

Name = Tariq Bilal

ID = 13588

Class = Bs(cs)

Q1 ) Distinguish between Classification and Regression with the help of relevant scenarios.?

## Classification vs Regression

Ans) Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data.

Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features.

### **Example of Regression:**

Predictive percentage of precipitation .

Prediction of diabetic patients.

Astronomy prediction.

### **Example of classification:**

Who will voters vote for?

Who will win most matches?

Hidden Patterns.

Q2) Perform Naïve Bayes or Decision tree classification for new instance where (SSN = 123-46-4455, Test1= 85, Test2= 31 and Final= 30) Find Grade.

Ans) Decision tree: predict the class label – Bayesian classifier: statistical classifier; predict class membership probabilities.

Naïve Bayesian classifier:

- Simple classifier that assumes attribute independence
- Efficient when applied to large databases
- Comparable in performance to decision trees

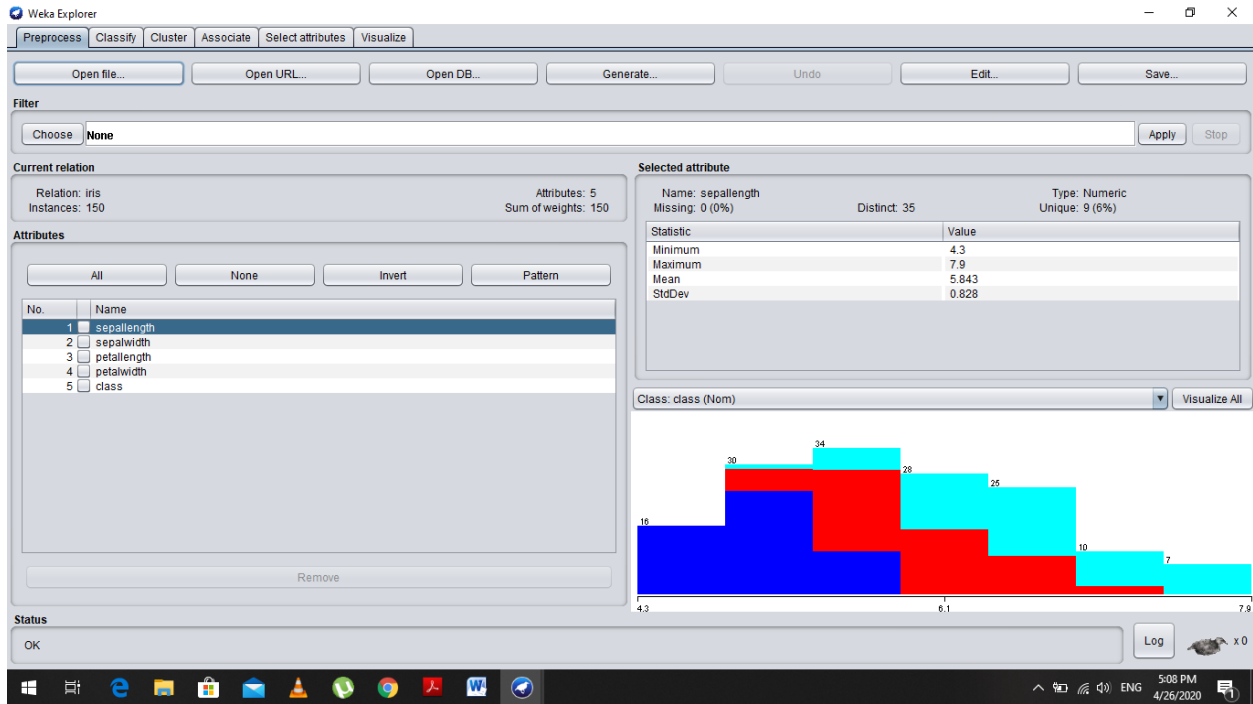
$$P(H_i | X) = \frac{P(X|H_i) P(H_i)}{P(X)}$$

Q3

Ans)

### Opening File

Adding filter (Numeric to Nominal) so Naïve Bayes could be applied.



