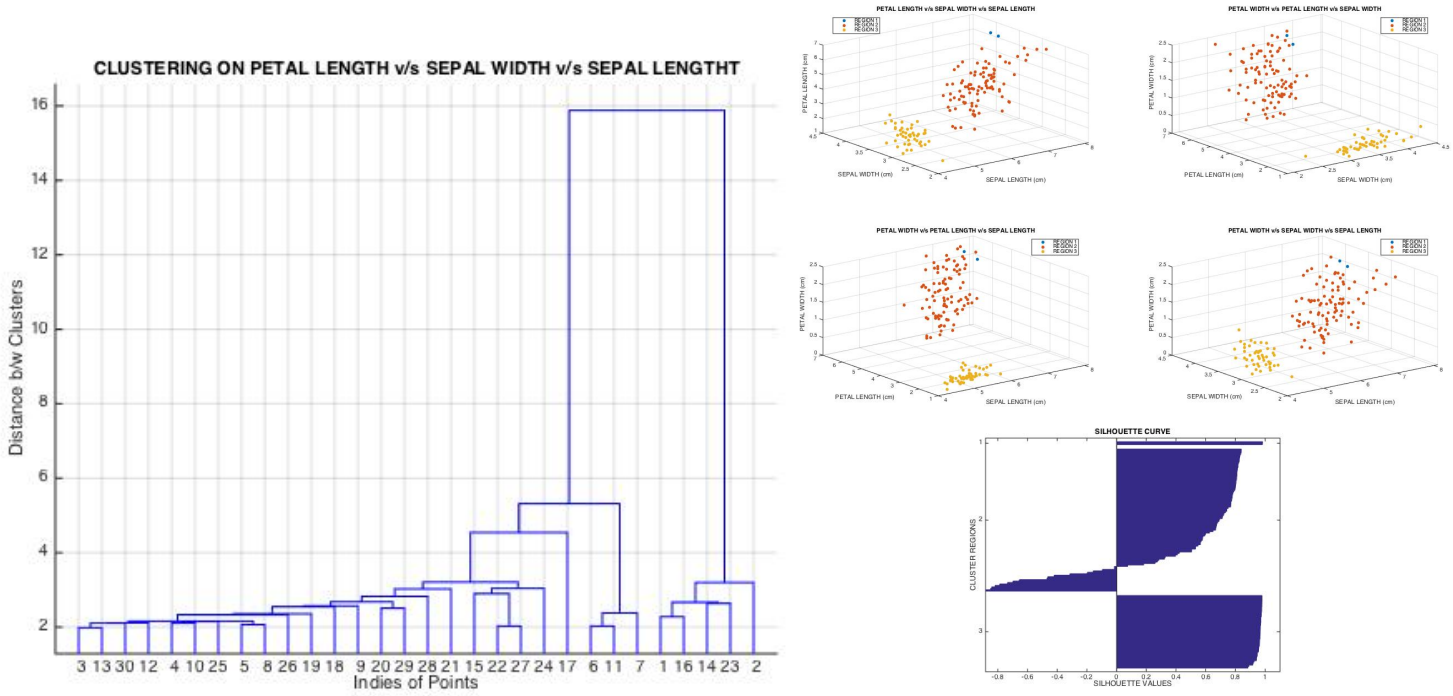






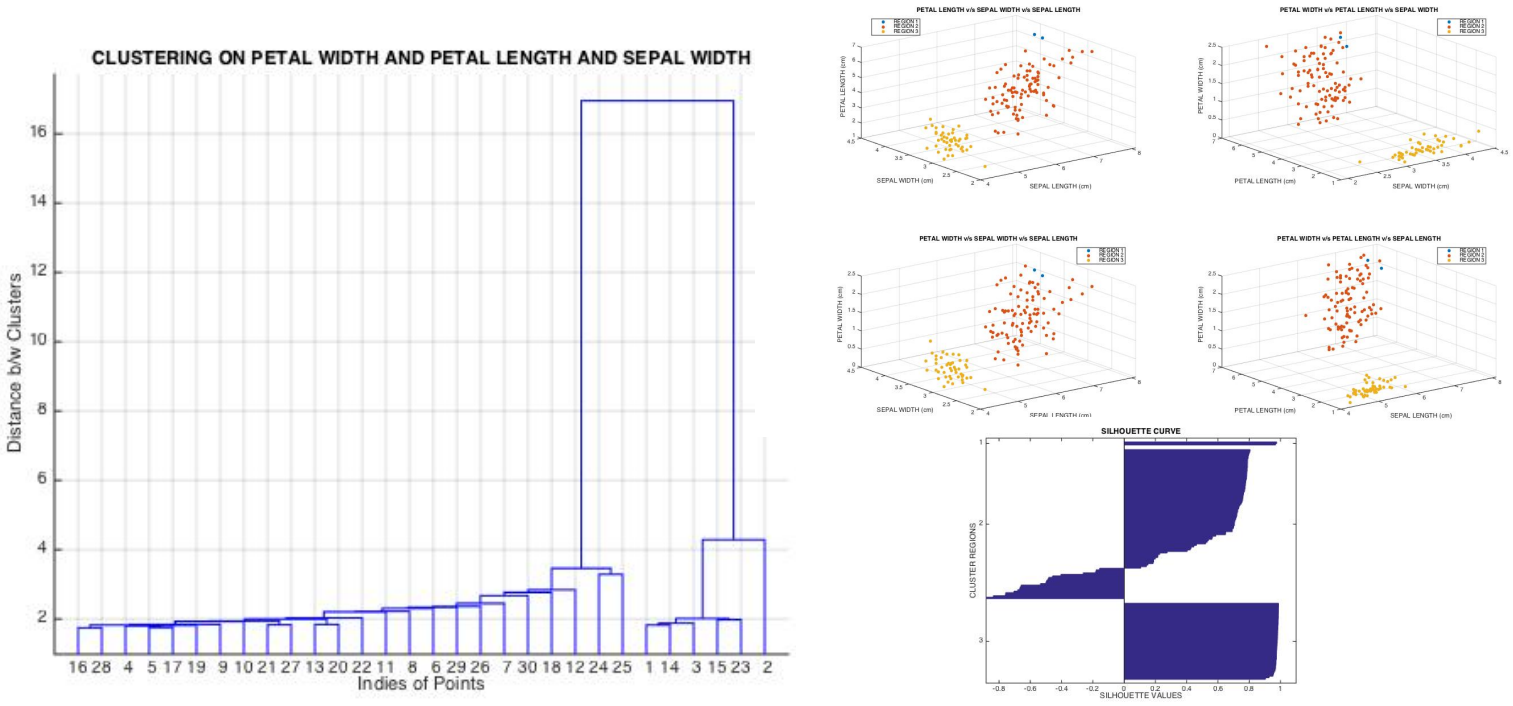
SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL LENGTH:

Figure I.46-Accuracy 68%, Silhouette Value 0.6433, Testing Time 0.010582s



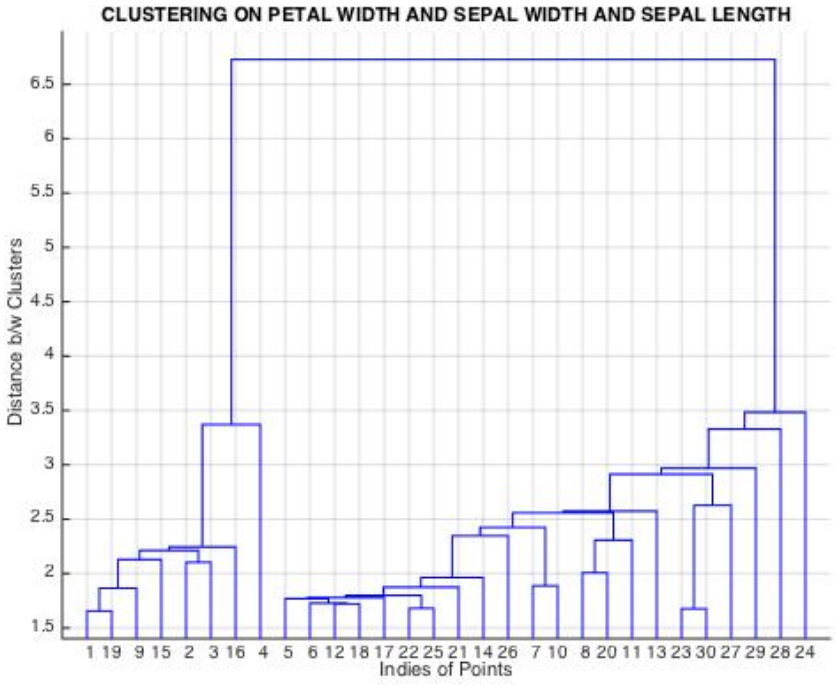
SEPAL WIDTH V/S PETAL LENGTH V/S PETAL WIDTH:

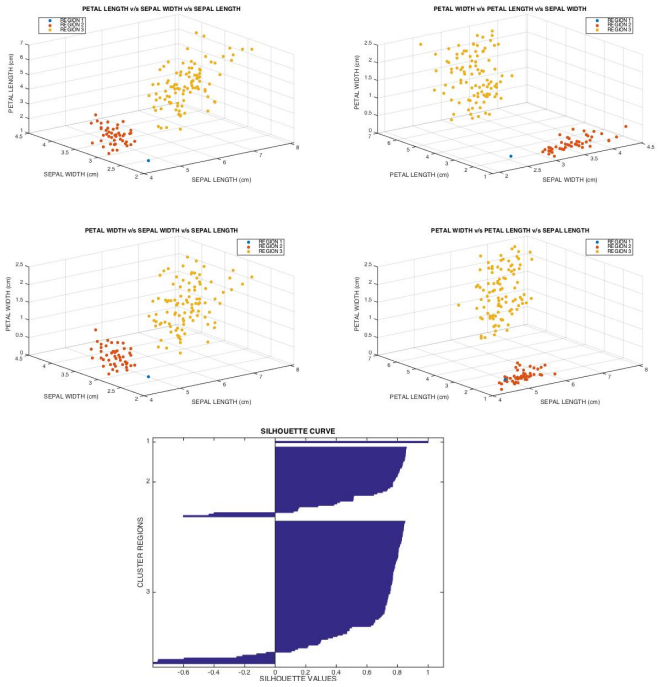
Figure I.47-Accuracy 68%, Silhouette Value 0.5985, Testing Time 0.007566s



SEPAL LENGTH V/S SEPAL WIDTH V/S PETAL WIDTH:

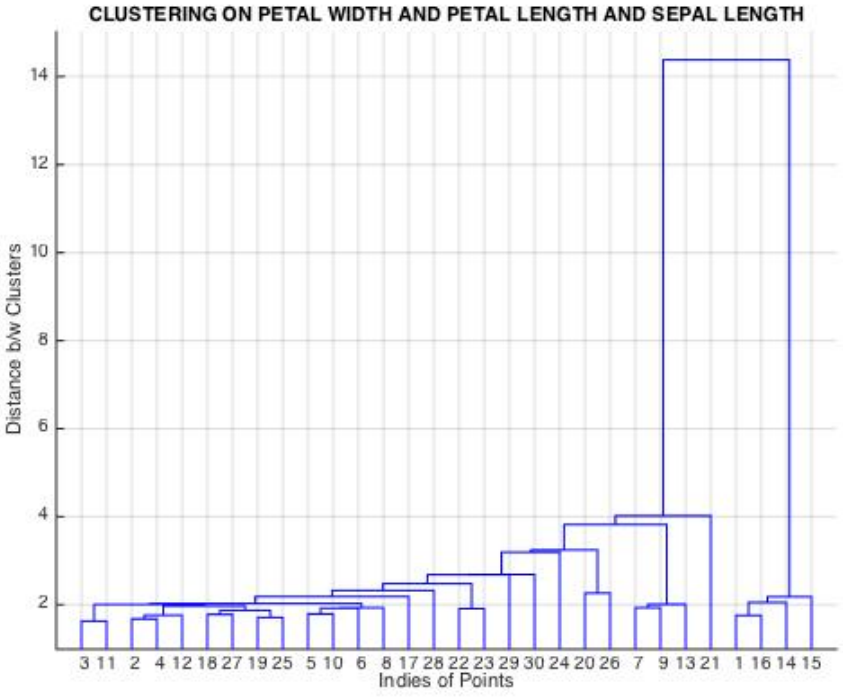
Figure I.48-Accuracy 66.67% , Silhouette Value 0.5999, Testing Time 0.004728s

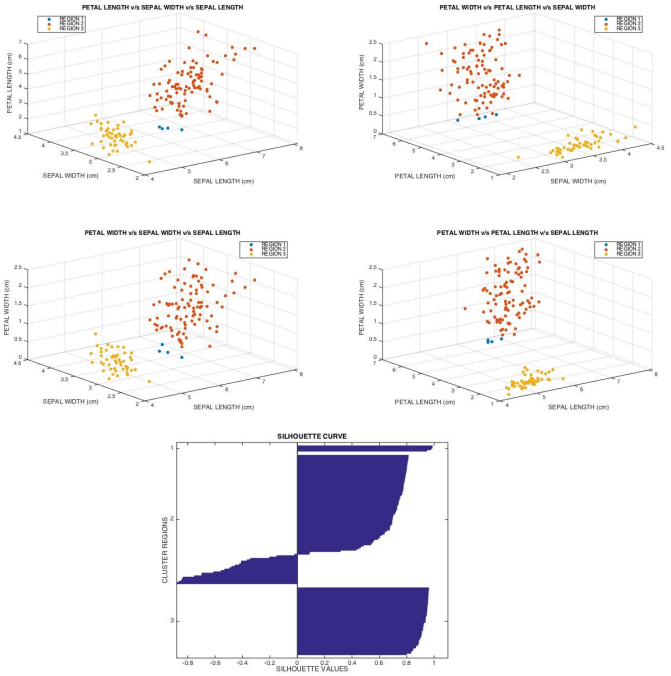




SEPAL LENGTH V/S PETAL LENGTH V/S PETAL WIDTH:

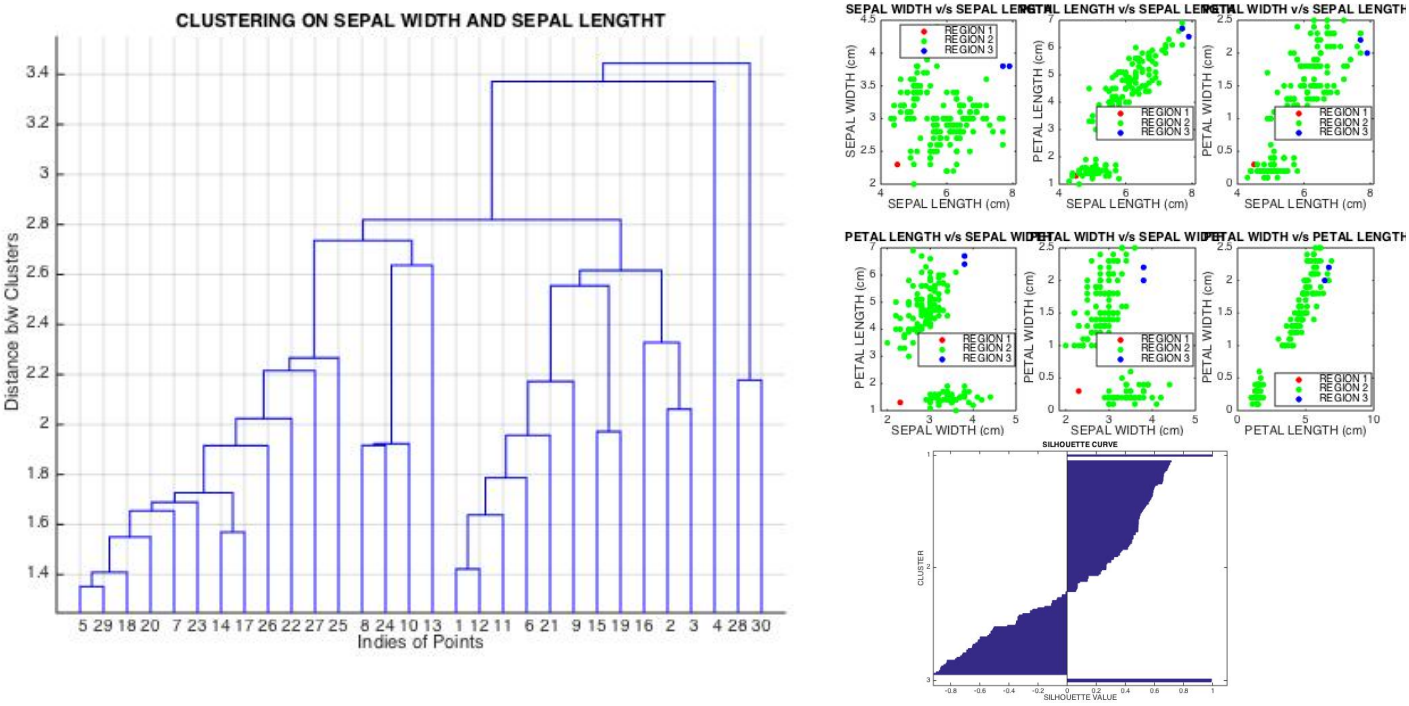
Figure I.49-Accuracy 69.33%, Silhouette Value 0.5934, Testing Time 0.005387s





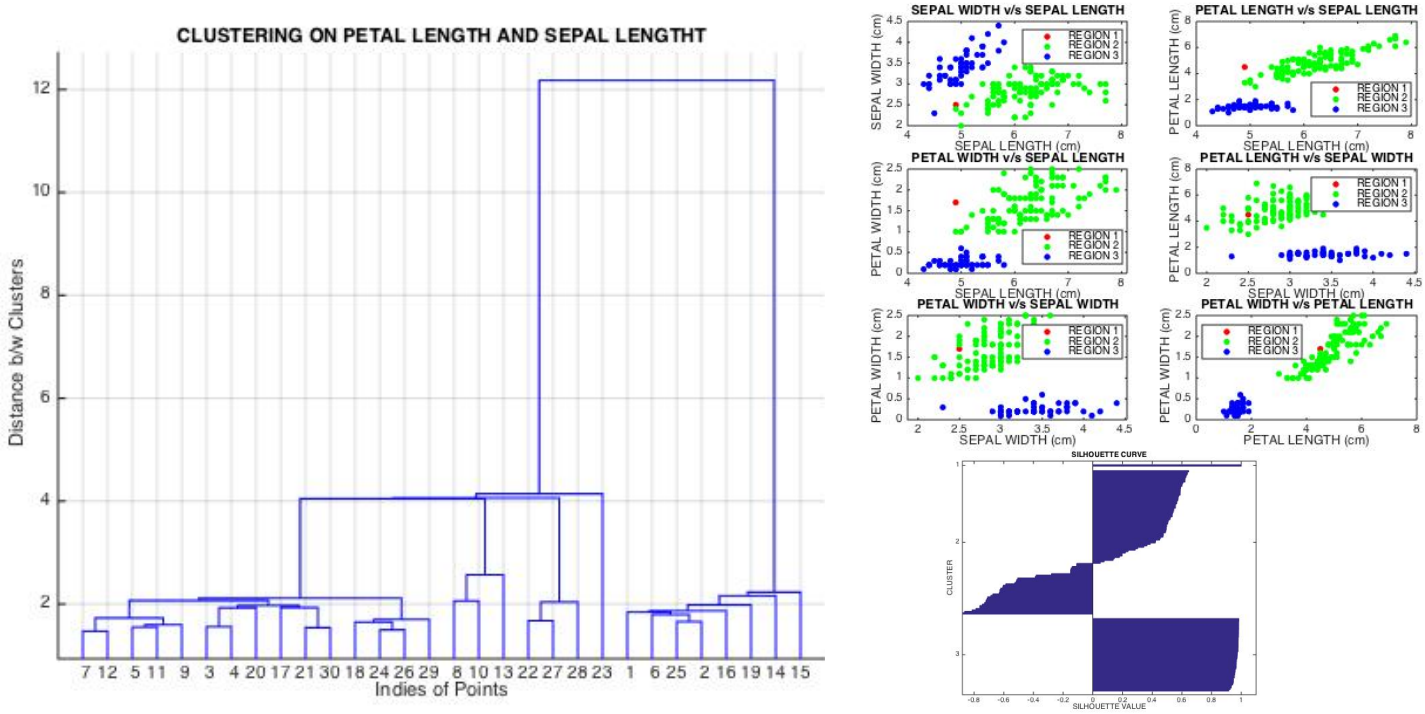
SEPAL LENGTH V/S SEPAL WIDTH:

Figure I.50-Accuracy 35.33%, Silhouette Value-0.1028, Testing Time 0.005569s



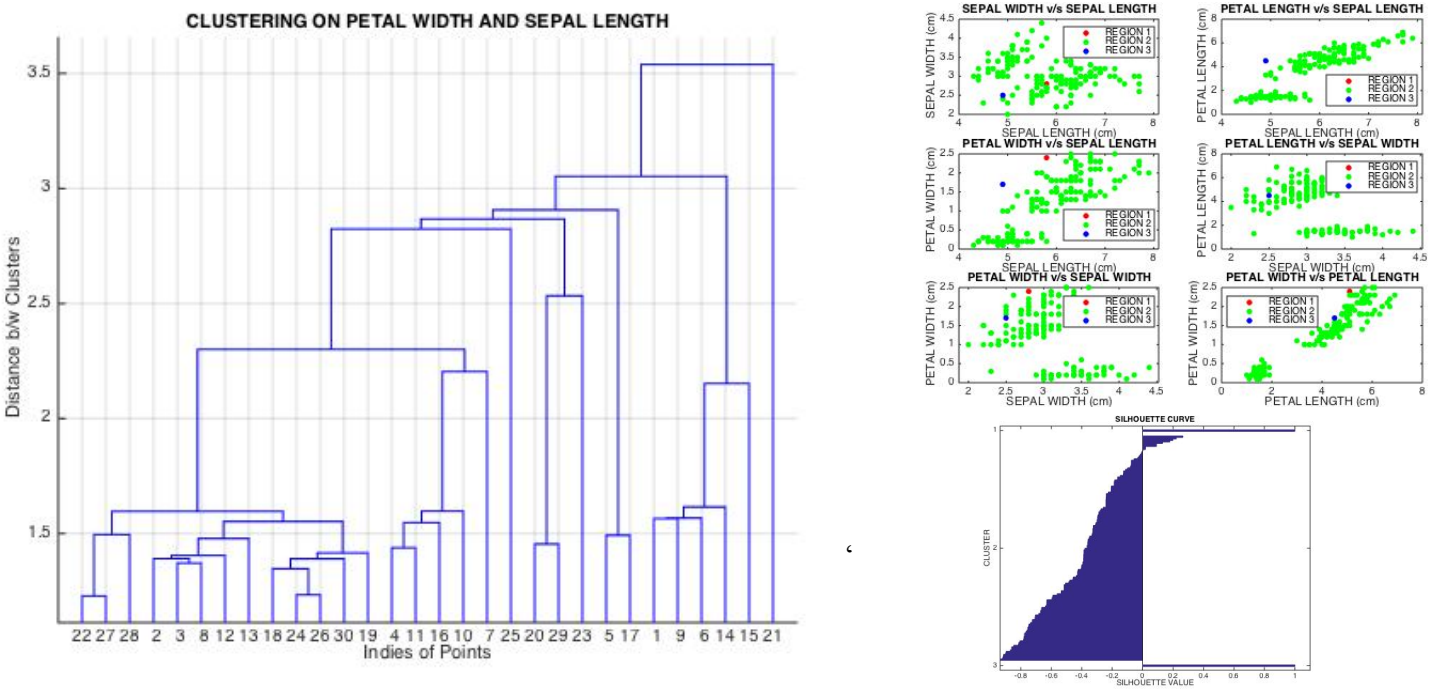
SEPAL LENGTH V/S PETAL LENGTH:

