
DATA
MINING
WEB CONNECTION
AGGREGATION INNOVATION BUSINESS
SOFTWARE INFORMATION
CODE
CONNECTIVITY
SOLUTION
COLLECT
BUG
ERROR
STORAGE
ENCRYPTION
MANAGEMENT DECISION
SECURITY COMPUTER
TECHNOLOGY
CONCEPT
TRANSFORMATION
EXTRACTION
NETWORK ANALYTICS
ELABORATION
INTERNET
BINARY
ANALYSIS
PATTERN
PROCESS
RECOGNITION
SYSTEM CALCULATION
MACHINE
LEARNING
CUSTOMER
TROUBLESHOOT

Author Names

C. M. Abdullah-----------------18-38631-2

HUMAYRA NAZIN TOMA----19-40196-1

Safinul Ibrat Sakib-------------19-40298-1

Tahmid hassan Ridoy---------19-40307-1

Faculty : AKINUL ISLAM JONY

American international university

Department of Computer Science and Engineering

Contents

AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH

Section 1: Project Overview

The project we are working on is to build a classifier application which will find out the predictive accuracy of a selected data set using k-NN, Decision tree and Naive bayes algorithm. After finding out the predictive accuracy we have to select which algorithm have found the more correct predictive accuracy among the three algorithms. For this project we have used weka tools to run the algorithm on our data set.

Section 2: Dataset Overview

Cancer cases registered and deaths for NZ population.

Source - Ministry of Health NZ

The file has cases registered by cancer codes (brain, breast, etc.) and the death (data slicers - Year, Gender, Cancer category)

This database contains 6 attributes.

Attribute Information

1) Year

2) Type (registered patients, death)

3) sex (All Sex, Male, Female)

4) numbers (total count of patients)

5) ICD codes (international cancer codes)

6) Cancer category (breast, prostrate, neck etc.)

Data Source : Kaggle.com

Url : https://www.kaggle.cancer-data-set-registered-vs-death-by-yearsex

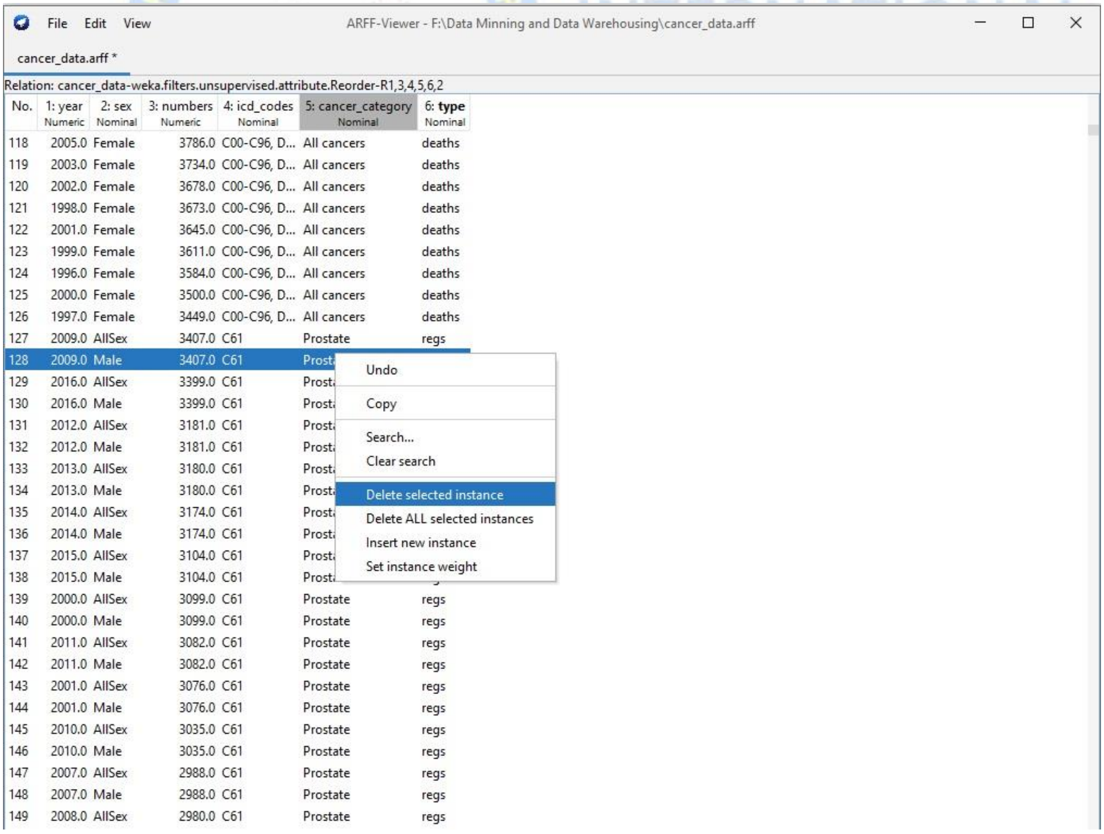
Section 3: Model Development

3.1 : Data prepossessing

Two strategies for dealing with missing attribute values were described:

• Discard Instances :

- This is the simplest strategy: delete all instances where there is at least one missing value and use the remainder.
- This strategy has the advantage of avoiding introducing any data errors.
- Its main disadvantage is that discarding data may damage the reliability of the resulting classifier.
- Together these weaknesses are quite substantial. Although the 'discard instances' strategy may be worth trying when the proportion of missing values is small, it is not recommended in general.



No.	1: year Numeric	2: sex Nominal	3: numbers Numeric	4: icd_codes Nominal	5: cancer_category Nominal	6: type Nominal
118	2005.0	Female	3786.0	C00-C96, D...	All cancers	deaths
119	2003.0	Female	3734.0	C00-C96, D...	All cancers	deaths
120	2002.0	Female	3678.0	C00-C96, D...	All cancers	deaths
121	1998.0	Female	3673.0	C00-C96, D...	All cancers	deaths
122	2001.0	Female	3645.0	C00-C96, D...	All cancers	deaths
123	1999.0	Female	3611.0	C00-C96, D...	All cancers	deaths
124	1996.0	Female	3584.0	C00-C96, D...	All cancers	deaths
125	2000.0	Female	3500.0	C00-C96, D...	All cancers	deaths
126	1997.0	Female	3449.0	C00-C96, D...	All cancers	deaths
127	2009.0	AllSex	3407.0	C61	Prostate	regs
128	2009.0	Male	3407.0	C61	Prost	
129	2016.0	AllSex	3399.0	C61	Prost	
130	2016.0	Male	3399.0	C61	Prost	
131	2012.0	AllSex	3181.0	C61	Prost	
132	2012.0	Male	3181.0	C61	Prost	
133	2013.0	AllSex	3180.0	C61	Prost	
134	2013.0	Male	3180.0	C61	Prost	
135	2014.0	AllSex	3174.0	C61	Prost	
136	2014.0	Male	3174.0	C61	Prost	
137	2015.0	AllSex	3104.0	C61	Prost	
138	2015.0	Male	3104.0	C61	Prost	
139	2000.0	AllSex	3099.0	C61	Prostate	regs
140	2000.0	Male	3099.0	C61	Prostate	regs
141	2011.0	AllSex	3082.0	C61	Prostate	regs
142	2011.0	Male	3082.0	C61	Prostate	regs
143	2001.0	AllSex	3076.0	C61	Prostate	regs
144	2001.0	Male	3076.0	C61	Prostate	regs
145	2010.0	AllSex	3035.0	C61	Prostate	regs
146	2010.0	Male	3035.0	C61	Prostate	regs
147	2007.0	AllSex	2988.0	C61	Prostate	regs
148	2007.0	Male	2988.0	C61	Prostate	regs
149	2008.0	AllSex	2980.0	C61	Prostate	regs

Figure 1: Discard Instance

• Replace by most frequent/average value :

- With this strategy any missing values of a categorical attribute are replaced by its most commonly occurring value in the training set.
- Any missing values of a continuous attribute are replaced by its average value in the training set.

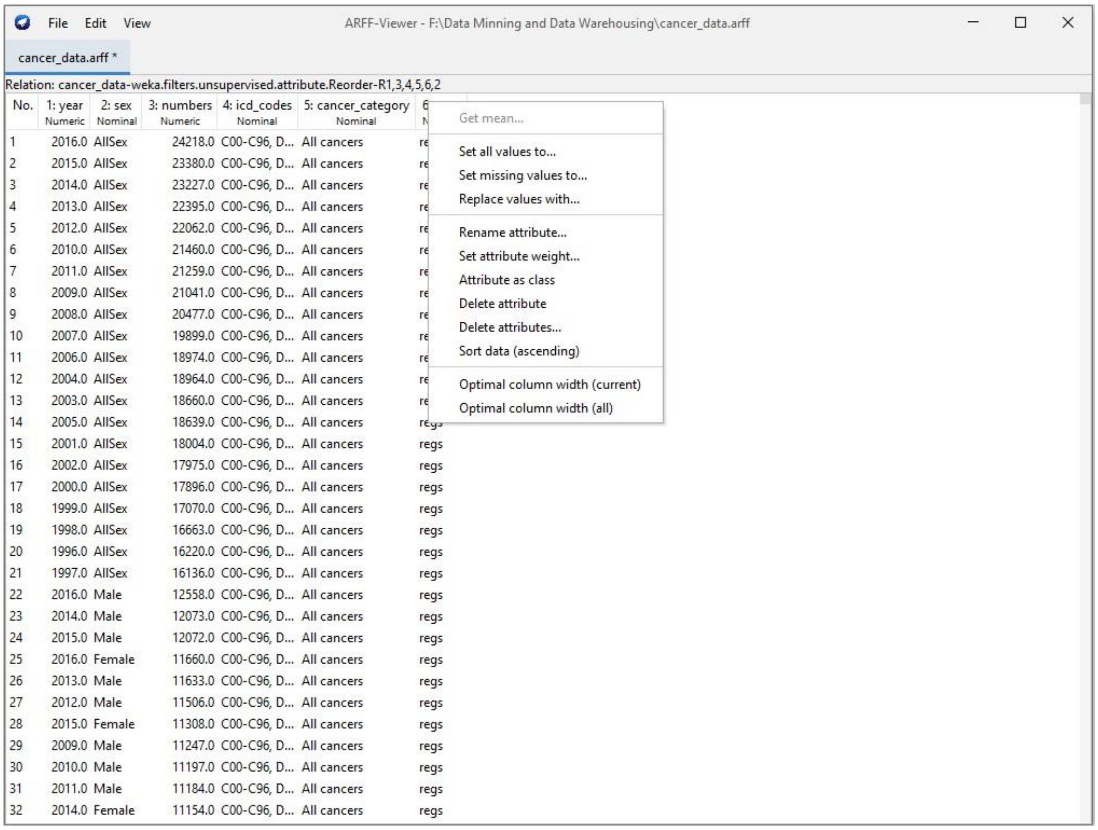


Figure 2: Replace by most frequent/average value

After Data prepossessing we can apply the models in our Data Set .

