DJILLALI LIABES UNIVERSITY OF SIDI BEL ABBES
FACULTY OF EXACT SCIENCES
DEPARTMENT OF COMPUTER SCIENCES

Module : Data Mining
1ST YEAR OF MASTER'S DEGEREE IN
NETWORKS,SYSTEMS & INFORMATION SECURITY(RSSI)
2021/2022

Ensemble Methods with Weka
TP-05

Student:
HADJAZI Mohammed
Hisham
Group: 01 / RSSI

Module Instructor:
Pr.ELBERRICHI Zakaria
TP Instructor:
Dr.FAHSI.Mahmoud

A paper submitted in fulfilment of the requirements for the
Data Mining TP-05

April 30, 2022

Contents

List of Figures

Chapter 1

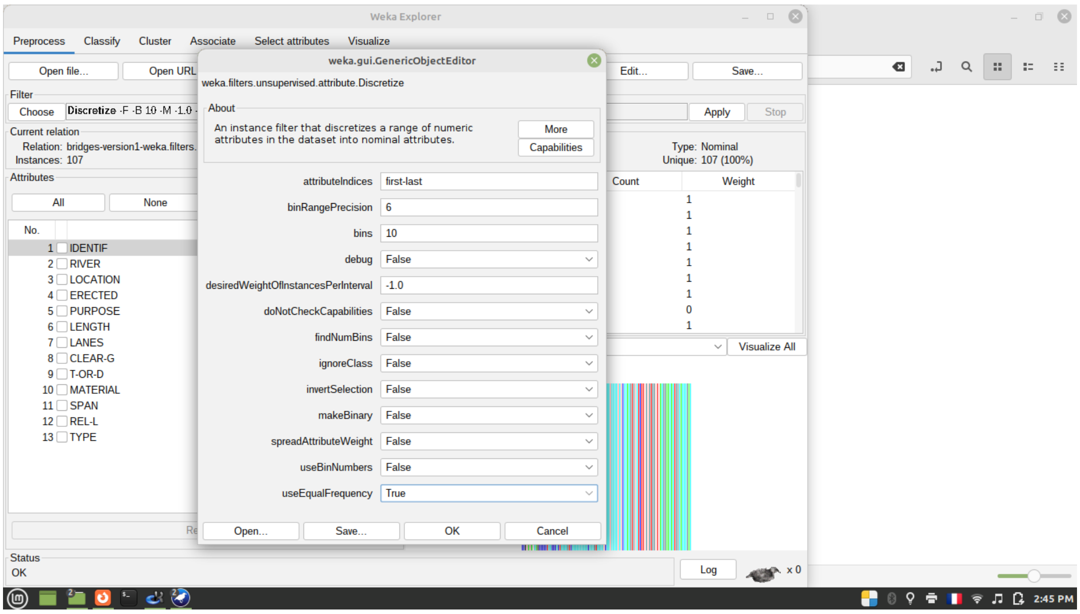
Dataset

1.1 Bridges dataset

```
 1. Title: Pittsburgh bridges

 2. Sources:
    -- Yoram Reich & Steven J. Fenves
       Department of Civil Engineering
       and
       Engineering Design Research Center
       Carnegie Mellon University
       Pittsburgh, PA 15213

       Compiled from various sources.

    -- Donor: Yoram Reich (yoram.reich@cs.cmu.edu)
    -- Date: 1 August 1990

 3. Past Usage:

    -- Reich & Fenves (1989). Incremental Learning for Capturing Design
       Expertise. Technical Report: EDRC 12-34-89, Engineering Design
       Research Center, Carnegie Mellon University, Pittsburgh, PA.
       -- Qualitative results and runs with original ordering of examples.
          using COBWEB.

    -- Reich (1989). Converging to ``Ideal'' Design Knowledge by Learning,
       Proceedings of the First International Workshop on Formal Methods in
       Engineering Design, pp: 330-349, Colorado Springs, CO, January 1990.
       -- Describes a new design method with Bridger (variant of COBWEB) using
   this domain. (Also an EDRC report: 12-35-89)

    -- Reich (1989) Combining Nominal and Continuous Properties in an
       Incremental Learning System for Design. Technical Report: EDRC 12-33-89.
       -- Comparison of performance of Bridger when running on both versions
   (V1 and V2) of the database

    -- Reich (1989) Incremental Concept Formation with Mixed Property Types
       Unpublished Manuscript.
       -- Results using 10 random 10-fold cross-validation test with Bridger
   (relative error rate):
   Version V1 of the database:
   MATERIAL 18.4%, REL-L 38.7%, SPAN 42.7%, T-OR-D 14.7%, TYPE 47.6%.
   Version V2 of the database:
   MATERIAL 24.2%, REL-L 41.7%, SPAN 39.9%, T-OR-D 14.7%, TYPE 56.5%.

    -- Quinlan (1989) Personal communication.
       -- Results of a 10-fold cross-validation test with C4.5, and with
          a separate decision tree for each design property obtained the
   following error rates on version V1 of the database:
   MATERIAL 15%, REL-L 32%, SPAN 32%, T-OR-D 15%, TYPE 44%.

 4. Number of instances: 108

 5. Relevant Information:

    There are two versions to the database:
       V1 contains the original examples and
       V2 contains descriptions after discretizing numeric properties.

    There are no ``classes'' in the domain. Rather this is a DESIGN domain where
    5 properties (design description) need to be predicted based on 7
    specification properties.

 6. Number of Attributes: 13: 7 specifications, 5 design description, and 1
    identifier (not used for the classification)

 7. Attribute Information:
   The type field state whether a property is continuous/integer (c)
       or nominal (n).
   For properties with c,n type, the range of continuous numbers is given
   first and the possible values of the nominal follow the semi-colon.


       name     type    possible values comments
   ----------------------------------------------------------------------
   1.  IDENTIF --identifier of the examples
   2.  RIVER n A, M, O
   3.  LOCATION n       1 to 52
   4.  ERECTED c,n 1818-1986 ; CRAFTS, EMERGING, MATURE, MODERN
   5.  PURPOSE n WALK, AQUEDUCT, RR, HIGHWAY
   6.  LENGTH c,n 804-4558 ; SHORT, MEDIUM, LONG
   7.  LANES c,n 1, 2, 4, 6 ; 1, 2, 4, 6
   8.  CLEAR-G n N, G
   9.  T-OR-D n THROUGH, DECK
   10. MATERIAL n WOOD, IRON, STEEL
   11. SPAN n SHORT, MEDUIM, LONG
   12. REL-L n S, S-F, F
   13. TYPE n WOOD, SUSPEN, SIMPLE-T, ARCH, CANTILEV, CONT-T


8. More complicated attributes:

   One can use a hierarchical structure for the Type property. There are two
   options.

 option 1 (use examples without modification)
        --------

 Type
      /      /      \     \
    /       /       \      \
 wood suspen arch truss
        /  |    \
      /    |      \
 cantilev  cont-t   simple


 option 2 (requires changes in the Type property - specified bellow)
 --------

 Type

  /      /        |          \
     /     /          |         \
     wood   suspen arch       truss
 / \          /  |  \    \
       /     \     /    |   \ \
 tied-a    not-tied  cantilev cont-t simple arch-t


 Change the Type  property of the following examples (in both V1 and V2):
 E28   ->  arch-t
 E91,E90,E84,E83,E73  -> tied-a
 E97,E78,E77,E75,E66,E64,E43  -> not-tied


9. Missing Attribute Values:
   Attribute #:  # instances with missing values:
    2  1
    6 27
    7 16
    8  2
    9  6
   10  2
   11 16
   12  5
   13  3

Information about the dataset
CLASSTYPE: nominal
CLASSINDEX: no
```

1.1.1 Appling Filters

Descritize filter with equal frequency binding



Filter: Discretize -F -B 10 -M -1.0

Current relation
Relation: bridges-version1-weka.filters.
Instances: 107

Type: Nominal
Unique: 107 (100%)

No.	Name
1	IDENTIF
2	RIVER
3	LOCATION
4	ERECTED
5	PURPOSE
6	LENGTH
7	LANES
8	CLEAR-G
9	T-OR-D
10	MATERIAL
11	SPAN
12	REL-L
13	TYPE

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

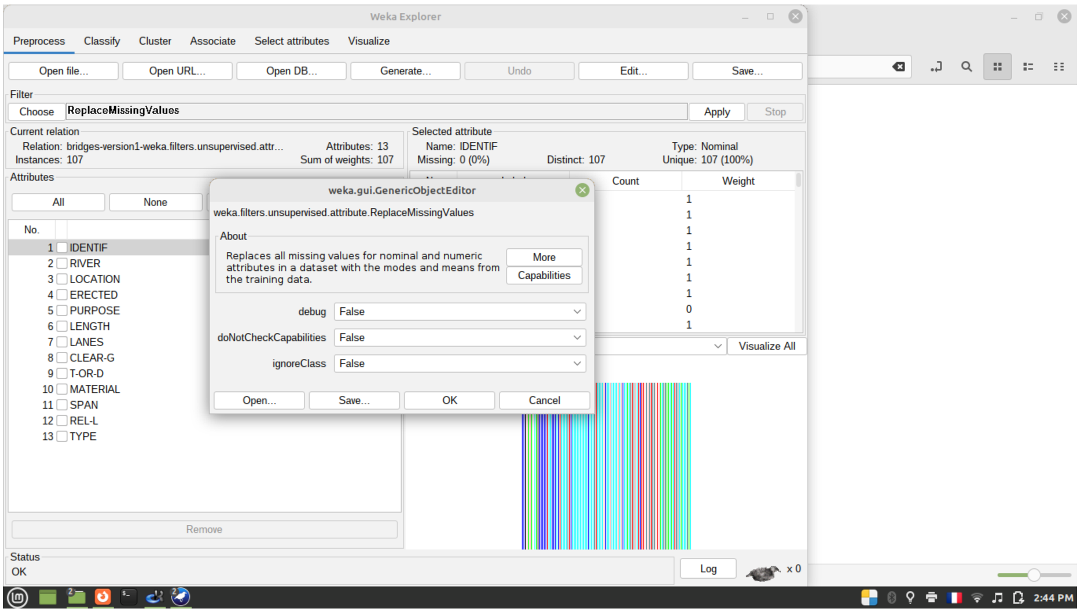
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.

attributeIndices	first-last
binRangePrecision	6
bins	10
debug	False
desiredWeightOfInstancesPerInterval	-1.0
doNotCheckCapabilities	False
findNumBins	False
ignoreClass	False
invertSelection	False
makeBinary	False
spreadAttributeWeight	False
useBinNumbers	False
useEqualFrequency	True

Status
OK

FIGURE 1.1: Descritize filter

Replace Missing Values filter



Filter: ReplaceMissingValues

Current relation
Relation: bridges-version1-weka.filters.unsupervised.attr... Attributes: 13
Instances: 107 Sum of weights: 107

Selected attribute
Name: IDENTIF Type: Nominal
Missing: 0 (0%) Distinct: 107 Unique: 107 (100%)

No.	Name
1	IDENTIF
2	RIVER
3	LOCATION
4	ERECTED
5	PURPOSE
6	LENGTH
7	LANES
8	CLEAR-G
9	T-OR-D
10	MATERIAL
11	SPAN
12	REL-L
13	TYPE

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.ReplaceMissingValues

About
Replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data.

debug	False
doNotCheckCapabilities	False
ignoreClass	False

Status
OK

FIGURE 1.2: Replace Missing Values filter

Chapter 2

Choosing Algorithms Process

2.1 Introduction

The choice of algorithms was totally random as it is not the point of this TP. it must be noted that all algorithms were run on default settings except KNN which had a k value of 3.