## Applying Naïve Bayes

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4" -seed 1

Test options

Use training set  
 Supplied test set   
 Cross-validation Folds 10  
 Percentage split % 66

(Nom) class

Result list (right-click for options)

- 17:06:47 - bayes NaiveBayes
- 17:07:05 - trees REPTree
- 17:07:14 - functions LibSVM

Classifier output

```

=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    iris
Instances:   150
Attributes:  5
  sepallength
  sepalwidth
  petallength
  petalwidth
  class
Test mode:   split 66.0% train, remainder test

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute      Class
              Iris-setosa Iris-versicolor Iris-virginica
-----
sepallength
  mean          4.9913      5.9379      6.5795
  std. dev.     0.355      0.5042      0.6353
  weight sum    50           50           50
  precision     0.1059      0.1059      0.1059
sepalwidth
  mean          3.4015      2.7687      2.9629
  std. dev.     0.3925      0.3038      0.3088
  weight sum    50           50           50
  precision     0.1091      0.1091      0.1091
petallength
  mean          1.4694      4.2452      5.5516
  std. dev.     0.1782      0.4712      0.5529
  weight sum    50           50           50
  precision     0.1405      0.1405      0.1405
petalwidth
  mean          0.2743      1.3097      2.0343
  std. dev.     0.1096      0.1915      0.2646
  weight sum    50           50           50
  precision     0.1143      0.1143      0.1143

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

```

Status

OK  x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4" -seed 1

Test options

Use training set  
 Supplied test set   
 Cross-validation Folds 10  
 Percentage split % 66

(Nom) class

Result list (right-click for options)

- 17:06:47 - bayes NaiveBayes
- 17:07:05 - trees REPTree
- 17:07:14 - functions LibSVM

Classifier output

```

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

```

Status

OK  x 0

## Naïve Bayes summary

The screenshot shows the Weka Explorer interface with the Naïve Bayes classifier selected. The classifier output window displays the following summary and detailed accuracy by class:

```

=== Summary ===
Correctly Classified Instances      48           94.1176 %
Incorrectly Classified Instances     3            5.8824 %
Kappa statistic                     0.9113
Mean absolute error                  0.0447
Root mean squared error              0.1722
Relative absolute error              10.0365 %
Root relative squared error          36.4196 %
Total Number of Instances           51

=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Iris-setosa      1.000   0.000    1.000     1.000   1.000     1.000    1.000    1.000    Iris-setosa
Iris-versicolor  0.947   0.063    0.900     0.947   0.923     0.876    0.988    0.980    Iris-versicolor
Iris-virginica   0.882   0.029    0.938     0.882   0.909     0.867    0.988    0.980    Iris-virginica
Weighted Avg.    0.941   0.033    0.942     0.941   0.941     0.909    0.992    0.986

=== Confusion Matrix ===
  a  b  c  <-- classified as
15  0  0 | a = Iris-setosa
 0 18  1 | b = Iris-versicolor
 0  2 15 | c = Iris-virginica
    
```

The result list on the left shows three classifiers: Naïve Bayes, REPTree, and LibSVM. The Naïve Bayes classifier is selected.

## Applying Decision Tree

We are going to select classifier and choose **REPTree** (which is also Decision Tree).

The screenshot shows the Weka Explorer interface with the REPTree classifier selected. The classifier output window displays the following run information and classifier model:

```

=== Run information ===
Scheme:      weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Relation:    iris
Instances:   150
Attributes:  5
  sepallength
  sepalwidth
  petallength
  petalwidth
  class
Test mode:   split 66.0% train, remainder test

=== Classifier model (full training set) ===

REPTree
=====
petallength < 2.5 : Iris-setosa (33/0) [17/0]
petallength >= 2.5
| petalwidth < 1.75 : Iris-versicolor (36/3) [18/2]
| petalwidth >= 1.75 : Iris-virginica (31/1) [15/0]

Size of the tree : 5

Time taken to build model: 0 seconds
    
```

The result list on the left shows three classifiers: Naïve Bayes, REPTree, and LibSVM. The REPTree classifier is selected.