Figure I.51-Accuracy 67.33%, Silhouette Value 0.4068, Testing Time 0.005797s



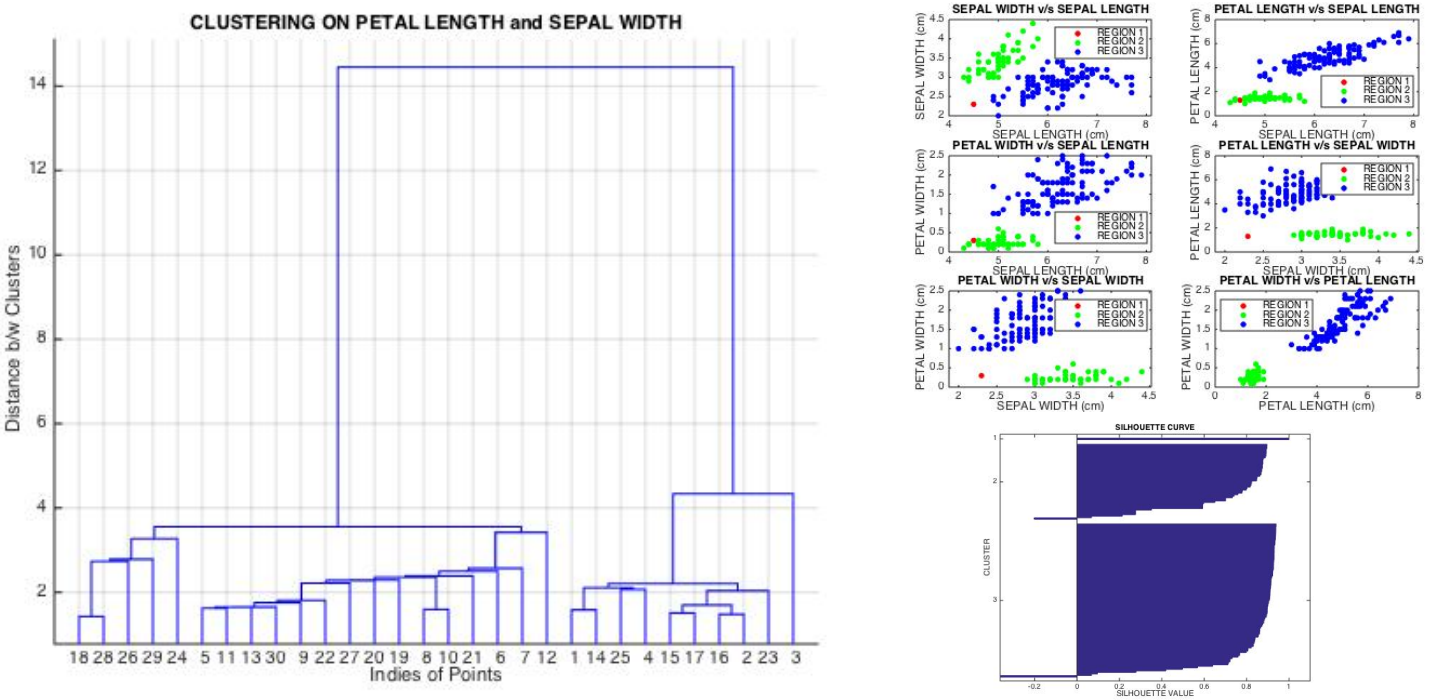
SEPAL LENGTH V/S PETAL WIDTH:

Figure I.52-Accuracy 34.67%, Silhouette Value -0.3799, Testing Time-0.00942s



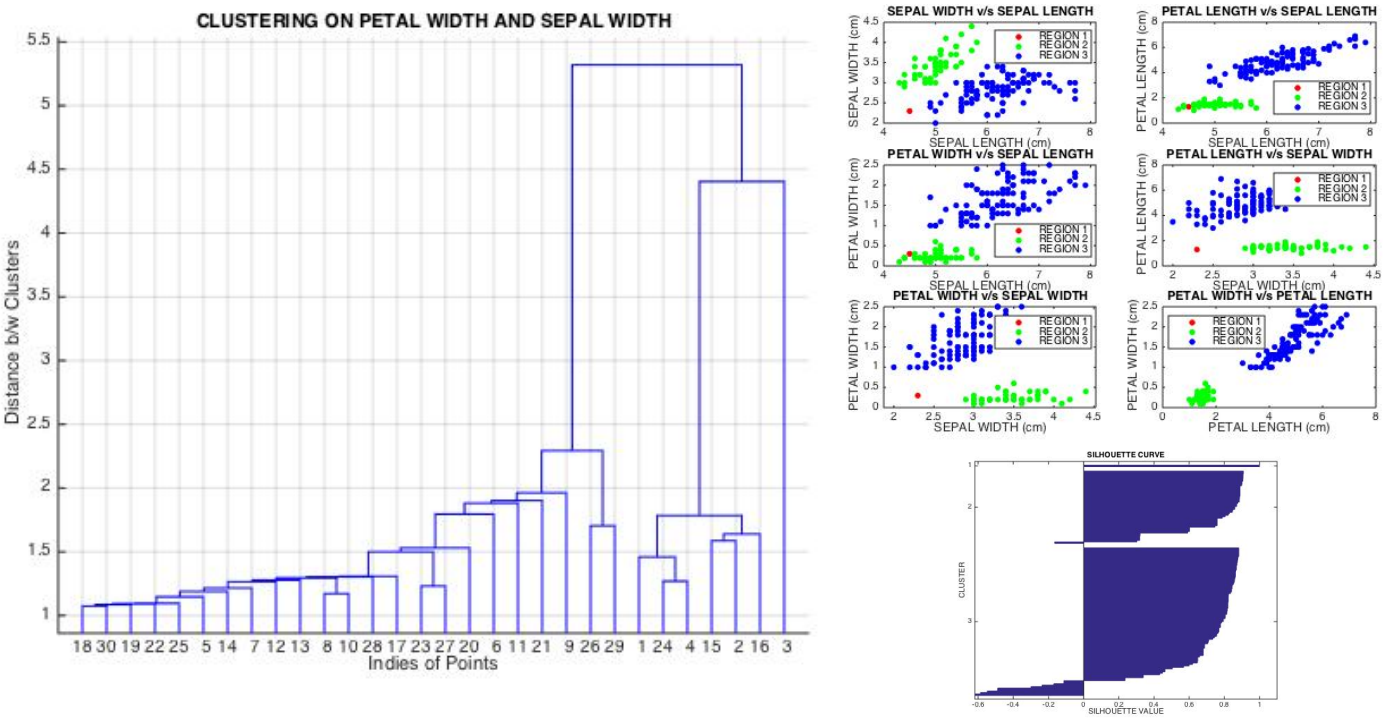
SEPAL WIDTH V/S PETAL LENGTH:

Figure I.53-Accuracy 66.67%, Silhouette Value 0.8019,Testing Time 0.008267s



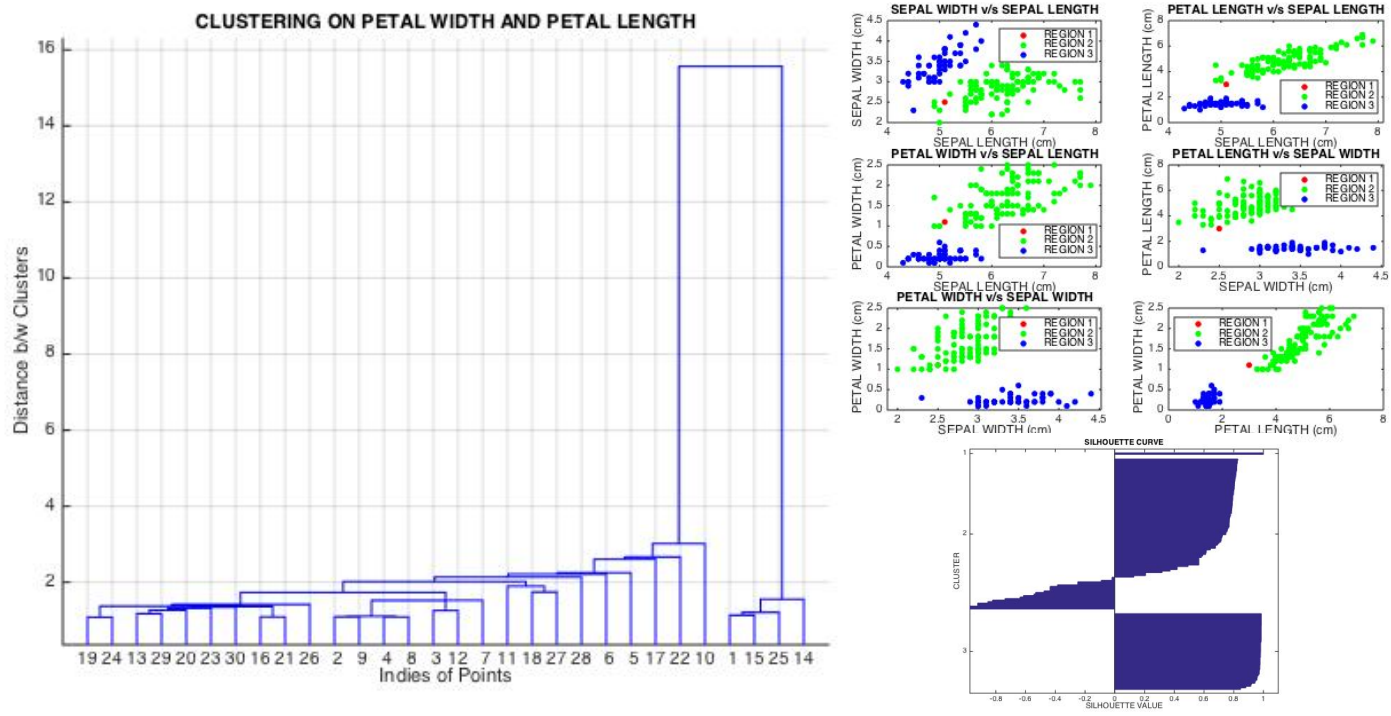
SEPAL WIDTH V/S PETAL WIDTH:

Figure I.54-Accuracy 66.67%, Silhouette Value 0.6777, Testing Time 0.02518s



PETAL LENGTH V/S PETAL WIDTH:

Figure I.55-Accuracy 67.33%, Silhouette Value 0.6221, Testing Time 0.008337s



Results:

The following section, shows the analysis results in more detail. All the graphs given, specify the time and accuracy of each classifier, as that done in the report. Additionally, classifiers here also show an attributional comparison. A comparison of attributes within the each classifier has been. In classification, bar graphs of accuracy and time have been represented along with an accuracy v/s time graph for all 3 attribute and 2 attribute classifiers.

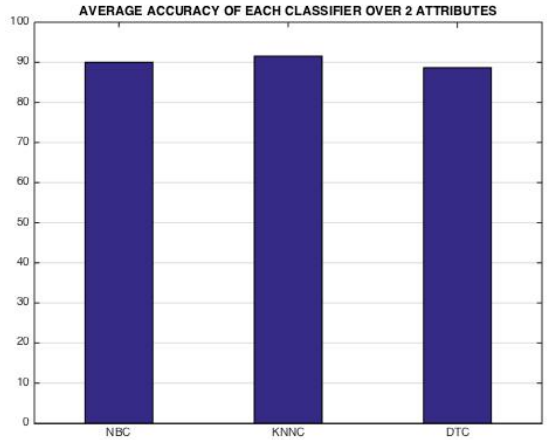
In clustering results, a bar graph of purity, silhouette value and time have been displayed, followed by a purity v/s time and a purity v/s silhouette graph of each classifiers.

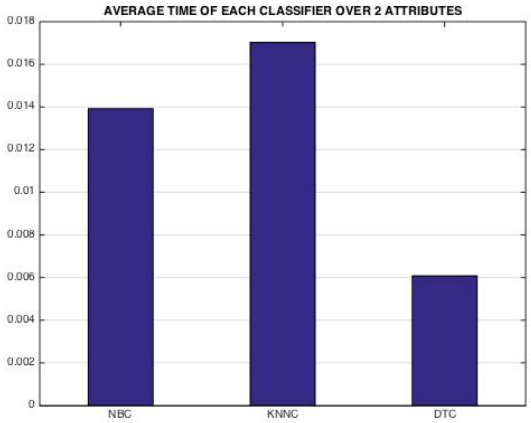
These graphs aid better understanding and an easier analysis of models within each algorithm.

*Note: All figures and tables in this section are self generated.

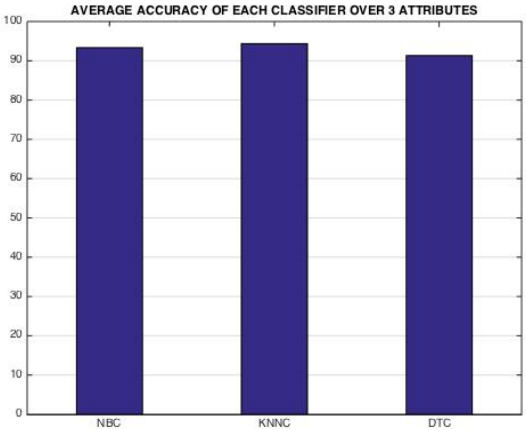
Classification:

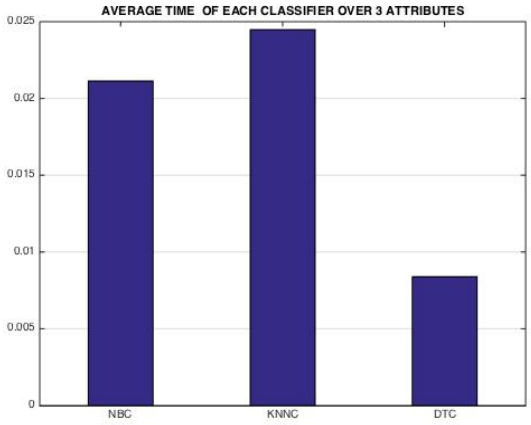
AVERAGE ACCURACY AND TIME FOR 2 ATTRIBUTE CLASSIFIERS:



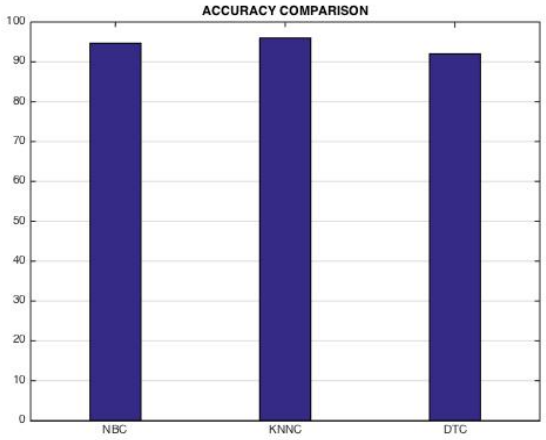


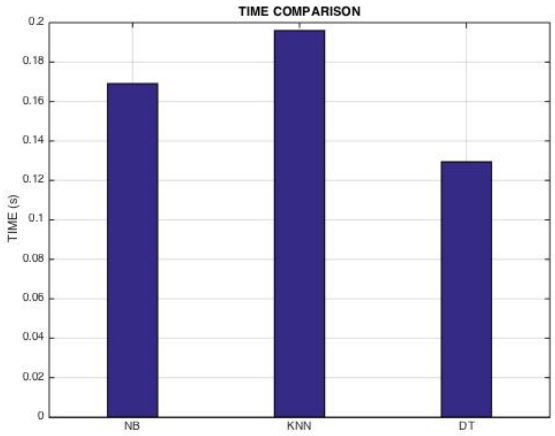
AVERAGE ACCURACY AND TIME FOR 3 ATTRIBUTE CLASSIFIERS:

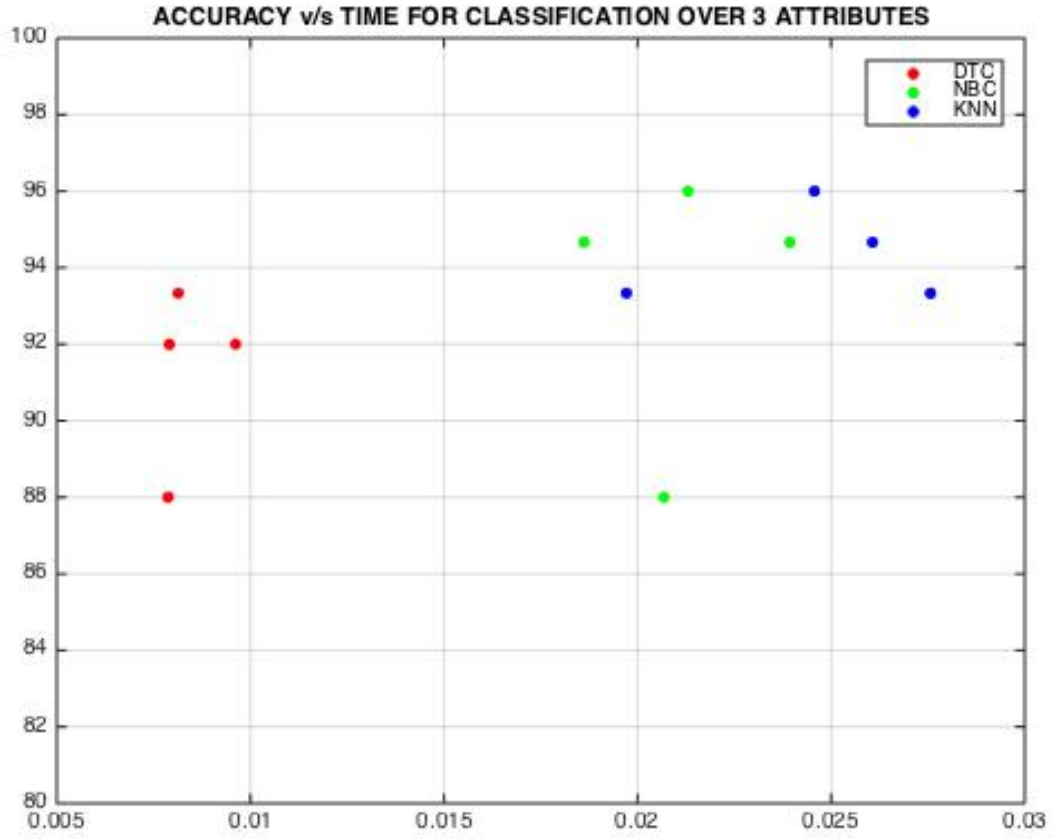


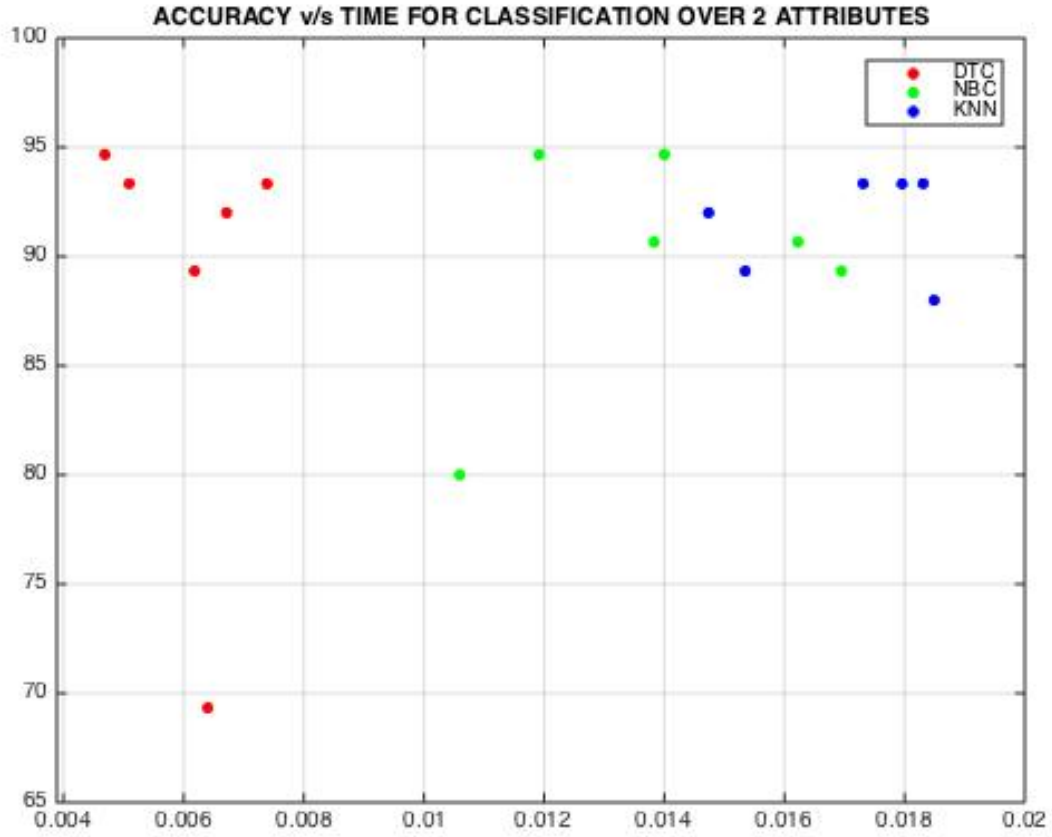


AVERAGE ACCURACY AND TIME FOR 2 ATTRIBUTE CLASSIFIERS:



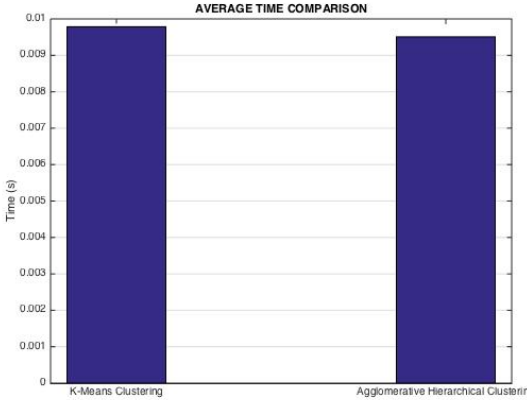


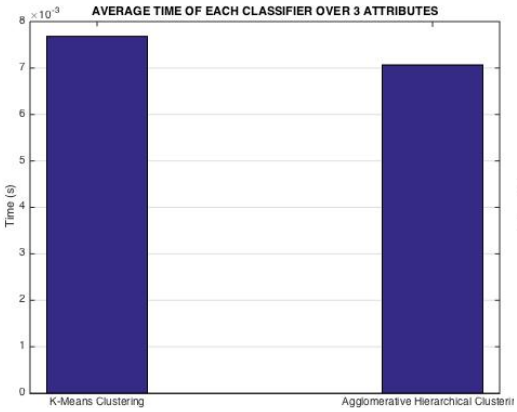

ACCURACY v/s TIME FOR CLASSIFICATION OVER 3 ATTRIBUTES
DTC
NBC
KNN
100
98
96
94
92
90
88
86
84
82
80
0.005
0.01
0.015
0.02
0.025
0.03

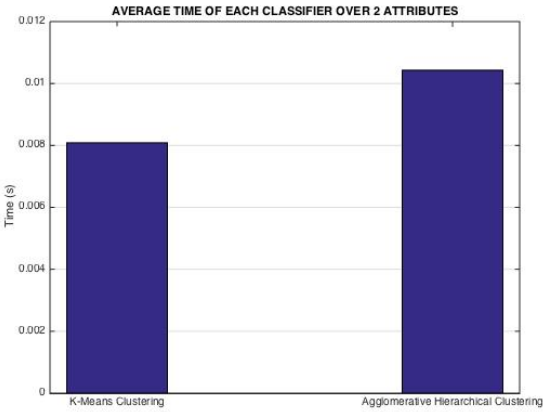

ACCURACY v/s TIME FOR CLASSIFICATION OVER 2 ATTRIBUTES
DTC
NBC
KNN
100
95
90
85
80
75
70
65
0.004
0.006
0.008
0.01
0.012
0.014
0.016
0.018
0.02

Clustering:

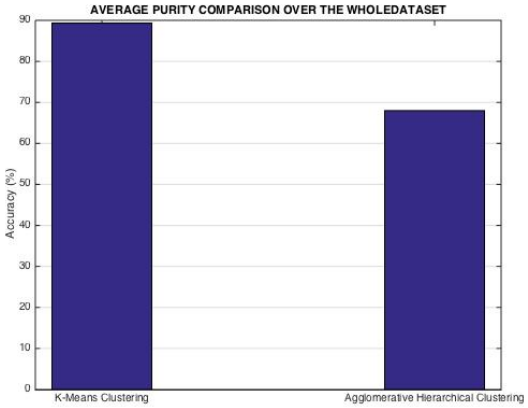
TIME COMPARISON:

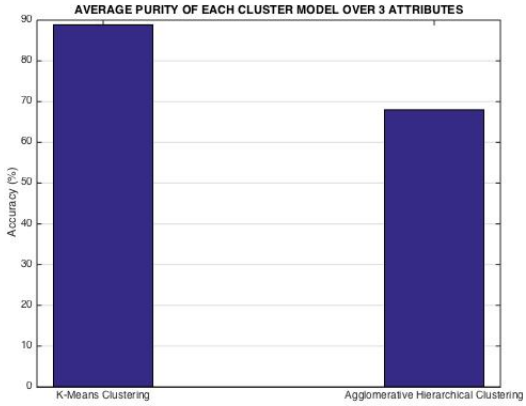


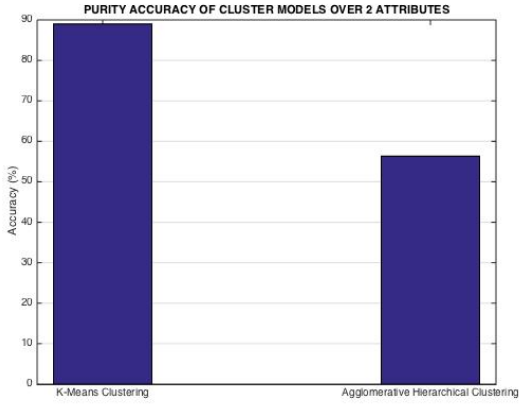




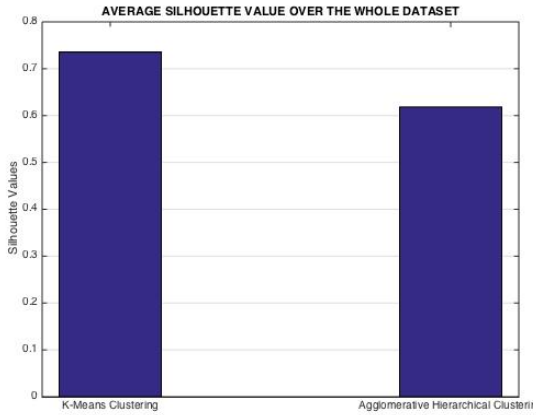
PURITY COMPARISON:

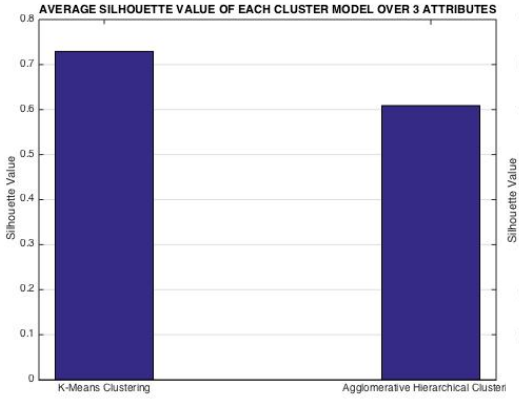


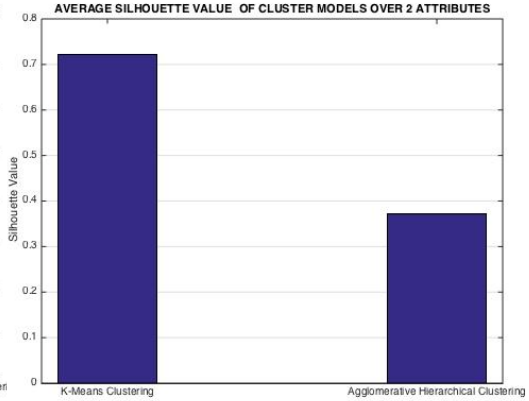


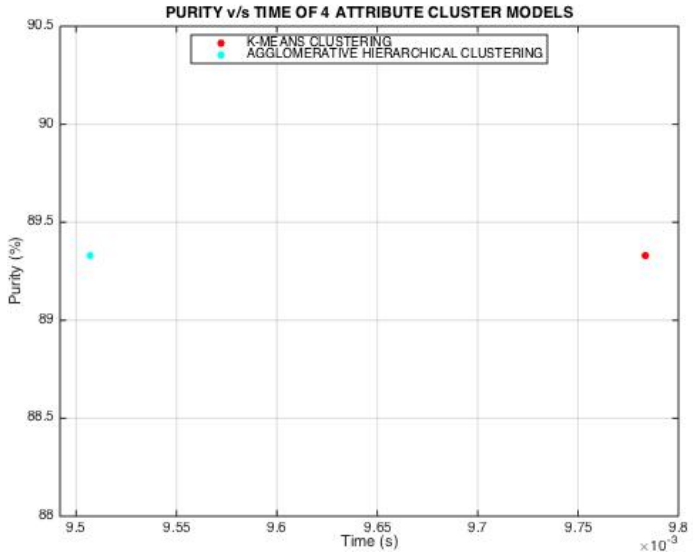


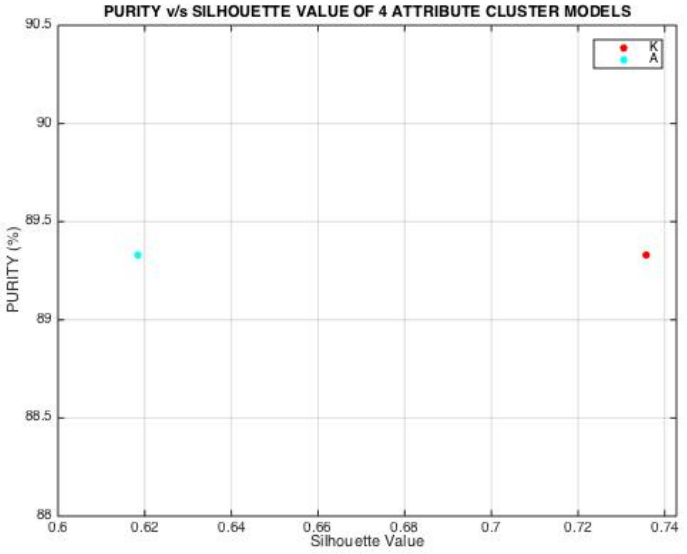
SILHOUETTE VALUE COMPARISON:

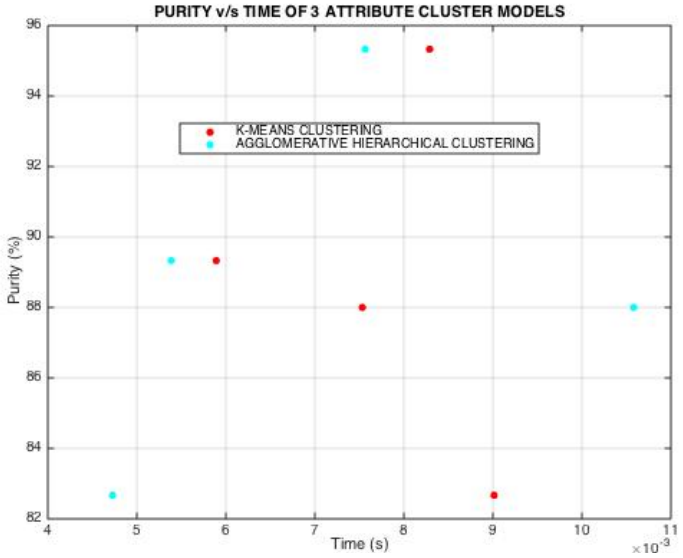


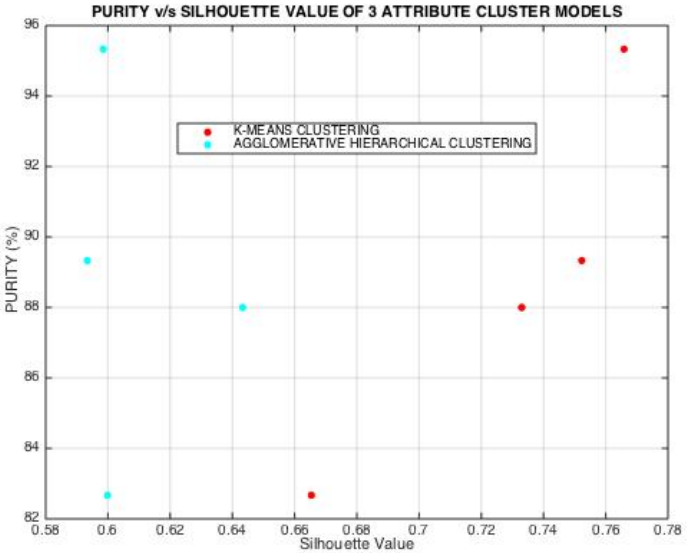


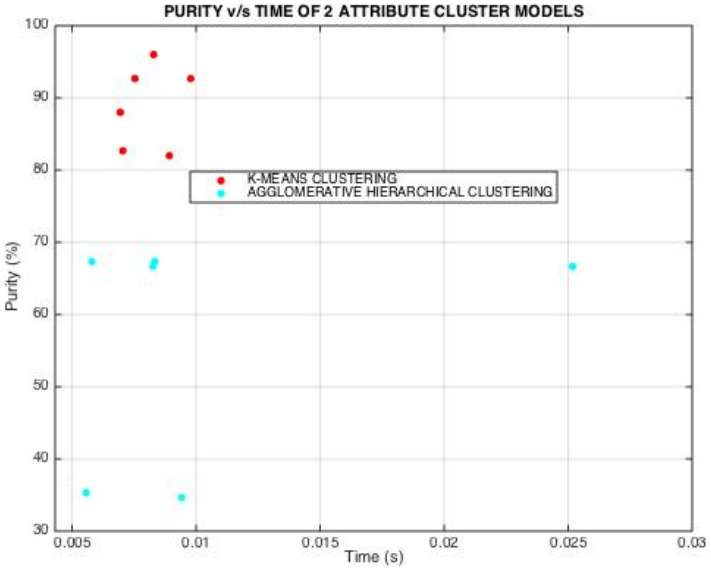


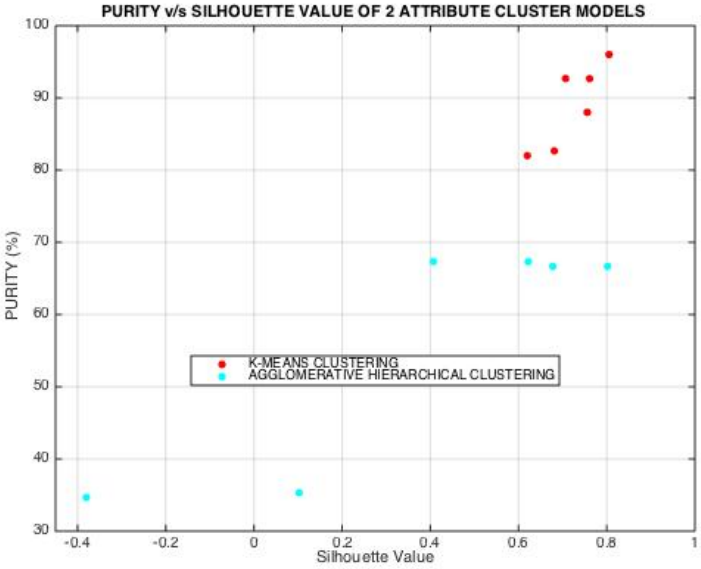

PURITY v/s TIME OF 4 ATTRIBUTE CLUSTER MODELS
K-MEANS CLUSTERING
AGGLOMERATIVE HIERARCHICAL CLUSTERING
Purity (%)
Time (s)
×10^-3


PURITY v/s SILHOUETTE VALUE OF 4 ATTRIBUTE CLUSTER MODELS
K
A
PURITY (%)
Silhouette Value


PURITY v/s TIME OF 3 ATTRIBUTE CLUSTER MODELS
K-MEANS CLUSTERING
AGGLOMERATIVE HIERARCHICAL CLUSTERING
Purity (%)
Time (s)
×10^-3


PURITY v/s SILHOUETTE VALUE OF 3 ATTRIBUTE CLUSTER MODELS
K-MEANS CLUSTERING
AGGLOMERATIVE HIERARCHICAL CLUSTERING
PURITY (%)
Silhouette Value


PURITY v/s TIME OF 2 ATTRIBUTE CLUSTER MODELS
K-MEANS CLUSTERING
AGGLOMERATIVE HIERARCHICAL CLUSTERING
Purity (%)
Time (s)


PURITY v/s SILHOUETTE VALUE OF 2 ATTRIBUTE CLUSTER MODELS
K-MEANS CLUSTERING
AGGLOMERATIVE HIERARCHICAL CLUSTERING
PURITY (%)
Silhouette Value

Appendix II

CODE:

DATASETS:

Code for partitioning Iris dataset into training and testing datasets.

```
load fisheriris;
setindices=rand(35,1)*50;
versiindices=rand(35,1)*100;
virginindices=rand(35,1)*150;
for i=1:35
 if  setindices(i) < 1
     setindices(i)=setindices(i)+1;
 end
if versiindices(i) < 50
    versiindices(i)=versiindices(i)+50;
end
if virginindices(i) < 50
   virginindices(i)=virginindices(i)+100;
end
if virginindices(i) < 100
   virginindices(i)=virginindices(i)+50;
end
end
setindices=floor(setindices);
versiindices=floor(versiindices);
virginindices=floor(virginindices);
trainset=[setindices;versiindices;virginindices];
trainset=unique(trainset);

for i=1:length(trainset)
traindata(i,:)=meas(trainset(i),:);
trainclass(i,:)=species(trainset(i),:);
meas(trainset(i),1)=0;
%species(trainset(i),1)={'none'};
end
k=1;
for i=1:length(meas)
    if meas(i,1) ~= 0
        testdata(k,:)=meas(i,:);
        testclass(k,:)=species(i,:);
        k=k+1;
    end
end
```

NAIVE BAYE'S ALGORITHM:

Code for implementation of naive Baye's algorithm over 4 attributes

```
%CODE FOR NAIVE BAYES ALGORITHM:
datafile;
%CLASSIFICATION OVER THE WHOLEDATASET
nb=fitcnb(traindata,trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass=predict(nb,testdata);
toc
k=1;
flag=0;
% code for getting indices of misclassified points.
for i=1:length(predclass)
if strcmp(predclass(i),testclass(i)) == 0
   diff(k,1)=i;
   k=k+1;
   flag=1;
end
end
%code for displaying the loss of the classifier.
nbloss=(size(diff)./size(testclass)).*100;
nbloss=nbloss(1,1);
str1='The loss percentage is: ';
str0=num2str(nbloss);
str2=' %';
str3=strcat(str0,str2);
loss=strcat(str1,str3);
display(loss);
%graph on attribute 1 and 2
figure('Name','NAIVE BAYE''S CLASSIFICATION OVER THE WHOLEDATASET');
subplot(6,2,1)
gscatter(testdata(:,1),testdata(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')

subplot(6,2,2)
gscatter(testdata(:,1),testdata(:,2),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')
%graph on attribute 1 and 3
subplot(6,2,3)
gscatter(testdata(:,1),testdata(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')

subplot(6,2,4)
gscatter(testdata(:,1),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')
%graph on attribute 1 and 4
subplot(6,2,5)
gscatter(testdata(:,1),testdata(:,4),testclass);
```

```
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')

subplot(6,2,6)
gscatter(testdata(:,1),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')
%graph on attribute 2 and 3
subplot(6,2,7)
gscatter(testdata(:,2),testdata(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)')
ylabel('PETAL LENGTH (cm)')

subplot(6,2,8)
gscatter(testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph on attribute 2 and 4
subplot(6,2,9)
gscatter(testdata(:,2),testdata(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,10)
gscatter(testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph on attribute 3 and 4
subplot(6,2,11);
gscatter(testdata(:,3),testdata(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,12);
gscatter(testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%Graph BASED ON attribute 1, 2&3
figure('Name','NAIVE BAYE''S CLASSIFICATION OVER THE WHOLEDATASET');
subplot(4,2,1)
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
```

```
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,2);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON 2, 3 & 4
subplot(4,2,3);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,4);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),predclass);
line(testdata(diff,2),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION';

%graph BASED ON 1, 3 & 4
subplot(4,2,5);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,6);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON 1, 2 & 4
subplot(4,2,7);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,8);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION'
```

```
%CONFUSION MATRIX:
conf=confusionmat(testclass,predclass);
display(conf);
```

Code for implementation of naive Baye's algorithm over 3 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3
nb1=fitcnb(traindata(:,[1 2 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass1=predict(nb1,testdata(:,[1 2 3]));
toc
%calculating the misclassified indices
k=1;
flag=0;
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff1(k,1)=i;
   k=k+1;
   flag=1;
end
end
%calculating loss
nbloss=(size(diff1)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(nbloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);
att1=testdata;
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3');
subplot(4,2,1)
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),testclass);
if flag==1
line(att1(diff1,1),att1(diff1,2),att1(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att1(:,2),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),testclass);
if flag==1
```

```
line(testdata(diff1,2),testdata(diff1,3),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),testclass);
if flag==1
line(att1(diff1,1),att1(diff1,3),att1(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),testclass);
flag==1
line(att1(diff1,1),att1(diff1,2),att1(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
display(conf1);

% CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4
att2=testdata;
nb2=fitcnb(traindata(:,[2 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(nb2,testdata(:,[2 3 4]));
toc
%calculating indices of misclassified points
j=1;
flag=0;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(j,1)=i;
   j=j+1;
flag=1;
```

```
end
end
%calculating loss
nbloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(nbloss(1,1));
str22=' %';
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);
%display(diff2);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),predclass2);
if flag==1
line(att2(diff2,2),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),predclass2);
```

```
if flag==1
line(att2(diff2,1),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

% CONFUSION MATRIX:
conf2=confusionmat(testclass,predclass2);
display(conf2);

%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4
att3=testdata;
nb3=fitcnb(traindata(:,[1 2 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(nb3,att3(:,[1 2 4]));
toc
%calculating misclassified indices
f=1;
flag=0;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
   diff3(f,1)=i;
   f=f+1;
flag=1;
end
end
%calculating loss
nbloss=(size(diff3)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(nbloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);
%display(diff3);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4');
subplot(4,2,1)
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
```

```
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,2),att3(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,2),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH - ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
gscatter3(att3(:,1),att3(:,2),att3(:,4),predclass3);
```

```
if flag==1
line(att3(diff3,1),att3(diff3,2),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
grid on;
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);

%CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4
att4=testdata;
nb4=fitcnb(traindata(:,[1 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(nb4,att4(:,[1 3 4]));
toc
%calculating misclassified indices
f=1;
flag=0;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(f,1)=i;
   f=f+1;
flag=1;
end
end
%calculating loss
nbloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(nbloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);
%display(diff4);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
```

```
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,2),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf4=confusionmat(testclass,predclass4);
display(conf4);
```

Code for implementation of naive Baye's algorithm over 2 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTE 1 & 2
att1=testdata;
nb1=fitcnb(traindata(:,[1 2]),trainclass,'classNames',{'setosa','versicolor','virginica'});
```

```
tic
predclass1=predict(nb1,testdata(:,[1 2]));
toc
%calculating misclassified indices
flag=0;
k=1;
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff1(k,1)=i;
   flag=1;
   k=k+1;
end
end
%calculating loss
nbloss=(size(diff1)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(nbloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE''S CLASSIFICATION BASED ON ATTRIBUTE 1 & 2');
subplot(6,2,1)
gscatter(att1(:,1),att1(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2)
gscatter(att1(:,1),att1(:,2),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3)
gscatter(att1(:,1),att1(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4)
gscatter(att1(:,1),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att1(:,1),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att1(:,1),att1(:,4),predclass1);
line(testdata(diff1,1),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none')
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att1(:,2),att1(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
```

```
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att1(:,2),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att1(:,2),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att1(:,2),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att1(:,3),att1(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att1(:,3),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,3),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
display(conf1);

% CLASSIFICATION BASED ON ATTRIBUTE 1 & 3
att2=testdata;
nb2=fitcnb(traindata(:,[1 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(nb2,testdata(:,[1 3]));
toc
%calculating misclassified indices
p=1;
flag=0;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(p,1)=i;
   p=p+1;
   flag=1;
end
end
%calculating loss percentage
nbloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(nbloss(1,1));
str22=' %';
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);
```

```
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE''S CLASSIFICATION BASED ON ATTRIBUTE 1 & 3');
subplot(6,2,1);
gscatter(att2(:,1),att2(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att2(:,1),att2(:,2),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att2(:,1),att2(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,4);
gscatter(att2(:,1),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att2(:,1),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att2(:,1),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att2(:,2),att2(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att2(:,2),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,2),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att2(:,2),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att2(:,2),att2(:,4),predclass2);
if flag==1
```

```
line(testdata(diff2,2),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att2(:,3),att2(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att2(:,3),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,3),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%CONFUSION MATRIX:
conf2=confusionmat(testclass,predclass2);
display(conf2);

% CLASSIFICATION BASED ON ATTRIBUTE 1 & 4
att3=testdata;
nb3=fitcnb(traindata(:,[1 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(nb3,testdata(:,[1 4]));
toc
%calculating misclassified indices
l=1;
flag=0;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
   diff3(l,1)=i;
   l=l+1;
   flag=1;
end
end
%claculating loss
nbloss=(size(diff3)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(nbloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE"S CLASSIFICATION BASED ON ATTRIBUTE 1 & 4');
subplot(6,2,1);
gscatter(att3(:,1),att3(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att3(:,1),att3(:,2),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att3(:,1),att3(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
```

```
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att3(:,1),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graoh wrt attribute 1,4
subplot(6,2,5);
gscatter(att3(:,1),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att3(:,1),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att3(:,2),att3(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att3(:,2),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt 2,4
subplot(6,2,9);
gscatter(att3(:,2),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att3(:,2),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt 3,4
subplot(6,2,11);
gscatter(att3(:,3),att3(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att3(:,3),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,3),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
```

```
% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 3
att4=testdata;
nb4=fitcnb(traindata(:,[2 3]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(nb4,testdata(:,[2 3]));
toc
%calculating misclassified indices
m=1;
flag=0;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(m,1)=i;
   m=m+1;
   flag=1;
end
end
%calculating loss
nbloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(nbloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE''S CLASSIFICATION BASED ON ATTRIBUTE 2 & 3')
subplot(6,2,1);
gscatter(att4(:,1),att4(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att4(:,1),att4(:,2),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att4(:,1),att4(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att4(:,1),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt to attribute 1,4
subplot(6,2,5);
gscatter(att4(:,1),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att4(:,1),att4(:,4),predclass4);
```

```
if flag==1
line(testdata(diff4,1),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt to attribute 2,3
subplot(6,2,7);
gscatter(att4(:,2),att4(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att4(:,2),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att4(:,2),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att4(:,2),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute attribute 3,4
subplot(6,2,11);
gscatter(att4(:,3),att4(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att4(:,3),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,3),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

% %CONFUSION MATRIX:
conf4=confusionmat(testclass,predclass4);
display(conf4);

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 4
att5=testdata;
nb5=fitcnb(traindata(:,[2 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass5=predict(nb5,testdata(:,[2 4]));
toc
%calculating misclassified indices
n=1;
flag=0;
for i=1:length(predclass5)
if strcmp(predclass5(i),testclass(i)) == 0
   diff5(n,1)=i;
   n=n+1;
```

```
flag=1;
end
end
%calculating loss percentage
nbloss=(size(diff5)./size(testclass)).*100;
str15='The loss percentage is: ';
str05=num2str(nbloss(1,1));
str25=' %';
str35=strcat(str05,str25);
loss=strcat(str15,str35);
display(loss);
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE"S CLASSIFICATION BASED ON ATTRIBUTE 2 & 4');
subplot(6,2,1);
gscatter(att5(:,1),att5(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att5(:,1),att5(:,2),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att5(:,1),att5(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att5(:,1),att5(:,3),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att5(:,1),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att5(:,1),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att5(:,2),att5(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att5(:,2),att5(:,3),predclass5);
if flag==1
line(testdata(diff5,2),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
```

```
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att5(:,2),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att5(:,2),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,2),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att5(:,3),att5(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att5(:,3),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,3),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
%
conf5=confusionmat(testclass,predclass5);
display(conf5);

%CLASSIFICATION BASED ON ATTRIBUTE 3 & 4
att6=testdata;
nb6=fitcnb(traindata(:,[3 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass6=predict(nb6,testdata(:,[3 4]));
toc
%calculating miscalssified indices
n=1;
flag=0;
for i=1:length(predclass6)
if strcmp(predclass6(i),testclass(i)) == 0
   diff6(n,1)=i;
   n=n+1;
flag=1;
end
end
%calculating loss percentage
nbloss=(size(diff6)./size(testclass)).*100;
str16='The loss percentage is: ';
str06=num2str(nbloss(1,1));
str26=' %';
str36=strcat(str06,str26);
loss=strcat(str16,str36);
display(loss);
%graph wrt attribute 1,2
figure('Name','NAIVE BAYE"S CLASSIFICATION BASED ON ATTRIBUTE 3 & 4');
subplot(6,2,1);
gscatter(att6(:,1),att6(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
```

```
subplot(6,2,2);
gscatter(att6(:,1),att6(:,2),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att6(:,1),att6(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att6(:,1),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att6(:,1),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att6(:,1),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att6(:,2),att6(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att6(:,2),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att6(:,2),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att6(:,2),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
```

```
gscatter(att6(:,3),att6(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att6(:,3),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,3),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

%CONFUSION MATRIX:
conf6=confusionmat(testclass,predclass6);
display(conf6);
```