2.1.1 Logistic Regression

2.1.2 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          70               66.6667 %
Incorrectly Classified Instances        35               33.3333 %
Kappa statistic                          0.5366
Mean absolute error                      0.1247
Root mean squared error                  0.2817
Relative absolute error                 49.2037 %
Root relative squared error             79.3308 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.091    0.011    0.500      0.091    0.154      0.180    0.759     0.258     SUSPEN
                 0.886    0.279    0.696      0.886    0.780      0.601    0.885     0.827     SIMPLE-T
                 0.538    0.043    0.636      0.538    0.583      0.532    0.848     0.606     ARCH
                 0.091    0.085    0.111      0.091    0.100      0.006    0.673     0.181     CANTILEV
                 0.600    0.053    0.545      0.600    0.571      0.525    0.888     0.702     CONT-T
Weighted Avg.    0.667    0.137    0.639      0.667    0.632      0.540    0.863     0.687

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  0  1  7  2  1  0 |  b = SUSPEN
  0  1 39  1  3  0 |  c = SIMPLE-T
  0  0  3  7  2  1 |  d = ARCH
  0  0  5  1  1  4 |  e = CANTILEV
  0  0  2  0  2  6 |  f = CONT-T
```

2.1.3 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          22               62.8571 %
Incorrectly Classified Instances        13               37.1429 %
Kappa statistic                          0.4939
Mean absolute error                      0.1214
Root mean squared error                  0.3107
Relative absolute error                 46.951  %
Root relative squared error             84.8568 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.652     0.218     SUSPEN
                 0.917    0.435    0.524      0.917    0.667      0.467    0.927     0.910     SIMPLE-T
                 0.200    0.033    0.500      0.200    0.286      0.251    0.929     0.573     ARCH
                 0.333    0.031    0.500      0.333    0.400      0.364    0.889     0.544     CANTILEV
                 0.667    0.031    0.667      0.667    0.667      0.635    0.980     0.867     CONT-T
Weighted Avg.    0.629    0.159    ?          0.629    ?          ?        0.904     0.746

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  0  0  4  1  0  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  4  1  0  0 |  d = ARCH
  0  0  2  0  1  0 |  e = CANTILEV
  0  0  0  0  1  2 |  f = CONT-T
```

2.1.4 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          70               66.6667 %
Incorrectly Classified Instances        35               33.3333 %
Kappa statistic                          0.5344
Mean absolute error                      0.1187
Root mean squared error                  0.2901
Relative absolute error                 46.4745 %
Root relative squared error             81.0366 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.091    0.011    0.500      0.091    0.154      0.180    0.712     0.222     SUSPEN
                 0.886    0.295    0.684      0.886    0.772      0.586    0.866     0.810     SIMPLE-T
                 0.538    0.065    0.538      0.538    0.538      0.473    0.889     0.646     ARCH
                 0.091    0.064    0.143      0.091    0.111      0.033    0.600     0.180     CANTILEV
                 0.600    0.042    0.600      0.600    0.600      0.558    0.948     0.680     CONT-T
Weighted Avg.    0.667    0.144    0.630      0.667    0.628      0.532    0.853     0.679

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  0  1  8  2  0  0 |  b = SUSPEN
  0  1 39  3  1  0 |  c = SIMPLE-T
  0  0  4  7  2  0 |  d = ARCH
  0  0  5  1  1  4 |  e = CANTILEV
  0  0  1  0  3  6 |  f = CONT-T
```

2.1.5 Naive Bayes Default Settings

2.1.6 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          74               70.4762 %
Incorrectly Classified Instances        31               29.5238 %
Kappa statistic                          0.6065
Mean absolute error                      0.1248
Root mean squared error                  0.2786
Relative absolute error                 49.239  %
Root relative squared error             78.4653 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.988     0.899     WOOD
                 0.273    0.032    0.500      0.273    0.353      0.318    0.684     0.344     SUSPEN
                 0.841    0.131    0.822      0.841    0.831      0.708    0.920     0.901     SIMPLE-T
                 0.462    0.076    0.462      0.462    0.462      0.385    0.867     0.515     ARCH
                 0.364    0.043    0.500      0.364    0.421      0.371    0.790     0.337     CANTILEV
                 0.800    0.053    0.615      0.800    0.696      0.666    0.929     0.580     CONT-T
Weighted Avg.    0.705    0.084    0.687      0.705    0.688      0.613    0.886     0.705
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  3  2  0  0 |  b = SUSPEN
  1  1 37  3  2  0 |  c = SIMPLE-T
  0  2  3  6  1  1 |  d = ARCH
  0  0  2  1  4  4 |  e = CANTILEV
  0  0  0  1  1  8 |  f = CONT-T
```

2.1.7 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          27               77.1429 %
Incorrectly Classified Instances         8               22.8571 %
Kappa statistic                          0.7086
Mean absolute error                      0.1159
Root mean squared error                  0.2539
Relative absolute error                 44.829  %
Root relative squared error             69.3254 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.036    0.875      1.000    0.933      0.919    0.980     0.909     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.561     0.178     SUSPEN
                 0.917    0.087    0.846      0.917    0.880      0.815    0.962     0.940     SIMPLE-T
                 0.800    0.033    0.800      0.800    0.800      0.767    0.968     0.835     ARCH
                 0.667    0.063    0.500      0.667    0.571      0.532    0.929     0.767     CANTILEV
                 1.000    0.063    0.600      1.000    0.750      0.750    1.000     1.000     CONT-T
Weighted Avg.    0.771    0.052    ?          0.771    ?          ?        0.910     0.800

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  1  0  2  1  1  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  0  4  1  0 |  d = ARCH
  0  0  0  0  2  1 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

2.1.8 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          73               69.5238 %
Incorrectly Classified Instances        32               30.4762 %
Kappa statistic                          0.5922
Mean absolute error                      0.1278
Root mean squared error                  0.2808
Relative absolute error                 50.0526 %
Root relative squared error             78.4417 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.990     0.929     WOOD
                 0.273    0.032    0.500      0.273    0.353      0.318    0.661     0.341     SUSPEN
                 0.864    0.131    0.826      0.864    0.844      0.728    0.917     0.893     SIMPLE-T
                 0.308    0.087    0.333      0.308    0.320      0.229    0.855     0.455     ARCH
                 0.364    0.053    0.444      0.364    0.400      0.340    0.770     0.389     CANTILEV
                 0.800    0.042    0.667      0.800    0.727      0.699    0.932     0.593     CONT-T
Weighted Avg.    0.695    0.085    0.672      0.695    0.677      0.602    0.880     0.705

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  3  2  0  0 |  b = SUSPEN
  1  0 38  3  2  0 |  c = SIMPLE-T
  0  3  3  4  2  1 |  d = ARCH
  0  0  2  2  4  3 |  e = CANTILEV
  0  0  0  1  1  8 |  f = CONT-T
```

2.1.9 KNN k-nearest neighbors with k = 3

2.1.10 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          64               60.9524 %
Incorrectly Classified Instances        41               39.0476 %
Kappa statistic                          0.4692
Mean absolute error                      0.1421
Root mean squared error                  0.2872
Relative absolute error                 56.0825 %
Root relative squared error             80.8694 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.981     0.827     WOOD
                 0.273    0.106    0.231      0.273    0.250      0.155    0.453     0.159     SUSPEN
                 0.841    0.213    0.740      0.841    0.787      0.620    0.908     0.833     SIMPLE-T
                 0.308    0.043    0.500      0.308    0.381      0.328    0.874     0.673     ARCH
                 0.091    0.074    0.125      0.091    0.105      0.019    0.795     0.243     CANTILEV
                 0.300    0.032    0.500      0.300    0.375      0.339    0.854     0.544     CONT-T
Weighted Avg.    0.610    0.123    0.579      0.610    0.585      0.484    0.850     0.652

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  4  0  1  0 |  b = SUSPEN
  1  2 37  2  2  0 |  c = SIMPLE-T
  0  4  2  4  2  1 |  d = ARCH
  0  2  5  1  1  2 |  e = CANTILEV
  0  2  2  1  2  3 |  f = CONT-T
```

2.1.11 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          22               62.8571 %
Incorrectly Classified Instances        13               37.1429 %
Kappa statistic                          0.5033
Mean absolute error                      0.1407
Root mean squared error                  0.2743
Relative absolute error                 54.4202 %
Root relative squared error             74.9086 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.036    0.875      1.000    0.933      0.919    0.998     0.982     WOOD
                 0.000    0.067    0.000      0.000    0.000      -0.101   0.374     0.125     SUSPEN
                 1.000    0.261    0.667      1.000    0.800      0.702    0.981     0.959     SIMPLE-T
                 0.200    0.067    0.333      0.200    0.250      0.167    0.942     0.758     ARCH
                 0.333    0.031    0.500      0.333    0.400      0.364    0.949     0.567     CANTILEV
                 0.333    0.031    0.500      0.333    0.400      0.364    0.924     0.758     CONT-T
Weighted Avg.    0.629    0.121    0.537      0.629    0.565      0.496    0.884     0.765

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  1  0  3  1  0  0 |  b = SUSPEN
  0  0 12  0  0  0 |  c = SIMPLE-T
  0  1  2  1  0  1 |  d = ARCH
  0  0  1  1  1  0 |  e = CANTILEV
  0  1  0  0  1  1 |  f = CONT-T
```

2.1.12 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          64               60.9524 %
Incorrectly Classified Instances        41               39.0476 %
Kappa statistic                          0.4669
Mean absolute error                      0.1437
Root mean squared error                  0.2858
Relative absolute error                 56.2786 %
```

```
Root relative squared error             79.8292 %
Total Number of Instances              105     
Ignored Class Unknown Instances                  2     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.981     0.836     WOOD
                 0.273    0.096    0.250      0.273    0.261      0.170    0.415     0.187     SUSPEN
                 0.841    0.230    0.725      0.841    0.779      0.604    0.912     0.864     SIMPLE-T
                 0.308    0.054    0.444      0.308    0.364      0.298    0.903     0.622     ARCH
                 0.091    0.064    0.143      0.091    0.111      0.033    0.754     0.226     CANTILEV
                 0.300    0.032    0.500      0.300    0.375      0.339    0.846     0.490     CONT-T
Weighted Avg.    0.610    0.129    0.570      0.610    0.582      0.477    0.847     0.657     

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  4  1  0  0 |  b = SUSPEN
  1  2 37  2  2  0 |  c = SIMPLE-T
  0  3  3  4  2  1 |  d = ARCH
  0  2  5  1  1  2 |  e = CANTILEV
  0  2  2  1  2  3 |  f = CONT-T
```

2.1.13 One Rule Default Settings

2.1.14 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          16               15.2381 %
Incorrectly Classified Instances        89               84.7619 %
Kappa statistic                          0     
Mean absolute error                      0.2825
Root mean squared error                  0.5315
Relative absolute error                111.4908 %
Root relative squared error            149.6859 %
Total Number of Instances              105     
Ignored Class Unknown Instances                  2     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.152      1.000    0.264      ?        0.500     0.150     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.103     SUSPEN
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.411     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.121     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.103     CANTILEV
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.093     CONT-T
Weighted Avg.    0.152    0.152    ?          0.152    ?          ?        0.500     0.241     