Then Click on Explorer . Then the tab will open then click on open file select the data set. (.arff , .csv)

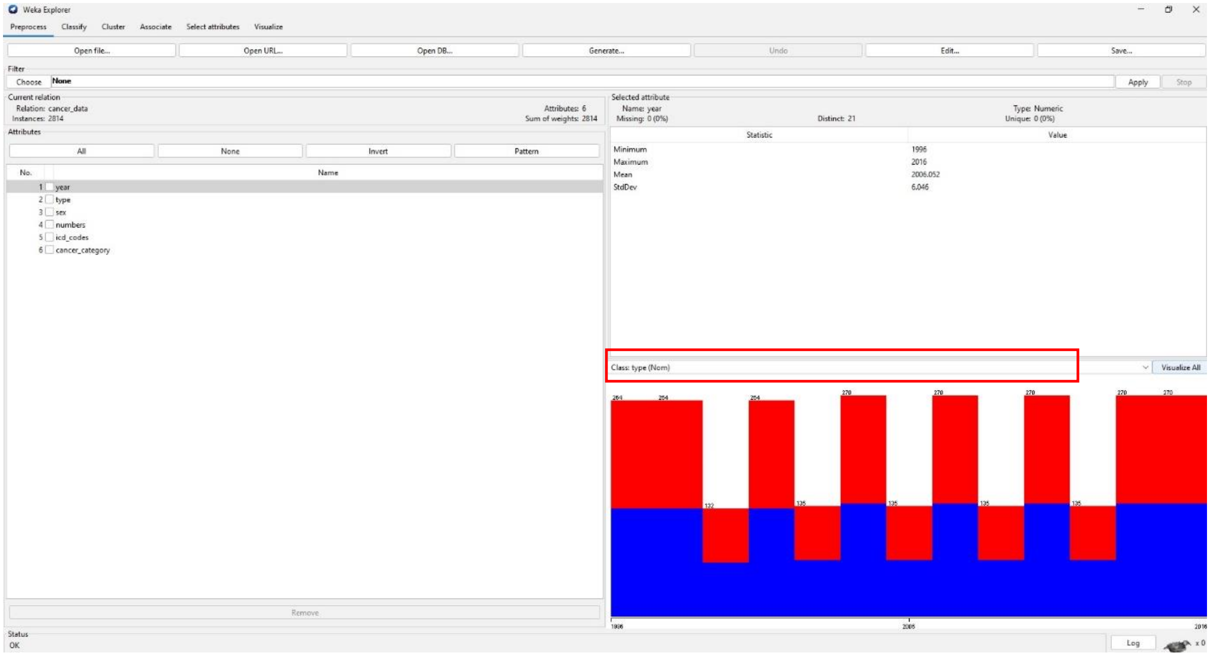


Figure - Class Selection

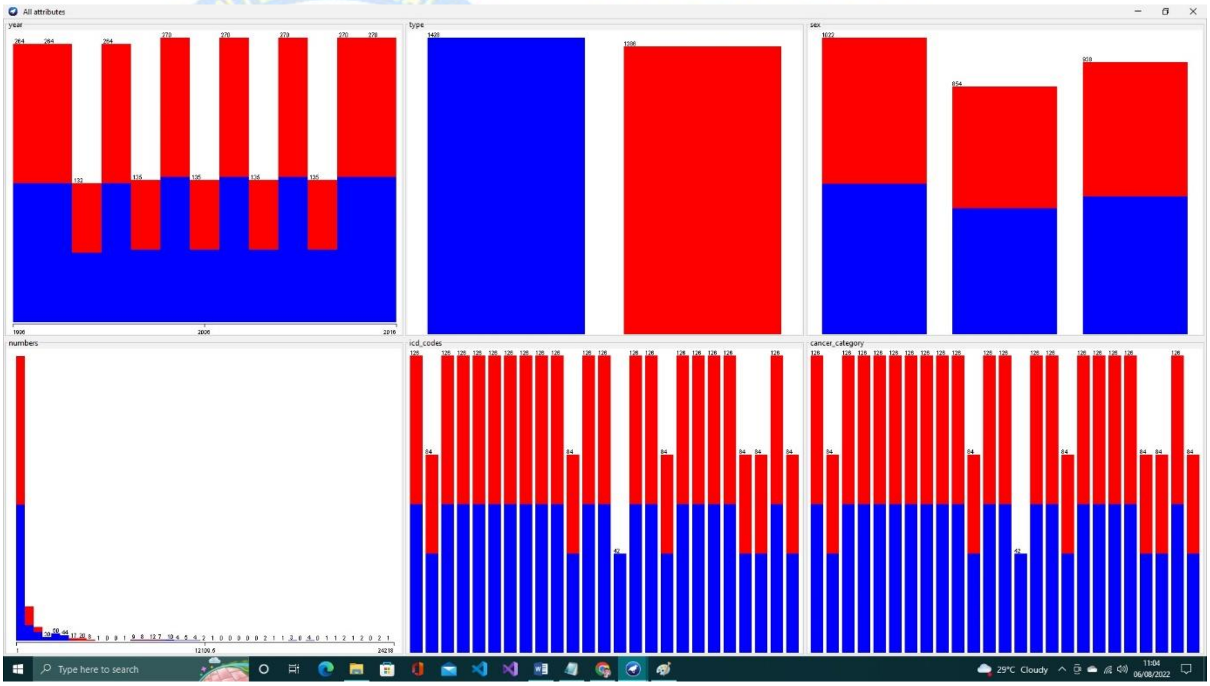
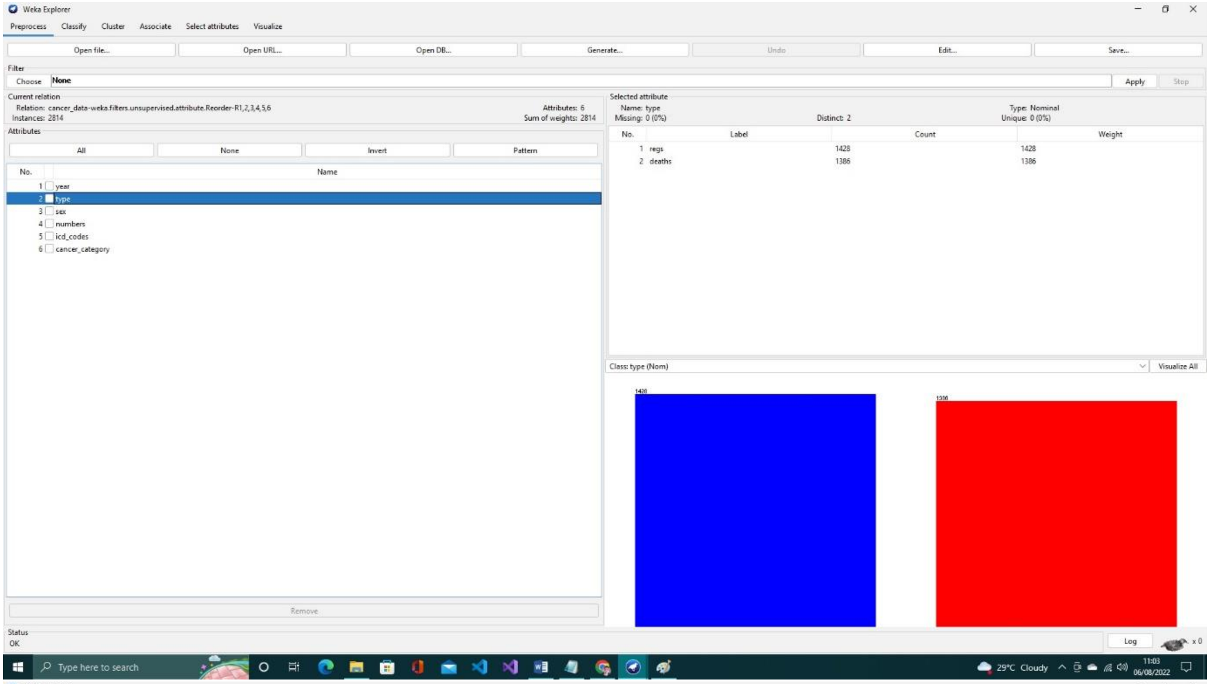


Figure - Graph for All Attributes.