KNN ALGORITHM:

Code for KNN Classification over 4 attributes:

```
%CODE FOR KNN MATRIX AND PRODUCING GRAPH:
datafile;
%CLASSIFICATION OVER THE WHOLEDATASET
knn=fitcknn(traindata,trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass=predict(knn,testdata);
toc
%code for getting the indices of misclassified points
k=1;
flag=0;
for i=1:length(predclass)
if strcmp(predclass(i),testclass(i)) == 0
   diff(k,1)=i;
   k=k+1;
   flag=1;
end
end
% code for calculating accuracy
knnloss=(size(diff)./size(testclass)).*100;
knnloss=knnloss(1,1);
str1='The loss percentage is: ';
str0=num2str(knnloss);
str2=' %';
str3=strcat(str0,str2);
loss=strcat(str1,str3);
display(loss);
%graph based on attribute 1, 2
figure('Name','KNN CLASSIFICATION OVER THE WHOLEDATASET');
subplot(6,2,1)
gscatter(testdata(:,1),testdata(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')

subplot(6,2,2)
gscatter(testdata(:,1),testdata(:,2),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')
%graph based on attribute 1, 3
subplot(6,2,3)
gscatter(testdata(:,1),testdata(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')

subplot(6,2,4)
gscatter(testdata(:,1),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')
%graph based on attribute 1, 4
subplot(6,2,5)
gscatter(testdata(:,1),testdata(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
```

```
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')

subplot(6,2,6)
gscatter(testdata(:,1),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')
%graph based on attribute 2, 3
subplot(6,2,7)
gscatter(testdata(:,2),testdata(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)')
ylabel('PETAL LENGTH (cm)')

subplot(6,2,8)
gscatter(testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph based on attribute 2, 4
subplot(6,2,9)
gscatter(testdata(:,2),testdata(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,10)
gscatter(testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph based on attribute 3, 4
subplot(6,2,11);
gscatter(testdata(:,3),testdata(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,12);
gscatter(testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%Graph BASED ON attribute 1, 2&3
figure('Name','KNN CLASSIFICATION OVER THE WHOLEDATASET');
subplot(4,2,1)
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
```

```
subplot(4,2,2);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 2, 3 & 4
subplot(4,2,3);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,4);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 1, 3 & 4
subplot(4,2,5);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,6);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 1, 2 & 4
subplot(4,2,7);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,8);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
```

```
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION'

% CONFUSION MATRIX:
conf=confusionmat(testclass,predclass);
display(conf);
```

Code for KNN Classification over 3 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3
knn1=fitcknn(traindata(:,[1 2 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass1=predict(knn1,testdata(:,[1 2 3]));
toc
%calculating indices of misclassified points
k=1;
flag=0;
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff(k,1)=i;
   k=k+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(knnloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);
att1=testdata;
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3');
subplot(4,2,1)
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,2);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),predclass1);
if flag==1
line(att1(diff,1),att1(diff,2),att1(diff,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att1(:,2),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,4);
grid on;
gscatter3(att1(:,2),att1(:,3),att1(:,4),predclass1);
if flag==1
```

```
line(att1(diff,2),att1(diff,3),att1(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,6);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),predclass1);
if flag==1
line(att1(diff,1),att1(diff,3),att1(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),predclass1);
if flag==1
line(att1(diff,1),att1(diff,2),att1(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
display(conf1);

% CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4
att2=testdata;
knn2=fitcknn(traindata(:,[2 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(knn2,testdata(:,[2 3 4]));
toc
%calculating miscalssified indices
j=1;
flag=0;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(j,1)=i;
   j=j+1;
   flag=1;
```

```
end
end
%calculating loss
knnloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(knnloss(1,1));
str22=' %';
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),predclass2);
if flag==1
line(att2(diff2,2),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),predclass2);
if flag==1
```

```
line(att2(diff2,1),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

% CONFUSION MATRIX:
conf2=confusionmat(testclass,predclass2);
display(conf2);


%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4
att3=testdata;
knn3=fitcknn(traindata(:,[1 2 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(knn3,att3(:,[1 2 4]));
toc
%calculating misclassified points
f=1;
flag=0;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
   diff3(f,1)=i;
   f=f+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff3)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(knnloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);

%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4');
subplot(4,2,1)
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
```

```
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,2);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,2),att3(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,2),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH - ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
gscatter3(att3(:,1),att3(:,2),att3(:,4),predclass3);
if flag==1
```

```
line(att3(diff3,1),att3(diff3,2),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
grid on;
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);


%CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4
att4=testdata;
knn4=fitcknn(traindata(:,[1 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(knn4,att4(:,[1 3 4]));
toc
%calculating indices of misclassified points
f=1;
flag=0;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(f,1)=i;
   f=f+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(knnloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);

%display(diff4);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),testclass);
```

```
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,2),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf4=confusionmat(testclass,predclass4);
display(conf4);
```

Code for KNN Classification over 2 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTE 1 & 2
att1=testdata;
```

```
knn1=fitcknn(traindata(:,[1 2]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass1=predict(knn1,testdata(:,[1 2]));
toc
%calculating indices of misclassified points
k=1;
flag=0;
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff1(k,1)=i;
   k=k+1;
   flag=1;
end
end
% loss percentage
knnloss=(size(diff1)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(knnloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);
%graph wrt attribute 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 1 & 2');
subplot(6,2,1)
gscatter(att1(:,1),att1(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2)
gscatter(att1(:,1),att1(:,2),predclass1);
 if flag==1
line(testdata(diff1,1),testdata(diff1,2),'marker','o','markersize',10,'linestyle','none');
 end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3)
gscatter(att1(:,1),att1(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4)
gscatter(att1(:,1),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att1(:,1),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att1(:,1),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none')
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
```

```
subplot(6,2,7);
gscatter(att1(:,2),att1(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att1(:,2),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att1(:,2),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att1(:,2),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att1(:,3),att1(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12)
gscatter(att1(:,3),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,3),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
display(conf1);

%CLASSIFICATION BASED ON ATTRIBUTE 1 & 3
att2=testdata;
knn2=fitcknn(traindata(:,[1 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(knn2,testdata(:,[1 3]));
toc
%calculating miscalssified indices
flag=0;
p=1;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(p,1)=i;
   p=p+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(knnloss(1,1));
str22=' %';
```

```
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);
%graph wrt att 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 1 & 3');
subplot(6,2,1);
gscatter(att2(:,1),att2(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att2(:,1),att2(:,2),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att2(:,1),att2(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,4);
gscatter(att2(:,1),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att2(:,1),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att2(:,1),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att2(:,2),att2(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att2(:,2),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,2),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att2(:,2),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
```

```
subplot(6,2,10);
gscatter(att2(:,2),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,2),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att2(:,3),att2(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att2(:,3),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,3),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%CONFUSION MATRIX:
 conf2=confusionmat(testclass,predclass2);
 display(conf2);

%CLASSIFICATION BASED ON ATTRIBUTE 1 & 4
att3=testdata;
knn3=fitcknn(traindata(:,[1 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(knn3,testdata(:,[1 4]));
toc
%calculating misclassified points
l=1;
flag=0;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
   diff3(l,1)=i;
   l=l+1;
   flag=1;
end
end
%calcualating loss
knnloss=(size(diff3)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(knnloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);
%graph wrt attribute 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 1 & 4');
subplot(6,2,1);
gscatter(att3(:,1),att3(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att3(:,1),att3(:,2),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
```

```
subplot(6,2,3);
gscatter(att3(:,1),att3(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att3(:,1),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att3(:,1),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att3(:,1),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att3(:,2),att3(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att3(:,2),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att3(:,2),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att3(:,2),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att3(:,3),att3(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att3(:,3),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,3),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
```

```
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 3
att4=testdata;
knn4=fitcknn(traindata(:,[2 3]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(knn4,testdata(:,[2 3]));
toc
%calculating misclassified points
m=1;
flag=0;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(m,1)=i;
   m=m+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(knnloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);
%graph wrt attribute 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 2 & 3')
subplot(6,2,1);
gscatter(att4(:,1),att4(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att4(:,1),att4(:,2),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att4(:,1),att4(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att4(:,1),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att4(:,1),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
```

```
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att4(:,1),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att4(:,2),att4(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att4(:,2),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att4(:,2),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att4(:,2),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att4(:,3),att4(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att4(:,3),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,3),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

% %CONFUSION MATRIX:
 conf4=confusionmat(testclass,predclass4);
 display(conf4);

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 4
att5=testdata;
knn5=fitcknn(traindata(:,[2 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass5=predict(knn5,testdata(:,[2 4]));
toc
%calculating misclassified points
n=1;
flag=0;
for i=1:length(predclass5)
```

```
if strcmp(predclass5(i),testclass(i)) == 0
   diff5(n,1)=i;
   n=n+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff5)./size(testclass)).*100;
str15='The loss percentage is: ';
str05=num2str(knnloss(1,1));
str25=' %';
str35=strcat(str05,str25);
loss=strcat(str15,str35);
display(loss);
%graph wrt attribute 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 2 & 4');
subplot(6,2,1);
gscatter(att5(:,1),att5(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att5(:,1),att5(:,2),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att5(:,1),att5(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att5(:,1),att5(:,3),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att5(:,1),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att5(:,1),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att5(:,2),att5(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att5(:,2),att5(:,3),predclass5);
```

```
if flag==1
line(testdata(diff5,2),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att5(:,2),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att5(:,2),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,2),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att5(:,3),att5(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att5(:,3),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,3),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
 conf5=confusionmat(testclass,predclass5);
display(conf5);

%CLASSIFICATION BASED ON ATTRIBUTE 3 & 4
att6=testdata;
knn6=fitcknn(traindata(:,[3 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass6=predict(knn6,testdata(:,[3 4]));
toc
%calculating indices of misclassified points
n=1;
flag=0;
for i=1:length(predclass6)
if strcmp(predclass6(i),testclass(i)) == 0
   diff6(n,1)=i;
   n=n+1;
   flag=1;
end
end
%calculating loss
knnloss=(size(diff6)./size(testclass)).*100;
str16='The loss percentage is: ';
str06=num2str(knnloss(1,1));
str26=' %';
str36=strcat(str06,str26);
loss=strcat(str16,str36);
display(loss);
%graph wrt attribute 1,2
figure('Name','KNN CLASSIFICATION BASED ON ATTRIBUTE 3 & 4');
subplot(6,2,1);
gscatter(att6(:,1),att6(:,2),testclass);
```

```
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att6(:,1),att6(:,2),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('SEPAL WIDTH (cm)');
%graph wrt attribute 1,3
subplot(6,2,3);
gscatter(att6(:,1),att6(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att6(:,1),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 1,4
subplot(6,2,5);
gscatter(att6(:,1),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att6(:,1),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 2,3
subplot(6,2,7);
gscatter(att6(:,2),att6(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att6(:,2),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL LENGTH (cm)');
%graph wrt attribute 2,4
subplot(6,2,9);
gscatter(att6(:,2),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att6(:,2),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('SEPAL WIDTH (cm)','fontsize',10);
```

```
ylabel('PETAL WIDTH (cm)');
%graph wrt attribute 3,4
subplot(6,2,11);
gscatter(att6(:,3),att6(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att6(:,3),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,3),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION','fontsize',10);
xlabel('PETAL LENGTH (cm)','fontsize',10);
ylabel('PETAL WIDTH (cm)');

%CONFUSION MATRIX:
conf6=confusionmat(testclass,predclass6);
display(conf6);
```

DECISION TREE CLASSIFICATION:

Code for decision tree classification over 4 attributes:

```
%CLASSIFICATION USING DECISION TREES
datafile;
%CLASSIFICATION OVER THE WHOLEDATASET
dt=fitctree(traindata,trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass=predict(dt,testdata);
toc
%code for calculating indices od misclassified points
k=1;
flag=0;
for i=1:length(predclass)
if strcmp(predclass(i),testclass(i)) == 0
   diff(k,1)=i;
   k=k+1;
   flag=1;
end
end
%code for calculating accuracy
dtloss=(size(diff)./size(testclass)).*100;
dtloss=dtloss(1,1);
str1='The loss percentage is: ';
str0=num2str(dtloss);
str2=' %';
str3=strcat(str0,str2);
loss=strcat(str1,str3);
display(loss);
%graph wrt attributes 1,2
figure('Name','DECISON TREE CLASSIFICATION OVER THE WHOLEDATASET');
subplot(6,2,1)
gscatter(testdata(:,1),testdata(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')

subplot(6,2,2)
gscatter(testdata(:,1),testdata(:,2),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('SEPAL WIDTH (cm)')
%graph wrt attributes 1,3
subplot(6,2,3)
gscatter(testdata(:,1),testdata(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')

subplot(6,2,4)
gscatter(testdata(:,1),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL LENGT (cm)')
%graph wrt attributes 1,4
subplot(6,2,5)
gscatter(testdata(:,1),testdata(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
```

```
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')

subplot(6,2,6)
gscatter(testdata(:,1),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)')
ylabel('PETAL WIDTH (cm)')
%graph wrt attributes 2,3
subplot(6,2,7)
gscatter(testdata(:,2),testdata(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)')
ylabel('PETAL LENGTH (cm)')

subplot(6,2,8)
gscatter(testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph wrt attributes 2,4
subplot(6,2,9)
gscatter(testdata(:,2),testdata(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,10)
gscatter(testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph wrt attributes 3,4
subplot(6,2,11);
gscatter(testdata(:,3),testdata(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

subplot(6,2,12);
gscatter(testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%Graph BASED ON attribute 1, 2&3
figure('Name','DECISION TREE CLASSIFICATION OVER THE WHOLEDATASET');
subplot(4,2,1)
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
```

```
subplot(4,2,2);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,3),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 2, 3 & 4
subplot(4,2,3);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,4);
grid on;
gscatter3(testdata(:,2),testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,2),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 1, 3 & 4
subplot(4,2,5);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,6);
grid on;
gscatter3(testdata(:,1),testdata(:,3),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,3),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph BASED ON attribute 1, 2 & 4
subplot(4,2,7);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
subplot(4,2,8);
grid on;
gscatter3(testdata(:,1),testdata(:,2),testdata(:,4),predclass);
if flag==1
line(testdata(diff,1),testdata(diff,2),testdata(diff,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
```

```
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION'

% CONFUSION MATRIX:
conf=confusionmat(testclass,predclass);
display(conf);
%VISUAL TREE
view(dt);
%CLASSIFICATION TREE
view(dt,'mode','graph');
```

Code for decision tree classification over 3 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3
dt1=fitctree(traindata(:,[1 2 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass1=predict(dt1,testdata(:,[1 2 3]));
toc
%calculating indices of misclassified points
flag=0;
k=1;
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff1(k,1)=i;
   k=k+1;
   flag=1;
end
end
att1=testdata;
%calculating loss
treeloss=(size(diff1)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(treeloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);

%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 3');
subplot(4,2,1)
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,3),predclass1);
if flag==1
line(att1(diff1,1),att1(diff1,2),att1(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att1(:,2),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
```

```
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att1(:,2),att1(:,3),att1(:,4),predclass1);
if flag==1
line(att1(diff1,2),att1(diff1,3),att1(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att1(:,1),att1(:,3),att1(:,4),predclass1);
if flag==1
line(att1(diff1,1),att1(diff1,3),att1(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att1(:,1),att1(:,2),att1(:,4),predclass1);
if flag==1
line(att1(diff1,1),att1(diff1,2),att1(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
display(conf1);
%VISUAL TREE
view(dt1);
%CLASSIFICATION TREE
view(dt1,'mode','graph');

% CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4
att2=testdata;
```

```
dt2=fitctree(traindata(:,[2 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(dt2,testdata(:,[2 3 4]));
toc
%calculating indices of misclassified points
flag=0;
j=1;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(j,1)=i;
   j=j+1;
   flag=1;
end
end
%calculating loss
treeloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(treeloss(1,1));
str22=' %';
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);

%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 2, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,3),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att2(:,2),att2(:,3),att2(:,4),predclass2);
if flag==1
line(att2(diff2,2),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';
```

```
%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att2(:,1),att2(:,3),att2(:,4),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,3),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att2(:,1),att2(:,2),att2(:,4),predclass2);
if flag==1
line(att2(diff2,1),att2(diff2,2),att2(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

% CONFUSION MATRIX:
conf2=confusionmat(testclass,predclass2);
display(conf2);
%VISUAL TREE
view(dt2);
%CLASSIFICATION TREE
view(dt2,'mode','graph');

%CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4
att3=testdata;
dt3=fitctree(traindata(:,[1 2 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(dt3,att3(:,[1 2 4]));
toc
%calculating indices of misclassified points
flag=0;
f=1;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
   diff3(f,1)=i;
   f=f+1;
   flag=1;
end
end
%calculating loss
```

```
treeloss=(size(diff1)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(treeloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 2 & 4');
subplot(4,2,1)
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,2);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,3),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,2),att3(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att3(:,2),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,2),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH - ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att3(:,1),att3(:,3),att3(:,4),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,3),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
```

```
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att3(:,1),att3(:,2),att3(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
gscatter3(att3(:,1),att3(:,2),att3(:,4),predclass3);
if flag==1
line(att3(diff3,1),att3(diff3,2),att3(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
grid on;
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);
%VISUAL TREE
view(dt3);
%CLASSIFICATION TREE
view(dt3,'mode','graph');

%CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4
att4=testdata;
dt4=fitctree(traindata(:,[1 3 4]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(dt4,att4(:,[1 3 4]));
toc
%calculating indices of misclassified points
flag=0;
f=1;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(f,1)=i;
   f=f+1;
   flag=1;
end
end
%calculating loss
treeloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(treeloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);

%display(diff4);
%graph wrt attribute 1 2 3
figure('Name','CLASSIFICATION BASED ON ATTRIBUTES 1, 3 & 4');
subplot(4,2,1);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';
```

```
subplot(4,2,2);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,3),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';



% GRAPH WRT TO ATTRIBUTE 2,3,4
subplot(4,2,3);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),testclass);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,4);
grid on;
gscatter3(att4(:,2),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,2),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

%graph wrt att 1 3 and 4
subplot(4,2,5);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,6);
grid on;
gscatter3(att4(:,1),att4(:,3),att4(:,4),predclass4);
if flag==1
line(att4(diff4,1),att4(diff4,3),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';


%graph wrt att 1 2 and 4
subplot(4,2,7);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),testclass);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION';

subplot(4,2,8);
grid on;
gscatter3(att4(:,1),att4(:,2),att4(:,4),predclass4);
```

```
if flag==1
line(att4(diff4,1),att4(diff4,2),att4(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION';

% CONFUSION MATRIX:
conf4=confusionmat(testclass,predclass4);
display(conf4);

%VISUAL TREE
view(dt4);
%CLASSIFICATION TREE
view(dt4,'mode','graph');
```

Code for decision tree classification over 2 attributes:

```
datafile;
%CLASSIFICATION BASED ON ATTRIBUTE 1 & 2
att1=testdata;
dt1=fitctree(traindata(:,[1 2]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass1=predict(dt1,testdata(:,[1 2]));
toc
k=1;
flag=0;
%code for calculating misclassified indices
for i=1:length(predclass1)
if strcmp(predclass1(i),testclass(i)) == 0
   diff1(k,1)=i;
   k=k+1;
   flag=1;
end
end
%code for calculating loss
dtloss=(size(diff1)./size(testclass)).*100;
str11='The loss percentage is: ';
str01=num2str(dtloss(1,1));
str21=' %';
str31=strcat(str01,str21);
loss=strcat(str11,str31);
display(loss);
%graph based on attribute 1,2
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 1 & 2');
subplot(6,2,1)
gscatter(att1(:,1),att1(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2)
gscatter(att1(:,1),att1(:,2),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
%graph based on attribute 1,3
subplot(6,2,3)
gscatter(att1(:,1),att1(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
```

```
subplot(6,2,4)
gscatter(att1(:,1),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
%graph based on attribute 1,4
subplot(6,2,5);
gscatter(att1(:,1),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att1(:,1),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,1),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none')
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph based on attribute 2,3
subplot(6,2,7);
gscatter(att1(:,2),att1(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att1(:,2),att1(:,3),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph based on attribute 2,4
subplot(6,2,9);
gscatter(att1(:,2),att1(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att1(:,2),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,2),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph based on attribute 3,4
subplot(6,2,11);
gscatter(att1(:,3),att1(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att1(:,3),att1(:,4),predclass1);
if flag==1
line(testdata(diff1,3),testdata(diff1,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
conf1=confusionmat(testclass,predclass1);
```

```
%VISUAL TREE
view(dt1);
%CLASSIFICATION TREE
view(dt1,'mode','graph');

% CLASSIFICATION BASED ON ATTRIBUTE 1 & 3
att2=testdata;
dt2=fitctree(traindata(:,[1 3]),trainclass,'classNames',{'setosa','versicolor','virginica'});
tic
predclass2=predict(dt2,testdata(:,[1 3]));
toc
%code for calculating misclassified points
p=1;
flag=0;
for i=1:length(predclass2)
if strcmp(predclass2(i),testclass(i)) == 0
   diff2(p,1)=i;
   p=p+1;
   flag=1;
end
end
% calculating loss
dtloss=(size(diff2)./size(testclass)).*100;
str12='The loss percentage is: ';
str02=num2str(dtloss(1,1));
str22=' %';
str32=strcat(str02,str22);
loss=strcat(str12,str32);
display(loss);
%graph based on attribute 1,2
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 1 & 3');
subplot(6,2,1);
gscatter(att2(:,1),att2(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att2(:,1),att2(:,2),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
%graph based on attribute 1,3
subplot(6,2,3);
gscatter(att2(:,1),att2(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,4);
gscatter(att2(:,1),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,1),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph based on attribute 1,4
subplot(6,2,5);
gscatter(att2(:,1),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att2(:,1),att2(:,4),predclass2);
```

```
if flag==1
line(testdata(diff2,1),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph based on attribute 2,3
subplot(6,2,7);
gscatter(att2(:,2),att2(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att2(:,2),att2(:,3),predclass2);
if flag==1
line(testdata(diff2,2),testdata(diff2,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%graph based on attribute 2,4
subplot(6,2,9);
gscatter(att2(:,2),att2(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att2(:,2),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,2),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%graph based on attribute 3,4
subplot(6,2,11);
gscatter(att2(:,3),att2(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att2(:,3),att2(:,4),predclass2);
if flag==1
line(testdata(diff2,3),testdata(diff2,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%CONFUSION MATRIX:
conf2=confusionmat(testclass,predclass2);
%VISUAL TREE
view(dt2);
%CLASSIFICATION TREE
view(dt2,'mode','graph');

% CLASSIFICATION BASED ON ATTRIBUTE 1 & 4
att3=testdata;
dt3=fitctree(traindata(:,[1 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass3=predict(dt3,testdata(:,[1 4]));
toc
l=1;
%cALCULATING MISCLASSIFIED POINTS
flag=0;
for i=1:length(predclass3)
if strcmp(predclass3(i),testclass(i)) == 0
```