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
 11  0  0  0  0  0 |  b = SUSPEN
 44  0  0  0  0  0 |  c = SIMPLE-T
 13  0  0  0  0  0 |  d = ARCH
 11  0  0  0  0  0 |  e = CANTILEV
 10  0  0  0  0  0 |  f = CONT-T
```

2.1.15 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances           7               20      %
Incorrectly Classified Instances        28               80      %
Kappa statistic                          0     
Mean absolute error                      0.2667
Root mean squared error                  0.5164
Relative absolute error                103.1508 %
Root relative squared error            141.0174 %
Total Number of Instances               35     
Ignored Class Unknown Instances                  1     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.200      1.000    0.333      ?        0.500     0.194     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.139     SUSPEN
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.333     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.139     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.083     CANTILEV
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.083     CONT-T
Weighted Avg.    0.200    0.200    ?          0.200    ?          ?        0.500     0.207

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  5  0  0  0  0  0 |  b = SUSPEN
 12  0  0  0  0  0 |  c = SIMPLE-T
  5  0  0  0  0  0 |  d = ARCH
  3  0  0  0  0  0 |  e = CANTILEV
  3  0  0  0  0  0 |  f = CONT-T
```

2.1.16 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          16               15.2381 %
Incorrectly Classified Instances        89               84.7619 %
Kappa statistic                          0
Mean absolute error                      0.2825
Root mean squared error                  0.5315
Relative absolute error                110.659  %
Root relative squared error            148.4891 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.152      1.000    0.264      ?        0.500     0.150     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.103     SUSPEN
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.411     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.121     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.103     CANTILEV
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.093     CONT-T
Weighted Avg.    0.152    0.152    ?          0.152    ?          ?        0.500     0.241

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
 11  0  0  0  0  0 |  b = SUSPEN
 44  0  0  0  0  0 |  c = SIMPLE-T
 13  0  0  0  0  0 |  d = ARCH
 11  0  0  0  0  0 |  e = CANTILEV
 10  0  0  0  0  0 |  f = CONT-T
```

2.1.17 PART Default Settings

2.1.18 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          61               58.0952 %
Incorrectly Classified Instances        44               41.9048 %
Kappa statistic                          0.3422
Mean absolute error                      0.187
Root mean squared error                  0.3111
Relative absolute error                 73.8067 %
Root relative squared error             87.5989 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.683     0.244     SUSPEN
                 0.977    0.672    0.512      0.977    0.672      0.376    0.675     0.565     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.608     0.158     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.634     0.142     CANTILEV
                 0.200    0.032    0.400      0.200    0.267      0.232    0.710     0.215     CONT-T
Weighted Avg.    0.581    0.285    ?          0.581    ?          ?        0.716     0.469

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
```

```
  0  0 11  0  0  0 |  b = SUSPEN
  0  0 43  0  0  1 |  c = SIMPLE-T
  0  0 12  0  0  1 |  d = ARCH
  0  0 10  0  0  1 |  e = CANTILEV
  0  0  8  0  0  2 |  f = CONT-T
```

2.1.19 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          21               60      %
Incorrectly Classified Instances        14               40      %
Kappa statistic                          0.4406
Mean absolute error                      0.1713
Root mean squared error                  0.3052
Relative absolute error                 66.2541 %
Root relative squared error             83.3438 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.726     0.260     SUSPEN
                 0.917    0.565    0.458      0.917    0.611      0.359    0.672     0.436     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.710     0.217     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.576     0.111     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    0.985     0.750     CONT-T
Weighted Avg.    0.600    0.196    ?          0.600    ?          ?        0.769     0.492

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  0  0  5  0  0  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  5  0  0  0 |  d = ARCH
  0  0  3  0  0  0 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

2.1.20 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          59               56.1905 %
Incorrectly Classified Instances        46               43.8095 %
Kappa statistic                          0.2987
Mean absolute error                      0.2007
Root mean squared error                  0.3202
Relative absolute error                 78.6194 %
Root relative squared error             89.4585 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.383     0.283     SUSPEN
                 0.977    0.738    0.489      0.977    0.652      0.321    0.430     0.497     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.487     0.205     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.177     0.094     CANTILEV
                 0.000    0.011    0.000      0.000    0.000      -0.032   0.410     0.128     CONT-T
Weighted Avg.    0.562    0.310    ?          0.562    ?          ?        0.490     0.437

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  0  0 11  0  0  0 |  b = SUSPEN
  0  0 43  0  0  1 |  c = SIMPLE-T
  0  0 13  0  0  0 |  d = ARCH
  0  0 11  0  0  0 |  e = CANTILEV
  0  0 10  0  0  0 |  f = CONT-T
```

2.1.21 C4.5 Default Settings

2.1.22 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          55               52.381  %
Incorrectly Classified Instances        50               47.619  %
Kappa statistic                          0.3199
Mean absolute error                      0.1739
Root mean squared error                  0.3341
Relative absolute error                 68.6043 %
Root relative squared error             94.0974 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.000    0.053    0.000      0.000    0.000      -0.076   0.706     0.205     SUSPEN
                 0.750    0.475    0.532      0.750    0.623      0.276    0.702     0.578     SIMPLE-T
                 0.308    0.120    0.267      0.308    0.286      0.177    0.641     0.251     ARCH
                 0.000    0.011    0.000      0.000    0.000      -0.034   0.590     0.123     CANTILEV
                 0.200    0.042    0.333      0.200    0.250      0.200    0.701     0.205     CONT-T
Weighted Avg.    0.524    0.225    0.440      0.524    0.472      0.297    0.728     0.480

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  0  0  7  3  1  0 |  b = SUSPEN
  0  3 33  6  0  2 |  c = SIMPLE-T
  0  1  7  4  0  1 |  d = ARCH
  0  1  8  1  0  1 |  e = CANTILEV
  0  0  7  1  0  2 |  f = CONT-T
```

2.1.23 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          21               60      %
Incorrectly Classified Instances        14               40      %
Kappa statistic                          0.4406
Mean absolute error                      0.1713
Root mean squared error                  0.3052
Relative absolute error                 66.2541 %
Root relative squared error             83.3438 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.726     0.260     SUSPEN
                 0.917    0.565    0.458      0.917    0.611      0.359    0.672     0.436     SIMPLE-T
                 0.000    0.000    ?          0.000    ?          ?        0.710     0.217     ARCH
                 0.000    0.000    ?          0.000    ?          ?        0.576     0.111     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    0.985     0.750     CONT-T
Weighted Avg.    0.600    0.196    ?          0.600    ?          ?        0.769     0.492

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  0  0  5  0  0  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  5  0  0  0 |  d = ARCH
  0  0  3  0  0  0 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

2.1.24 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          46               43.8095 %
Incorrectly Classified Instances        59               56.1905 %
Kappa statistic                          0.1708
Mean absolute error                      0.1774
Root mean squared error                  0.3569
Relative absolute error                 69.4986 %
```

```
Root relative squared error             99.7122 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== 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     WOOD
                 0.000    0.064    0.000      0.000    0.000      -0.084   0.623     0.189     SUSPEN
                 0.682    0.656    0.429      0.682    0.526      0.027    0.653     0.591     SIMPLE-T
                 0.000    0.076    0.000      0.000    0.000      -0.100   0.618     0.197     ARCH
                 0.000    0.011    0.000      0.000    0.000      -0.034   0.652     0.172     CANTILEV
                 0.000    0.053    0.000      0.000    0.000      -0.073   0.564     0.156     CONT-T
Weighted Avg.    0.438    0.297    0.332      0.438    0.373      0.132    0.690     0.477

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  0  0  9  1  1  0 |  b = SUSPEN
  0  4 30  6  0  4 |  c = SIMPLE-T
  0  1 11  0  0  1 |  d = ARCH
  0  1 10  0  0  0 |  e = CANTILEV
  0  0 10  0  0  0 |  f = CONT-T
```

2.1.25 Conclusion

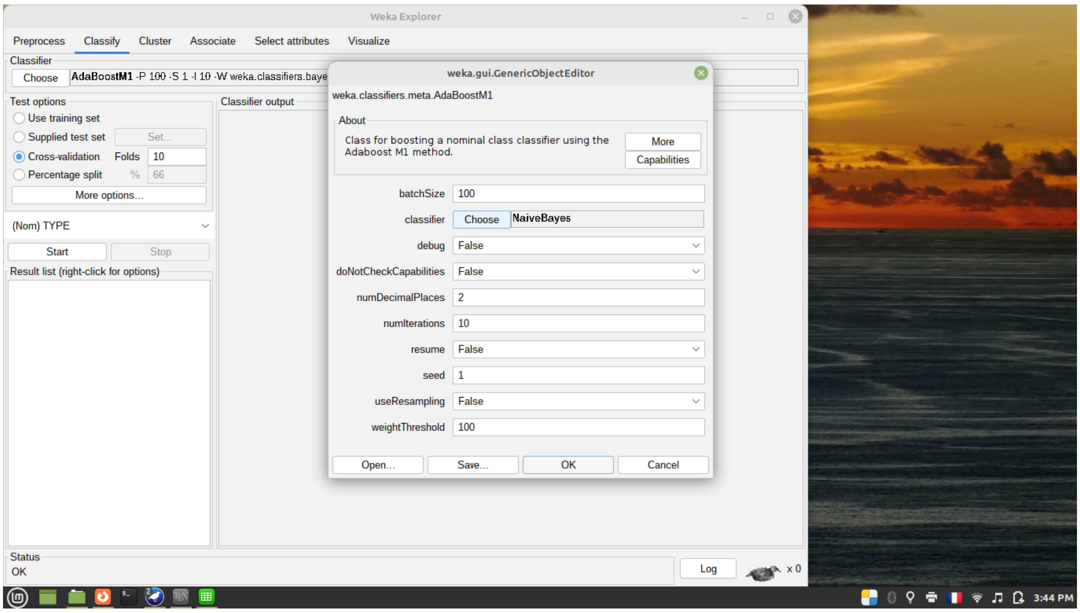
Correctly Classified Instances by Algorithm					
Evaluation Process	Cross Validation 10 Folds	Percentage Split 66%	Leave One Out Fold	AVG Algorithms	Rank Algorithms
Logistic Regression	66.6667%	62.8571%	66.6667%	65.3968%	2
Naïve Bayes	70.4762%	77.1429%	69.5238%	72.3810%	1
KNN k=3	60.9524%	62.8571%	60.9524%	61.5873%	3
One Rule	15.2381%	20.0000%	15.2381%	16.8254%	6
PART	58.0952%	60.0000%	56.1905%	58.0952%	4
C4.5	52.3810%	60.0000%	43.8095%	52.0635%	5

We find out that Naive Bayes is the best performer here in all evaluation tests with best result in Percentage Split 66% with 77.1429% Correctly Classified Instances and an average of 72.3810%. so we will be comparing ensemble methods to Naive Bayes and try to improve it.

Chapter 3

Meta Learning (Ensemble Algorithms)

3.1 Boosting Default Settings



Classifier: AdaBoostM1 -P 100 -S 1 -I 10 -W weka.classifiers.baye

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

(Nom) TYPE

Status
OK

weka.classifiers.meta.AdaBoostM1

About

Class for boosting a nominal class classifier using the Adaboost M1 method.

batchSize	100
classifier	Choose NaiveBayes
debug	False
doNotCheckCapabilities	False
numDecimalPlaces	2
numIterations	10
resume	False
seed	1
useResampling	False
weightThreshold	100

FIGURE 3.1: Boosting

3.1.1 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          67               63.8095 %
Incorrectly Classified Instances        38               36.1905 %
Kappa statistic                          0.5255
Mean absolute error                      0.1181
Root mean squared error                  0.3332
Relative absolute error                 46.6022 %
Root relative squared error             93.8417 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.938    0.034    0.833      0.938    0.882      0.862    0.985     0.870     WOOD
                 0.364    0.064    0.400      0.364    0.381      0.313    0.691     0.335     SUSPEN
                 0.727    0.148    0.780      0.727    0.753      0.586    0.851     0.822     SIMPLE-T
                 0.462    0.076    0.462      0.462    0.462      0.385    0.892     0.446     ARCH
                 0.455    0.106    0.333      0.455    0.385      0.305    0.735     0.350     CANTILEV
                 0.500    0.032    0.625      0.500    0.556      0.518    0.907     0.683     CONT-T
```

```
Weighted Avg.    0.638    0.097    0.648      0.638    0.640      0.539    0.853     0.669

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 15  0  1  0  0  0 |  a = WOOD
  2  4  3  2  0  0 |  b = SUSPEN
  1  3 32  4  4  0 |  c = SIMPLE-T
  0  2  3  6  2  0 |  d = ARCH
  0  1  2  0  5  3 |  e = CANTILEV
  0  0  0  1  4  5 |  f = CONT-T
```