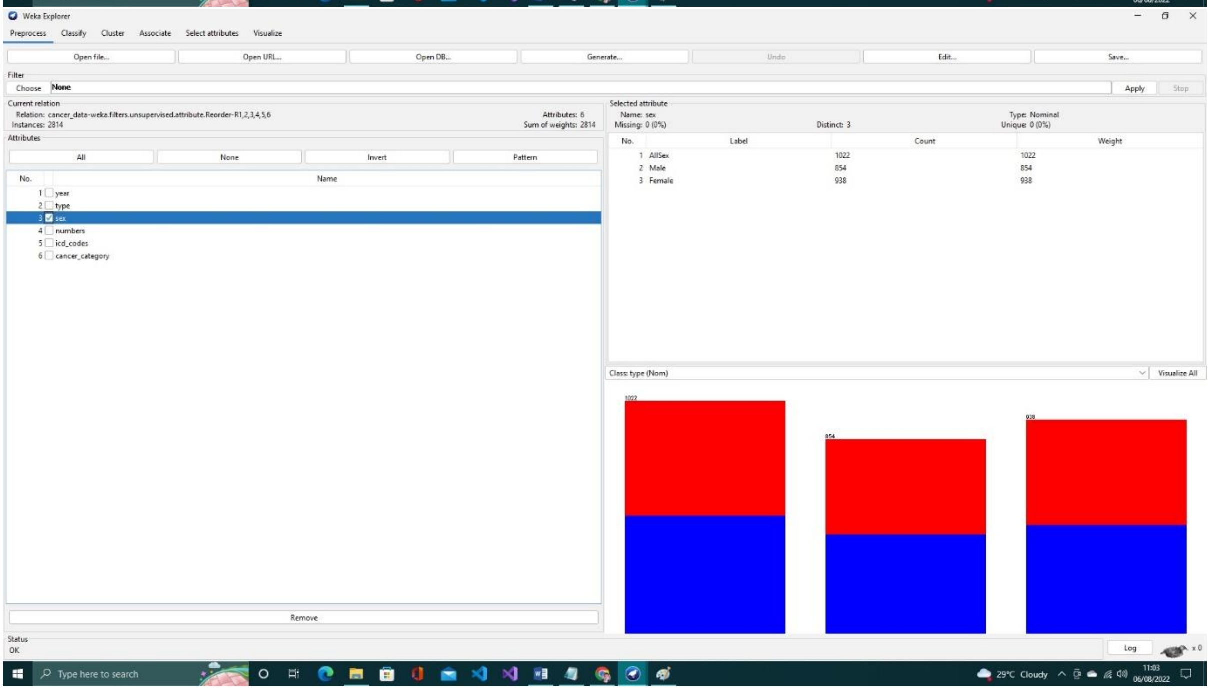


Figure - Detail About type and Sex Attribute

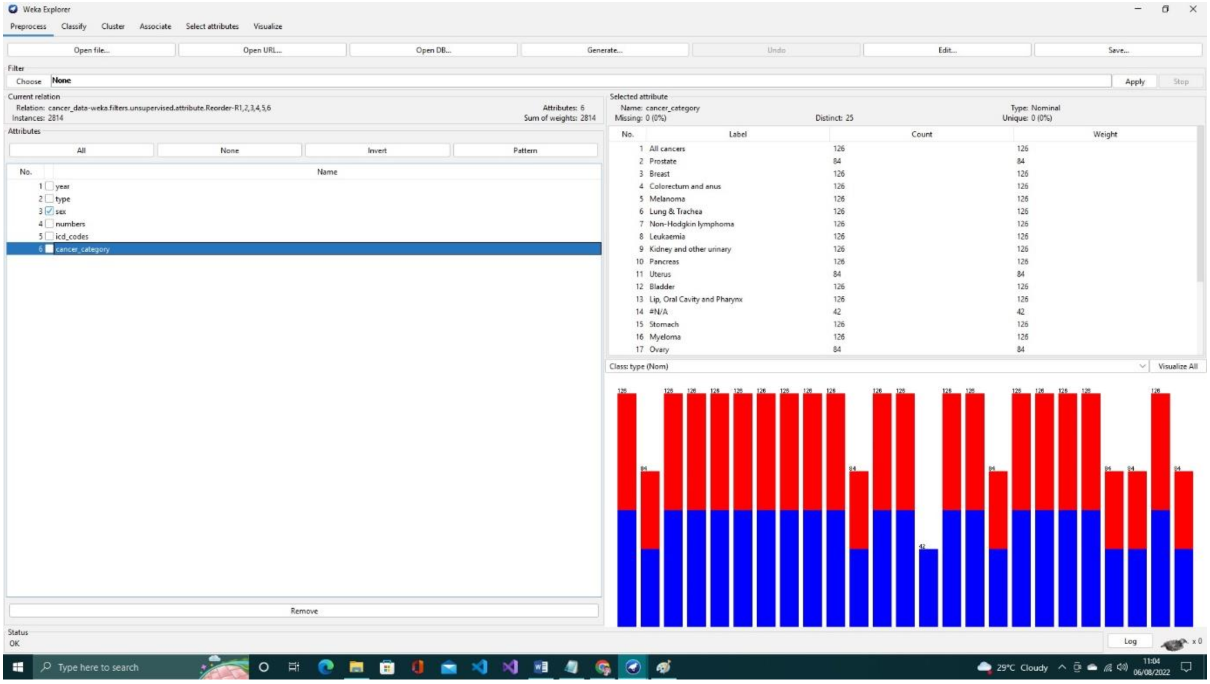
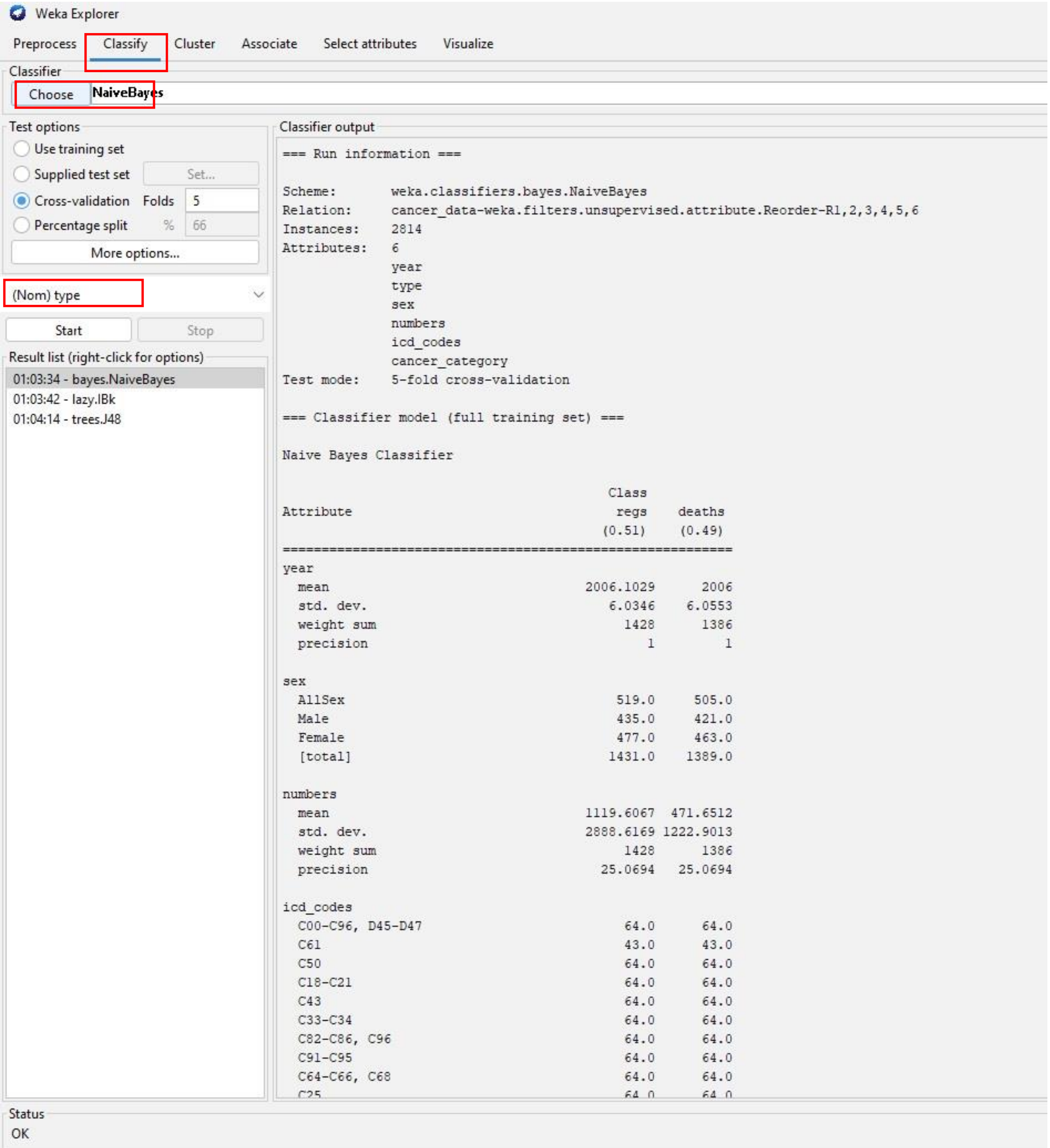


Figure - It is showing the cancer catagory by Class Type from dataset

3.2 : Naïve Bayes

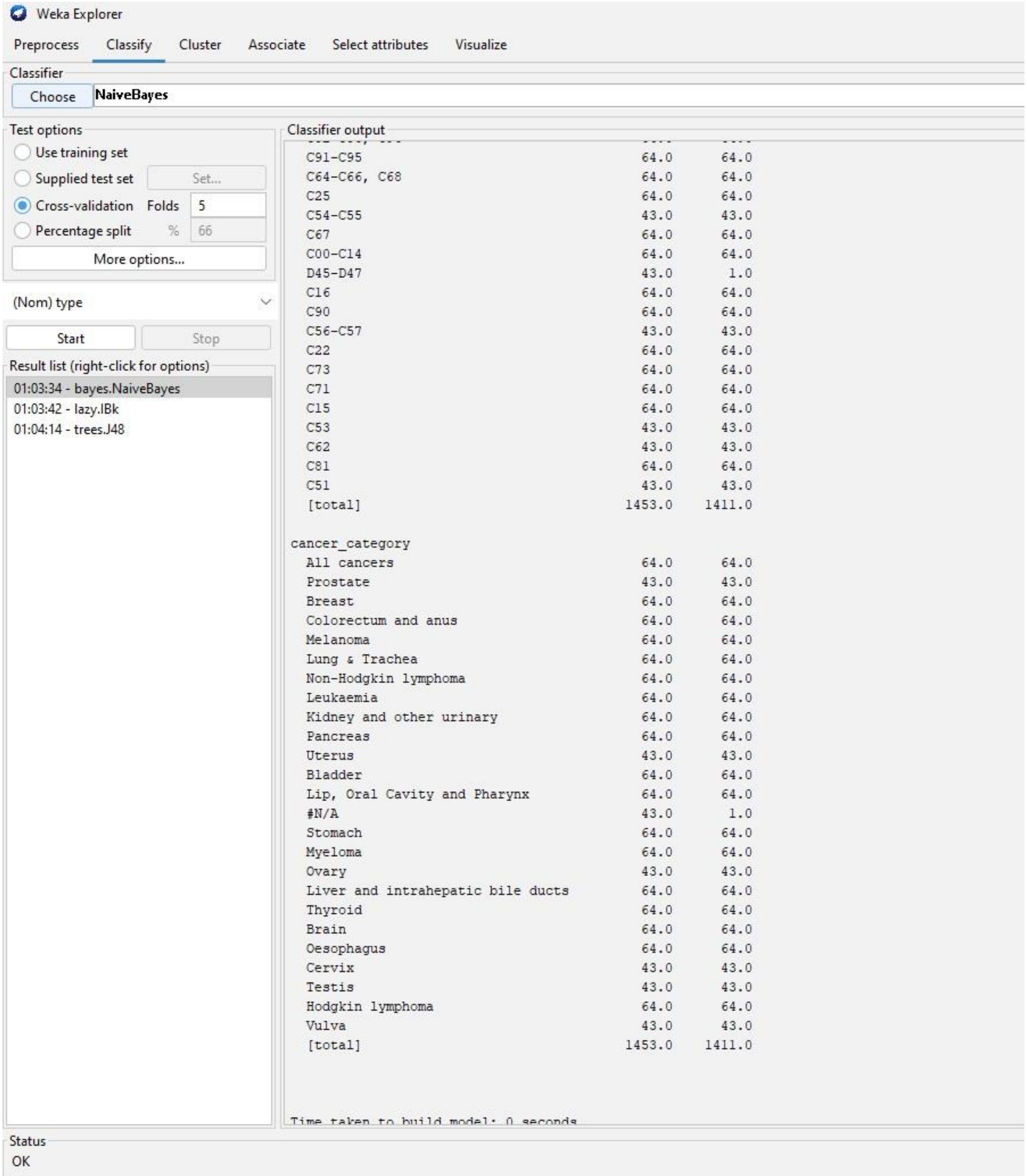
Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification



Steps For development process of Naïve bayes :

First click on **classify** on the top menu bar then select **Choose** after that click on **Bayes folder** then select **Naïve bayes**. Select and Input **5- Fold Cross Validation .**

After that select the **class** (top of the start button) here we have selected **(Nom) type** from our data set. Then click on **Start button**. The result will show.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 5
Percentage split % 66
More options...
(Nom) type
Start
Stop
Result list (right-click for options)
01:03:34 - bayes.NaiveBayes
01:03:42 - lazy.IBk
01:04:14 - trees.J48
Classifier output
C91-C95 64.0 64.0
C64-C66, C68 64.0 64.0
C25 64.0 64.0
C54-C55 43.0 43.0
C67 64.0 64.0
C00-C14 64.0 64.0
D45-D47 43.0 1.0
C16 64.0 64.0
C90 64.0 64.0
C56-C57 43.0 43.0
C22 64.0 64.0
C73 64.0 64.0
C71 64.0 64.0
C15 64.0 64.0
C53 43.0 43.0
C62 43.0 43.0
C81 64.0 64.0
C51 43.0 43.0
[total] 1453.0 1411.0
cancer_category
All cancers 64.0 64.0
Prostate 43.0 43.0
Breast 64.0 64.0
Colorectum and anus 64.0 64.0
Melanoma 64.0 64.0
Lung & Trachea 64.0 64.0
Non-Hodgkin lymphoma 64.0 64.0
Leukaemia 64.0 64.0
Kidney and other urinary 64.0 64.0
Pancreas 64.0 64.0
Uterus 43.0 43.0
Bladder 64.0 64.0
Lip, Oral Cavity and Pharynx 64.0 64.0
#N/A 43.0 1.0
Stomach 64.0 64.0
Myeloma 64.0 64.0
Ovary 43.0 43.0
Liver and intrahepatic bile ducts 64.0 64.0
Thyroid 64.0 64.0
Brain 64.0 64.0
Oesophagus 64.0 64.0
Cervix 43.0 43.0
Testis 43.0 43.0
Hodgkin lymphoma 64.0 64.0
Vulva 43.0 43.0
[total] 1453.0 1411.0
Time taken to build model: 0 seconds
Status
OK