```
    diff3(l,1)=i;
    l=l+1;
    flag=1;
end
end
%CALCULATING LOSS
dtloss=(size(diff3)./size(testclass)).*100;
str13='The loss percentage is: ';
str03=num2str(dtloss(1,1));
str23=' %';
str33=strcat(str03,str23);
loss=strcat(str13,str33);
display(loss);
%GRAPH WRT attribute 1,2
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 1 & 4');
subplot(6,2,1);
gscatter(att3(:,1),att3(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att3(:,1),att3(:,2),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
%GRAPH WRT attribute 1,3
subplot(6,2,3);
gscatter(att3(:,1),att3(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att3(:,1),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
%GRAPH WRT attribute 1,4
subplot(6,2,5);
gscatter(att3(:,1),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att3(:,1),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,1),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 2,3
subplot(6,2,7);
gscatter(att3(:,2),att3(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att3(:,2),att3(:,3),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,3),'marker','o','markersize',10,'linestyle','none');
```

```
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%GRAPH WRT attribute 2,4
subplot(6,2,9);
gscatter(att3(:,2),att3(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att3(:,2),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,2),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 3,4
subplot(6,2,11);
gscatter(att3(:,3),att3(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att3(:,3),att3(:,4),predclass3);
if flag==1
line(testdata(diff3,3),testdata(diff3,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

% CONFUSION MATRIX:
conf3=confusionmat(testclass,predclass3);
display(conf3);
%VISUAL TREE
view(dt3);
%CLASSIFICATION TREE
view(dt3,'mode','graph');

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 3
att4=testdata;
dt4=fitctree(traindata(:,[2 3]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass4=predict(dt4,testdata(:,[2 3]));
toc
%calculating misclassified indices
m=1;
flag=0;
for i=1:length(predclass4)
if strcmp(predclass4(i),testclass(i)) == 0
   diff4(m,1)=i;
   m=m+1;
   flag=1;
end
end
%calculating loss
dtloss=(size(diff4)./size(testclass)).*100;
str14='The loss percentage is: ';
str04=num2str(dtloss(1,1));
str24=' %';
str34=strcat(str04,str24);
loss=strcat(str14,str34);
display(loss);
%GRAPH WRT attribute 1,2
```

```
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 2 & 3')
subplot(6,2,1);
gscatter(att4(:,1),att4(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att4(:,1),att4(:,2),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
%GRAPH WRT attribute 1,3
subplot(6,2,3);
gscatter(att4(:,1),att4(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att4(:,1),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
%GRAPH WRT attribute 1,4
subplot(6,2,5);
gscatter(att4(:,1),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att4(:,1),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,1),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 2,3
subplot(6,2,7);
gscatter(att4(:,2),att4(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att4(:,2),att4(:,3),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%GRAPH WRT attribute 2,4
subplot(6,2,9);
gscatter(att4(:,2),att4(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att4(:,2),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,2),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
```

```
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 3,4
subplot(6,2,11);
gscatter(att4(:,3),att4(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att4(:,3),att4(:,4),predclass4);
if flag==1
line(testdata(diff4,3),testdata(diff4,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%%CONFUSION MATRIX:
conf4=confusionmat(testclass,predclass4);
display(conf4);
%VISUAL TREE
view(dt4);
%CLASSIFICATION TREE
view(dt4,'mode','graph');

%CLASSIFICATION BASED ON ATTRIBUTE 2 & 4
att5=testdata;
dt5=fitctree(traindata(:,[2 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass5=predict(dt5,testdata(:,[2 4]));
toc
%calculate the misclassified indices
n=1;
flag=0;
for i=1:length(predclass5)
if strcmp(predclass5(i),testclass(i)) == 0
   diff5(n,1)=i;
   n=n+1;
   flag=1;
end
end
%calculate the loss
dtloss=(size(diff5)./size(testclass)).*100;
str15='The loss percentage is: ';
str05=num2str(dtloss(1,1));
str25=' %';
str35=strcat(str05,str25);
loss=strcat(str15,str35);
display(loss);
%GRAPH WRT attribute 1,2
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 2 & 4');
subplot(6,2,1);
gscatter(att5(:,1),att5(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att5(:,1),att5(:,2),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
```

```
%GRAPH WRT attribute 1,3
subplot(6,2,3);
gscatter(att5(:,1),att5(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att5(:,1),att5(:,3),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
%GRAPH WRT attribute 1,4
subplot(6,2,5);
gscatter(att5(:,1),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att5(:,1),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,1),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%GRAPH WRT attribute 2,3
subplot(6,2,7);
gscatter(att5(:,2),att5(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att5(:,2),att5(:,3),predclass5);
if flag==1
line(testdata(diff5,2),testdata(diff5,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%GRAPH WRT attribute 2,4
subplot(6,2,9);
gscatter(att5(:,2),att5(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att5(:,2),att5(:,4),predclass5);
if flag==1
line(testdata(diff5,2),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 3,4
subplot(6,2,11);
gscatter(att5(:,3),att5(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att5(:,3),att5(:,4),predclass5);
if flag==1
```

```
line(testdata(diff5,3),testdata(diff5,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
% CONFUSION MATRIX:
%
conf5=confusionmat(testclass,predclass5);
display(conf5);
%VISUAL TREE
view(dt5);
%CLASSIFICATION TREE
view(dt5,'mode','graph');

%CLASSIFICATION BASED ON ATTRIBUTE 3 & 4
att6=testdata;
dt6=fitctree(traindata(:,[3 4]),trainclass,'ClassNames',{'setosa','versicolor','virginica'});
tic
predclass6=predict(dt6,testdata(:,[3 4]));
toc
%calculating miscalssified points
n=1;
flag=0;
for i=1:length(predclass6)
if strcmp(predclass6(i),testclass(i)) == 0
   diff6(n,1)=i;
   n=n+1;
   flag=1;
end
end
% calculating loss
dtloss=(size(diff6)./size(testclass)).*100;
str16='The loss percentage is: ';
str06=num2str(dtloss(1,1));
str26=' %';
str36=strcat(str06,str26);
loss=strcat(str16,str36);
display(loss);
%GRAPH WRT attribute 1,2
figure('Name','DECISION TREE CLASSIFICATION BASED ON ATTRIBUTE 3 & 4');
subplot(6,2,1);
gscatter(att6(:,1),att6(:,2),testclass);
title('SEPAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
subplot(6,2,2);
gscatter(att6(:,1),att6(:,2),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,2),'marker','o','markersize',10,'linestyle','none');
end
title('SEPAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('SEPAL WIDTH (cm)');
%GRAPH WRT attribute 1,3
subplot(6,2,3);
gscatter(att6(:,1),att6(:,3),testclass);
title(' PETAL LENGTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL LENGT (cm)');
subplot(6,2,4);
gscatter(att6(:,1),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title(' PETAL LENGTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
```

```
ylabel('PETAL LENGT (cm)');
%GRAPH WRT attribute 1,4
subplot(6,2,5);
gscatter(att6(:,1),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,6);
gscatter(att6(:,1),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,1),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 2,3
subplot(6,2,7);
gscatter(att6(:,2),att6(:,3),testclass);
title('PETAL LENGTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
subplot(6,2,8);
gscatter(att6(:,2),att6(:,3),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,3),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL LENGTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL LENGTH (cm)');
%GRAPH WRT attribute 2,4
subplot(6,2,9);
gscatter(att6(:,2),att6(:,4),testclass);
title('PETAL WIDTH v/s SEPAL WIDTH-ACTUAL CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,10);
gscatter(att6(:,2),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,2),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s SEPAL WIDTH-PREDICTED CLASSIFICATION');
xlabel('SEPAL WIDTH (cm)');
ylabel('PETAL WIDTH (cm)');
%GRAPH WRT attribute 3,4
subplot(6,2,11);
gscatter(att6(:,3),att6(:,4),testclass);
title('PETAL WIDTH v/s PETAL LENGTH-ACTUAL CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');
subplot(6,2,12);
gscatter(att6(:,3),att6(:,4),predclass6);
if flag==1
line(testdata(diff6,3),testdata(diff6,4),'marker','o','markersize',10,'linestyle','none');
end
title('PETAL WIDTH v/s PETAL LENGTH-PREDICTED CLASSIFICATION');
xlabel('PETAL LENGTH (cm)');
ylabel('PETAL WIDTH (cm)');

%CONFUSION MATRIX:
conf6=confusionmat(testclass,predclass6);
display(conf6);
%VISUAL TREE
view(dt6);
%CLASSIFICATION TREE
view(dt6,'mode','graph');
```

K-MEANS CLUSTERING:

Code for k-means clustering over 4 attributes:

```
%clustering using K-MEANS
load fisheriris;
att=meas;
class=species;
tic
[res,cent]=kmeans(att,3,'Replicates',4);
toc
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y
xlim=min(att(:,1)):0.01:max(att(:,1));
ylim=min(att(:,2)):0.01:max(att(:,2));
[xgrid,ygrid]=meshgrid(xlim,ylim);
plotgrid=[xgrid(:) ygrid(:)];
figure('Name','CLUSTERING ON ENTIRE DATASET');
subplot(3,2,1)
 regions=kmeans(plotgrid,3,'Start',cent(:,[1 2]),'MaxIter',6);
 gscatter(plotgrid(:,1),plotgrid(:,2),regions);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
ylabel 'SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
xlim1=min(att(:,1)):0.01:max(att(:,1));
ylim1=min(att(:,3)):0.01:max(att(:,3));
[xgrid1,ygrid1]=meshgrid(xlim1,ylim1);
plotgrid1=[xgrid1(:) ygrid1(:)];
subplot(3,2,2)
 regions1=kmeans(plotgrid1,3,'MaxIter',6,'start',cent(:,[1 3]));
 gscatter(plotgrid1(:,1),plotgrid1(:,2),regions1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
ylabel 'PETAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
xlim3=min(att(:,1)):0.01:max(att(:,1));
ylim3=min(att(:,4)):0.01:max(att(:,4));
[xgrid3,ygrid3]=meshgrid(xlim3,ylim3);
plotgrid3=[xgrid3(:) ygrid3(:)];
subplot(3,2,3)
 regions3=kmeans(plotgrid3,3,'MaxIter',6,'start',cent(:,[1 2]));
 gscatter(plotgrid3(:,1),plotgrid3(:,2),regions3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
xlim4=min(att(:,2)):0.01:max(att(:,2));
ylim4=min(att(:,3)):0.01:max(att(:,3));
[xgrid4,ygrid4]=meshgrid(xlim4,ylim4);
plotgrid4=[xgrid4(:) ygrid4(:)];
subplot(3,2,4)
 regions4=kmeans(plotgrid4,3,'MaxIter',6,'start',cent(:,[2 3]));
 gscatter(plotgrid4(:,1),plotgrid4(:,2),regions4);
hold on
```

```
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH';
ylabel 'PETAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
xlim5=min(att(:,2)):0.01:max(att(:,2));
ylim5=min(att(:,4)):0.01:max(att(:,4));
[xgrid5,ygrid5]=meshgrid(xlim5,ylim5);
plotgrid5=[xgrid5(:) ygrid5(:)];
subplot(3,2,5)
 regions5=kmeans(plotgrid5,3,'MaxIter',6,'start',cent(:,[2 3]));
 gscatter(plotgrid5(:,1),plotgrid5(:,2),regions5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
xlim6=min(att(:,3)):0.01:max(att(:,3));
ylim6=min(att(:,4)):0.01:max(att(:,4));
[xgrid6,ygrid6]=meshgrid(xlim6,ylim6);
plotgrid6=[xgrid6(:) ygrid6(:)];
subplot(3,2,6)
 regions6=kmeans(plotgrid6,3,'MaxIter',6,'start',cent(:,[2 3]));
 gscatter(plotgrid6(:,1),plotgrid6(:,2),regions6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;

%CLUSTERING GRAPH WRT TO  ATT1-X,ATT2-Y AND ATT3-Z
figure('Name','CLUSTERING ON ENTIRE DATASET');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),res);
line(cent(:,1),cent(:,2),cent(:,3),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT2-X,ATT3-Y AND ATT4-Z
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),res);
line(cent(:,2),cent(:,3),cent(:,4),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT1-X,ATT2-Y AND ATT4-Z
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,2),att(:,4),res);
line(cent(:,1),cent(:,2),cent(:,4),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
```

```
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT1-X,ATT3-Y AND ATT4-Z
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),res);
line(cent(:,1),cent(:,3),cent(:,4),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
figure('Name','CLUSTERING ON ENTIRE DATASET');
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%silhouette curve
[silcurve,h] = silhouette(att,res);
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';
display(mean(silcurve));
%code for calculating purity
conf=zeros(3,3);
for i=1:length(res)
   if strcmp(species(i),'setosa')==1
      if res(i)==2
         conf(1,1)=conf(1,1)+1;
      end
      if res(i)==3
         conf(3,1)=conf(3,1)+1;
      end
      if res(i)==1
         conf(2,1)=conf(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res(i)==2
         conf(1,2)=conf(1,2)+1;
      end
      if res(i)==3
         conf(3,2)=conf(3,2)+1;
      end
      if res(i)==1
         conf(2,2)=conf(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res(i)==2
         conf(1,3)=conf(1,3)+1;
      end
      if res(i)==3
         conf(3,3)=conf(3,3)+1;
      end
      if res(i)==1
         conf(2,3)=conf(2,3)+1;
      end

   end
end
purity=(max(conf(1,:))+max(conf(2,:))+max(conf(3,:)))./150
```

Code for k-means clustering over 3 attributes:

```
load fisheriris;
att=meas;
class=species;
```

```
%CLUSTER BASED ON ATTRIBUTE 1,2 & 3
tic
[res_123,cent_123]=kmeans(att(:,[1 2 3]),3,'Replicates',4);
toc
%purity
conf=zeros(3,3);
for i=1:length(res_123)
   if strcmp(species(i),'setosa')==1
      if res_123(i)==2
         conf(1,1)=conf(1,1)+1;
      end
      if res_123(i)==3
         conf(3,1)=conf(3,1)+1;
      end
      if res_123(i)==1
         conf(2,1)=conf(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_123(i)==2
         conf(1,2)=conf(1,2)+1;
      end
      if res_123(i)==3
         conf(3,2)=conf(3,2)+1;
      end
      if res_123(i)==1
         conf(2,2)=conf(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_123(i)==2
         conf(1,3)=conf(1,3)+1;
      end
      if res_123(i)==3
         conf(3,3)=conf(3,3)+1;
      end
      if res_123(i)==1
         conf(2,3)=conf(2,3)+1;
      end

   end
end
purity=(max(conf(1,:))+max(conf(2,:))+max(conf(3,:)))./150
%graph wrt attribute 1,2 and 3
figure ('Name','CLUSTER BASED ON ATTRIBUTE 1,2 & 3');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),res_123);
line(cent_123(:,1),cent_123(:,2),cent_123(:,3),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%graph wrt attribute 2,3 and 4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),res_123);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attribute 1,3,4
```

```
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),res_123);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attribute 1,2,4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,2),att(:,4),res_123);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%silhouette curve
figure ('Name','CLUSTER BASED ON ATTRIBUTE 1,2 & 3');
[silcurve_123,h_123]=silhouette(att(:,[1 2 3]),res_123);
mean(silcurve_123)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
conf2=zeros(3,3);
%CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4
tic
[res_234,cent_234]=kmeans(att(:,[2 3 4]),3,'Replicates',4);
toc
%purity
for i=1:length(res_234)
   if strcmp(species(i),'setosa')==1
      if res_234(i)==2
         conf2(1,1)=conf2(1,1)+1;
      end
      if res_234(i)==3
         conf2(3,1)=conf2(3,1)+1;
      end
      if res_234(i)==1
         conf2(2,1)=conf2(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_234(i)==2
         conf2(1,2)=conf2(1,2)+1;
      end
      if res_234(i)==3
         conf2(3,2)=conf2(3,2)+1;
      end
      if res_234(i)==1
         conf2(2,2)=conf2(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_234(i)==2
         conf2(1,3)=conf2(1,3)+1;
      end
      if res_234(i)==3
         conf2(3,3)=conf2(3,3)+1;
      end
      if res_234(i)==1
         conf2(2,3)=conf2(2,3)+1;
      end
```

```
    end
end
purity=(max(conf2(1,:))+max(conf2(2,:))+max(conf2(3,:))))./150

%graph wrt attribute 1,2,3
 figure ('Name','CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4)');
 subplot(2,2,1);
 grid on;
 gscatter3(att(:,1),att(:,2),att(:,3),res_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between attribute 2,3,4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),res_234);
line(cent_234(:,1),cent_234(:,2),cent_234(:,3),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%graph between attribute 1,2,4
subplot(2,2,3);
grid on
gscatter3(att(:,1),att(:,2),att(:,4),res_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between attribute 1,3,4
subplot(2,2,4);
 grid on;
gscatter3(att(:,1),att(:,3),att(:,4),res_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%silhouette curve
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4');
[silcurve_234,h_234]=silhouette(att(:,[2 3 4]),res_234);
mean(silcurve_234)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
conf3=zeros(3,3);
%CLUSTERING BETWEEN 1,2 AND 4
tic
[res_124,cent_124]=kmeans(att(:,[1 2 4]),3,'Replicates',4);
toc
%purity
for i=1:length(res_124)
   if strcmp(species(i),'setosa')==1
      if res_124(i)==2
         conf3(1,1)=conf3(1,1)+1;
      end
      if res_124(i)==3
         conf3(3,1)=conf3(3,1)+1;
      end
      if res_124(i)==1
         conf3(2,1)=conf3(2,1)+1;
      end
```

```
    end
    if strcmp(species(i),'versicolor')==1
        if res_124(i)==2
            conf3(1,2)=conf3(1,2)+1;
        end
        if res_124(i)==3
            conf3(3,2)=conf3(3,2)+1;
        end
        if res_124(i)==1
            conf3(2,2)=conf3(2,2)+1;
        end

    end
    if strcmp(species(i),'virginica')==1
        if res_124(i)==2
            conf3(1,3)=conf3(1,3)+1;
        end
        if res_124(i)==3
            conf3(3,3)=conf3(3,3)+1;
        end
        if res_124(i)==1
            conf3(2,3)=conf3(2,3)+1;
        end

    end
end
purity=(max(conf3(1,:))+max(conf3(2,:))+max(conf3(3,:)))./150

%graph wrt attribute 1 2 3
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1,2 AND 4d');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),res_124);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attribute 2 3 4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),res_124);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attribute 1 2 4
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,2),att(:,4),res_124);
line(cent_124(:,1),cent_124(:,2),cent_124(:,3),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%graph wrt attribute 1 3 4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),res_124);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
```

```
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%silhouette curve
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1,2 AND 4');
[silcurve_124,h_124]=silhouette(att(:,[1 2 4]),res_124);
mean(silcurve_124)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
conf4=zeros(3,3);
% %CLUSTERING BASED ON 1, 3 AND 4
tic
[res_134,cent_134]=kmeans(att(:,[1 3 4]),3,'Replicates',4);
toc
%purity
for i=1:length(res_134)
   if strcmp(species(i),'setosa')==1
      if res_134(i)==2
         conf4(1,1)=conf4(1,1)+1;
      end
      if res_134(i)==3
         conf4(3,1)=conf4(3,1)+1;
      end
      if res_134(i)==1
         conf4(2,1)=conf4(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_134(i)==2
         conf4(1,2)=conf4(1,2)+1;
      end
      if res_134(i)==3
         conf4(3,2)=conf4(3,2)+1;
      end
      if res_134(i)==1
         conf4(2,2)=conf4(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_134(i)==2
         conf4(1,3)=conf4(1,3)+1;
      end
      if res_134(i)==3
         conf4(3,3)=conf4(3,3)+1;
      end
      if res_134(i)==1
         conf4(2,3)=conf4(2,3)+1;
      end

   end
end
purity=(max(conf4(1,:))+max(conf4(2,:))+max(conf4(3,:)))./150

%graph wrt attribute 1 2 3
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1, 3 AND 4');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),res_134);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attribute 2 3 4
subplot(2,2,2);
```

```
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),res_134);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between wrt attribute 1,2,4
subplot(2,2,3);
grid on
gscatter3(att(:,1),att(:,2),att(:,4),res_134);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between wrt attribute 1,3,4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),res_134);
line(cent_134(:,1),cent_134(:,2),cent_134(:,3),'marker','x','markersize',12,'color','k','linewidth',4,'linestyle','none');
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','CENTROID','LOCATION','BEST');
%silhouette curve
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1, 3 AND 4');
[silcurve_134,h_134]=silhouette(att(:,[1 3 4]),res_134);
mean(silcurve_134)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
```

Code for k-means clustering over 2 attributes:

```
%clustering using K-MEANS
load fisheriris;
att=meas;
class=species;
%CLUSTERING BASED ON ATTRIBUTE 1 AND 2
tic
[res_12,cent_12]=kmeans(att(:,[1 2]),3,'Replicates',4);
toc
%PURITY
conf1=zeros(3,3);
for i=1:length(res_12)
   if strcmp(species(i),'setosa')==1
      if res_12(i)==2
         conf1(1,1)=conf1(1,1)+1;
      end
      if res_12(i)==3
         conf1(3,1)=conf1(3,1)+1;
      end
      if res_12(i)==1
         conf1(2,1)=conf1(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_12(i)==2
         conf1(1,2)=conf1(1,2)+1;
      end
      if res_12(i)==3
         conf1(3,2)=conf1(3,2)+1;
      end
```

```
        if res_12(i)==1
            conf1(2,2)=conf1(2,2)+1;
        end

    end
    if strcmp(species(i),'virginica')==1
        if res_12(i)==2
            conf1(1,3)=conf1(1,3)+1;
        end
        if res_12(i)==3
            conf1(3,3)=conf1(3,3)+1;
        end
        if res_12(i)==1
            conf1(2,3)=conf1(2,3)+1;
        end

    end
end
purity=(max(conf1(1,:))+max(conf1(2,:))+max(conf1(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y
xlim_12=min(att(:,1)):0.01:max(att(:,1));
ylim_12=min(att(:,2)):0.01:max(att(:,2));
[xgrid_12,ygrid_12]=meshgrid(xlim_12,ylim_12);
plotgrid_12=[xgrid_12(:) ygrid_12(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 2');
subplot(3,2,1)
 regions_12=kmeans(plotgrid_12,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12(:,1),plotgrid_12(:,2),regions_12);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
line(cent_12(:,1),cent_12(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_12_1=min(att(:,1)):0.01:max(att(:,1));
ylim_12_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_12_1,ygrid_12_1]=meshgrid(xlim_12_1,ylim_12_1);
plotgrid_12_1=[xgrid_12_1(:) ygrid_12_1(:)];
subplot(3,2,2)
 regions_12_1=kmeans(plotgrid_12_1,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12_1(:,1),plotgrid_12_1(:,2),regions_12_1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_12_3=min(att(:,1)):0.01:max(att(:,1));
ylim_12_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_12_3,ygrid_12_3]=meshgrid(xlim_12_3,ylim_12_3);
plotgrid_12_3=[xgrid_12_3(:) ygrid_12_3(:)];
subplot(3,2,3)
 regions_12_3=kmeans(plotgrid_12_3,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12_3(:,1),plotgrid_12_3(:,2),regions_12_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
```

```
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_12_4=min(att(:,2)):0.01:max(att(:,2));
ylim_12_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_12_4,ygrid_12_4]=meshgrid(xlim_12_4,ylim_12_4);
plotgrid_12_4=[xgrid_12_4(:) ygrid_12_4(:)];
subplot(3,2,4)
 regions_12_4=kmeans(plotgrid_12_4,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12_4(:,1),plotgrid_12_4(:,2),regions_12_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_12_5=min(att(:,2)):0.01:max(att(:,2));
ylim_12_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_12_5,ygrid_12_5]=meshgrid(xlim_12_5,ylim_12_5);
plotgrid_12_5=[xgrid_12_5(:) ygrid_12_5(:)];
subplot(3,2,5)
 regions_12_5=kmeans(plotgrid_12_5,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12_5(:,1),plotgrid_12_5(:,2),regions_12_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y

xlim_12_6=min(att(:,3)):0.01:max(att(:,3));
ylim_12_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_12_6,ygrid_12_6]=meshgrid(xlim_12_6,ylim_12_6);
plotgrid_12_6=[xgrid_12_6(:) ygrid_12_6(:)];
subplot(3,2,6)
 regions_12_6=kmeans(plotgrid_12_6,3,'MaxIter',6,'start',cent_12(:,[1 2]));
 gscatter(plotgrid_12_6(:,1),plotgrid_12_6(:,2),regions_12_6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 2');
[silcurve,h] = silhouette(att(:,[1 2]),res_12);
mean(silcurve)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';