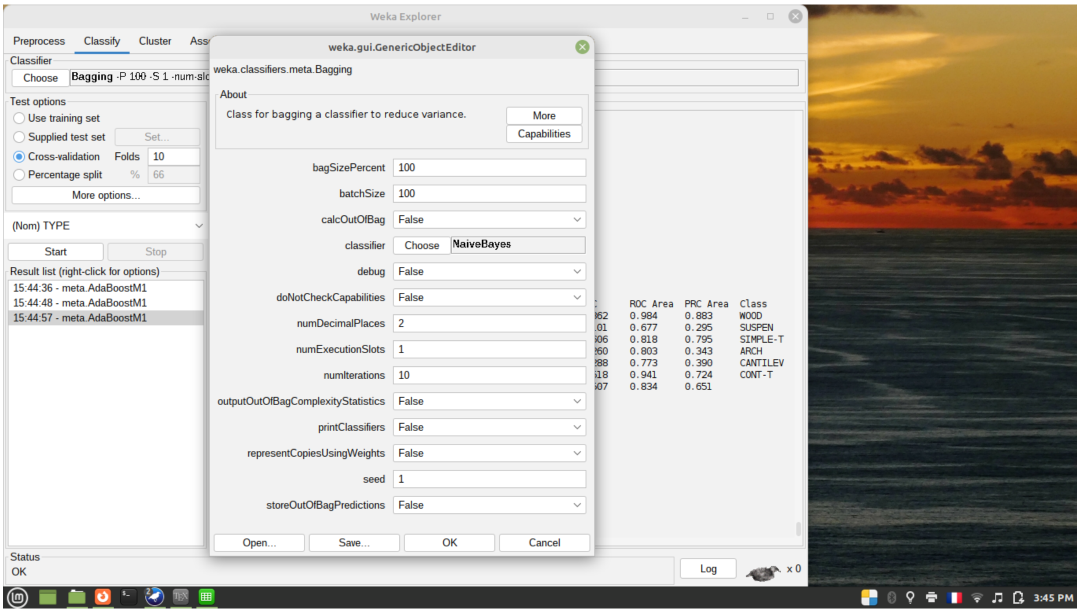
3.1.2 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          25               71.4286 %
Incorrectly Classified Instances        10               28.5714 %
Kappa statistic                          0.6392
Mean absolute error                      0.0951
Root mean squared error                  0.3016
Relative absolute error                 36.787  %
Root relative squared error             82.3474 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.857    0.036    0.857      0.857    0.857      0.821    0.973     0.802     WOOD
                 0.000    0.033    0.000      0.000    0.000      -0.070   0.545     0.167     SUSPEN
                 0.833    0.087    0.833      0.833    0.833      0.746    0.931     0.899     SIMPLE-T
                 0.800    0.067    0.667      0.800    0.727      0.681    0.916     0.750     ARCH
                 0.667    0.094    0.400      0.667    0.500      0.458    0.949     0.698     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    1.000     1.000     CONT-T
Weighted Avg.    0.714    0.062    0.651      0.714    0.677      0.620    0.889     0.745

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  6  1  0  0  0  0 |  a = WOOD
  1  0  2  2  0  0 |  b = SUSPEN
  0  0 10  0  2  0 |  c = SIMPLE-T
  0  0  0  4  1  0 |  d = ARCH
  0  0  0  0  2  1 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

3.1.3 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          63               60      %
Incorrectly Classified Instances        42               40      %
Kappa statistic                          0.4814
Mean absolute error                      0.129
Root mean squared error                  0.3469
Relative absolute error                 50.5412 %
Root relative squared error             96.9025 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.938    0.034    0.833      0.938    0.882      0.862    0.984     0.883     WOOD
                 0.182    0.085    0.200      0.182    0.190      0.101    0.677     0.295     SUSPEN
                 0.705    0.115    0.816      0.705    0.756      0.606    0.818     0.795     SIMPLE-T
                 0.385    0.109    0.333      0.385    0.357      0.260    0.803     0.343     ARCH
                 0.455    0.117    0.313      0.455    0.370      0.288    0.773     0.390     CANTILEV
                 0.500    0.032    0.625      0.500    0.556      0.518    0.941     0.724     CONT-T
Weighted Avg.    0.600    0.091    0.623      0.600    0.607      0.507    0.834     0.651

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 15  0  1  0  0  0 |  a = WOOD
  2  2  2  4  0  1 |  b = SUSPEN
  1  2 31  4  6  0 |  c = SIMPLE-T
  0  5  2  5  1  0 |  d = ARCH
  0  1  2  1  5  2 |  e = CANTILEV
  0  0  0  1  4  5 |  f = CONT-T
```

3.2 Bagging Default Settings



Classifier

Choose Bagging -P 100 -S 1 -num-slc

Test options

- ○ Use training set
- ○ Supplied test set Set...
- ● Cross-validation Folds 10
- ○ Percentage split % 66

More options...

(Nom) TYPE

Result list (right-click for options)

- 15:44:36 - meta.AdaBoostM1
- 15:44:48 - meta.AdaBoostM1
- 15:44:57 - meta.AdaBoostM1

weka.gui.GenericObjectEditor

weka.classifiers.meta.Bagging

About

Class for bagging a classifier to reduce variance. More Capabilities

bagSizePercent	100
batchSize	100
calcOutOfBag	False
classifier	Choose NaiveBayes
debug	False
doNotCheckCapabilities	False
numDecimalPlaces	2
numExecutionSlots	1
numIterations	10
outputOutOfBagComplexityStatistics	False
printClassifiers	False
representCopiesUsingWeights	False
seed	1
storeOutOfBagPredictions	False

Open... Save... OK Cancel

```
:      ROC Area  PRC Area  Class
362    0.984     0.883     WOOD
.01    0.677     0.295     SUSPEN
506    0.818     0.795     SIMPLE-T
260    0.803     0.343     ARCH
288    0.773     0.390     CANTILEV
518    0.941     0.724     CONT-T
507    0.834     0.651
```

FIGURE 3.2: Bagging

3.2.1 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          74               70.4762 %
Incorrectly Classified Instances        31               29.5238 %
Kappa statistic                          0.608
Mean absolute error                      0.1315
Root mean squared error                  0.2759
Relative absolute error                 51.8958 %
Root relative squared error             77.7009 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.990     0.929     WOOD
                 0.273    0.032    0.500      0.273    0.353      0.318    0.683     0.390     SUSPEN
                 0.841    0.115    0.841      0.841    0.841      0.726    0.918     0.896     SIMPLE-T
                 0.462    0.087    0.429      0.462    0.444      0.363    0.867     0.511     ARCH
                 0.364    0.043    0.500      0.364    0.421      0.371    0.803     0.343     CANTILEV
                 0.800    0.053    0.615      0.800    0.696      0.666    0.936     0.599     CONT-T
Weighted Avg.    0.705    0.079    0.691      0.705    0.690      0.618    0.888     0.714

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  2  2  1  0 |  b = SUSPEN
  1  1 37  3  1  1 |  c = SIMPLE-T
  0  2  3  6  1  1 |  d = ARCH
  0  0  2  2  4  3 |  e = CANTILEV
  0  0  0  1  1  8 |  f = CONT-T
```

3.2.2 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          22               62.8571 %
Incorrectly Classified Instances        13               37.1429 %
Kappa statistic                          0.5295
Mean absolute error                      0.1316
Root mean squared error                  0.2753
Relative absolute error                 50.9168 %
```

```
Root relative squared error             75.1768 %
Total Number of Instances               35     
Ignored Class Unknown Instances                  1     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.036    0.875      1.000    0.933      0.919    0.990     0.962     WOOD
                 0.000    0.033    0.000      0.000    0.000      -0.070   0.600     0.217     SUSPEN
                 0.917    0.087    0.846      0.917    0.880      0.815    0.958     0.938     SIMPLE-T
                 0.000    0.033    0.000      0.000    0.000      -0.070   0.942     0.612     ARCH
                 0.667    0.125    0.333      0.667    0.444      0.402    0.899     0.567     CANTILEV
                 0.667    0.125    0.333      0.667    0.444      0.402    0.980     0.806     CONT-T
Weighted Avg.    0.629    0.068    0.522      0.629    0.565      0.512    0.908     0.750     

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  1  0  2  1  1  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  1  0  0  2  2 |  d = ARCH
  0  0  0  0  2  1 |  e = CANTILEV
  0  0  0  0  1  2 |  f = CONT-T
```

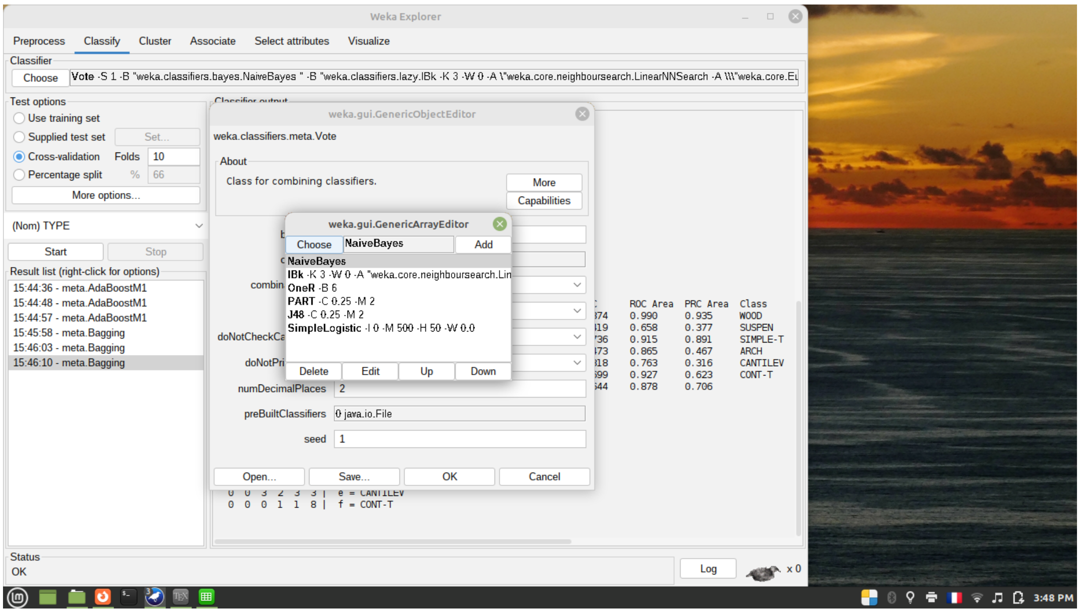
3.2.3 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          77               73.3333 %
Incorrectly Classified Instances        28               26.6667 %
Kappa statistic                          0.6373
Mean absolute error                      0.1336
Root mean squared error                  0.2776
Relative absolute error                 52.323  %
Root relative squared error             77.5381 %
Total Number of Instances              105     
Ignored Class Unknown Instances                  2     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.990     0.935     WOOD
                 0.273    0.011    0.750      0.273    0.400      0.419    0.658     0.377     SUSPEN
                 0.909    0.164    0.800      0.909    0.851      0.736    0.915     0.891     SIMPLE-T
                 0.538    0.065    0.538      0.538    0.538      0.473    0.865     0.467     ARCH
                 0.273    0.032    0.500      0.273    0.353      0.318    0.763     0.316     CANTILEV
                 0.800    0.042    0.667      0.800    0.727      0.699    0.927     0.623     CONT-T
Weighted Avg.    0.733    0.092    0.718      0.733    0.707      0.644    0.878     0.706     

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  3  4  1  0  0 |  b = SUSPEN
  1  0 40  2  1  0 |  c = SIMPLE-T
  0  1  3  7  1  1 |  d = ARCH
  0  0  3  2  3  3 |  e = CANTILEV
  0  0  0  1  1  8 |  f = CONT-T
```