Decision Tree summary:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4" -seed 1

Test options

Use training set  
 Supplied test set (Set...)  
 Cross-validation (Folds: 10)  
 Percentage split (%: 66)

(Nom) class

Start Stop

Result list (right-click for options)

- 17:06:47 - bayes.NaiveBayes
- 17:07:05 - trees.REPTree
- 17:07:14 - functions.LibSVM

Classifier output

```

time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      47           92.1569 %
Incorrectly Classified Instances     4           7.8431 %
Kappa statistic                     0.9824
Mean absolute error                  0.058
Root mean squared error              0.1963
Relative absolute error              13.0214 %
Root relative squared error          41.5318 %
Total Number of Instances           51

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MOC      ROC Area  ERC Area  Class
Iris-setosa      1.000   0.000   1.000     1.000   1.000     1.000   1.000   1.000   Iris-setosa
Iris-versicolor  0.789   0.000   1.000     0.789   0.882     0.838   0.972   0.950   Iris-versicolor
Iris-virginica   1.000   0.118   0.810     1.000   0.895     0.845   0.971   0.895   Iris-virginica
Weighted Avg.    0.922   0.039   0.937     0.922   0.921     0.888   0.980   0.946

=== Confusion Matrix ===
 a  b  c  <-- classified as
15  0  0 | a = Iris-setosa
 0 15  4 | b = Iris-versicolor
 0  0 17 | c = Iris-virginica
  
```

Status

OK Log

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0

Test options

Use training set  
 Supplied test set (Set...)  
 Cross-validation (Folds: 10)  
 Percentage split (%: 66)

(Nom) class

Start Stop

Result list (right-click for options)

- 17:06:47 - bayes.NaiveBayes
- 17:07:05 - trees.REPTree
- 17:07:14 - functions.LibSVM

Weka Classifier Tree Visualizer: 17:07:05 - trees.REPTree (iris)

Tree View

```

graph TD
    Node1((1: petal length)) -- "< 2.5" --> Node2[2: Iris-setosa (33/0) [17/0]]
    Node1 -- "≥ 2.5" --> Node3((3: petal width))
    Node3 -- "< 1.75" --> Node4[4: Iris-versicolor (36/3) [18/2]]
    Node3 -- "≥ 1.75" --> Node5[5: Iris-virginica (31/1) [15/0]]
  
```

Status

OK Log

SVM:

**Weka Explorer**

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

**Classifier**

Choose **LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4"-seed 1**

**Test options**

Use training set  
 Supplied test set   
 Cross-validation Folds: 10  
 Percentage split %: 66

(Nom) class

**Result list (right-click for options)**

- 17:06:47 - bayes.NaiveBayes
- 17:07:05 - trees.REPTree
- 17:07:14 - functions.LibSVM

**Classifier output**

```

=== Run information ===

Scheme:      weka.classifiers.functions.LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-
Relation:    iris
Instances:   150
Attributes:  5
             sepallength
             sepalwidth
             petallength
             petalwidth
             class
Test mode:   split 66.0% train, remainder test

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances      51          100 %
Incorrectly Classified Instances     0           0 %
Kappa statistic                      1
Mean absolute error                   0
Root mean squared error               0
Relative absolute error               0 %
Root relative squared error           0 %
Total Number of Instances           51
  
```

Status: OK

## SVM summary:

**Weka Explorer**

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

**Classifier**

Choose **LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-8-4"-seed 1**

**Test options**

Use training set  
 Supplied test set   
 Cross-validation Folds: 10  
 Percentage split %: 66

(Nom) class

**Result list (right-click for options)**

- 17:06:47 - bayes.NaiveBayes
- 17:07:05 - trees.REPTree
- 17:07:14 - functions.LibSVM

**Classifier output**

```

=== Summary ===

Correctly Classified Instances      51          100 %
Incorrectly Classified Instances     0           0 %
Kappa statistic                      1
Mean absolute error                   0
Root mean squared error               0
Relative absolute error               0 %
Root relative squared error           0 %
Total Number of Instances           51

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Iris-setosa
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Iris-versicolor
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Iris-virginica
Weighted Avg.  1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000

=== Confusion Matrix ===

 a b c <-- Classified as
15 0 0 | a = Iris-setosa
 0 19 0 | b = Iris-versicolor
 0 0 17 | c = Iris-virginica
  
```

Status: OK

Comparing the results:

Naïve Bayes:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
0.900	0.947	0.923	0.876
0.938	0.882	0.909	0.867
0.942	0.941	0.941	0.909

Decision tree:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
1.000	0.789	0.882	0.838
0.810	1.000	0.895	0.845
0.937	0.922	0.921	0.888

NVM:

Precision	Recall	Mare	MMRE
1.000	1.000	1.000	1.000
0.959	0.940	0.949	0.925
0.941	0.960	0.950	0.925
0.967	0.967	0.967	0.950