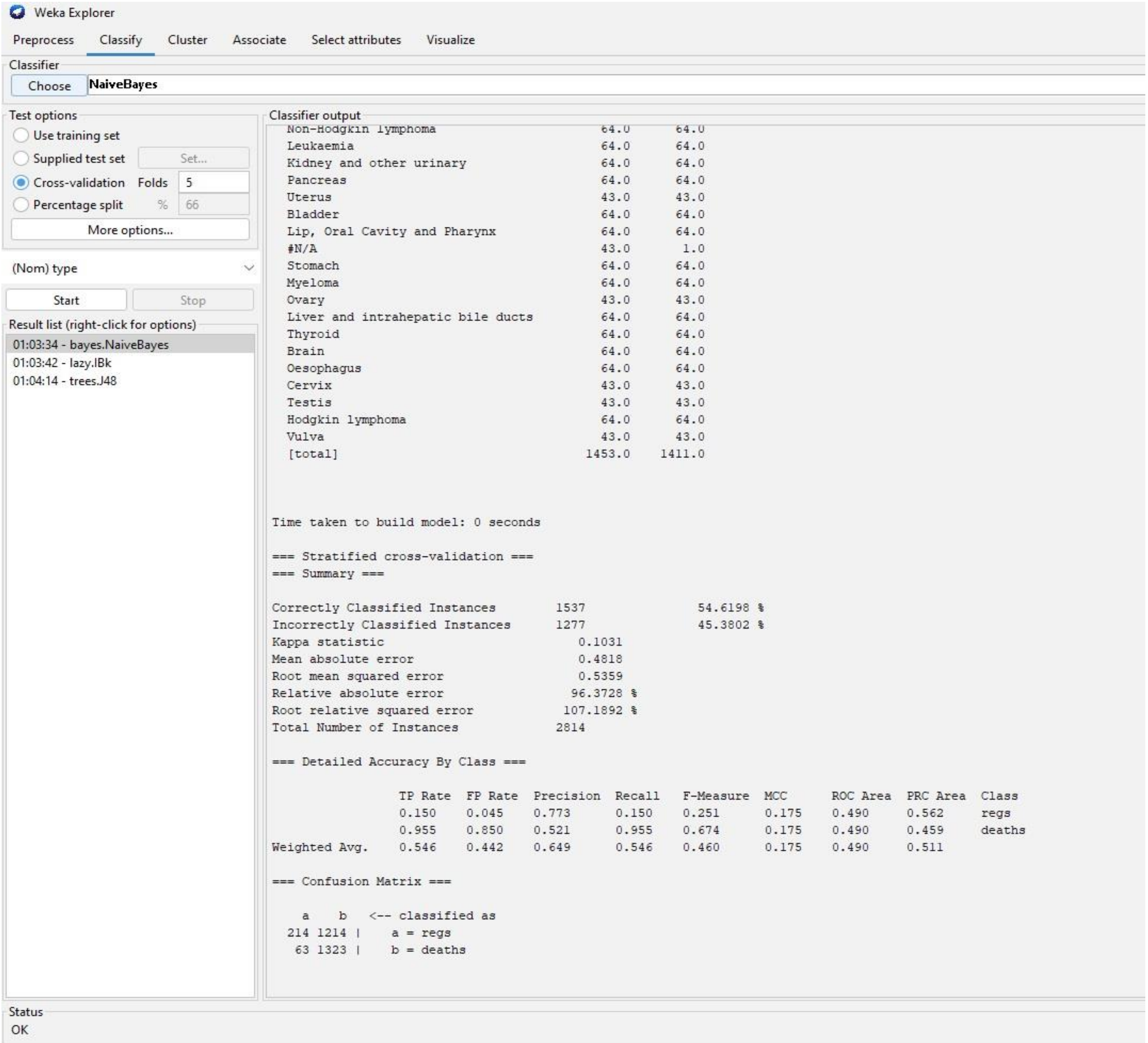


Figure 3: Naive Bayes Model developed using weka Tool

3.3 : K-NN

The k-nearest neighbours' algorithm (k-NN) is a non-parametric supervised learning method. It is used for classification and regression. In both cases, the input consists of the k closest training examples in a data set. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbour.

In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbours.

k-NN is a type of classification where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, if the features represent different physical units or come in vastly different scales then normalizing the training data can improve its accuracy dramatically.

A peculiarity of the k-NN algorithm is that it is sensitive to the local structure of the data.

Steps For development process of K –nearest neighbours :

First click on **classify** on the top menu bar then select **Choose** after that click on **lazy folder** then select **IBk**. it is **K –nearest neighbours.** Select and input **5 - Fold Cross Validation .**

After that select the **class** (top of the start button) here we have selected **(Nom) type** from our data set. Then click on **Start button**. The result will show.

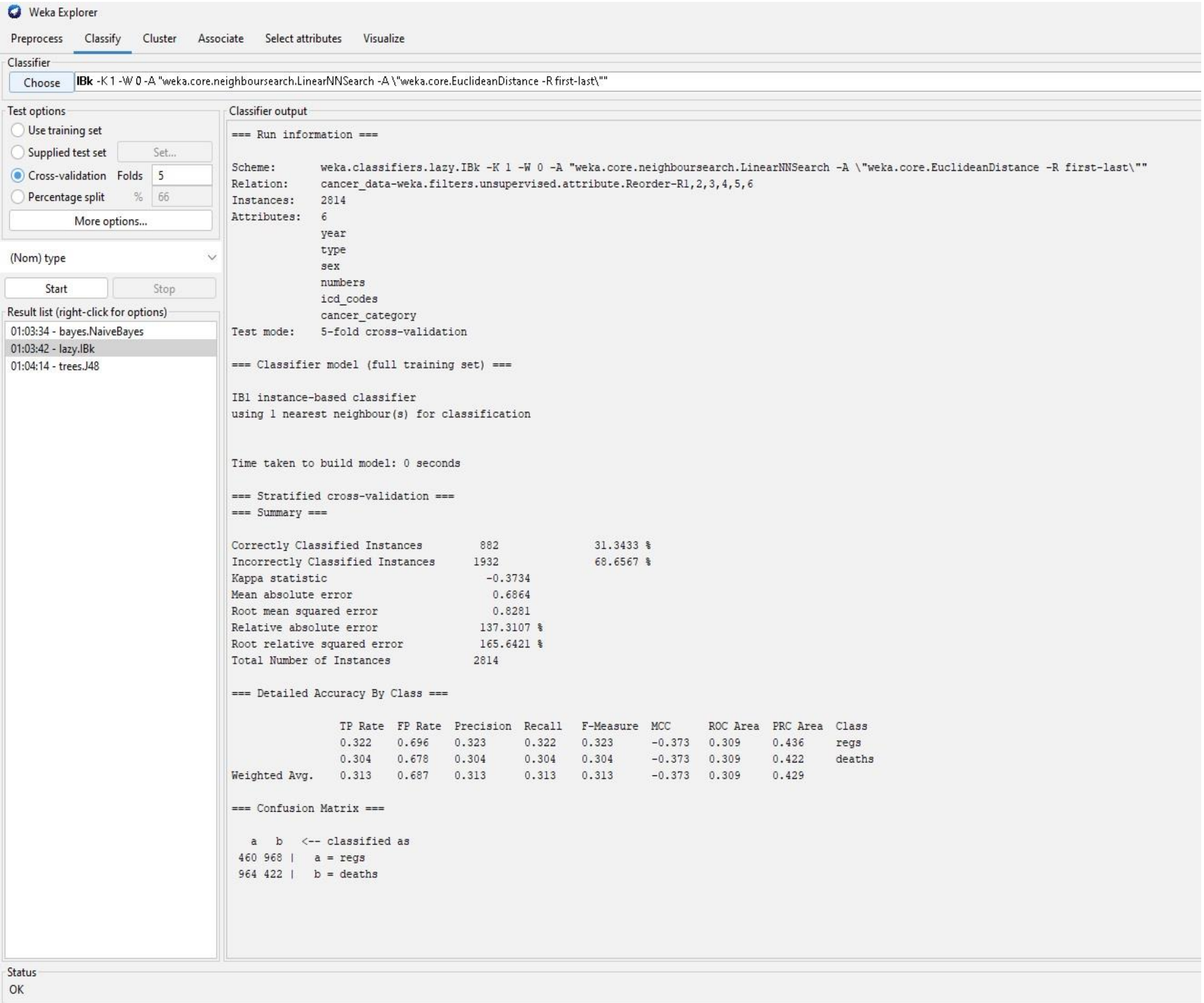


Figure 4 : K-NN Model developed using weka Tool

3.4 : Decision Tree

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

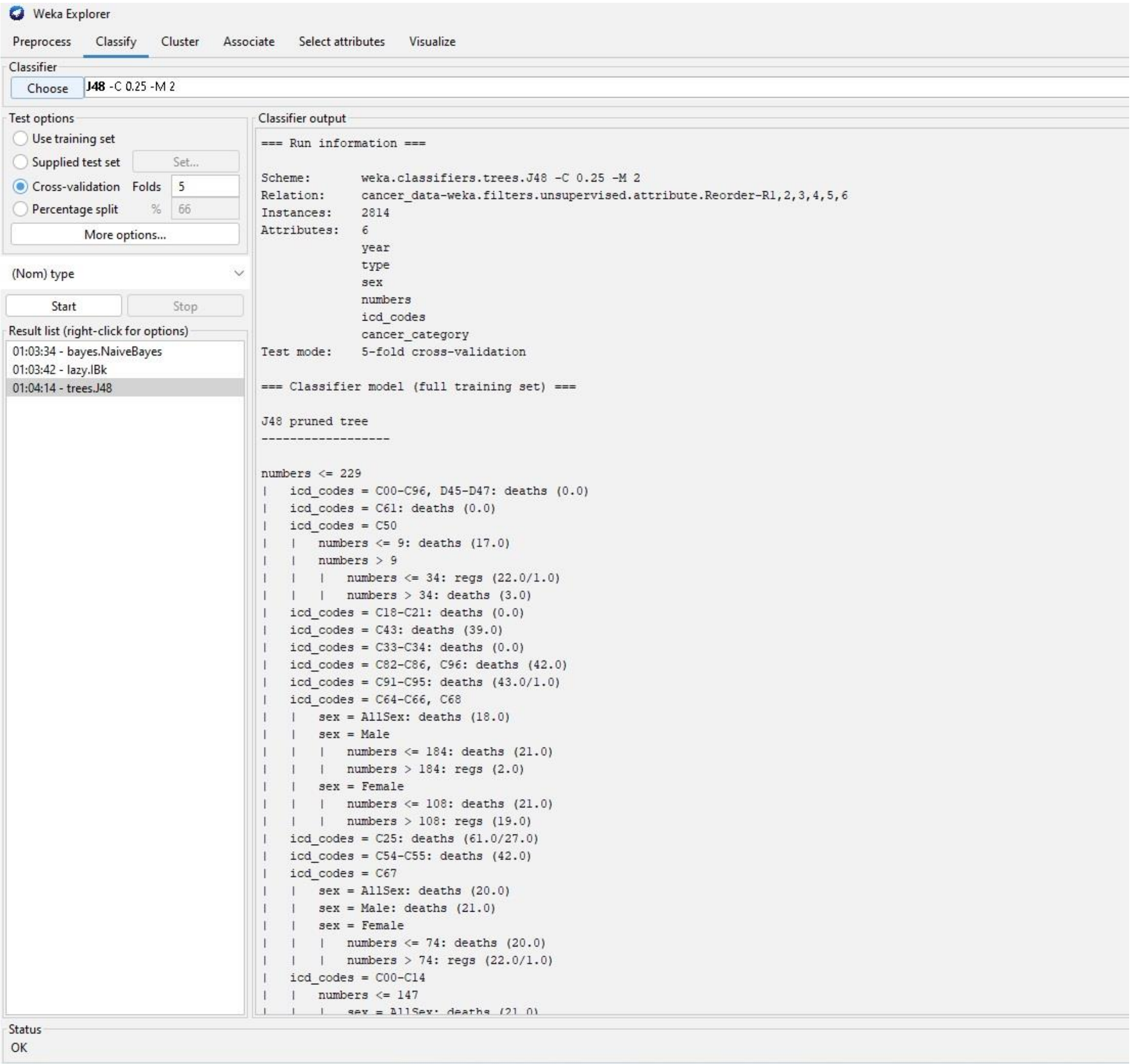
In decision analysis, a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated.