3.3 Voting (Using Majority Vote option)

In Voting I have decided to use Majority Vote option rather than the default setting, the order of the classifiers used in the voting process didn't change the results in any way an the classifiers used in voting were as follow:

1. Naive Bayes (default settings)
2. KNN (k=3)
3. One Rule (default settings)
4. PART (default settings)
5. J48 (default settings)
6. Simple Logistic (default settings)



Classifier

```
Vote -S 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.lazy.IBk -K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.Eu
```

Test options

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

(Nom) TYPE

Result list (right-click for options)

```
15:44:36 - meta.AdaBoostM1
15:44:48 - meta.AdaBoostM1
15:44:57 - meta.AdaBoostM1
15:45:58 - meta.Bagging
15:46:03 - meta.Bagging
15:46:10 - meta.Bagging
```

Status
OK

weka.gui.GenericObjectEditor

weka.classifiers.meta.Vote

About

Class for combining classifiers.

combin
doNotCheckCa
doNotPri
numDecimalPlaces 2
preBuiltClassifiers 0 java.io.File
seed 1

weka.gui.GenericArrayEditor

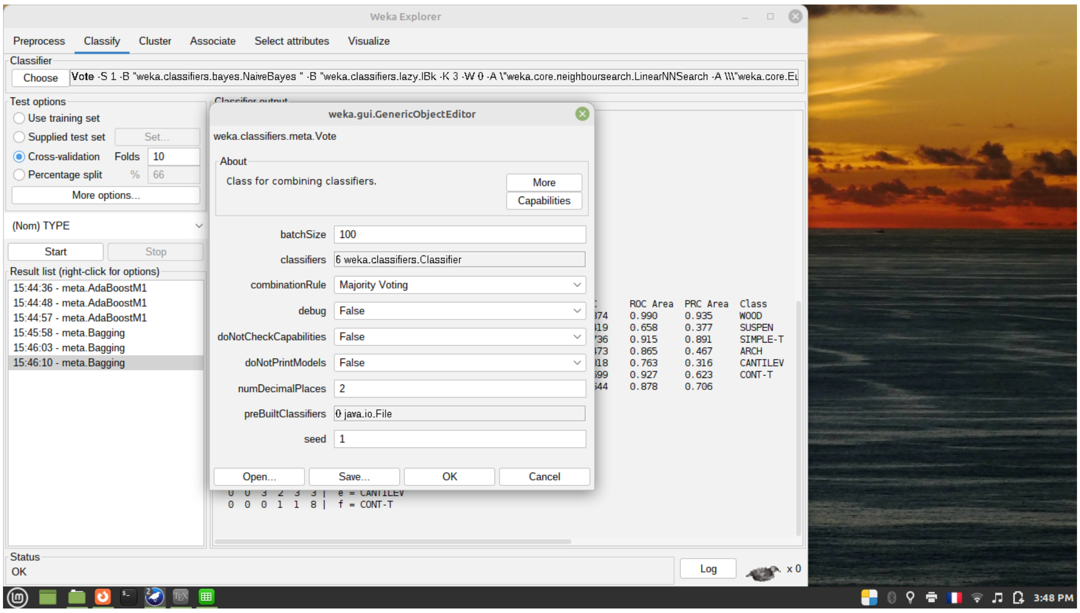
NaiveBayes

```
NaiveBayes
IBk -K 3 -W 0 -A "weka.core.neighboursearch.Lin
OneR -B 6
PART -C 0.25 -M 2
J48 -C 0.25 -M 2
SimpleLogistic -I 0 -M 500 -H 50 -W 0.0
```

Classifier output

```
 ROC Area  PRC Area  Class
374  0.990     0.935     WOOD
419  0.658     0.377     SUSPEN
736  0.915     0.891     SIMPLE-T
473  0.865     0.467     ARCH
318  0.763     0.316     CANTILEV
599  0.927     0.623     CONT-T
544  0.878     0.706

 0  0  3  2  3  3 |  e = CANTILEV
 0  0  0  1  1  8 |  f = CONT-T
```

FIGURE 3.3: Voting 1



Classifier: Vote -S 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.lazy.IBk -K 3 -W 0 -A \"weka.core.neighboursearch.LinearNNSearch -A \\\"weka.core.Eu

Test options
○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) TYPE

Result list (right-click for options)
15:44:36 - meta.AdaBoostM1
15:44:48 - meta.AdaBoostM1
15:44:57 - meta.AdaBoostM1
15:45:58 - meta.Bagging
15:46:03 - meta.Bagging
15:46:10 - meta.Bagging

weka.gui.GenericObjectEditor

weka.classifiers.meta.Vote

About
Class for combining classifiers.

batchSize	100
classifiers	6 weka.classifiers.Classifier
combinationRule	Majority Voting
debug	False
doNotCheckCapabilities	False
doNotPrintModels	False
numDecimalPlaces	2
preBuiltClassifiers	0 java.io.File
seed	1

Classifier output

```
    ROC Area  PRC Area  Class
374  0.990    0.935     WOOD
419  0.658    0.377     SUSPEN
736  0.915    0.891     SIMPLE-T
473  0.865    0.467     ARCH
318  0.763    0.316     CANTILEV
599  0.927    0.623     CONT-T
544  0.878    0.706

 0  0  3  2  3  3 |  e = CANTILEV
 0  0  0  1  1  8 |  f = CONT-T
```

Status
OK

FIGURE 3.4: Voting 2

3.3.1 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          72               68.5714 %
Incorrectly Classified Instances        33               31.4286 %
Kappa statistic                          0.5676
Mean absolute error                      0.1048
Root mean squared error                  0.3237
Relative absolute error                 41.3393 %
Root relative squared error             91.1471 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.978     0.800     WOOD
                 0.182    0.011    0.667      0.182    0.286      0.315    0.586     0.205     SUSPEN
                 0.886    0.230    0.736      0.886    0.804      0.648    0.816     0.675     SIMPLE-T
                 0.538    0.033    0.700      0.538    0.609      0.568    0.753     0.433     ARCH
                 0.091    0.053    0.167      0.091    0.118      0.050    0.519     0.109     CANTILEV
                 0.700    0.063    0.538      0.700    0.609      0.568    0.819     0.405     CONT-T
Weighted Avg.    0.686    0.120    0.656      0.686    0.648      0.567    0.778     0.530

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  2  5  0  1  0 |  b = SUSPEN
  1  0 39  2  1  1 |  c = SIMPLE-T
  0  0  3  7  2  1 |  d = ARCH
  0  1  4  1  1  4 |  e = CANTILEV
  0  0  2  0  1  7 |  f = CONT-T
```

3.3.2 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          22               62.8571 %
Incorrectly Classified Instances        13               37.1429 %
Kappa statistic                          0.4899
Mean absolute error                      0.1238
Root mean squared error                  0.3519
Relative absolute error                 47.8914 %
Root relative squared error             96.0871 %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.139     SUSPEN
                 0.917    0.478    0.500      0.917    0.647      0.431    0.708     0.466     SIMPLE-T
                 0.000    0.033    0.000      0.000    0.000      -0.070   0.484     0.139     ARCH
                 0.333    0.000    1.000      0.333    0.500      0.560    0.667     0.389     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    0.985     0.750     CONT-T
Weighted Avg.    0.629    0.171    ?          0.629    ?          ?        0.725     0.497

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  0  0  4  1  0  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  5  0  0  0 |  d = ARCH
  0  0  2  0  1  0 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

3.3.3 Leave One Out Fold)

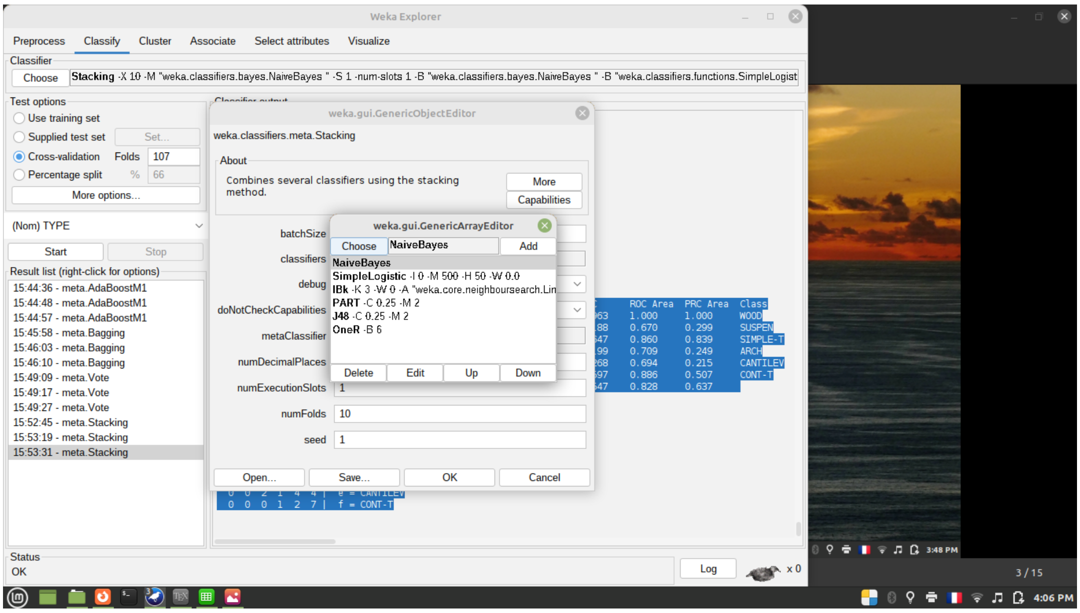
```
=== Summary ===

Correctly Classified Instances          69               65.7143 %
Incorrectly Classified Instances        36               34.2857 %
Kappa statistic                          0.5261
Mean absolute error                      0.1143
Root mean squared error                  0.3381
Relative absolute error                 44.7609 %
Root relative squared error             94.4389 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.045    0.800      1.000    0.889      0.874    0.978     0.800     WOOD
                 0.091    0.000    1.000      0.091    0.167      0.287    0.545     0.184     SUSPEN
                 0.864    0.262    0.704      0.864    0.776      0.594    0.789     0.642     SIMPLE-T
                 0.538    0.054    0.583      0.538    0.560      0.501    0.743     0.370     ARCH
                 0.091    0.053    0.167      0.091    0.118      0.050    0.519     0.109     CANTILEV
                 0.600    0.063    0.500      0.600    0.545      0.495    0.769     0.337     CONT-T
Weighted Avg.    0.657    0.135    0.659      0.657    0.611      0.526    0.756     0.500

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 16  0  0  0  0  0 |  a = WOOD
  3  1  6  1  0  0 |  b = SUSPEN
  1  0 38  3  1  1 |  c = SIMPLE-T
  0  0  3  7  2  1 |  d = ARCH
  0  0  5  1  1  4 |  e = CANTILEV
  0  0  2  0  2  6 |  f = CONT-T
```

3.4 Stacking 1

In stacking we will be using Naive Bayes as the Meta Classifier and the choice of stacked classifiers is as follows :

1. Naive Bayes (default settings)
2. Simple Logistic (default settings)
3. KNN (k=3)
4. PART (default settings)
5. J48 (default settings)
6. One Rule (default settings)

the order doesn't affect the results of the evaluation.



Classifier: Stacking -X 10 -M "weka.classifiers.bayes.NaiveBayes " -S 1 -num-slots 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.functions.SimpleLogist

Test options: Cross-validation, Folds 107

weka.gui.GenericObjectEditor

weka.classifiers.meta.Stacking

About: Combines several classifiers using the stacking method.

weka.gui.GenericArrayEditor