%CLUSTERING BASED ON ATTRIBUTE 1 AND 3
tic
[res_13,cent_13]=kmeans(att(:,[1 3]),3,'Replicates',4);
toc
%purity
conf2=zeros(3,3);
```

```
for i=1:length(res_13)
   if strcmp(species(i),'setosa')==1
      if res_13(i)==2
         conf2(1,1)=conf2(1,1)+1;
      end
      if res_13(i)==3
         conf2(3,1)=conf2(3,1)+1;
      end
      if res_13(i)==1
         conf2(2,1)=conf2(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_13(i)==2
         conf2(1,2)=conf2(1,2)+1;
      end
      if res_13(i)==3
         conf2(3,2)=conf2(3,2)+1;
      end
      if res_13(i)==1
         conf2(2,2)=conf2(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_13(i)==2
         conf2(1,3)=conf2(1,3)+1;
      end
      if res_13(i)==3
         conf2(3,3)=conf2(3,3)+1;
      end
      if res_13(i)==1
         conf2(2,3)=conf2(2,3)+1;
      end

   end
end
purity=(max(conf2(1,:))+max(conf2(2,:))+max(conf2(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

xlim_13=min(att(:,1)):0.01:max(att(:,1));
ylim_13=min(att(:,2)):0.01:max(att(:,2));
[xgrid_13,ygrid_13]=meshgrid(xlim_13,ylim_13);
plotgrid_13=[xgrid_13(:) ygrid_13(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 3');
subplot(3,2,1)
 regions_13=kmeans(plotgrid_13,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13(:,1),plotgrid_13(:,2),regions_13);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_13_1=min(att(:,1)):0.01:max(att(:,1));
ylim_13_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_13_1,ygrid_13_1]=meshgrid(xlim_13_1,ylim_13_1);
plotgrid_13_1=[xgrid_13_1(:) ygrid_13_1(:)];
subplot(3,2,2)
 regions_13_1=kmeans(plotgrid_13_1,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13_1(:,1),plotgrid_13_1(:,2),regions_13_1);
hold on
```

```
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
line(cent_13(:,1),cent_13(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_13_3=min(att(:,1)):0.01:max(att(:,1));
ylim_13_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_13_3,ygrid_13_3]=meshgrid(xlim_13_3,ylim_13_3);
plotgrid_13_3=[xgrid_13_3(:) ygrid_13_3(:)];
subplot(3,2,3)
 regions_13_3=kmeans(plotgrid_13_3,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13_3(:,1),plotgrid_13_3(:,2),regions_13_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_13_4=min(att(:,2)):0.01:max(att(:,2));
ylim_13_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_13_4,ygrid_13_4]=meshgrid(xlim_13_4,ylim_13_4);
plotgrid_13_4=[xgrid_13_4(:) ygrid_13_4(:)];
subplot(3,2,4)
 regions_13_4=kmeans(plotgrid_13_4,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13_4(:,1),plotgrid_13_4(:,2),regions_13_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_13_5=min(att(:,2)):0.01:max(att(:,2));
ylim_13_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_13_5,ygrid_13_5]=meshgrid(xlim_13_5,ylim_13_5);
plotgrid_13_5=[xgrid_13_5(:) ygrid_13_5(:)];
subplot(3,2,5)
 regions_13_5=kmeans(plotgrid_13_5,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13_5(:,1),plotgrid_13_5(:,2),regions_13_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y

xlim_13_6=min(att(:,3)):0.01:max(att(:,3));
ylim_13_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_13_6,ygrid_13_6]=meshgrid(xlim_13_6,ylim_13_6);
plotgrid_13_6=[xgrid_13_6(:) ygrid_13_6(:)];
subplot(3,2,6)
 regions_13_6=kmeans(plotgrid_13_6,3,'MaxIter',6,'start',cent_13(:,[1 2]));
 gscatter(plotgrid_13_6(:,1),plotgrid_13_6(:,2),regions_13_6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
```

```
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 3');
[silcurve,h] = silhouette(att(:,[1 3]),res_13);
mean(silcurve)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

%CLUSTERING BASED ON ATTRIBUTE 1 AND 4
tic
[res_14,cent_14]=kmeans(att(:,[1 4]),3,'Replicates',4);
toc
%purity
conf3=zeros(3,3);
for i=1:length(res_14)
   if strcmp(species(i),'setosa')==1
      if res_14(i)==2
         conf3(1,1)=conf3(1,1)+1;
      end
      if res_14(i)==3
         conf3(3,1)=conf3(3,1)+1;
      end
      if res_14(i)==1
         conf3(2,1)=conf3(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_14(i)==2
         conf3(1,2)=conf3(1,2)+1;
      end
      if res_14(i)==3
         conf3(3,2)=conf3(3,2)+1;
      end
      if res_14(i)==1
         conf3(2,2)=conf3(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_14(i)==2
         conf3(1,3)=conf3(1,3)+1;
      end
      if res_14(i)==3
         conf3(3,3)=conf3(3,3)+1;
      end
      if res_14(i)==1
         conf3(2,3)=conf3(2,3)+1;
      end

   end
end
purity=(max(conf3(1,:))+max(conf3(2,:))+max(conf3(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

xlim_14=min(att(:,1)):0.01:max(att(:,1));
ylim_14=min(att(:,2)):0.01:max(att(:,2));
[xgrid_14,ygrid_14]=meshgrid(xlim_14,ylim_14);
plotgrid_14=[xgrid_14(:) ygrid_14(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 4');
subplot(3,2,1)
```

```
 regions_14=kmeans(plotgrid_14,3,'MaxIter',6,'start',cent_14(:,[1 2]));
 gscatter(plotgrid_14(:,1),plotgrid_14(:,2),regions_14);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_14_1=min(att(:,1)):0.01:max(att(:,1));
ylim_14_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_14_1,ygrid_14_1]=meshgrid(xlim_14_1,ylim_14_1);
plotgrid_14_1=[xgrid_14_1(:) ygrid_14_1(:)];
subplot(3,2,2)
 regions_14_1=kmeans(plotgrid_14_1,3,'MaxIter',6,'start',cent_14(:,[1 2]));
 gscatter(plotgrid_14_1(:,1),plotgrid_14_1(:,2),regions_14_1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_14_3=min(att(:,1)):0.01:max(att(:,1));
ylim_14_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_14_3,ygrid_14_3]=meshgrid(xlim_14_3,ylim_14_3);
plotgrid_14_3=[xgrid_14_3(:) ygrid_14_3(:)];
subplot(3,2,3)
 regions_14_3=kmeans(plotgrid_14_3,3,'MaxIter',6,'start',cent_14(:,[1 2]));
 gscatter(plotgrid_14_3(:,1),plotgrid_14_3(:,2),regions_14_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
line(cent_14(:,1),cent_14(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_14_4=min(att(:,2)):0.01:max(att(:,2));
ylim_14_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_14_4,ygrid_14_4]=meshgrid(xlim_14_4,ylim_14_4);
plotgrid_14_4=[xgrid_14_4(:) ygrid_14_4(:)];
subplot(3,2,4)
 regions_14_4=kmeans(plotgrid_14_4,3,'MaxIter',6,'start',cent_14(:,[1 2]));
 gscatter(plotgrid_14_4(:,1),plotgrid_14_4(:,2),regions_14_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_14_5=min(att(:,2)):0.01:max(att(:,2));
ylim_14_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_14_5,ygrid_14_5]=meshgrid(xlim_14_5,ylim_14_5);
plotgrid_14_5=[xgrid_14_5(:) ygrid_14_5(:)];
subplot(3,2,5)
 regions_14_5=kmeans(plotgrid_14_5,3,'MaxIter',6,'start',cent_14(:,[1 2]));
```

```
 gscatter(plotgrid_14_5(:,1),plotgrid_14_5(:,2),regions_14_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y

xlim_14_6=min(att(:,3)):0.01:max(att(:,3));
ylim_14_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_14_6,ygrid_14_6]=meshgrid(xlim_14_6,ylim_14_6);
plotgrid_14_6=[xgrid_14_6(:) ygrid_14_6(:)];
subplot(3,2,6)
 regions_14_6=kmeans(plotgrid_14_6,3,'MaxIter',6,'start',cent_14(:,[1 2]));
 gscatter(plotgrid_14_6(:,1),plotgrid_14_6(:,2),regions_14_6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 4');
[silcurve_14,h] = silhouette(att(:,[1 4]),res_14);
mean(silcurve_14)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

% CLUSTERING BASED ON ATTRIBUTE 2 AND 3
tic
[res_23,cent_23]=kmeans(att(:,[2 3]),3,'Replicates',4);
toc
%purity
conf4=zeros(3,3);
for i=1:length(res_23)
   if strcmp(species(i),'setosa')==1
      if res_23(i)==2
         conf4(1,1)=conf4(1,1)+1;
      end
      if res_23(i)==3
         conf4(3,1)=conf4(3,1)+1;
      end
      if res_23(i)==1
         conf4(2,1)=conf4(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_23(i)==2
         conf4(1,2)=conf4(1,2)+1;
      end
      if res_23(i)==3
         conf4(3,2)=conf4(3,2)+1;
      end
      if res_23(i)==1
         conf4(2,2)=conf4(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_23(i)==2
         conf4(1,3)=conf4(1,3)+1;
```

```
        end
        if res_23(i)==3
            conf4(3,3)=conf4(3,3)+1;
        end
        if res_23(i)==1
            conf4(2,3)=conf4(2,3)+1;
        end

    end
end
purity=(max(conf4(1,:))+max(conf4(2,:))+max(conf4(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

xlim_23=min(att(:,1)):0.01:max(att(:,1));
ylim_23=min(att(:,2)):0.01:max(att(:,2));
[xgrid_23,ygrid_23]=meshgrid(xlim_23,ylim_23);
plotgrid_23=[xgrid_23(:) ygrid_23(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 3');
subplot(3,2,1)
 regions_23=kmeans(plotgrid_23,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23(:,1),plotgrid_23(:,2),regions_23);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_23_1=min(att(:,1)):0.01:max(att(:,1));
ylim_23_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_23_1,ygrid_23_1]=meshgrid(xlim_23_1,ylim_23_1);
plotgrid_23_1=[xgrid_23_1(:) ygrid_23_1(:)];
subplot(3,2,2)
 regions_23_1=kmeans(plotgrid_23_1,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23_1(:,1),plotgrid_23_1(:,2),regions_23_1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_23_3=min(att(:,1)):0.01:max(att(:,1));
ylim_23_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_23_3,ygrid_23_3]=meshgrid(xlim_23_3,ylim_23_3);
plotgrid_23_3=[xgrid_23_3(:) ygrid_23_3(:)];
subplot(3,2,3)
 regions_23_3=kmeans(plotgrid_23_3,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23_3(:,1),plotgrid_23_3(:,2),regions_23_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_23_4=min(att(:,2)):0.01:max(att(:,2));
ylim_23_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_23_4,ygrid_23_4]=meshgrid(xlim_23_4,ylim_23_4);
plotgrid_23_4=[xgrid_23_4(:) ygrid_23_4(:)];
```

```
subplot(3,2,4)
 regions_23_4=kmeans(plotgrid_23_4,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23_4(:,1),plotgrid_23_4(:,2),regions_23_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
line(cent_23(:,1),cent_23(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_23_5=min(att(:,2)):0.01:max(att(:,2));
ylim_23_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_23_5,ygrid_23_5]=meshgrid(xlim_23_5,ylim_23_5);
plotgrid_23_5=[xgrid_23_5(:) ygrid_23_5(:)];
subplot(3,2,5)
 regions_23_5=kmeans(plotgrid_23_5,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23_5(:,1),plotgrid_23_5(:,2),regions_23_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y

xlim_23_6=min(att(:,3)):0.01:max(att(:,3));
ylim_23_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_23_6,ygrid_23_6]=meshgrid(xlim_23_6,ylim_23_6);
plotgrid_23_6=[xgrid_23_6(:) ygrid_23_6(:)];
subplot(3,2,6)
 regions_23_6=kmeans(plotgrid_23_6,3,'MaxIter',6,'start',cent_23(:,[1 2]));
 gscatter(plotgrid_23_6(:,1),plotgrid_23_6(:,2),regions_23_6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 3');
[silcurve_23,h_23] = silhouette(att(:,[2 3]),res_23);
mean(silcurve_23)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

% CLUSTERING BASED ON ATTRIBUTE 2 AND 4
tic
[res_24,cent_24]=kmeans(att(:,[2 4]),3,'Replicates',4);
toc
%purity
conf5=zeros(3,3);
for i=1:length(res_24)
   if strcmp(species(i),'setosa')==1
      if res_24(i)==2
         conf5(1,1)=conf5(1,1)+1;
      end
      if res_24(i)==3
         conf5(3,1)=conf5(3,1)+1;
      end
      if res_24(i)==1
```

```
            conf5(2,1)=conf5(2,1)+1;
        end

    end
    if strcmp(species(i),'versicolor')==1
        if res_24(i)==2
            conf5(1,2)=conf5(1,2)+1;
        end
        if res_24(i)==3
            conf5(3,2)=conf5(3,2)+1;
        end
        if res_24(i)==1
            conf5(2,2)=conf5(2,2)+1;
        end

    end
    if strcmp(species(i),'virginica')==1
        if res_24(i)==2
            conf5(1,3)=conf5(1,3)+1;
        end
        if res_24(i)==3
            conf5(3,3)=conf5(3,3)+1;
        end
        if res_24(i)==1
            conf5(2,3)=conf5(2,3)+1;
        end

    end
end
purity=(max(conf5(1,:))+max(conf5(2,:))+max(conf5(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

xlim_24=min(att(:,1)):0.01:max(att(:,1));
ylim_24=min(att(:,2)):0.01:max(att(:,2));
[xgrid_24,ygrid_24]=meshgrid(xlim_24,ylim_24);
plotgrid_24=[xgrid_24(:) ygrid_24(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 4');
subplot(3,2,1)
 regions_24=kmeans(plotgrid_24,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24(:,1),plotgrid_24(:,2),regions_24);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_24_1=min(att(:,1)):0.01:max(att(:,1));
ylim_24_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_24_1,ygrid_24_1]=meshgrid(xlim_24_1,ylim_24_1);
plotgrid_24_1=[xgrid_24_1(:) ygrid_24_1(:)];
subplot(3,2,2)
 regions_24_1=kmeans(plotgrid_24_1,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24_1(:,1),plotgrid_24_1(:,2),regions_24_1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_24_3=min(att(:,1)):0.01:max(att(:,1));
```

```
ylim_24_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_24_3,ygrid_24_3]=meshgrid(xlim_24_3,ylim_24_3);
plotgrid_24_3=[xgrid_24_3(:) ygrid_24_3(:)];
subplot(3,2,3)
 regions_24_3=kmeans(plotgrid_24_3,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24_3(:,1),plotgrid_24_3(:,2),regions_24_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_24_4=min(att(:,2)):0.01:max(att(:,2));
ylim_24_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_24_4,ygrid_24_4]=meshgrid(xlim_24_4,ylim_24_4);
plotgrid_24_4=[xgrid_24_4(:) ygrid_24_4(:)];
subplot(3,2,4)
 regions_24_4=kmeans(plotgrid_24_4,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24_4(:,1),plotgrid_24_4(:,2),regions_24_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_24_5=min(att(:,2)):0.01:max(att(:,2));
ylim_24_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_24_5,ygrid_24_5]=meshgrid(xlim_24_5,ylim_24_5);
plotgrid_24_5=[xgrid_24_5(:) ygrid_24_5(:)];
subplot(3,2,5)
 regions_24_5=kmeans(plotgrid_24_5,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24_5(:,1),plotgrid_24_5(:,2),regions_24_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
line(cent_24(:,1),cent_24(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y

xlim_24_6=min(att(:,3)):0.01:max(att(:,3));
ylim_24_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_24_6,ygrid_24_6]=meshgrid(xlim_24_6,ylim_24_6);
plotgrid_24_6=[xgrid_24_6(:) ygrid_24_6(:)];
subplot(3,2,6)
 regions_24_6=kmeans(plotgrid_24_6,3,'MaxIter',6,'start',cent_24(:,[1 2]));
 gscatter(plotgrid_24_6(:,1),plotgrid_24_6(:,2),regions_24_6);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 4');
[silcurve_24,h_24] = silhouette(att(:,[2 4]),res_24);
mean(silcurve_24)
```

```
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';


% CLUSTERING BASED ON ATTRIBUTE 3 AND 4
tic
[res_34,cent_34]=kmeans(att(:,[3 4]),3,'Replicates',4);
toc
%purity
conf6=zeros(3,3);
for i=1:length(res_34)
   if strcmp(species(i),'setosa')==1
      if res_34(i)==2
         conf6(1,1)=conf6(1,1)+1;
      end
      if res_34(i)==3
         conf6(3,1)=conf6(3,1)+1;
      end
      if res_34(i)==1
         conf6(2,1)=conf6(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if res_34(i)==2
         conf6(1,2)=conf6(1,2)+1;
      end
      if res_34(i)==3
         conf6(3,2)=conf6(3,2)+1;
      end
      if res_34(i)==1
         conf6(2,2)=conf6(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if res_34(i)==2
         conf6(1,3)=conf6(1,3)+1;
      end
      if res_34(i)==3
         conf6(3,3)=conf6(3,3)+1;
      end
      if res_34(i)==1
         conf6(2,3)=conf6(2,3)+1;
      end

   end
end
purity=(max(conf6(1,:))+max(conf6(2,:))+max(conf6(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

xlim_34=min(att(:,1)):0.01:max(att(:,1));
ylim_34=min(att(:,2)):0.01:max(att(:,2));
[xgrid_34,ygrid_34]=meshgrid(xlim_34,ylim_34);
plotgrid_34=[xgrid_34(:) ygrid_34(:)];
figure('Name','CLUSTERING BASED ON ATTRIBUTE 3 AND 4');
subplot(3,2,1)
 regions_34=kmeans(plotgrid_34,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34(:,1),plotgrid_34(:,2),regions_34);
hold on
plot(att(:,1),att(:,2),'marker','*','markersize',10,'linestyle','none');
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
```

```
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y

xlim_34_1=min(att(:,1)):0.01:max(att(:,1));
ylim_34_1=min(att(:,3)):0.01:max(att(:,3));
[xgrid_34_1,ygrid_34_1]=meshgrid(xlim_34_1,ylim_34_1);
plotgrid_34_1=[xgrid_34_1(:) ygrid_34_1(:)];
subplot(3,2,2)
 regions_34_1=kmeans(plotgrid_34_1,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34_1(:,1),plotgrid_34_1(:,2),regions_34_1);
hold on
plot(att(:,1),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y

xlim_34_3=min(att(:,1)):0.01:max(att(:,1));
ylim_34_3=min(att(:,4)):0.01:max(att(:,4));
[xgrid_34_3,ygrid_34_3]=meshgrid(xlim_34_3,ylim_34_3);
plotgrid_34_3=[xgrid_34_3(:) ygrid_34_3(:)];
subplot(3,2,3)
 regions_34_3=kmeans(plotgrid_34_3,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34_3(:,1),plotgrid_34_3(:,2),regions_34_3);
hold on
plot(att(:,1),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y

xlim_34_4=min(att(:,2)):0.01:max(att(:,2));
ylim_34_4=min(att(:,3)):0.01:max(att(:,3));
[xgrid_34_4,ygrid_34_4]=meshgrid(xlim_34_4,ylim_34_4);
plotgrid_34_4=[xgrid_34_4(:) ygrid_34_4(:)];
subplot(3,2,4)
 regions_34_4=kmeans(plotgrid_34_4,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34_4(:,1),plotgrid_34_4(:,2),regions_34_4);
hold on
plot(att(:,2),att(:,3),'marker','*','markersize',10,'linestyle','none');
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y

xlim_34_5=min(att(:,2)):0.01:max(att(:,2));
ylim_34_5=min(att(:,4)):0.01:max(att(:,4));
[xgrid_34_5,ygrid_34_5]=meshgrid(xlim_34_5,ylim_34_5);
plotgrid_34_5=[xgrid_34_5(:) ygrid_34_5(:)];
subplot(3,2,5)
 regions_34_5=kmeans(plotgrid_34_5,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34_5(:,1),plotgrid_34_5(:,2),regions_34_5);
hold on
plot(att(:,2),att(:,4),'marker','*','markersize',10,'linestyle','none');
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
```

```
xlim_34_6=min(att(:,3)):0.01:max(att(:,3));
ylim_34_6=min(att(:,4)):0.01:max(att(:,4));
[xgrid_34_6,ygrid_34_6]=meshgrid(xlim_34_6,ylim_34_6);
plotgrid_34_6=[xgrid_34_6(:) ygrid_34_6(:)];
subplot(3,2,6)
 regions_34_6=kmeans(plotgrid_34_6,3,'MaxIter',6,'start',cent_34(:,[1 2]));
 gscatter(plotgrid_34_6(:,1),plotgrid_34_6(:,2),regions_34_6);
hold on
plot(att(:,3),att(:,4),'marker','*','markersize',10,'linestyle','none');
line(cent_34(:,1),cent_34(:,2),'marker','x','color','k','markersize',12,'linewidth',4,'linestyle','none');
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','DATA','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 3 AND 4');
[silcurve_34,h] = silhouette(att(:,[3 4]),res_34);
mean(silcurve_34)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';
```

AGGLOMERATIVE HIERARCHICAL CLUSTERING:

Code for agglomerative hierarchical clustering over 4 attributes:

```
load fisheriris;
att=meas;
class=species;
tic
agg=clusterdata(att,'maxclust',3);
toc
figure;
%dendrogram
dendrogram(linkage(squareform(pdist(att))));
grid on;
title 'CLUSTERING ON ENTIRE DATASET';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf=zeros(3,3);
for i=1:length(agg)
   if strcmp(species(i),'setosa')==1
      if agg(i)==2
         conf(1,1)=conf(1,1)+1;
      end
      if agg(i)==3
         conf(3,1)=conf(3,1)+1;
      end
      if agg(i)==1
         conf(2,1)=conf(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg(i)==2
         conf(1,2)=conf(1,2)+1;
      end
      if agg(i)==3
         conf(3,2)=conf(3,2)+1;
      end
      if agg(i)==1
         conf(2,2)=conf(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg(i)==2
         conf(1,3)=conf(1,3)+1;
      end
      if agg(i)==3
         conf(3,3)=conf(3,3)+1;
      end
      if agg(i)==1
         conf(2,3)=conf(2,3)+1;
      end

   end
end
purity=(max(conf(1,:))+max(conf(2,:))+max(conf(3,:)))./150


%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y
figure('Name','CLUSTERING ON ENTIRE DATASET');
subplot(3,2,1)
gscatter(att(:,1),att(:,2),agg);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
```

```
ylabel 'SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2);
gscatter(att(:,1),att(:,3),agg);
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
ylabel 'PETAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3);
gscatter(att(:,1),att(:,4),agg);
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4);
gscatter(att(:,2),att(:,3),agg);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH';
ylabel 'PETAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5);
gscatter(att(:,2),att(:,4),agg);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6);
gscatter(att(:,3),att(:,4),agg);
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH';
ylabel 'PETAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;

%CLUSTERING GRAPH WRT TO  ATT1-X,ATT2-Y AND ATT3-Z
figure('Name','CLUSTERING ON ENTIRE DATASET');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),agg);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT2-X,ATT3-Y AND ATT4-Z
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),agg);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT1-X,ATT2-Y AND ATT4-Z
subplot(2,2,4);
grid on;
```