A decision tree consists of three types of nodes: [1] Decision nodes – typically represented by squares

Chance nodes – typically represented by circles

End nodes – typically represented by triangles

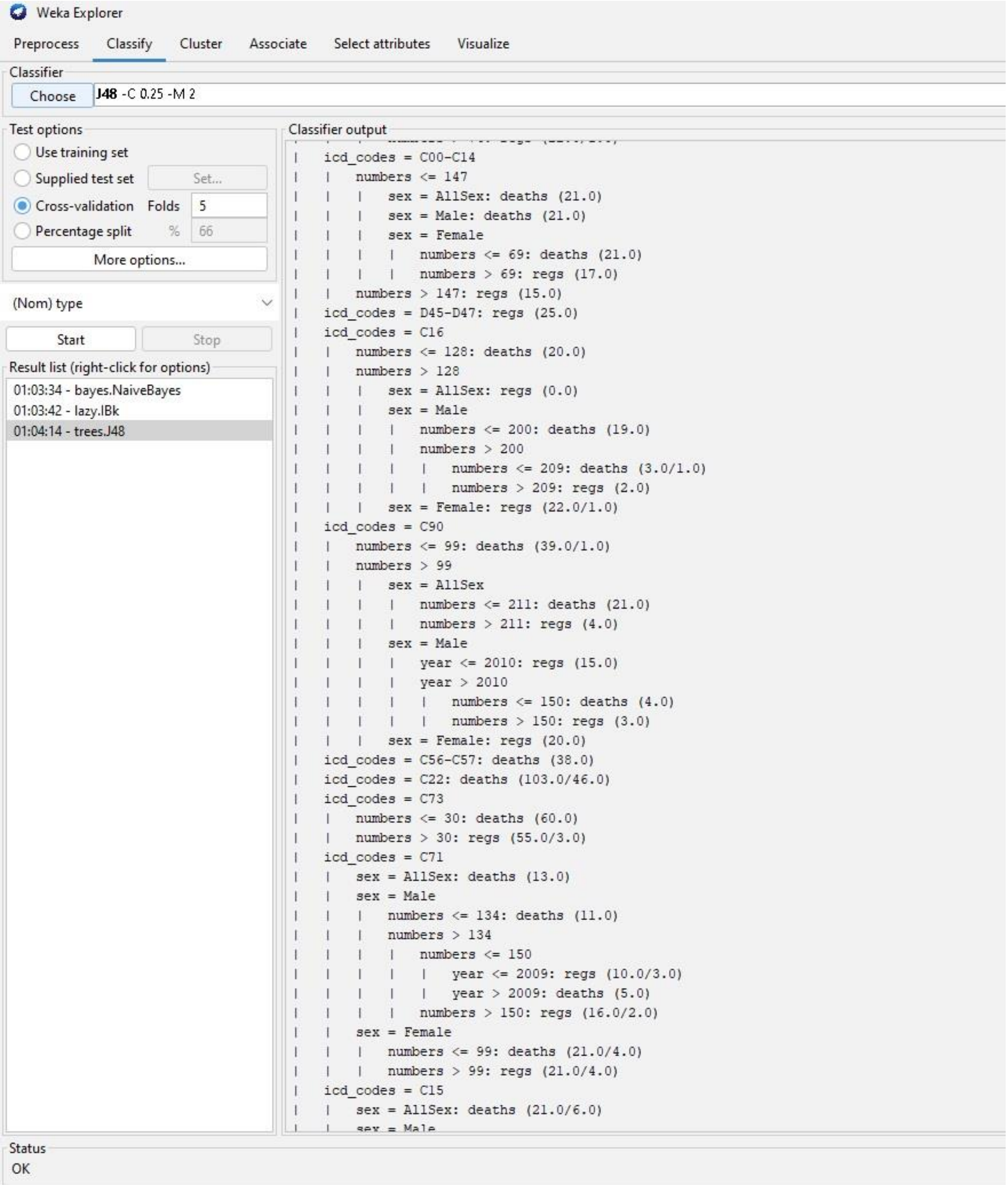
Decision trees are commonly used in operations research and operations management. If, in practice, decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm.

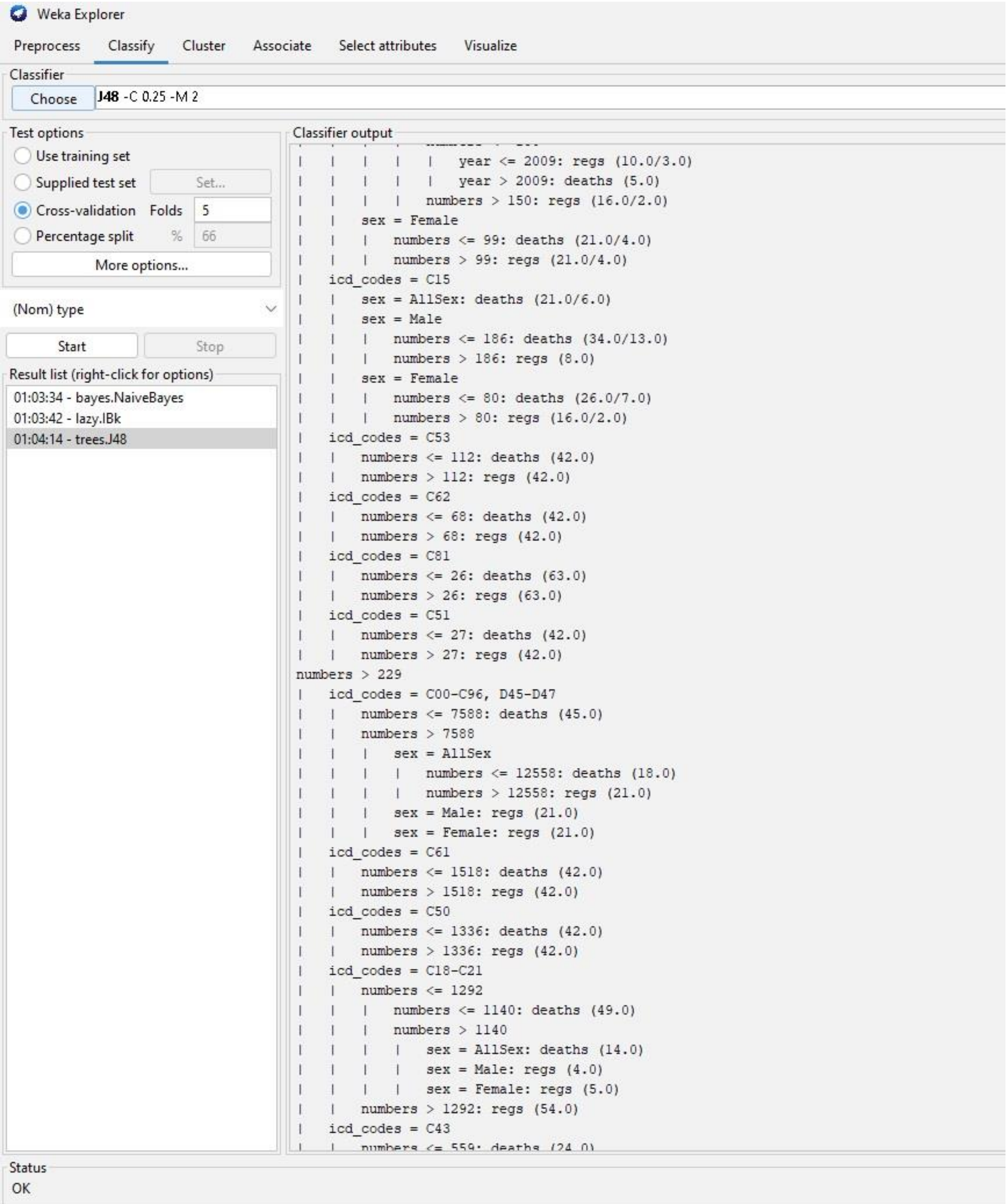


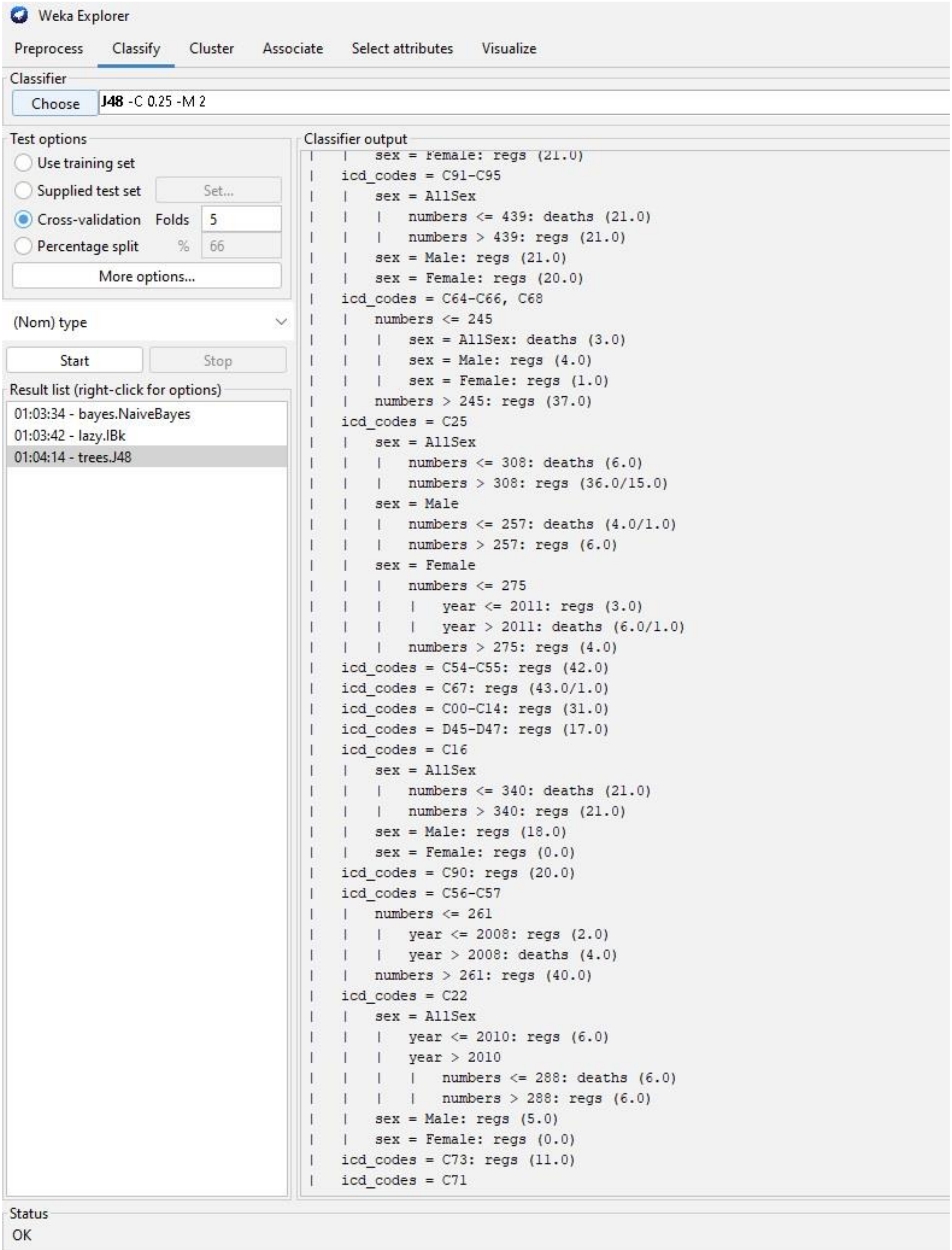
Steps For development process of Decision Tree:

First click on **classify** on the top menu bar then select **Choose** after that click on **tree folder** then select **J48**. it is**Decision Tree.** Select and input **5 - Fold Cross Validation .**

After that select the **class** (top of the **start** button) here we have selected **(Nom) type** from our data set. Then click on **Start button**. The result will show.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 5
Percentage split % 66
More options...
(Nom) type
Start
Stop
Result list (right-click for options)
01:03:34 - bayes.NaiveBayes
01:03:42 - lazy.IBk
01:04:14 - trees.J48
Classifier output
| icd_codes = C00-C14
| | numbers <= 147
| | | sex = AllSex: deaths (21.0)
| | | sex = Male: deaths (21.0)
| | | sex = Female
| | | | numbers <= 69: deaths (21.0)
| | | | numbers > 69: regs (17.0)
| | numbers > 147: regs (15.0)
| icd_codes = D45-D47: regs (25.0)
| icd_codes = C16
| | numbers <= 128: deaths (20.0)
| | numbers > 128
| | | sex = AllSex: regs (0.0)
| | | sex = Male
| | | | numbers <= 200: deaths (19.0)
| | | | numbers > 200
| | | | | numbers <= 209: deaths (3.0/1.0)
| | | | | numbers > 209: regs (2.0)
| | | sex = Female: regs (22.0/1.0)
| icd_codes = C90
| | numbers <= 99: deaths (39.0/1.0)
| | numbers > 99
| | | sex = AllSex
| | | | numbers <= 211: deaths (21.0)
| | | | numbers > 211: regs (4.0)
| | | sex = Male
| | | | year <= 2010: regs (15.0)
| | | | year > 2010
| | | | | numbers <= 150: deaths (4.0)
| | | | | numbers > 150: regs (3.0)
| | | sex = Female: regs (20.0)
| icd_codes = C56-C57: deaths (38.0)
| icd_codes = C22: deaths (103.0/46.0)
| icd_codes = C73
| | numbers <= 30: deaths (60.0)
| | numbers > 30: regs (55.0/3.0)
| icd_codes = C71
| | sex = AllSex: deaths (13.0)
| | sex = Male
| | | numbers <= 134: deaths (11.0)
| | | numbers > 134
| | | | numbers <= 150
| | | | | year <= 2009: regs (10.0/3.0)
| | | | | year > 2009: deaths (5.0)
| | | | numbers > 150: regs (16.0/2.0)
| | sex = Female
| | | numbers <= 99: deaths (21.0/4.0)
| | | numbers > 99: regs (21.0/4.0)
| icd_codes = C15
| | sex = AllSex: deaths (21.0/6.0)
Status
OK


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 5
Percentage split % 66
More options...
(Nom) type
Start
Stop
Result list (right-click for options)
01:03:34 - bayes.NaiveBayes
01:03:42 - lazy.IBk
01:04:14 - trees.J48
Classifier output
| | | | | year <= 2009: regs (10.0/3.0)
| | | | | year > 2009: deaths (5.0)
| | | | numbers > 150: regs (16.0/2.0)
| | sex = Female
| | | numbers <= 99: deaths (21.0/4.0)
| | | numbers > 99: regs (21.0/4.0)
| icd_codes = C15
| | sex = AllSex: deaths (21.0/6.0)
| | sex = Male
| | | numbers <= 186: deaths (34.0/13.0)
| | | numbers > 186: regs (8.0)
| | sex = Female
| | | numbers <= 80: deaths (26.0/7.0)
| | | numbers > 80: regs (16.0/2.0)
| icd_codes = C53
| | numbers <= 112: deaths (42.0)
| | numbers > 112: regs (42.0)
| icd_codes = C62
| | numbers <= 68: deaths (42.0)
| | numbers > 68: regs (42.0)
| icd_codes = C81
| | numbers <= 26: deaths (63.0)
| | numbers > 26: regs (63.0)
| icd_codes = C51
| | numbers <= 27: deaths (42.0)
| | numbers > 27: regs (42.0)
numbers > 229
| icd_codes = C00-C96, D45-D47
| | numbers <= 7588: deaths (45.0)
| | numbers > 7588
| | | sex = AllSex
| | | | numbers <= 12558: deaths (18.0)
| | | | numbers > 12558: regs (21.0)
| | | sex = Male: regs (21.0)
| | | sex = Female: regs (21.0)
| icd_codes = C61
| | numbers <= 1518: deaths (42.0)
| | numbers > 1518: regs (42.0)
| icd_codes = C50
| | numbers <= 1336: deaths (42.0)
| | numbers > 1336: regs (42.0)
| icd_codes = C18-C21
| | numbers <= 1292
| | | numbers <= 1140: deaths (49.0)
| | | numbers > 1140
| | | | sex = AllSex: deaths (14.0)
| | | | sex = Male: regs (4.0)
| | | | sex = Female: regs (5.0)
| | numbers > 1292: regs (54.0)
| icd_codes = C43
Status
OK



Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation Folds 5
Percentage split % 66
More options...
(Nom) type
Start
Stop
Result list (right-click for options)
01:03:34 - bayes.NaiveBayes
01:03:42 - lazy.IBk
01:04:14 - trees.J48
Classifier output
| | sex = Female: regs (21.0)
| icd_codes = C91-C95
| | sex = AllSex
| | | numbers <= 439: deaths (21.0)
| | | numbers > 439: regs (21.0)
| | sex = Male: regs (21.0)
| | sex = Female: regs (20.0)
| icd_codes = C64-C66, C68
| | numbers <= 245
| | | sex = AllSex: deaths (3.0)
| | | sex = Male: regs (4.0)
| | | sex = Female: regs (1.0)
| | numbers > 245: regs (37.0)
| icd_codes = C25
| | sex = AllSex
| | | numbers <= 308: deaths (6.0)
| | | numbers > 308: regs (36.0/15.0)
| | sex = Male
| | | numbers <= 257: deaths (4.0/1.0)
| | | numbers > 257: regs (6.0)
| | sex = Female
| | | numbers <= 275
| | | | year <= 2011: regs (3.0)
| | | | year > 2011: deaths (6.0/1.0)
| | | numbers > 275: regs (4.0)
| icd_codes = C54-C55: regs (42.0)
| icd_codes = C67: regs (43.0/1.0)
| icd_codes = C00-C14: regs (31.0)
| icd_codes = D45-D47: regs (17.0)
| icd_codes = C16
| | sex = AllSex
| | | numbers <= 340: deaths (21.0)
| | | numbers > 340: regs (21.0)
| | sex = Male: regs (18.0)
| | sex = Female: regs (0.0)
| icd_codes = C90: regs (20.0)
| icd_codes = C56-C57
| | numbers <= 261
| | | year <= 2008: regs (2.0)
| | | year > 2008: deaths (4.0)
| | numbers > 261: regs (40.0)
| icd_codes = C22
| | sex = AllSex
| | | year <= 2010: regs (6.0)
| | | year > 2010
| | | | numbers <= 288: deaths (6.0)
| | | | numbers > 288: regs (6.0)
| | sex = Male: regs (5.0)
| | sex = Female: regs (0.0)
| icd_codes = C73: regs (11.0)
| icd_codes = C71
Status
OK