```
NaiveBayes
SimpleLogistic -I 0 -M 500 -H 50 -W 0.0
IBk -K 3 -W 0 -A "weka.core.neighboursearch.Lin
PART -C 0.25 -M 2
J48 -C 0.25 -M 2
OneR -B 6
```

numExecutionSlots: 1

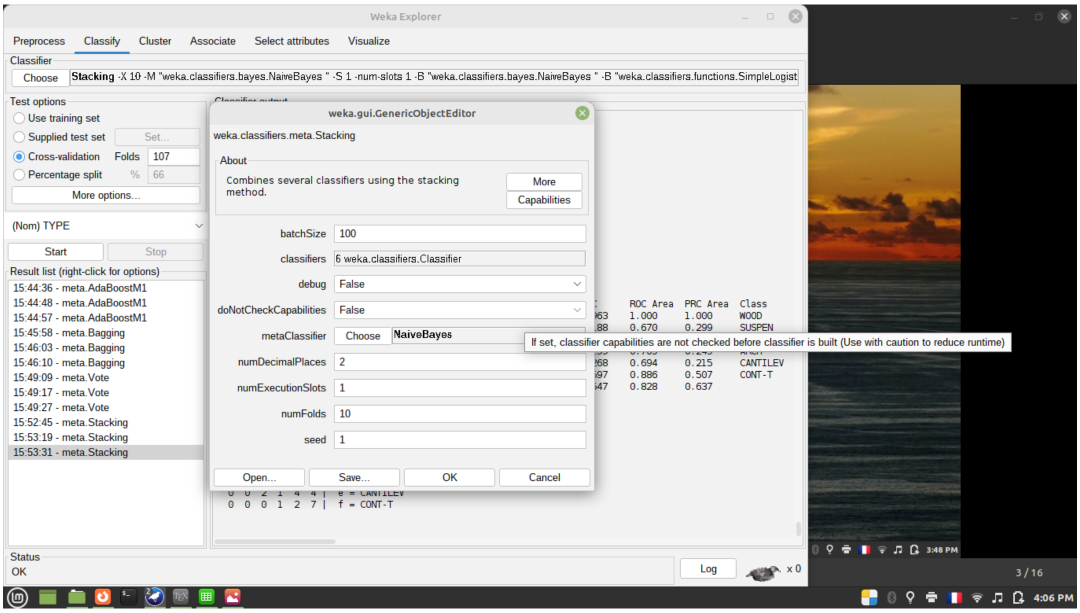
numFolds: 10

seed: 1

Result list (right-click for options)

```
15:44:36 - meta.AdaBoostM1
15:44:48 - meta.AdaBoostM1
15:44:57 - meta.AdaBoostM1
15:45:58 - meta.Bagging
15:46:03 - meta.Bagging
15:46:10 - meta.Bagging
15:49:09 - meta.Vote
15:49:17 - meta.Vote
15:49:27 - meta.Vote
15:52:45 - meta.Stacking
15:53:19 - meta.Stacking
15:53:31 - meta.Stacking
```

FIGURE 3.5: Stacking 1



Classifier: Stacking -X 10 -M "weka.classifiers.bayes.NaiveBayes " -S 1 -num-slots 1 -B "weka.classifiers.bayes.NaiveBayes " -B "weka.classifiers.functions.SimpleLogist

weka.gui.GenericObjectEditor

weka.classifiers.meta.Stacking

About: Combines several classifiers using the stacking method.

batchSize: 100

classifiers: 6 weka.classifiers.Classifier

debug: False

doNotCheckCapabilities: False

metaClassifier: NaiveBayes

If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime)

numDecimalPlaces: 2

numExecutionSlots: 1

numFolds: 10

seed: 1

FIGURE 3.6: Stacking 1.2

3.4.1 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          63               60      %
Incorrectly Classified Instances        42               40      %
Kappa statistic                          0.4763
Mean absolute error                      0.1364
Root mean squared error                  0.3621
Relative absolute error                 53.8379 %
Root relative squared error            101.9726 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.875    0.000    1.000      0.875    0.933      0.925    1.000     1.000     WOOD
                 0.182    0.064    0.250      0.182    0.211      0.136    0.656     0.258     SUSPEN
                 0.750    0.131    0.805      0.750    0.776      0.626    0.865     0.841     SIMPLE-T
                 0.538    0.120    0.389      0.538    0.452      0.366    0.782     0.438     ARCH
                 0.182    0.128    0.143      0.182    0.160      0.049    0.600     0.145     CANTILEV
                 0.500    0.053    0.500      0.500    0.500      0.447    0.792     0.415     CONT-T
Weighted Avg.    0.600    0.095    0.627      0.600    0.610      0.511    0.819     0.641

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 14  2  0  0  0  0 |  a = WOOD
  0  2  3  4  2  0 |  b = SUSPEN
  0  2 33  5  4  0 |  c = SIMPLE-T
  0  2  2  7  2  0 |  d = ARCH
  0  0  3  1  2  5 |  e = CANTILEV
  0  0  0  1  4  5 |  f = CONT-T
```

3.4.2 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          25               71.4286 %
Incorrectly Classified Instances        10               28.5714 %
Kappa statistic                          0.6461
Mean absolute error                      0.11
Root mean squared error                  0.2985
Relative absolute error                 42.546  %
Root relative squared error             81.513  %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.200    0.000    1.000      0.200    0.333      0.420    0.658     0.363     SUSPEN
                 0.750    0.000    1.000      0.750    0.857      0.815    0.910     0.904     SIMPLE-T
                 0.800    0.200    0.400      0.800    0.533      0.465    0.871     0.471     ARCH
                 0.333    0.094    0.250      0.333    0.286      0.211    0.566     0.142     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    0.985     0.750     CONT-T
Weighted Avg.    0.714    0.039    0.829      0.714    0.716      0.697    0.863     0.705

=== Confusion Matrix ===

 a b c d e f   <-- classified as
 7 0 0 0 0 0 | a = WOOD
 0 1 0 3 1 0 | b = SUSPEN
 0 0 9 1 1 1 | c = SIMPLE-T
 0 0 0 4 1 0 | d = ARCH
 0 0 0 2 1 0 | e = CANTILEV
 0 0 0 0 0 3 | f = CONT-T
```

3.4.3 Leave One Out Fold)

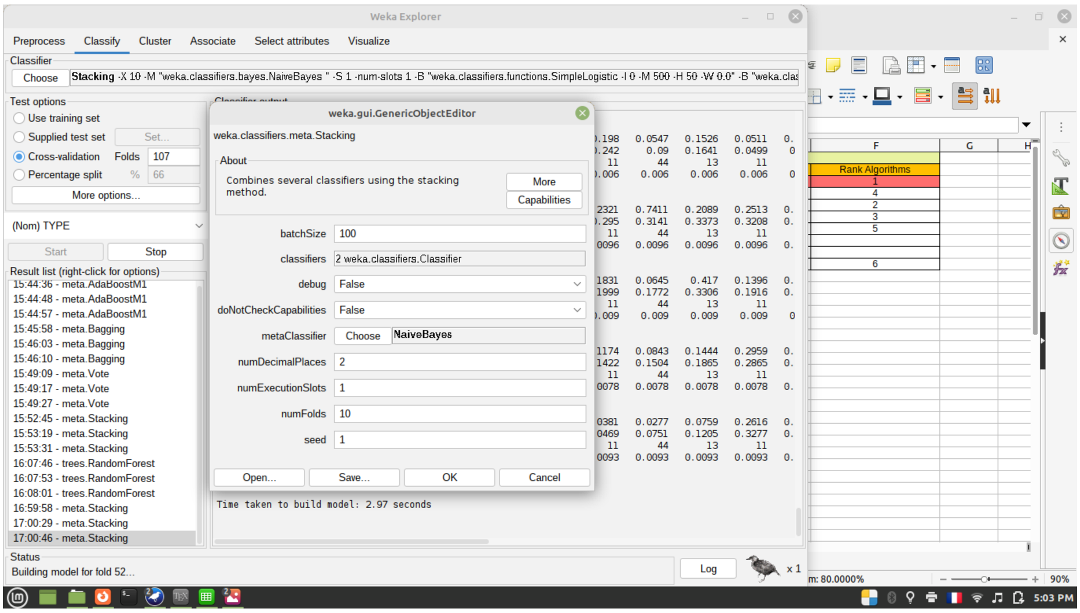
```
=== Summary ===

Correctly Classified Instances          65               61.9048 %
Incorrectly Classified Instances        40               38.0952 %
Kappa statistic                          0.5105
Mean absolute error                      0.1276
Root mean squared error                  0.3432
Relative absolute error                 49.9794 %
Root relative squared error             95.8625 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.938    0.000    1.000      0.938    0.968      0.963    1.000     1.000     WOOD
                 0.273    0.085    0.273      0.273    0.273      0.188    0.670     0.299     SUSPEN
                 0.705    0.082    0.861      0.705    0.775      0.647    0.860     0.839     SIMPLE-T
                 0.385    0.152    0.263      0.385    0.313      0.199    0.709     0.249     ARCH
                 0.364    0.085    0.333      0.364    0.348      0.268    0.694     0.215     CANTILEV
                 0.700    0.053    0.583      0.700    0.636      0.597    0.886     0.507     CONT-T
Weighted Avg.    0.619    0.076    0.665      0.619    0.637      0.547    0.828     0.637