```
gscatter3(att(:,1),att(:,2),att(:,4),agg);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%CLUSTERING GRAPH WRT TO  ATT1-X,ATT3-Y AND ATT4-Z
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),agg);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
figure('Name','CLUSTERING ON ENTIRE DATASET');
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%silhouette curve
[silcurve,h] = silhouette(att,agg);
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';
display(mean(silcurve));
```

Code for agglomerative hierarchical clustering over 3 attributes:

```
load fisheriris;
att=meas;
class=species;
%CLUSTER BASED ON ATTRIBUTE 1,2 & 3
tic
agg_123=clusterdata(att(:,[1 2 3]),'maxclust',3);
figure;
%dendrogram
dendrogram(linkage(squareform(pdist(att))));
grid on;
title 'CLUSTERING ON PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTHT';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
toc
%purity
conf=zeros(3,3);
for i=1:length(agg_123)
   if strcmp(species(i),'setosa')==1
      if agg_123(i)==2
         conf(1,1)=conf(1,1)+1;
      end
      if agg_123(i)==3
         conf(3,1)=conf(3,1)+1;
      end
      if agg_123(i)==1
         conf(2,1)=conf(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_123(i)==2
         conf(1,2)=conf(1,2)+1;
      end
      if agg_123(i)==3
         conf(3,2)=conf(3,2)+1;
      end
      if agg_123(i)==1
         conf(2,2)=conf(2,2)+1;
      end

   end
```

```
   if strcmp(species(i),'virginica')==1
      if agg_123(i)==2
         conf(1,3)=conf(1,3)+1;
      end
      if agg_123(i)==3
         conf(3,3)=conf(3,3)+1;
      end
      if agg_123(i)==1
         conf(2,3)=conf(2,3)+1;
      end

   end
end
purity=(max(conf(1,:))+max(conf(2,:))+max(conf(3,:)))./150
%graph wrt attributes 1,2 and 3
figure ('Name','CLUSTER BASED ON ATTRIBUTE 1,2 & 3');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),agg_123);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),agg_123);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt attributes 1,3,4
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),agg_123);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph  wrt attributes 1,2,4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,2),att(:,4),agg_123);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
figure ('Name','CLUSTER BASED ON ATTRIBUTE 1,2 & 3');
%silhouette curve
[silcurve_123,h_123]=silhouette(att(:,[1 2 3]),agg_123);
mean(silcurve_123)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
%CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4
tic
agg_234=clusterdata(att(:,[2 3 4]),'maxclust',3);
figure;
%dendrogram
dendrogram(linkage(squareform(pdist(att(:,[2 3 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND PETAL LENGTH AND SEPAL WIDTH';
xlabel 'Indies of Points';
```

```
ylabel 'Distance b/w Clusters';
toc
%purity
conf2=zeros(3,3);
for i=1:length(agg_234)
   if strcmp(species(i),'setosa')==1
      if agg_234(i)==2
         conf2(1,1)=conf2(1,1)+1;
      end
      if agg_234(i)==3
         conf2(3,1)=conf2(3,1)+1;
      end
      if agg_234(i)==1
         conf2(2,1)=conf2(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_234(i)==2
         conf2(1,2)=conf2(1,2)+1;
      end
      if agg_234(i)==3
         conf2(3,2)=conf2(3,2)+1;
      end
      if agg_234(i)==1
         conf2(2,2)=conf2(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_234(i)==2
         conf2(1,3)=conf2(1,3)+1;
      end
      if agg_234(i)==3
         conf2(3,3)=conf2(3,3)+1;
      end
      if agg_234(i)==1
         conf2(2,3)=conf2(2,3)+1;
      end

   end
end
purity=(max(conf2(1,:))+max(conf2(2,:))+max(conf2(3,:)))./150

%graph between 1,2,3
 figure ('Name','CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4');
 subplot(2,2,1);
 grid on;
 gscatter3(att(:,1),att(:,2),att(:,3),agg_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between 2,3,4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),agg_234);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between 1,2,4
subplot(2,2,3);
grid on
```

```
gscatter3(att(:,1),att(:,2),att(:,4),agg_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between 1,3,4
subplot(2,2,4);
 grid on;
gscatter3(att(:,1),att(:,3),att(:,4),agg_234);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%silhouette curve
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 2,3 AND 4');
[silcurve_234,h_234]=silhouette(att(:,[2 3 4]),agg_234);
mean(silcurve_234)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
%CLUSTERING BETWEEN 1,2 AND 4
tic
agg_124=clusterdata(att(:,[1 2 4]),'maxclust',3);
toc
figure;
%dendrogram
dendrogram(linkage(squareform(pdist(att(:,[1 2 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND SEPAL WIDTH AND SEPAL LENGTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf3=zeros(3,3);
for i=1:length(agg_124)
   if strcmp(species(i),'setosa')==1
      if agg_124(i)==2
         conf3(1,1)=conf3(1,1)+1;
      end
      if agg_124(i)==3
         conf3(3,1)=conf3(3,1)+1;
      end
      if agg_124(i)==1
         conf3(2,1)=conf3(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_124(i)==2
         conf3(1,2)=conf3(1,2)+1;
      end
      if agg_124(i)==3
         conf3(3,2)=conf3(3,2)+1;
      end
      if agg_124(i)==1
         conf3(2,2)=conf3(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_124(i)==2
         conf3(1,3)=conf3(1,3)+1;
      end
      if agg_124(i)==3
         conf3(3,3)=conf3(3,3)+1;
```

```
        end
        if agg_124(i)==1
            conf3(2,3)=conf3(2,3)+1;
        end

    end
end
purity=(max(conf3(1,:))+max(conf3(2,:))+max(conf3(3,:)))./150

%graph wrt 1 2 3
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1,2 AND 4');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),agg_124);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt 2 3 4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),agg_124);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt 1 2 4
subplot(2,2,3);
grid on;
gscatter3(att(:,1),att(:,2),att(:,4),agg_124);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt 1 3 4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),agg_124);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1,2 AND 4');
%silhouette curve
[silcurve_124,h_124]=silhouette(att(:,[1 2 4]),agg_124);
mean(silcurve_124)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';

% CLUSTERING BASED ON 1, 3 AND 4
tic
agg_134=clusterdata(att(:,[1 3 4]),'maxclust',3);
figure;
%denderogram
dendrogram(linkage(squareform(pdist(att(:,[1 3 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND PETAL LENGTH AND SEPAL LENGTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
```

```
conf4=zeros(3,3);
toc
for i=1:length(agg_134)
   if strcmp(species(i),'setosa')==1
      if agg_134(i)==2
         conf4(1,1)=conf4(1,1)+1;
      end
      if agg_134(i)==3
         conf4(3,1)=conf4(3,1)+1;
      end
      if agg_134(i)==1
         conf4(2,1)=conf4(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_134(i)==2
         conf4(1,2)=conf4(1,2)+1;
      end
      if agg_134(i)==3
         conf4(3,2)=conf4(3,2)+1;
      end
      if agg_134(i)==1
         conf4(2,2)=conf4(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_134(i)==2
         conf4(1,3)=conf4(1,3)+1;
      end
      if agg_134(i)==3
         conf4(3,3)=conf4(3,3)+1;
      end
      if agg_134(i)==1
         conf4(2,3)=conf4(2,3)+1;
      end

   end
end
purity=(max(conf4(1,:))+max(conf4(2,:))+max(conf4(3,:)))./150
%graph wrt 1 2 3
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1, 3 AND 4');
subplot(2,2,1);
grid on;
gscatter3(att(:,1),att(:,2),att(:,3),agg_134);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
zlabel 'PETAL LENGTH (cm)';
title 'PETAL LENGTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt 2 3 4
subplot(2,2,2);
grid on;
gscatter3(att(:,2),att(:,3),att(:,4),agg_134);
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL WIDTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph wrt between 1,2,4
subplot(2,2,3);
grid on
gscatter3(att(:,1),att(:,2),att(:,4),agg_134);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
```

```
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s SEPAL WIDTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
%graph between wrt 1,3,4
subplot(2,2,4);
grid on;
gscatter3(att(:,1),att(:,3),att(:,4),agg_134);
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
zlabel 'PETAL WIDTH (cm)';
title 'PETAL WIDTH v/s PETAL LENGTH v/s SEPAL LENGTH';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
figure ('Name','CLUSTERING BASED ON ATTRIBUTE 1, 3 AND 4');
%silhouette curve
[silcurve_134,h_134]=silhouette(att(:,[1 3 4]),agg_134);
mean(silcurve_134)
xlabel 'SILHOUETTE VALUES';
ylabel 'CLUSTER REGIONS';
title 'SILHOUETTE CURVE';
```

Code for agglomerative hierarchical clustering over 2 attributes:

```
%clustering using agglomerative hierarchical clustering
load fisheriris;
att=meas;
class=species;
%CLUSTERING BASED ON ATTRIBUTE 1 AND 2
tic
agg_12=clusterdata(att(:,[1 2]),'maxclust',3);
toc
%denderogram
figure;
dendrogram(linkage(squareform(pdist(att(:,[1 2])))));
grid on;
title 'CLUSTERING ON SEPAL WIDTH AND SEPAL LENGTHT';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf1=zeros(3,3);
for i=1:length(agg_12)
   if strcmp(species(i),'setosa')==1
      if agg_12(i)==2
         conf1(1,1)=conf1(1,1)+1;
      end
      if agg_12(i)==3
         conf1(3,1)=conf1(3,1)+1;
      end
      if agg_12(i)==1
         conf1(2,1)=conf1(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_12(i)==2
         conf1(1,2)=conf1(1,2)+1;
      end
      if agg_12(i)==3
         conf1(3,2)=conf1(3,2)+1;
      end
      if agg_12(i)==1
         conf1(2,2)=conf1(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_12(i)==2
```

```
            conf1(1,3)=conf1(1,3)+1;
        end
        if agg_12(i)==3
            conf1(3,3)=conf1(3,3)+1;
        end
        if agg_12(i)==1
            conf1(2,3)=conf1(2,3)+1;
        end

    end
end
purity=(max(conf1(1,:))+max(conf1(2,:))+max(conf1(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y
figure;
subplot(2,3,1);
gscatter(att(:,1),att(:,2),agg_12);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(2,3,2);
gscatter(att(:,1),att(:,3),agg_12);
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(2,3,3);
gscatter(att(:,1),att(:,4),agg_12);
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(2,3,4);
gscatter(att(:,2),att(:,3),agg_12);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(2,3,5);
gscatter(att(:,2),att(:,4),agg_12);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(2,3,6);
gscatter(att(:,3),att(:,4),agg_12);
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 2');
[silcurve,h] = silhouette(att(:,[1 2]),agg_12);
mean(silcurve)
ylabel 'CLUSTER';
```

```
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';



%CLUSTERING BASED ON ATTRIBUTE 1 AND 3
tic
agg_13=clusterdata(att(:,[1 3]),'maxclust',3);
toc
%dendrograph
figure;
dendrogram(linkage(squareform(pdist(att(:,[1 3])))));
grid on;
title 'CLUSTERING ON PETAL LENGTH AND SEPAL LENGTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf2=zeros(3,3);
for i=1:length(agg_13)
   if strcmp(species(i),'setosa')==1
      if agg_13(i)==2
         conf2(1,1)=conf2(1,1)+1;
      end
      if agg_13(i)==3
         conf2(3,1)=conf2(3,1)+1;
      end
      if agg_13(i)==1
         conf2(2,1)=conf2(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_13(i)==2
         conf2(1,2)=conf2(1,2)+1;
      end
      if agg_13(i)==3
         conf2(3,2)=conf2(3,2)+1;
      end
      if agg_13(i)==1
         conf2(2,2)=conf2(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_13(i)==2
         conf2(1,3)=conf2(1,3)+1;
      end
      if agg_13(i)==3
         conf2(3,3)=conf2(3,3)+1;
      end
      if agg_13(i)==1
         conf2(2,3)=conf2(2,3)+1;
      end

   end
end
purity=(max(conf2(1,:))+max(conf2(2,:))+max(conf2(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 3');
subplot(3,2,1);
 gscatter(att(:,1),att(:,2),agg_13);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
```

```
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2);
gscatter(att(:,1),att(:,3),agg_13);
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3);
gscatter(att(:,1),att(:,4),agg_13);
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4);
gscatter(att(:,2),att(:,3),agg_13);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5);
gscatter(att(:,2),att(:,4),agg_13);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6);
gscatter(att(:,3),att(:,4),agg_13);
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 3');
[silcurve,h] = silhouette(att(:,[1 3]),agg_13);
mean(silcurve)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

%CLUSTERING BASED ON ATTRIBUTE 1 AND 4
tic
agg_14=clusterdata(att(:,[1 4]),'maxclust',3);
toc
%dendrogram
figure;
dendrogram(linkage(squareform(pdist(att(:,[1 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND SEPAL LENGTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf3=zeros(3,3);
for i=1:length(agg_14)
   if strcmp(species(i),'setosa')==1
      if agg_14(i)==2
         conf3(1,1)=conf3(1,1)+1;
```

```
        end
        if agg_14(i)==3
            conf3(3,1)=conf3(3,1)+1;
        end
        if agg_14(i)==1
            conf3(2,1)=conf3(2,1)+1;
        end

    end
    if strcmp(species(i),'versicolor')==1
        if agg_14(i)==2
            conf3(1,2)=conf3(1,2)+1;
        end
        if agg_14(i)==3
            conf3(3,2)=conf3(3,2)+1;
        end
        if agg_14(i)==1
            conf3(2,2)=conf3(2,2)+1;
        end

    end
    if strcmp(species(i),'virginica')==1
        if agg_14(i)==2
            conf3(1,3)=conf3(1,3)+1;
        end
        if agg_14(i)==3
            conf3(3,3)=conf3(3,3)+1;
        end
        if agg_14(i)==1
            conf3(2,3)=conf3(2,3)+1;
        end

    end
end
purity=(max(conf3(1,:))+max(conf3(2,:))+max(conf3(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y


figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 4');
subplot(3,2,1)
 gscatter(att(:,1),att(:,2),agg_14);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2)
gscatter(att(:,1),att(:,3),agg_14);

title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3)
gscatter(att(:,1),att(:,4),agg_14);

title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4)
```

```
gscatter(att(:,2),att(:,3),agg_14);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5)
gscatter(att(:,2),att(:,4),agg_14);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6)
gscatter(att(:,3),att(:,4),agg_14);
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 1 AND 4');
[silcurve_14,h] = silhouette(att(:,[1 4]),agg_14);
mean(silcurve_14)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

% CLUSTERING BASED ON ATTRIBUTE 2 AND 3
tic
agg_23=clusterdata(att(:,[2 3]),'maxclust',3);
toc
%dendogram
figure;
dendrogram(linkage(squareform(pdist(att(:,[2 3])))));
grid on;
title 'CLUSTERING ON PETAL LENGTH and SEPAL WIDTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf4=zeros(3,3);
for i=1:length(agg_23)
   if strcmp(species(i),'setosa')==1
      if agg_23(i)==2
         conf4(1,1)=conf4(1,1)+1;
      end
      if agg_23(i)==3
         conf4(3,1)=conf4(3,1)+1;
      end
      if agg_23(i)==1
         conf4(2,1)=conf4(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_23(i)==2
         conf4(1,2)=conf4(1,2)+1;
      end
      if agg_23(i)==3
         conf4(3,2)=conf4(3,2)+1;
      end
      if agg_23(i)==1
         conf4(2,2)=conf4(2,2)+1;
      end
```

```
    end
    if strcmp(species(i),'virginica')==1
        if agg_23(i)==2
            conf4(1,3)=conf4(1,3)+1;
        end
        if agg_23(i)==3
            conf4(3,3)=conf4(3,3)+1;
        end
        if agg_23(i)==1
            conf4(2,3)=conf4(2,3)+1;
        end

    end
end
purity=(max(conf4(1,:))+max(conf4(2,:))+max(conf4(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y

figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 3');
subplot(3,2,1)
 gscatter(att(:,1),att(:,2),agg_23);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2)
gscatter(att(:,1),att(:,3),agg_23);
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3)
gscatter(att(:,1),att(:,4),agg_23);
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4)
gscatter(att(:,2),att(:,3),agg_23);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5)
gscatter(att(:,2),att(:,4),agg_23);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6)
gscatter(att(:,3),att(:,4),agg_23);
title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
```

```
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 3');
[silcurve_23,h_23] = silhouette(att(:,[2 3]),agg_23);
mean(silcurve_23)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';

% CLUSTERING BASED ON ATTRIBUTE 2 AND 4
tic
agg_24=clusterdata(att(:,[2 4]),'maxclust',3);
toc
%dendogram
figure;
dendrogram(linkage(squareform(pdist(att(:,[2 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND SEPAL WIDTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
%purity
conf5=zeros(3,3);
for i=1:length(agg_24)
   if strcmp(species(i),'setosa')==1
      if agg_24(i)==2
         conf5(1,1)=conf5(1,1)+1;
      end
      if agg_24(i)==3
         conf5(3,1)=conf5(3,1)+1;
      end
      if agg_24(i)==1
         conf5(2,1)=conf5(2,1)+1;
      end

   end
   if strcmp(species(i),'versicolor')==1
      if agg_24(i)==2
         conf5(1,2)=conf5(1,2)+1;
      end
      if agg_24(i)==3
         conf5(3,2)=conf5(3,2)+1;
      end
      if agg_24(i)==1
         conf5(2,2)=conf5(2,2)+1;
      end

   end
   if strcmp(species(i),'virginica')==1
      if agg_24(i)==2
         conf5(1,3)=conf5(1,3)+1;
      end
      if agg_24(i)==3
         conf5(3,3)=conf5(3,3)+1;
      end
      if agg_24(i)==1
         conf5(2,3)=conf5(2,3)+1;
      end

   end
end
purity=(max(conf5(1,:))+max(conf5(2,:))+max(conf5(3,:)))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y


figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 4');
subplot(3,2,1)
 gscatter(att(:,1),att(:,2),agg_24);
```

```
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2)
gscatter(att(:,1),att(:,3),agg_24);

title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3)
gscatter(att(:,1),att(:,4),agg_24);
title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4)
gscatter(att(:,2),att(:,3),agg_24);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5)
gscatter(att(:,2),att(:,4),agg_24);
title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6)
gscatter(att(:,3),att(:,4),agg_24);

title 'PETAL WIDTH v/s PETAL LENGTH';
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
figure('Name','CLUSTERING BASED ON ATTRIBUTE 2 AND 4');
[silcurve_24,h_24] = silhouette(att(:,[2 4]),agg_24);
mean(silcurve_24)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';


% CLUSTERING BASED ON ATTRIBUTE 3 AND 4
tic
agg_34=clusterdata(att(:,[3 4]),'maxclust',3);
toc
%dendrogram
figure;
dendrogram(linkage(squareform(pdist(att(:,[3 4])))));
grid on;
title 'CLUSTERING ON PETAL WIDTH AND PETAL LENGTH';
xlabel 'Indies of Points';
ylabel 'Distance b/w Clusters';
```

```
%purity
conf6=zeros(3,3);
for i=1:length(agg_34)
    if strcmp(species(i),'setosa')==1
        if agg_34(i)==2
            conf6(1,1)=conf6(1,1)+1;
        end
        if agg_34(i)==3
            conf6(3,1)=conf6(3,1)+1;
        end
        if agg_34(i)==1
            conf6(2,1)=conf6(2,1)+1;
        end

    end
    if strcmp(species(i),'versicolor')==1
        if agg_34(i)==2
            conf6(1,2)=conf6(1,2)+1;
        end
        if agg_34(i)==3
            conf6(3,2)=conf6(3,2)+1;
        end
        if agg_34(i)==1
            conf6(2,2)=conf6(2,2)+1;
        end

    end
    if strcmp(species(i),'virginica')==1
        if agg_34(i)==2
            conf6(1,3)=conf6(1,3)+1;
        end
        if agg_34(i)==3
            conf6(3,3)=conf6(3,3)+1;
        end
        if agg_34(i)==1
            conf6(2,3)=conf6(2,3)+1;
        end

    end
end
purity=(max(conf6(1,:))+max(conf6(2,:))+max(conf6(3,:))))./150
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT2-Y


figure('Name','CLUSTERING BASED ON ATTRIBUTE 3 AND 4');
subplot(3,2,1)
gscatter(att(:,1),att(:,2),agg_34);
title 'SEPAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'SEPAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT3-Y
subplot(3,2,2)
gscatter(att(:,1),att(:,3),agg_34);
title 'PETAL LENGTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT1-X AND ATT4-Y
subplot(3,2,3)
gscatter(att(:,1),att(:,4),agg_34);

title 'PETAL WIDTH v/s SEPAL LENGTH';
xlabel 'SEPAL LENGTH (cm)';
```

```
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT3-Y
subplot(3,2,4)
gscatter(att(:,2),att(:,3),agg_34);
title 'PETAL LENGTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL LENGTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT2-X AND ATT4-Y
subplot(3,2,5)
gscatter(att(:,2),att(:,4),agg_34);

title 'PETAL WIDTH v/s SEPAL WIDTH';
xlabel 'SEPAL WIDTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%CLUSTERING GRAPH WRT TO ATT3-X AND ATT4-Y
subplot(3,2,6)
gscatter(att(:,3),att(:,4),agg_34);
xlabel 'PETAL LENGTH (cm)';
ylabel 'PETAL WIDTH (cm)';
legend('REGION 1','REGION 2','REGION 3','LOCATION','BEST');
hold off;
%silhouette curve
figure('Name','CLUSTERING BASED ON ATTRIBUTE 3 AND 4');
[silcurve_34,h] = silhouette(att(:,[3 4]),agg_34);
mean(silcurve_34)
ylabel 'CLUSTER';
xlabel 'SILHOUETTE VALUE';
title 'SILHOUETTE CURVE';
```

RESULTS-CLASSIFICATION:

Bar chart of accuracy of classifiers over 4 attributes:

```
nbacc=100-5.3333;
knnacc=100-4;
dtacc=100-8;
wholeset=[nbacc knnacc dtacc];
figure('Name','ACCURACY COMPARISON');
ylim([0 100]);
bar(wholeset,0.3);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('ACCURACY COMPARISON');
```

Bar chart of accuracy of classifiers over 3 attributes and average accuracy of 3 attribute classifiers:

```
nbacc3=[88,96,94.6667,94.6667];
knnacc3=[94.6667,93.3333,93.3333,96];
dtacc3=[92,93.3333,88,92];
figure('Name','ACCURACY COMPARISON');
bar([nbacc3;knnacc3;dtacc3]);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
legend('PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH-ACTUAL','location','best');
title('ACCURACY COMPARISON');
 figure('Name','ACCURACY COMPARISON');
nbavg=mean(nbacc3);
knnavg=mean(knnacc3);
dtavg=mean(dtacc3);
bar([nbavg;knnavg;dtavg],0.4);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE ACCURACY OF EACH CLASSIFIER OVER 3 ATTRIBUTES');
```

Bar chart of accuracy of classifiers over 2 attributes and average accuracy of 2 attribute classifiers:

```
nbacc2=[80,89.3333,90.6667,90.6667,94.6667,94.6667];
knnacc2=[93.3333,89.3333,93.3333,88,92,93.3333];
dtacc2=[69.3333,92,94.6667,93.3333,89.3333,93.3333];
figure('Name','ACCURACY COMPARISON');
bar([nbacc2;knnacc2;dtacc2]);
ylabel 'Accuracy (%)';
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('ACCURACY COMPARISON');
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
figure('Name','ACCURACY COMPARISON');
nbavg=mean(nbacc2);
knnavg=mean(knnacc2);
dtavg=mean(dtacc2);
bar([nbavg knnavg dtavg],0.4);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE ACCURACY OF EACH CLASSIFIER OVER 2 ATTRIBUTES');
```

Bar chart of accuracy of all classifiers and average of each algorithm:

```
nbaccfull=[94.6667,88,96,94.6667,94.6667,80,89.3333,90.6667,90.6667,94.6667,94.6667];
knnaccfull=[96,94.6667,93.3333,93.3333,96,93.3333,89.3333,93.3333,88,92,93.3333];
dtaccfull=[92,92,93.3333,88,92,69.3333,92,94.6667,93.3333,89.3333,93.3333];
figure('Name','ACCURACY COMPARISON');
bar([nbaccfull;knnaccfull;dtaccfull]);
```