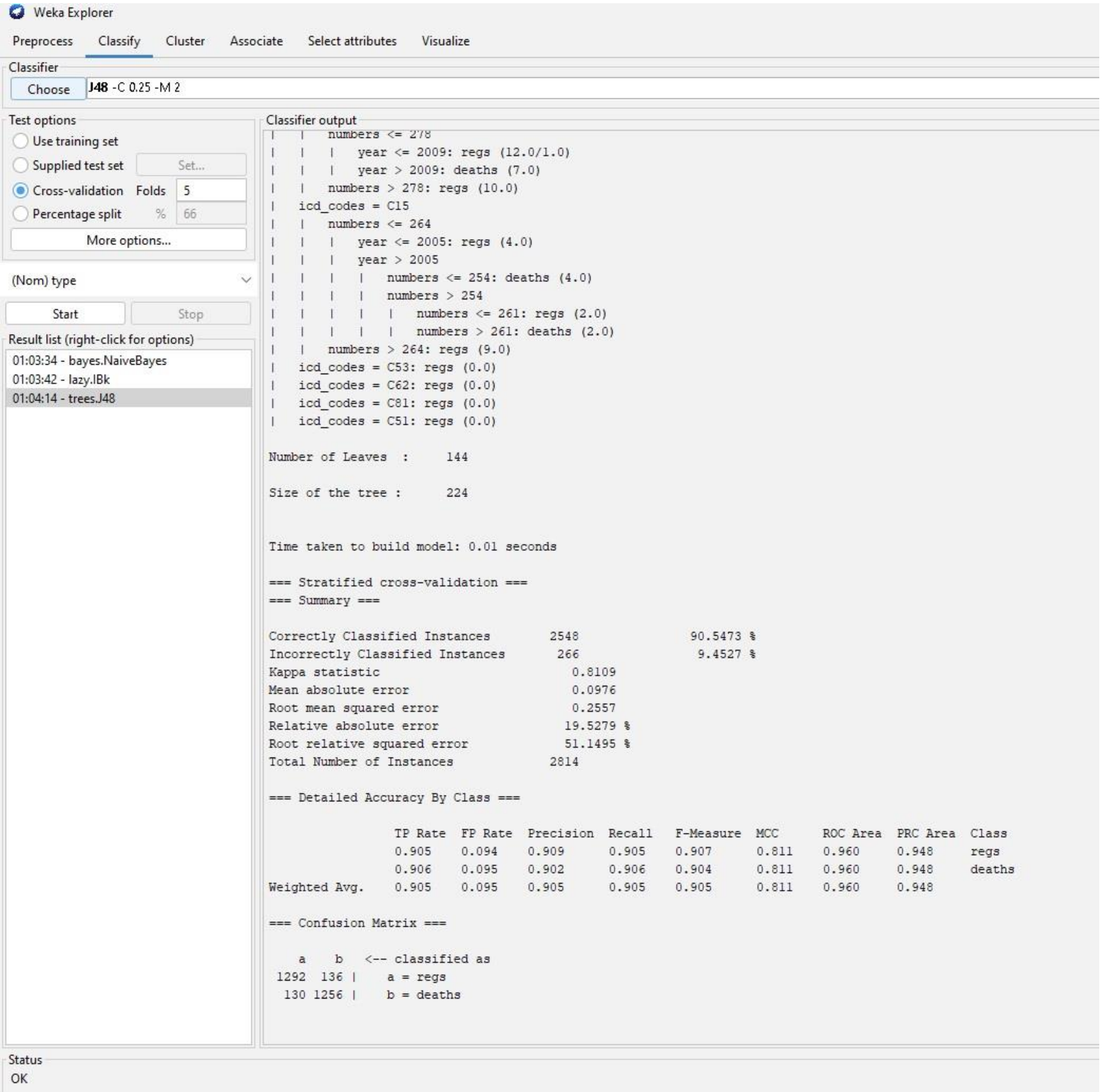


Figure 5 : Decision Tree Model developed using weka Tool

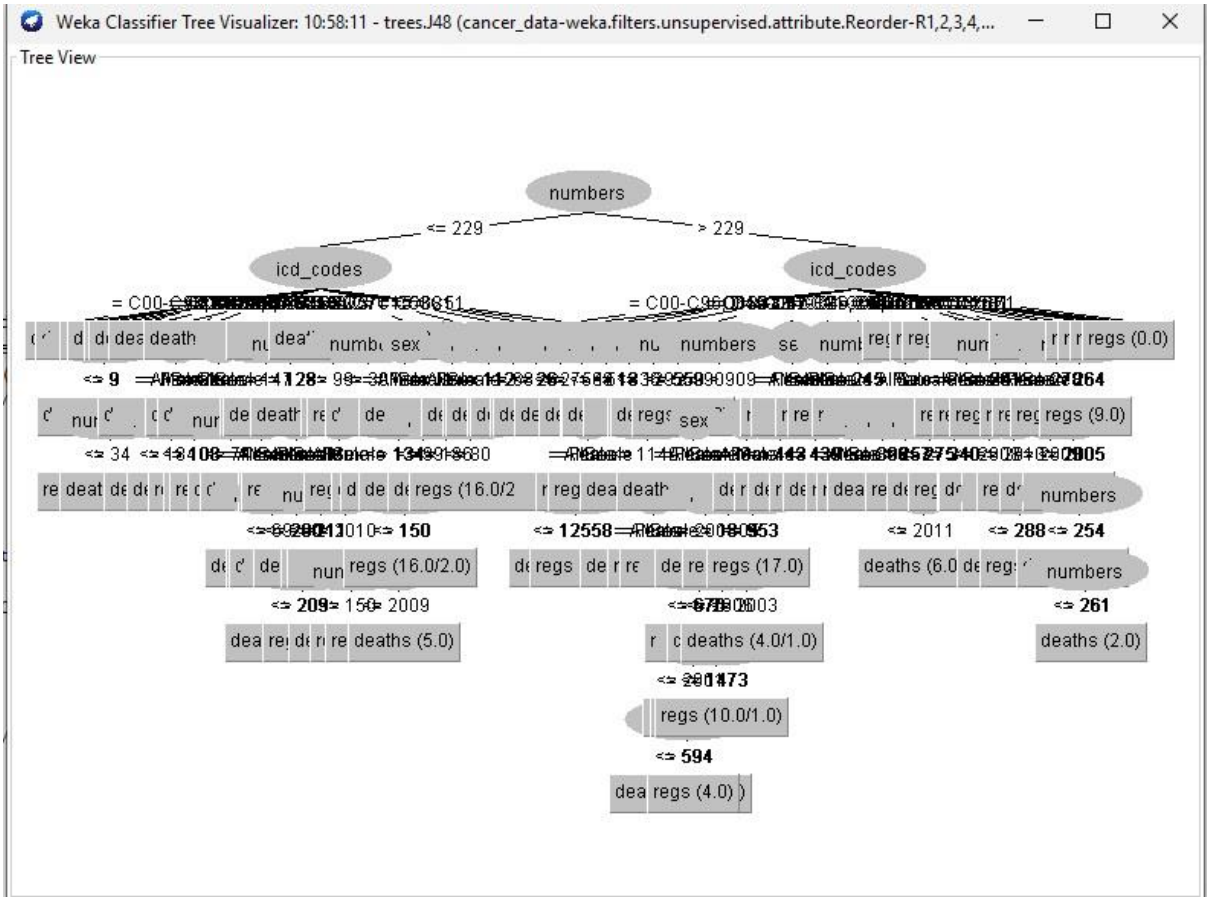


Figure 6: Decision Tree

For this Visual Tree ,Select the **trees J48** in the **Result List** then **Right Click** . After that Select **Visualize Tree .**

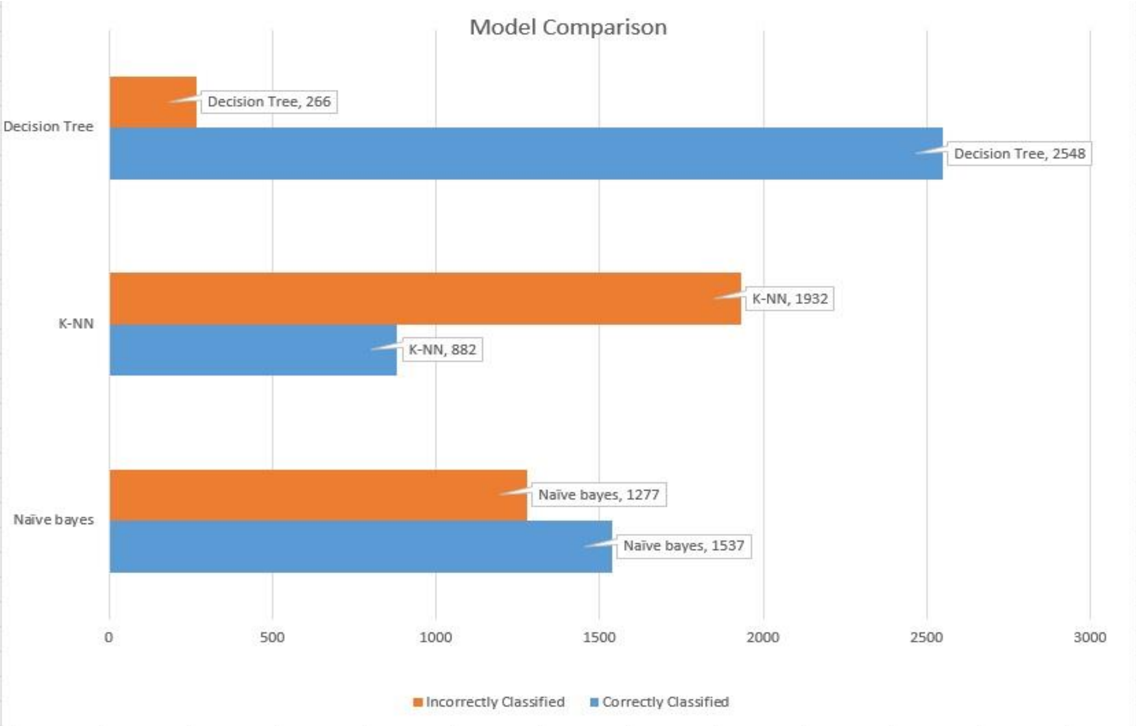


Figure 5 : The Bar chart we can see that Decision Tree is more efficient than other models , It can correctly classify 2548 from 2814 instances. Naïve Bayes can correctly classify 1537 from 2814 instances .K-NN can correctly classify 882 from 2814 instances

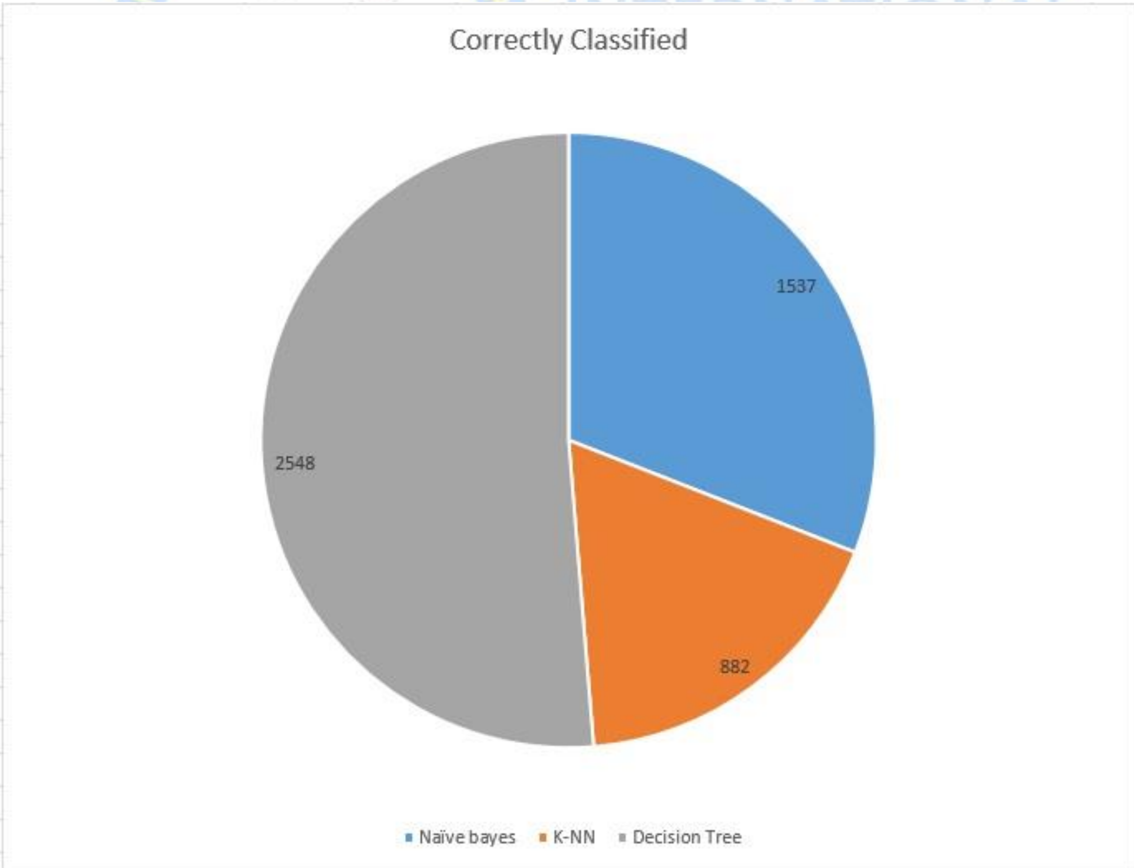


Figure 6: Pie chart showing the Amount of Correctly Classified Instances.

Section 4: Discussion And Conclusion

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         882               31.3433 %
Incorrectly Classified Instances      1932               68.6567 %
Kappa statistic                         -0.3734
Mean absolute error                      0.6864
Root mean squared error                  0.8281
Relative absolute error                137.3107 %
Root relative squared error            165.6421 %
Total Number of Instances             2814

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.322    0.696    0.323      0.322    0.323      -0.373   0.309     0.436     regs
                 0.304    0.678    0.304      0.304    0.304      -0.373   0.309     0.422     deaths
Weighted Avg.    0.313    0.687    0.313      0.313    0.313      -0.373   0.309     0.429