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 15  1  0  0  0  0 |  a = WOOD
  0  3  2  6  0  0 |  b = SUSPEN
  0  4 31  6  2  1 |  c = SIMPLE-T
  0  3  1  5  4  0 |  d = ARCH
  0  0  2  1  4  4 |  e = CANTILEV
  0  0  0  1  2  7 |  f = CONT-T
```

3.5 Stacking 2

In stacking we will be using Naive Bayes as the Meta Classifier and the choice of stacked classifiers is as follows :

1. Simple Logistic (default settings)
2. Naive Bayes (default settings)

The order doesn't affect the results of the evaluation. what is weird from the results is that sometimes less is better as we can see that using only 2 classifiers resulted in better performance compared to the previous iteration where we have used 6 classifiers.



Classifier: Stacking -X 10 -M "weka.classifiers.bayes.NaiveBayes " -S 1 -num-slots 1 -B "weka.classifiers.functions.SimpleLogistic -I 0 -M 500 -H 50 -W 0.0" -B "weka.clas

Test options
○ Use training set
○ Supplied test set Set...
● Cross-validation Folds 107
○ Percentage split % 66
More options...

(Nom) TYPE

Start Stop

Result list (right-click for options)
15:44:36 - meta.AdaBoostM1
15:44:48 - meta.AdaBoostM1
15:44:57 - meta.AdaBoostM1
15:45:58 - meta.Bagging
15:46:03 - meta.Bagging
15:46:10 - meta.Bagging
15:49:09 - meta.Vote
15:49:17 - meta.Vote
15:49:27 - meta.Vote
15:52:45 - meta.Stacking
15:53:19 - meta.Stacking
15:53:31 - meta.Stacking
16:07:46 - trees.RandomForest
16:07:53 - trees.RandomForest
16:08:01 - trees.RandomForest
16:59:58 - meta.Stacking
17:00:29 - meta.Stacking
17:00:46 - meta.Stacking

weka.gui.GenericObjectEditor
weka.classifiers.meta.Stacking
About
Combines several classifiers using the stacking method.
More
Capabilities

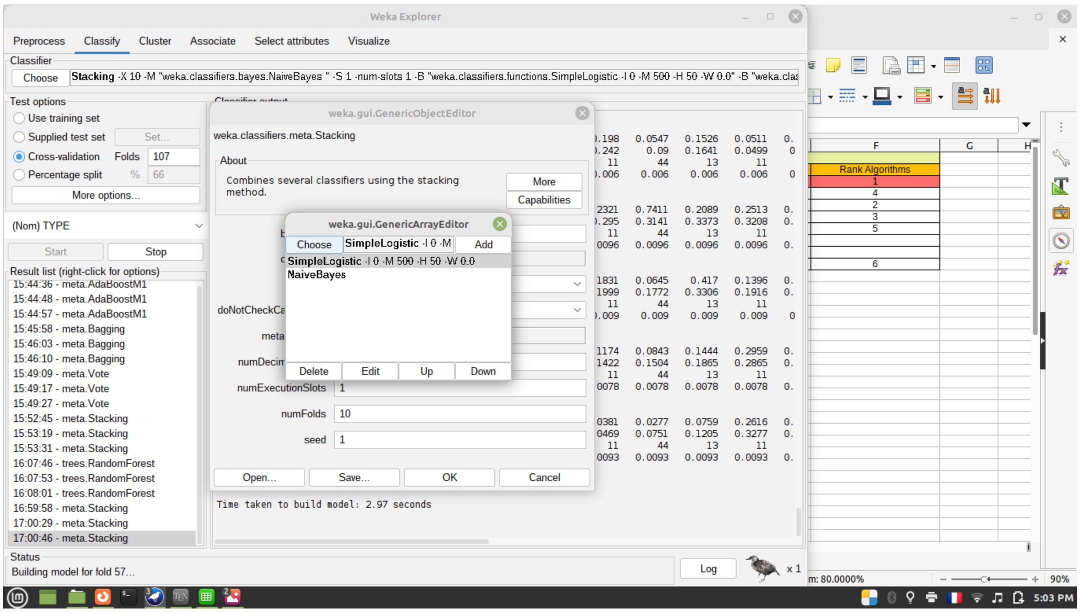
batchSize 100
classifiers 2 weka.classifiers.Classifier
debug False
doNotCheckCapabilities False
metaClassifier Choose NaiveBayes
numDecimalPlaces 2
numExecutionSlots 1
numFolds 10
seed 1

Open... Save... OK Cancel

Time taken to build model: 2.97 seconds

Status
Building model for fold 52...

FIGURE 3.7: Stacking 2



Weka Explorer

Classifier: Stacking -X 10 -M "weka.classifiers.bayes.NaiveBayes " -S 1 -num-slots 1 -B "weka.classifiers.functions.SimpleLogistic -I 0 -M 500 -H 50 -W 0.0" -B "weka.clas

Test options: ● Cross-validation Folds 107; ○ Percentage split % 66

(Nom) TYPE

weka.gui.GenericObjectEditor: weka.classifiers.meta.Stacking

About: Combines several classifiers using the stacking method.

weka.gui.GenericArrayEditor: SimpleLogistic -I 0 -M 500 -H 50 -W 0.0; NaiveBayes

numExecutionSlots 1
numFolds 10
seed 1

Result list (right-click for options):
15:44:36 - meta.AdaBoostM1
15:44:48 - meta.AdaBoostM1
15:44:57 - meta.AdaBoostM1
15:45:58 - meta.Bagging
15:46:03 - meta.Bagging
15:46:10 - meta.Bagging
15:49:09 - meta.Vote
15:49:17 - meta.Vote
15:49:27 - meta.Vote
15:52:45 - meta.Stacking
15:53:19 - meta.Stacking
15:53:31 - meta.Stacking
16:07:46 - trees.RandomForest
16:07:53 - trees.RandomForest
16:08:01 - trees.RandomForest
16:59:58 - meta.Stacking
17:00:29 - meta.Stacking
17:00:46 - meta.Stacking

```
Time taken to build model: 2.97 seconds
```

Status: Building model for fold 57...

FIGURE 3.8: Stacking 2.1

3.5.1 Cross Validation (10 Folds)

```
=== Summary ===

Correctly Classified Instances          72               68.5714 %
Incorrectly Classified Instances        33               31.4286 %
Kappa statistic                          0.5855
Mean absolute error                      0.1097
Root mean squared error                  0.3187
Relative absolute error                 43.287  %
Root relative squared error             89.7341 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.875    0.000    1.000      0.875    0.933      0.925    0.999     0.996     WOOD
                 0.364    0.053    0.444      0.364    0.400      0.340    0.746     0.345     SUSPEN
                 0.795    0.131    0.814      0.795    0.805      0.667    0.878     0.839     SIMPLE-T
                 0.615    0.098    0.471      0.615    0.533      0.463    0.794     0.361     ARCH
                 0.364    0.064    0.400      0.364    0.381      0.313    0.589     0.217     CANTILEV
                 0.700    0.053    0.583      0.700    0.636      0.597    0.844     0.502     CONT-T
Weighted Avg.    0.686    0.084    0.696      0.686    0.688      0.603    0.839     0.655

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 14  2  0  0  0  0 |  a = WOOD
  0  4  3  4  0  0 |  b = SUSPEN
  0  2 35  3  3  1 |  c = SIMPLE-T
  0  1  3  8  1  0 |  d = ARCH
  0  0  2  1  4  4 |  e = CANTILEV
  0  0  0  1  2  7 |  f = CONT-T
```

3.5.2 Percentage Split (66%)

```
=== Summary ===

Correctly Classified Instances          28               80      %
Incorrectly Classified Instances         7               20      %
Kappa statistic                          0.7416
Mean absolute error                      0.085
Root mean squared error                  0.2592
Relative absolute error                 32.8804 %
Root relative squared error             70.78   %
Total Number of Instances               35
Ignored Class Unknown Instances                  1

=== 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     WOOD
                 0.200    0.000    1.000      0.200    0.333      0.420    0.652     0.361     SUSPEN
                 0.917    0.130    0.786      0.917    0.846      0.762    0.896     0.907     SIMPLE-T
                 1.000    0.100    0.625      1.000    0.769      0.750    0.948     0.696     ARCH
                 0.333    0.000    1.000      0.333    0.500      0.560    0.515     0.396     CANTILEV
                 1.000    0.031    0.750      1.000    0.857      0.852    0.980     0.806     CONT-T
Weighted Avg.    0.800    0.062    0.852      0.800    0.764      0.749    0.864     0.765

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
  7  0  0  0  0  0 |  a = WOOD
  0  1  3  1  0  0 |  b = SUSPEN
  0  0 11  0  0  1 |  c = SIMPLE-T
  0  0  0  5  0  0 |  d = ARCH
  0  0  0  2  1  0 |  e = CANTILEV
  0  0  0  0  0  3 |  f = CONT-T
```

3.5.3 Leave One Out Fold)

```
=== Summary ===

Correctly Classified Instances          71               67.619  %
Incorrectly Classified Instances        34               32.381  %
Kappa statistic                          0.5789
Mean absolute error                      0.1106
Root mean squared error                  0.3159
Relative absolute error                 43.3071 %
Root relative squared error             88.2497 %
Total Number of Instances              105
Ignored Class Unknown Instances                  2

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.938    0.000    1.000      0.938    0.968      0.963    1.000     1.000     WOOD
                 0.273    0.043    0.429      0.273    0.333      0.283    0.635     0.251     SUSPEN
                 0.727    0.115    0.821      0.727    0.771      0.625    0.837     0.758     SIMPLE-T
                 0.615    0.141    0.381      0.615    0.471      0.390    0.746     0.313     ARCH
                 0.455    0.074    0.417      0.455    0.435      0.366    0.656     0.317     CANTILEV
                 0.800    0.032    0.727      0.800    0.762      0.737    0.896     0.516     CONT-T
Weighted Avg.    0.676    0.081    0.701      0.676    0.682      0.595    0.816     0.617

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 15  1  0  0  0  0 |  a = WOOD
  0  3  3  5  0  0 |  b = SUSPEN
  0  3 32  6  3  0 |  c = SIMPLE-T
  0  0  2  8  3  0 |  d = ARCH
  0  0  2  1  5  3 |  e = CANTILEV
  0  0  0  1  1  8 |  f = CONT-T
```

3.6 Conclusion

Correctly Classified Instances by Algorithm					
Evaluation Process	Cross Validation 10 Folds	Percentage Split 66%	Leave One Out Fold	AVG Algorithms	Rank
Naïve Bayes	70.4762%	77.1429%	69.5238%	72.3810%	2
Boosting	63.8095%	71.4286%	60.0000%	65.0794%	5
Bagging	70.4762%	62.8571%	73.3333%	68.8889%	3
Voting	68.5714%	62.8571%	65.7143%	65.7143%	4
Stacking 1	60.0000%	71.4286%	61.9048%	64.4445%	6
Stacking 2	68.5714%	80.0000%	67.6190%	72.0635%	1
RandomForest	51.4286%	42.8571%	52.3810%	48.8889%	7

It was very hard to improve upon Naive Bayes with any of the ensemble methods. as if we take the average of the tests Naive Bayes still performs better with an average of 72.3810% Correctly Classified Instances followed by Stacking 2 that has (Linear Regression and Naive Bayes as builder classifiers) with 72.0635% Correctly Classified Instances. However the best result was by Stacking 2 in the Percentage Split 66% evaluation method with a score of 80.0000% Correctly Classified Instances.

It is also known that Naive Bayes in general doesn't benefit from ensemble methods because of its low variance so combining it with other methods doesn't help.

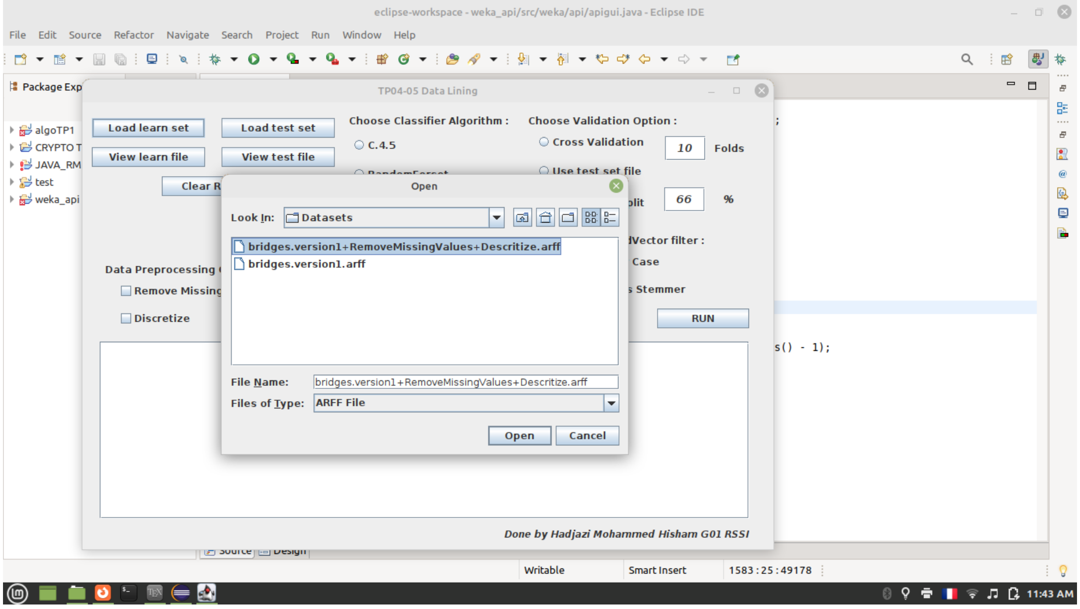
Chapter 4

Coding part

Again not enough time to do what i wanted to do, it must be noted that the code was taken directly from the video provided in the FicheTP4 part 2.

4.1 Naive Bayes result

Here we are executing with Naive Bayes to take a reference of the performance.



○ Cross Validation 10 Folds

66 %

Look In: Datasets

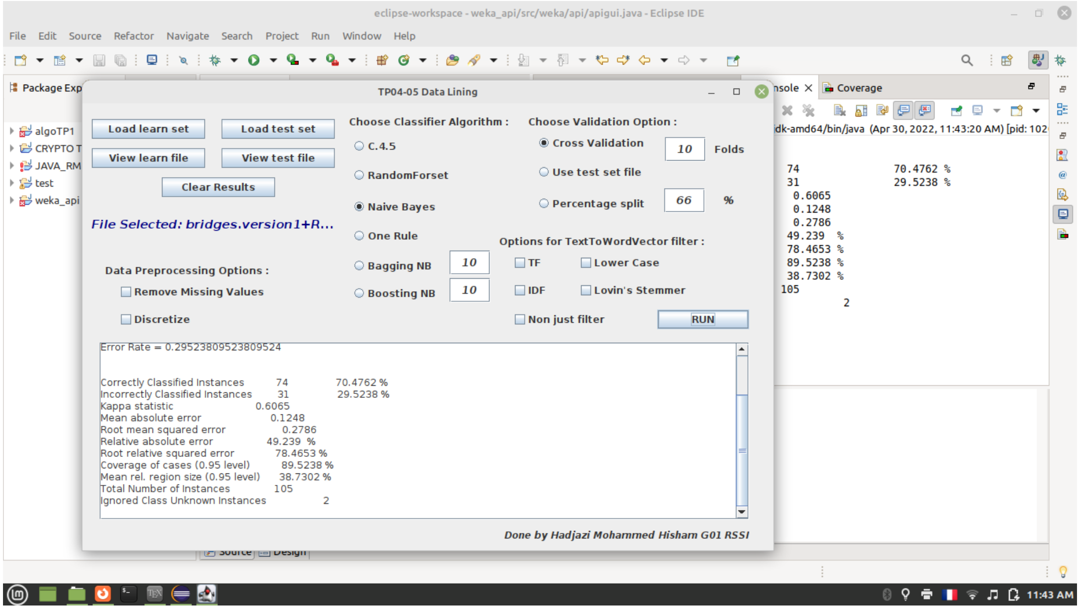
bridges.version1+RemoveMissingValues+Descritize.arff
bridges.version1.arff

File Name: bridges.version1+RemoveMissingValues+Descritize.arff

Files of Type: ARFF File

1583 : 25 : 49178

FIGURE 4.1: Loading ARFF file



File Selected: bridges.version1+R...

Choose Classifier Algorithm :
○ C.4.5
○ RandomForset
● Naive Bayes
○ One Rule
○ Bagging NB 10
○ Boosting NB 10

Choose Validation Option :
● Cross Validation 10 Folds
○ Use test set file
○ Percentage split 66 %

Data Preprocessing Options :
□ Remove Missing Values
□ Discretize

Options for TextToWordVector filter :
□ TF □ Lower Case
□ IDF □ Lovin's Stemmer
□ Non just filter