```
ylabel 'Accuracy (%)';
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('ACCURACY COMPARISON');
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
%average for each algorithm
nbfullaccavg=mean(nbaccfull);
knnfullaccavg=mean(knnaccfull);
dtfullaccavg=mean(dtaccfull);
figure('Name','ACCURACY COMPARISON');
bar([nbfullaccavg;knnfullaccavg;dtfullaccavg],0.3);
grid on;
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE ACCURACY COMPARISON OF ALL CLASSIFIERS FOR EACH ALGORITHM');
```

Bar chart of time of classifiers over 4 attributes:

```
nbtime=0.1691;
att={'NB','KNN','DT'};
knntime=0.19601;
dttime=0.129402;
wholeset=[nbtime knntime dttime];
figure('Name','TIME COMPARISON');
bar(wholeset,0.3);
grid on;
ylabel 'TIME (s)';
set(gca,'XTickLabel',att,'YGrid','on');
title('TIME COMPARISON');
```

Bar chart of time of classifiers over 3 attributes and average time of 3 attribute classifiers for each algorithm:

```
nbtime3=[0.020681,0.021303,0.023927,0.018614];
knntime3=[0.026067,0.027566,0.019711,0.024565];
dttime3=[0.007909,0.008138,0.007875,0.009617];
figure('Name','TIME COMPARISON');
bar([nbtime3;knntime3;dttime3]);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
legend('PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','location','best');
title('TIME COMPARISON');
 figure('Name','TIME COMPARISON');
nbavg=mean(nbtime3);
knnavg=mean(knntime3);
dtavg=mean(dttime3);
bar([nbavg;knnavg;dtavg],0.4);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE TIME  OF EACH CLASSIFIER OVER 3 ATTRIBUTES');
```

Bar chart of time of classifiers over 2 attributes and average time of 3 attribute classifiers for each algorithm:

```
nbtime2=[0.010596,0.016948,0.013831,0.016223,0.013999,0.01191];
knntime2=[0.017959,0.015347,0.018308,0.018493,0.014737,0.01731];
dttime2=[0.0064,0.006715,0.004689,0.007387,0.006183,0.005095];
figure('Name','TIME COMPARISON');
bar([nbtime2;knntime2;dttime2]);
grid on;
legend('SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
```

```
ylabel 'TIME (s)';
title('TIME COMPARISON');
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
figure('Name','TIME COMPARISON');
nbavg=mean(nbtime2);
knnavg=mean(knntime2);
dtavg=mean(dttime2);
bar([nbavg knnavg dtavg],0.4);
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE TIME OF EACH CLASSIFIER OVER 2 ATTRIBUTES');
```

Bar chart of time of all classifiers and average of each algorithm:

```
nbtimefull=[0.1691,0.020681,0.021303,0.023927,0.018614,0.010596,0.016948,0.013831,0.016223,0.013999,0.01191];
knntimefull=[0.19601,0.026067,0.027566,0.019711,0.024565,0.017959,0.015347,0.018308,0.018493,0.014737,0.01731];
dttimefull=[0.129402,0.007909,0.008138,0.007875,0.009617,0.0064,0.006715,0.004689,0.007387,0.006183,0.005095];
figure('Name','TIME COMPARISON');
bar([nbtimefull;knntimefull;dttimefull]);
ylabel 'Time (s)';
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH &
SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL
WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('TIME COMPARISON');
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
%average for each algorithm
nbfulltime=mean(nbtimefull);
knnfulltime=mean(knntimefull);
dtfulltime=mean(dttimefull);
figure('Name','TIME COMPARISON');
bar([nbfulltime;knnfulltime;dtfulltime],0.3);
grid on;
ylabel 'Time (s)';
set(gca,'XTickLabel',{'NBC','KNNC','DTC'},'YGrid','on');
title('AVERAGE TIME COMPARISON OF ALL CLASSIFIERS FOR EACH ALGORITHM');set(gca,'XTickLabel',
{'NBC','KNNC','DTC'},'YGrid','on');
```

Accuracy v/s Time graphs of 4 attribute classifiers, 3 attribute classifiers, 2 attribute classifiers and all classifiers

```
data0=[dtacc2',dttime2'];
data0class=['DTC';'DTC';'DTC';'DTC';'DTC';'DTC'];
data0type=['ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2'];
data1=[dtacc3',dttime3'];
data1class=['DTC';'DTC';'DTC';'DTC'];
data1type=['ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3'];
data2=[nbacc2',nbtime2'];
data2class=['NBC';'NBC';'NBC';'NBC';'NBC';'NBC'];
data2type=['ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2'];
data3=[nbacc3',nbtime3'];
data3class=['NBC';'NBC';'NBC';'NBC'];
data3type=['ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3'];
data4=[knnacc2',knntime2'];
data4class=['KNN';'KNN';'KNN';'KNN';'KNN';'KNN'];
data4type=['ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2';'ATTRIBUTE 2'];
data5=[knnacc3',knntime3'];
data5type=['ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3';'ATTRIBUTE 3'];
data5class=['KNN';'KNN';'KNN';'KNN'];
data6=[nbacc nbtime];
data6class='NBC';
data6type='ATTRIBUTE 4';
data7=[knnacc knntime];
data7class='KNN';
data7type='ATTRIBUTE 4';
data8=[dtacc dttime];
```

```
data8class='DTC';
data8type='ATTRIBUTE 4';
%wholedataset accuracy time
accuracy_time_data=[data0;data1;data2;data3;data4;data5;data6;data7;data8];
accuracy_time_datatype=[data0type;data1type;data2type;data3type;data4type;data5type;data6type;data7type;data8type];
accuracy_time_dataclass=[data0class;data1class;data2class;data3class;data4class;data5class;data6class;data7class;data8class];
%4 attributes accuracy time
accuracy_time4=[data6;data7;data8];
accuracy_time4class=[data6class;data7class;data8class];
% 2 attribute accuracy time
accuracy_time2=[data0;data2;data4];
accuracy_time2class=[data0class;data2class;data4class];
%3 attribute accuracy time
accuracy_time3=[data1;data3;data5];
accuracy_time3class=[data1class;data3class;data5class];
%accuracy time for 2 attributes
figure('Name','ACCURACY v/s TIME FOR CLASSIFICATION OVER 2 ATTRIBUTES');
gscatter(accuracy_time2(:,2),accuracy_time2(:,1),accuracy_time2class);
grid on;
ylim([65 100]);
title('ACCURACY v/s TIME FOR CLASSIFICATION OVER 2 ATTRIBUTES');
%accuracy time for 3 attributes
figure('Name','ACCURACY v/s TIME FOR CLASSIFICATION OVER 3 ATTRIBUTES');
gscatter(accuracy_time3(:,2),accuracy_time3(:,1),accuracy_time3class);
grid on;
ylim([80 100]);
title('ACCURACY v/s TIME FOR CLASSIFICATION OVER 3 ATTRIBUTES');
% %accuracy time for 4 attributes
figure('Name','ACCURACY v/s TIME FOR CLASSIFICATION OVER 4 ATTRIBUTES');
gscatter(accuracy_time4(:,2),accuracy_time4(:,1),accuracy_time4class);
grid on;
ylim([90 100]);
title('ACCURACY v/s TIME FOR CLASSIFICATION OVER 4 ATTRIBUTES');

accuracy over entire dataset
figure('Name','ACCURACY v/s TIME FOR ALL CLASSIFIERS');
gscatter(accuracy_time_data(:,2),accuracy_time_data(:,1),accuracy_time_dataclass);
line(data7(:,2),data7(:,1),'marker','p','color','k','linewidth',1);
line(data0(3,2),data0(3,1),'marker','d','color','k','linewidth',1);
line(data0(3,2),data0(3,1),'marker','s','color','k','linewidth',1);
line(data5(4,2),data5(4,1),'marker','p','color','k','linewidth',1);
line(data3(2,2),data3(2,1),'marker','p','color','k','linewidth',1);
line(data8(:,2),data8(:,1),'marker','o','color','k','linewidth',1,'linestyle','none');
line(data3(4,2),data3(4,1),'marker','o','color','k','linewidth',1,'linestyle','none');
line(data6(:,2),data6(:,1),'marker','o','color','k','linewidth',1,'linestyle','none');
line(data1(4,2),data1(4,1),'marker','o','color','k','linewidth',1,'linestyle','none');
x1=[data7(:,2),data5(4,2)];
y1=[data7(:,1),data5(4,1)];
x2=[data6(:,2),data3(4,2)];
y2=[data6(:,1),data3(4,1)];
x3=[data8(:,2),data1(4,2)];
y3=[data8(:,1),data1(4,1)];
line(x1,y1,'linewidth',1,'color','k');
line(x2,y2,'linewidth',1,'color','k');
line(x3,y3,'linewidth',1,'color','k');
grid on;
%line(accuracy_time_data(:,2),accuracy_time_data(:,1),'marker','*','color','k','linestyle','none');
ylim([65 100]);
title('ACCURACY v/s TIME FOR ALL CLASSIFIERS');
xlabel 'Time (s)';
ylabel 'Accuracy (%)';
legend('DECISION TREE CLASSIFIER','NAIVE BAYE''S CLASSIFIER','K-NEAREST NEIGHBOUR CLASSIFIER','MOST
ACCURATE CLASSIFIER','FASTEST CLASSIFIER','PREFERRED CLASSIFIER','LOCATION','BEST');
```

RESULTS-CLUSTERING:

Bar chart of purity and silhouette values of cluster models over 4 attributes:

```
kp=89.33;
ahp=68;
wholeset=[kp ahp];
figure('Name','ACCURACY COMPARISON');
ylim([0 100]);
bar(wholeset,0.3);
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE PURITY COMPARISON OVER THE WHOLEDATASET');
%silhouette graphs
ks=0.7357;
ahs=0.6184;
wholeset_s=[ks ahs];
figure('Name','SILHOUETTE VALUE COMPARISON');
ylim([0 100]);
bar(wholeset_s,0.3);
ylabel 'Silhouette Values';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE SILHOUETTE VALUE OVER THE WHOLE DATASET');
```

Bar chart of purity and silhouette values of cluster models over 3 attributes:

```
kp3=[88,95.33,82.67,89.33];
ahp3=[68,68,66.67,69.33];
figure('Name','AVERAGE PURITY COMPARISON');
bar([kp3;ahp3]);
grid on;
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('PURITY OF CLUSTER MODELS OVER 3 ATTRIBUTES');
%averaged purity
 figure('Name','PURITY COMPARISON');
kpavg=mean(kp3);
ahpavg=mean(ahp3);
bar([kpavg;ahpavg],0.3);
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
ylabel 'Accuracy (%)';
title('AVERAGE PURITY OF EACH CLUSTER MODEL OVER 3 ATTRIBUTES');

%silhouette graphs
ks3=[0.7330,0.7659,0.6654,0.7523];
ahs3=[0.6433,0.5985,0.5999,0.5934];
ks2=[0.6201,0.7560,0.6813,0.7612,0.7067,0.8055];
ahs2=[0.1028,0.4068,-0.3799,0.8019,0.6777,0.6221];
figure('Name','SILHOUETTE VALUE COMPARISON');
bar([ks3;ahs3]);
grid on;
ylabel 'Silhouette Value';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('SILHOUETTE VALUES OF CLUSTER MODELS OVER 3 ATTRIBUTES');
%averaged silvalue
figure('Name','SILHOUETTE VALUE COMPARISON');
ksavg=mean(ks3);
ahsavg=mean(ahs3);
bar([ksavg;ahsavg],0.3);
ylabel 'Silhouette Value';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE SILHOUETTE VALUE OF EACH CLUSTER MODEL OVER 3 ATTRIBUTES');
```

Bar chart of purity and silhouette values of cluster models over 2 attributes:

```
kp2=[82, 88,82.67,92.67,92.67,96];
ahp2=[35.33,67.33,34.67,66.67,66.67,67.33];
```

```
figure('Name','AVERAGE PURITY COMPARISON');
bar([kp2;ahp2]);
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('PURITY OF CLUSTER MODELS OVER 2 ATTRIBUTES');
%averaged purity
figure('Name','PURITY COMPARISON');
kpavg=mean(kp2);
ahpavg=mean(ahp2);
bar([kpavg;ahpavg],0.3);
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE PURITY OF CLUSTER MODELS OVER 2 ATTRIBUTES');

%silhouette graphs

ks2=[0.6201,0.7560,0.6813,0.7612,0.7067,0.8055];
ahs2=[0.1028,0.4068,-0.3799,0.8019,0.6777,0.6221];
figure('Name','SILHOUETTE VALUE COMPARISON');
bar([ks2;ahs2]);
ylabel 'Silhouette Value';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('SILHOUETTE VALUES OF CLUSTER MODELS OVER 2 ATTRIBUTES');
 %averaged silvalue
figure('Name','AVERAGED SILHOUETTE VALUE COMPARISON');
ksavg=mean(ks2);
ahsavg=mean(ahs2);
bar([ksavg ahsavg],0.3);
ylabel 'Silhouette Value';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE SILHOUETTE VALUE  OF CLUSTER MODELS OVER 2 ATTRIBUTES');
```

Bar chart of purity and silhouette values of cluster models of all cluster models and averaged purity and silhouette values of all cluster models for each algorithm:

```
k_purity=[89.33,88,95.33,82.67,89.33,82,88,82.67,92.67,92.67,96];
a_purity=[68,68,68,66.67,69.33,35.33,67.33,34.67,66.67,66.67,67.33];
k_sil=[0.7357,0.7330,0.7659,0.6654,0.7523,0.6201,0.7560,0.6813,0.7612,0.7067,0.8055];
a_sil=[0.6184,0.6433,0.5985,0.5999,0.5934,0.1028,0.4068,-0.3799,0.8019,0.6777,0.6221];
clustbar=[k_purity;a_purity];
silbar=[k_sil;a_sil];
figure;
bar(clustbar);
grid on;
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means','Agglomerative Hierarchical Clustering'},'YGrid','on');
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH &
SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL
WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('ACCURACY COMPARISON');
%averaged bar graph
figure;
kpavg=mean(k_purity);
ahpavg=mean(a_purity);
bar([kpavg;ahpavg],0.3);
grid on;
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE ACCURACY COMPARISON');
%silhouette graphs
figure;
bar(clustbar);
grid on;
ylabel 'Silhouette Value';
```

```
set(gca,'XTickLabel',{'K-Means','Agglomerative Hierarchical Clustering'},'YGrid','on');
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('SILHOUETTE VALUE COMPARISON');
%averaged bar graph
figure;
ksavg=mean(k_sil);
ahsavg=mean(a_sil);
bar([ksavg;ahsavg],0.3);
grid on;
ylabel 'Accuracy (%)';
set(gca,'XTickLabel',{'K-Means','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE ACCURACY COMPARISON');
```

Bar chart of time of cluster models over 4 attributes:

```
ktime=0.0097835;
ahtime=0.009507;
wholeset=[ktime ahtime];
figure('Name','TIME COMPARISON');
ylim([0 100]);
bar(wholeset,0.3);
ylabel 'Time (s)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE TIME COMPARISON');
```

Bar chart of time of cluster models over 3 attributes and average time of 3 attribute cluster models for each algorithm:

```
ktime3=[0.0075335,0.00829,0.0090135,0.00589225];
ahtime3=[0.010582,0.007566,0.004728,0.005387];

figure('Name','TIME COMPARISON');
bar([kttime3;ahttime3]);
ylabel 'Accuracy (%)';
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('ACCURACY COMPARISON');
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
%average timings
figure('Name','TIME COMPARISON');
ktavg=mean(ktime3);
ahtavg=mean(ahtime3);
bar([ktavg;ahtavg],0.3);
ylabel 'Time (s)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE TIME OF EACH CLASSIFIER OVER 3 ATTRIBUTES');
```

Bar chart of time of cluster models over 2 attributes and average time of 2 attribute cluster models for each algorithm:

```
kt2=[0.00891975,0.00694425,0.0070455,0.0097835,0.0075335,0.00829];
aht2=[0.005569,0.005797,0.00942,0.008267,0.02518,0.008337];
figure('Name','TIME COMPARISON');
ktavg=mean(kt2);
ahtavg=mean(aht2);
bar([ktavg ahtavg],0.3);
ylabel 'Time (s)';
```

```
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE TIME OF EACH CLASSIFIER OVER 2 ATTRIBUTES');
figure('Name','TIME COMPARISON');
%average timings
ktavg=mean(kt2);
ahtavg=mean(aht2);
bar([ktavg;ahtavg],0.3);
ylabel 'Time (s)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE TIME OF EACH CLASSIFIER OVER 3 ATTRIBUTES');
```

Bar chart of time of all cluster models and average of each algorithm:

```
k_time=[0.0097835,0.0075335,0.00829,0.0090135,0.00589225,0.00891975,0.00694425,0.0070455,0.0097835,0.0075335,0.0082
9];
a_time=[0.009507,0.010582,0.007566,0.004728,0.005387,0.005569,0.005797,0.00942,0.008267,0.02518,0.008337];
timebar=[k_time;a_time];
figure;
bar(timebar);
grid on;
ylabel 'Time (s)';
set(gca,'XTickLabel',{'K-Means','Agglomerative Hierarchical Clustering'},'YGrid','on');
legend('PETAL WIDTH,PETAL LENGTH, SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','PETAL WIDTH, SEPAL WIDTH & SEPAL
LENGTH','PETAL WIDTH, PETAL LENGTH & SEPAL LENGTH','SEPAL WIDTH & SEPAL LENGTH','PETAL LENGTH &
SEPAL LENGTH','PETAL WIDTH & SEPAL LENGTH','PETAL LENGTH & SEPAL WIDTH','PETAL WIDTH & SEPAL
WIDTH','PETAL WIDTH & PETAL LENGTH','location','best');
title('TIME COMPARISON')
%average timings
ktavg=mean(k_time);
ahtavg=mean(a_time);
bar([ktavg;ahtavg],0.3);
ylabel 'Time (s)';
set(gca,'XTickLabel',{'K-Means Clustering','Agglomerative Hierarchical Clustering'},'YGrid','on');
title('AVERAGE TIME OF EACH CLASSIFIER OF EACH ALGORITHM OVER THE WHOLEDATASET');
```

Purity v/s Time and Purity v/s Silhouette plots for all cluster models, 4 attribute cluster models, 3 attribute cluster models and 2 attribute cluster models

```
%graphs
k_time=[0.0097835,0.0075335,0.00829,0.0090135,0.00589225,0.00891975,0.00694425,0.0070455,0.0097835,0.0075335,0.0082
9];
k2_time=k_time(:,[6 7 8 9 10 11]);
k3_time=k_time(:,[2 3 4 5]);
k4_time=k_time(:,1);
k_class=['K','K','K','K','K','K','K','K','K','K','K'];
k4_class=k_class(:,1);
k3_class=k_class(:,[2 3 4 5]);
k2_class=k_class(:,[6 7 8 9 10 11]);
a_class=['A','A','A','A','A','A','A','A','A','A','A'];
a_time=[0.009507,0.010582,0.007566,0.004728,0.005387,0.005569,0.005797,0.00942,0.008267,0.02518,0.008337];
a4_class=a_class(:,1);
a3_class=a_class(:,[2 3 4 5]);
a2_class=a_class(:,[6 7 8 9 10 11]);
a2_time=a_time(:,[6 7 8 9 10 11]);
a3_time=a_time(:,[2 3 4 5]);
a4_time=a_time(:,1);
k_purity=[89.33,88,95.33,82.67,89.33,82,88,82.67,92.67,92.67,96];
k2_purity=k_purity(:,[6 7 8 9 10 11]);
k3_purity=k_purity(:,[2 3 4 5]);
k4_purity=k_purity(:,1);
k_sil=[0.7357,0.7330,0.7659,0.6654,0.7523,0.6201,0.7560,0.6813,0.7612,0.7067,0.8055];
k2_sil=k_sil(:,[6 7 8 9 10 11]);
k3_sil=k_sil(:,[2 3 4 5]);
k4_sil=k_sil(:,1);
```

```
a_purity=[68,68,68,66.67,69.33,35.33,67.33,34.67,66.67,66.67,67.33];
a2_purity=a_purity(:,[6 7 8 9 10 11]);
a3_purity=k_purity(:,[2 3 4 5]);
a4_purity=k_purity(:,1);
a_sil=[0.6184,0.6433,0.5985,0.5999,0.5934,0.1028,0.4068,-0.3799,0.8019,0.6777,0.6221];
a2_sil=a_sil(:,[6 7 8 9 10 11]);
a3_sil=a_sil(:,[2 3 4 5]);
a4_sil=a_sil(:,1);
clust=[k_purity a_purity];
clust2=[k2_purity a2_purity];
clust3=[k3_purity a3_purity];
clust4=[k4_purity a4_purity];
sil=[k_sil a_sil];
sil2=[k2_sil a2_sil];
sil3=[k3_sil a3_sil];
sil4=[k4_sil a4_sil];
time=[k_time a_time];
time2=[k2_time a2_time];
time3=[k3_time a3_time];
time4=[k4_time a4_time];
class=[k_class a_class];
class2=[k2_class a2_class];
class3=[k3_class a3_class];
class4=[k4_class a4_class];
purgra=[clust' time'];
purgra2=[clust2' time2'];
purgra3=[clust3' time3'];
purgra4=[clust4' time4'];
silgra=[sil',clust'];
silgra2=[sil2' clust2'];
silgra3=[sil3' clust3'];
silgra4=[sil4' clust4'];

%purity vs time graph_allpoints
figure;
gscatter(purgra(:,2),purgra(:,1),class');
line(purgra(11,2),purgra(11,1),'marker','s','color','k','linewidth',1);
line(purgra(11,2),purgra(11,1),'marker','p','color','k','linewidth',1);
line(purgra(15,2),purgra(15,1),'marker','o','color','k','linewidth',1);
grid on;
title('PURITY v/s TIME');
xlabel 'Time (s)';
ylabel('Purity (%)');
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','PUREST CLUSTERING','PREFERRED CLUSTERING','FASTEST CLUSTERING','location','best');
%purity vs time graph_4attributes
figure;
gscatter(purgra4(:,2),purgra4(:,1),class4');
grid on;
title('PURITY v/s TIME OF 4 ATTRIBUTE CLUSTER MODELS');
xlabel 'Time (s)';
ylabel('Purity (%)');
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','location','best');
%purity vs time graph_3attribute
figure;
gscatter(purgra3(:,2),purgra3(:,1),class3');
grid on;
title('PURITY v/s TIME OF 3 ATTRIBUTE CLUSTER MODELS');
xlabel 'Time (s)';
ylabel('Purity (%)');
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','location','best');
%purity vs time graph_2attribute
figure;
gscatter(purgra2(:,2),purgra2(:,1),class2');
grid on;
title('PURITY v/s TIME OF 2 ATTRIBUTE CLUSTER MODELS');
```

```
xlabel 'Time (s)';
ylabel('Purity (%)');
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','location','best');
%purity vs silhouette_wholedataset
figure;
gscatter(silgra(:,1),silgra(:,2),class');
line(silgra(11,1),silgra(11,2),'marker','o','color','k','linewidth',1);
grid on;
title('PURITY v/s SILHOUETTE VALUE');
xlabel 'Silhouette Value';
ylabel 'PURITY (%)';
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','MOST EFFICIENT
CLUSTERING','location','best');
%purity vs silhouette_4attribute
figure;
gscatter(silgra4(:,1),silgra4(:,2),class4');
grid on;
title('PURITY v/s SILHOUETTE VALUE OF 4 ATTRIBUTE CLUSTER MODELS');
xlabel 'Silhouette Value';
ylabel 'PURITY (%)';
%purity vs silhouette_3attribute
figure;
gscatter(silgra3(:,1),silgra3(:,2),class3');
grid on;
title('PURITY v/s SILHOUETTE VALUE OF 3 ATTRIBUTE CLUSTER MODELS');
xlabel 'Silhouette Value';
ylabel 'PURITY (%)';
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','location','best');
%purity vs silhouette_2attribute
figure;
gscatter(silgra2(:,1),silgra2(:,2),class2');
grid on;
title('PURITY v/s SILHOUETTE VALUE OF 2 ATTRIBUTE CLUSTER MODELS ');
xlabel 'Silhouette Value';
ylabel 'PURITY (%)';
legend('K-MEANS CLUSTERING','AGGLOMERATIVE HIERARCHICAL CLUSTERING','location','best');
```