=== Confusion Matrix ===

   a   b   <-- classified as
 460 968 |   a = regs
 964 422 |   b = deaths
```

Status
OK

Figure 8: Confusion Matrix of K-NN

In this dataset(K-NN) there are only two classes, one is often regarded as “a=regs” and the other as “b=deaths”. This case the entries in the two rows and columns of the confusion matrix are referred to as regs and deaths. As there are only two classes in Cancer case dataset, the revised confusion matrix for cancer case test set according to regs and deaths true and false is given below

1. Of 1424 the instances classified as regs, 460 genuinely are regs (true regs) and the other 964 are really deaths (false deaths).
2. Of the 1390 instances classified as deaths, 968 are really regs (false deaths) and the other 422 are genuinely deaths (true deaths).

	a= regs	b=deaths
a= regs	460 (32.22%)	968(67.78%)
b=deaths	964 (69.56%)	422(30.44 %)

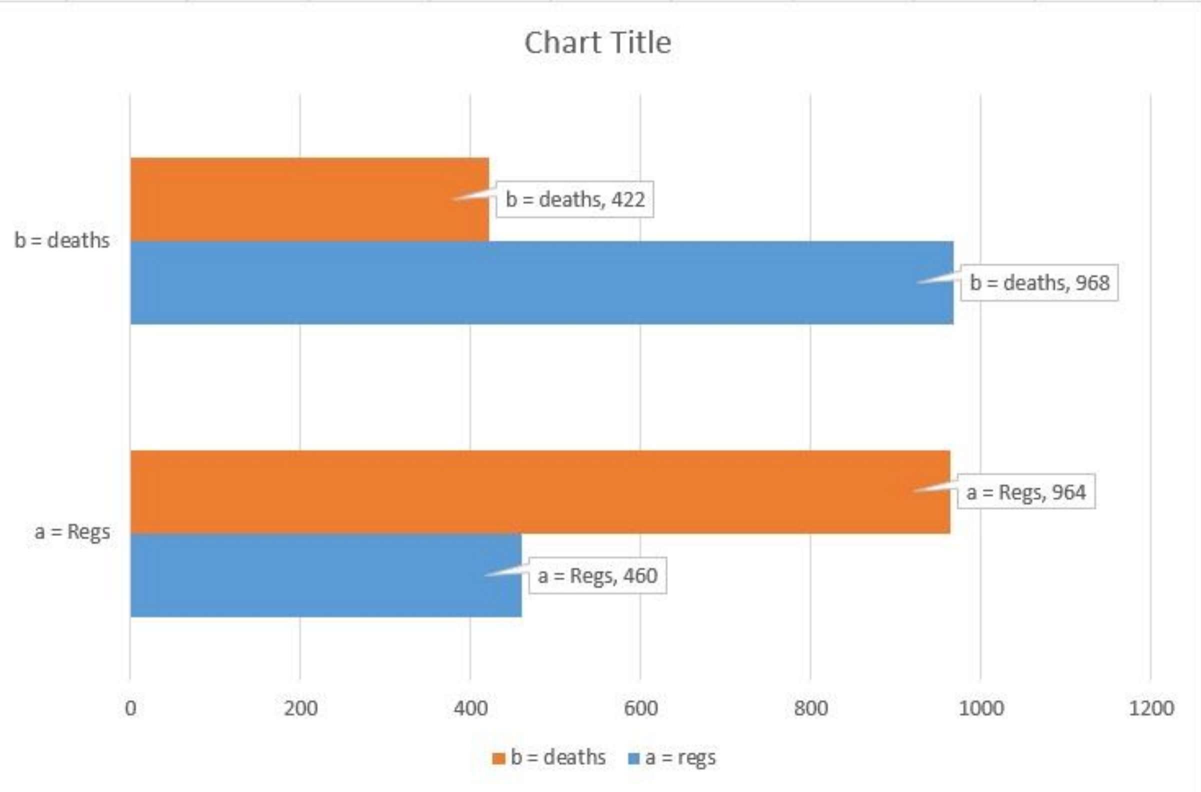


Figure: Chart of K-NN confusion matrix


```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2548               90.5473 %
Incorrectly Classified Instances       266                9.4527 %
Kappa statistic                          0.8109
Mean absolute error                      0.0976
Root mean squared error                  0.2557
Relative absolute error                 19.5279 %
Root relative squared error             51.1495 %
Total Number of Instances             2814

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.905    0.094    0.909      0.905    0.907      0.811    0.960     0.948     regs
                 0.906    0.095    0.902      0.906    0.904      0.811    0.960     0.948     deaths
Weighted Avg.    0.905    0.095    0.905      0.905    0.905      0.811    0.960     0.948

=== Confusion Matrix ===

    a    b   <-- classified as
 1292  136 |    a = regs
  130 1256 |    b = deaths
```

Figure 9: Confusion Matrix of Decision Tree

In this dataset (Decision Tree) there are only two classes, one is often regarded as "a=regs" and the other as "b=deaths". This case the entries in the two rows and columns of the confusion matrix are referred to as regs and deaths. As there are only two classes in Cancer case dataset, the revised confusion matrix for cancer case test set according to regs and deaths true and false is given below

1. Of 1422 the instances classified as regs, 1292 genuinely are regs (true regs) and the other 130 are really deaths (false deaths).
2. Of the 1392 instances classified as deaths, 136 are really regs (false deaths) and the other 1256 are genuinely deaths (true deaths).

	a= regs	b=deaths
a= regs	1292 (90.5 %)	136 (9.5%)
b=deaths	130 (9.4%)	1256 (90.6%)

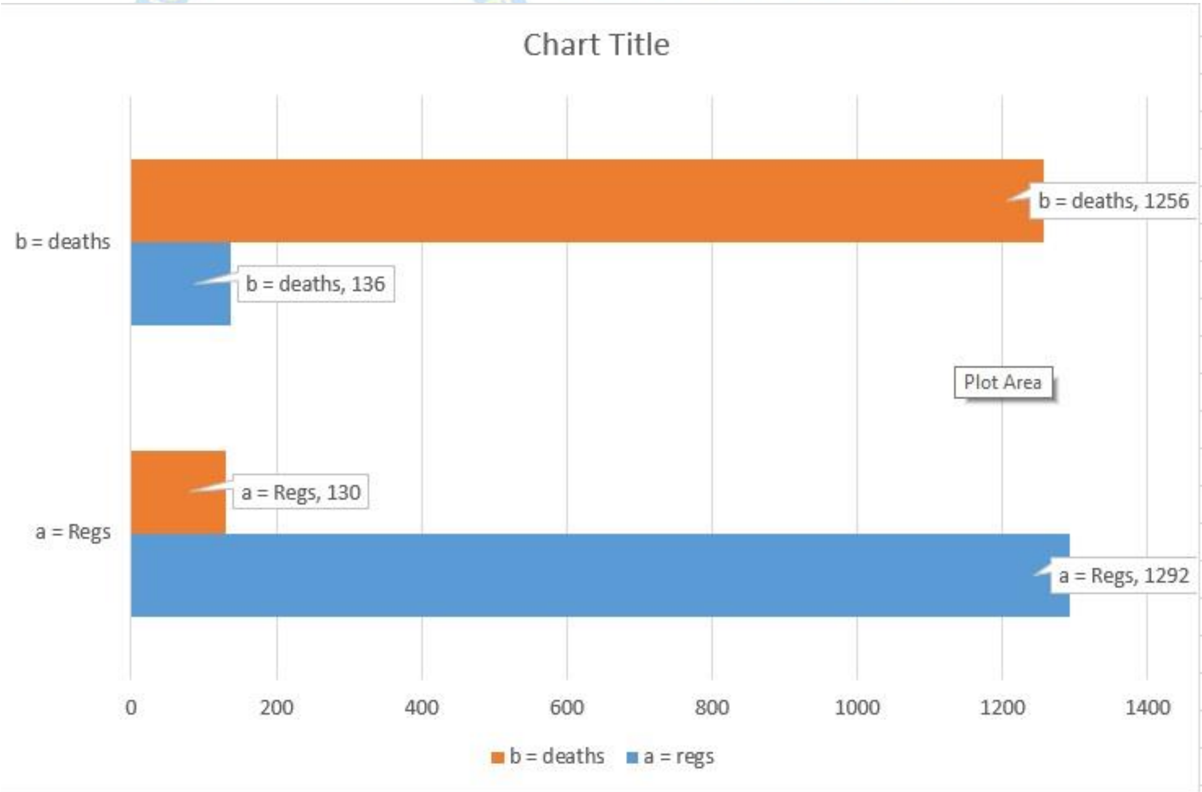


Figure: Chart of Decision tree confusion matrix

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1537              54.6198 %
Incorrectly Classified Instances      1277              45.3802 %
Kappa statistic                          0.1031
Mean absolute error                      0.4818
Root mean squared error                  0.5359
Relative absolute error                 96.3728 %
Root relative squared error            107.1892 %
Total Number of Instances             2814

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.150    0.045    0.773      0.150    0.251      0.175    0.490     0.562     regs
                 0.955    0.850    0.521      0.955    0.674      0.175    0.490     0.459     deaths
Weighted Avg.    0.546    0.442    0.649      0.546    0.460      0.175    0.490     0.511

=== Confusion Matrix ===

    a    b   <-- classified as
  214 1214 |    a = regs
   63 1323 |    b = deaths

tatus
)K
```

Figure 10 : Confusion Matrix of Naive bayes .

In this dataset (naïve Bayes) there are only two classes, one is often regarded as "a=regs" and the other as "b=deaths". This case the entries in the two rows and columns of the confusion matrix are referred to as regs and deaths. As there are only two classes in Cancer case dataset, the revised confusion matrix for cancer case test set according to regs and deaths true and false is given below

1. Of 277 the instances classified as regs, 214 genuinely are regs (true regs) and the other 63 are really deaths (false deaths).
2. Of the 2537 instances classified as deaths, 1214 are really regs (false deaths) and the other 1323 are genuinely deaths (true deaths).

	a= regs	b=deaths
a= regs	217 (15.16%)	1214 (84.84%)
b=deaths	63 (4.5 %)	1323 (95.5 %)

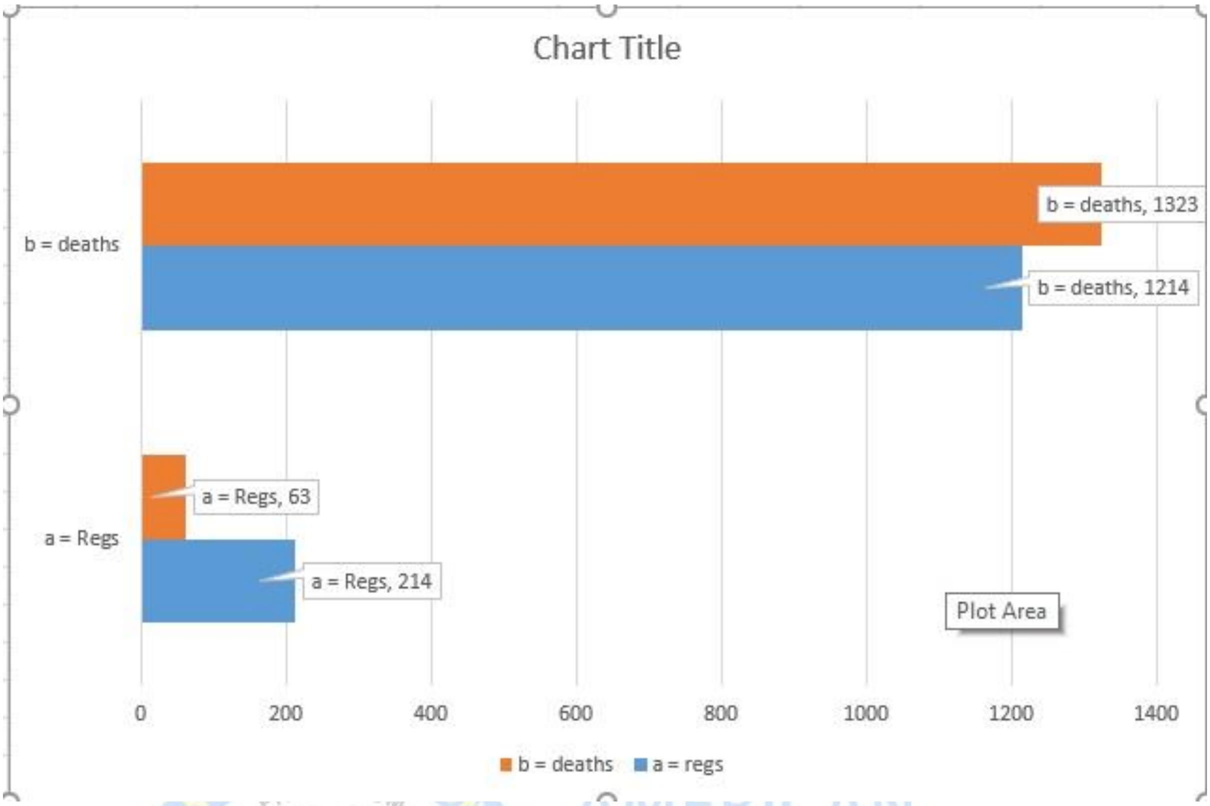


Figure: Chart of Naive Bayes confusion matrix

So for this specific Dataset Decision tree is Suitable. Because only Decision tree can correctly classify 2548 instance's from 2814 instances. Also from the confusion matrix we can see that in decision tree the false regs and false deaths are minimum then other models. With a perfect classifier there would be no false regs or false deaths. So we can say that Decision tree Model is suitable for this Data set. In this data set we have used 5-fold cross-validation for accuracy prediction.