```
Error Rate = 0.29523809523809524


Correctly Classified Instances        74          70.4762 %
Incorrectly Classified Instances      31          29.5238 %
Kappa statistic                    0.6065
Mean absolute error                  0.1248
Root mean squared error                0.2786
Relative absolute error             49.239  %
Root relative squared error           78.4653 %
Coverage of cases (0.95 level)          89.5238 %
Mean rel. region size (0.95 level)     38.7302 %
Total Number of Instances           105
Ignored Class Unknown Instances              2
```

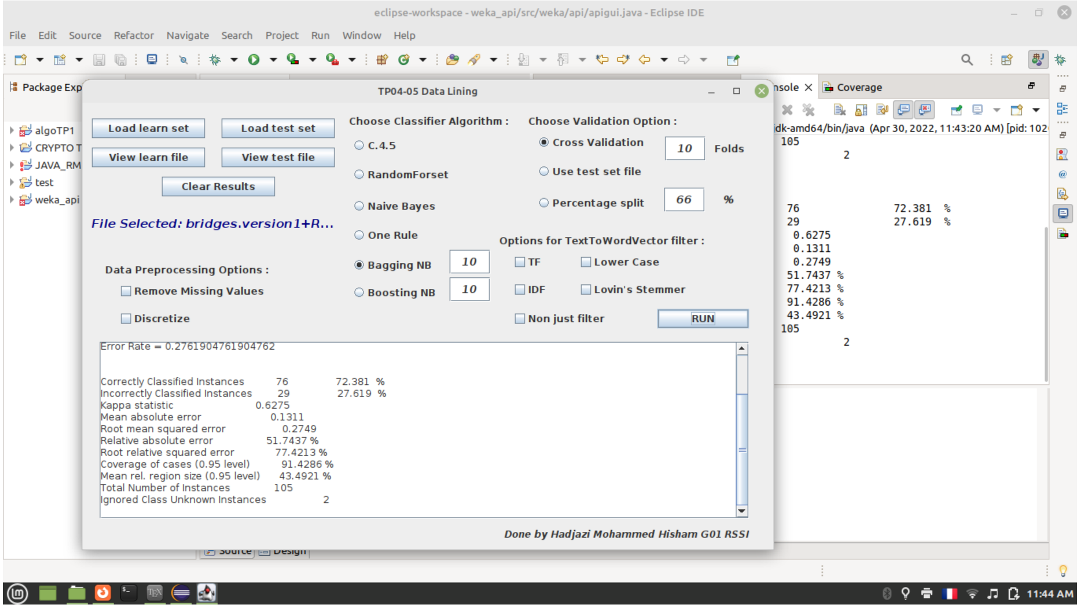
Done by Hadjazi Mohammed Hisham G01 RSSI

```
dk-amd64/bin/java (Apr 30, 2022, 11:43:20 AM) [pid: 102
74                  70.4762 %
31                  29.5238 %
 0.6065
 0.1248
 0.2786
49.239  %
78.4653 %
89.5238 %
38.7302 %
105
          2
```

FIGURE 4.2: Naive Bayes

4.2 Bagging

```
Bagging bagger = new Bagging();
bagger.setClassifier(new NaiveBayes());
bagger.setNumIterations(Integer.parseInt(input3.getText()));
bagger.buildClassifier(datasetInstances);
```



File Selected: bridges.version1+R...

Choose Classifier Algorithm :
○ C.4.5
○ RandomForset
○ Naive Bayes
○ One Rule
● Bagging NB 10
○ Boosting NB 10

Choose Validation Option :
● Cross Validation 10 Folds
○ Use test set file
○ Percentage split 66 %

Data Preprocessing Options :
□ Remove Missing Values
□ Discretize

Options for TextToWordVector filter :
□ TF □ Lower Case
□ IDF □ Lovin's Stemmer
□ Non just filter

```
Error Rate = 0.2761904761904762


Correctly Classified Instances        76          72.381  %
Incorrectly Classified Instances      29          27.619  %
Kappa statistic                    0.6275
Mean absolute error                  0.1311
Root mean squared error                0.2749
Relative absolute error             51.7437 %
Root relative squared error           77.4213 %
Coverage of cases (0.95 level)          91.4286 %
Mean rel. region size (0.95 level)     43.4921 %
Total Number of Instances           105
Ignored Class Unknown Instances              2
```

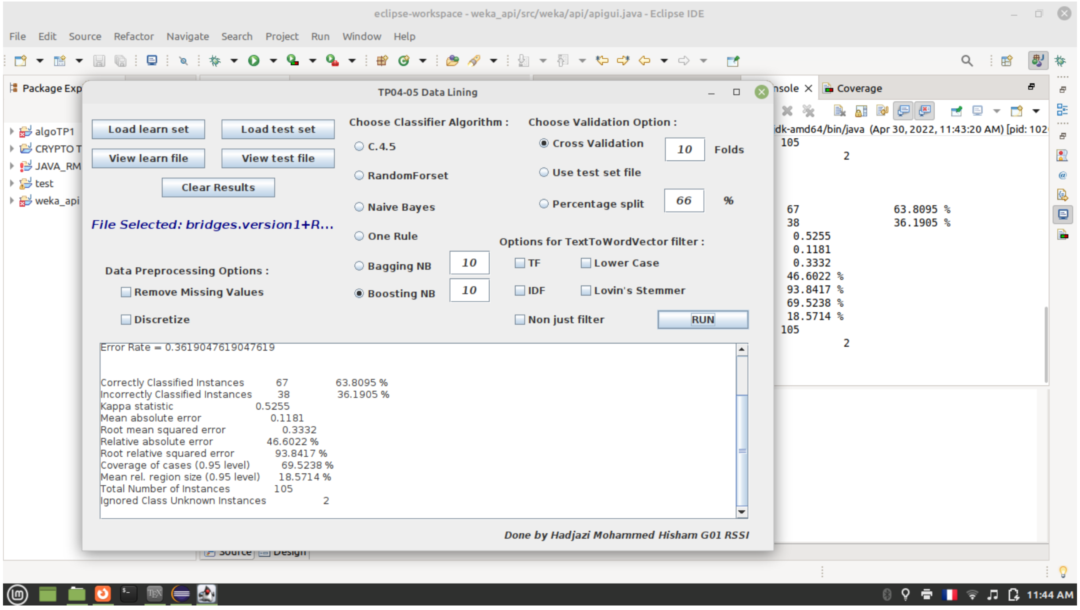
Done by Hadjazi Mohammed Hisham G01 RSSI

```
dk-amd64/bin/java (Apr 30, 2022, 11:43:20 AM) [pid: 102
105
          2

76                  72.381  %
29                  27.619  %
 0.6275
 0.1311
 0.2749
51.7437 %
77.4213 %
91.4286 %
43.4921 %
105
          2
```

FIGURE 4.3: Bagging

4.3 Boosting

```
            AdaBoostM1 booster = new AdaBoostM1();
            booster.setClassifier(new NaiveBayes());
            booster.setNumIterations(Integer.parseInt(input4.
                getText()));
            booster.buildClassifier(datasetInstances);
```



File Selected: bridges.version1+R...

Data Preprocessing Options :

□ Remove Missing Values

□ Discretize

Choose Classifier Algorithm :

○ C.4.5

○ RandomForset

○ Naive Bayes

○ One Rule

○ Bagging NB 10

● Boosting NB 10

Choose Validation Option :

● Cross Validation 10 Folds

○ Use test set file

○ Percentage split 66 %

Options for TextToWordVector filter :

□ TF □ Lower Case

□ IDF □ Lovin's Stemmer

□ Non just filter

```
Error Rate = 0.3619047619047619


Correctly Classified Instances        67              63.8095 %
Incorrectly Classified Instances      38              36.1905 %
Kappa statistic                    0.5255
Mean absolute error                  0.1181
Root mean squared error                0.3332
Relative absolute error            46.6022 %
Root relative squared error          93.8417 %
Coverage of cases (0.95 level)         69.5238 %
Mean rel. region size (0.95 level)    18.5714 %
Total Number of Instances            105
Ignored Class Unknown Instances                2
```

Done by Hadjazi Mohammed Hisham G01 RSSI

```
idk-amd64/bin/java  (Apr 30, 2022, 11:43:20 AM) [pid: 102
105
           2



 67                  63.8095 %
 38                  36.1905 %
  0.5255
  0.1181
  0.3332
 46.6022 %
 93.8417 %
 69.5238 %
 18.5714 %
105
           2
```

FIGURE 4.4: Boosting