



INTRODUCTION TO PATTERN RECOGNITION

FINAL PROJECT REPORT

TOPIC
FORENSICS INVESTIGATION FROM FINGERPRINT
MICROBES

Yusuf SEYİTOĞLU
1030516739

AMAÇ

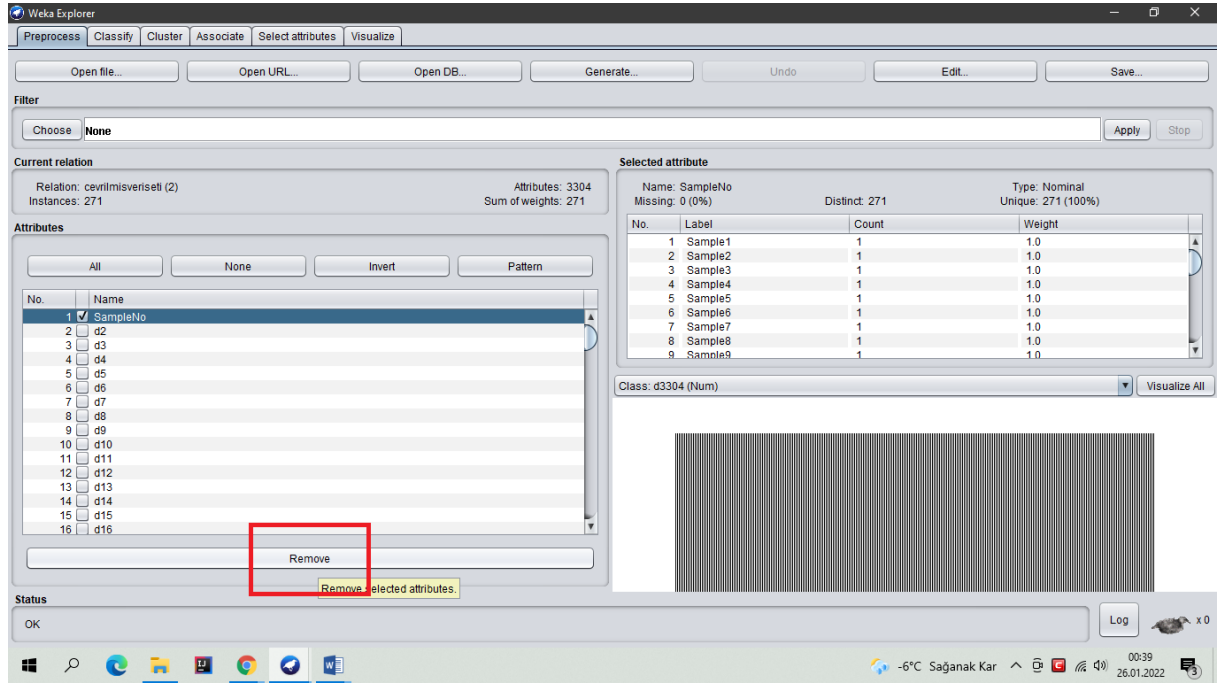
Projedeki amaç bilgisayar klavyesinden toplanan klinik verilere göre sınıflandırma algoritmalarını kullanarak klavyeye dokunulan elin hangi el olduğunu en yüksek oranda tespit etmektir. Kullanılan veri setinde 271 örnek ve her bir örnek için 3302 özellik bulunmaktadır. Örneklerin 136 tanesi sağ, 135 tanesi sol elden alınan örneklerdir

YÖNTEM VE METHODLAR

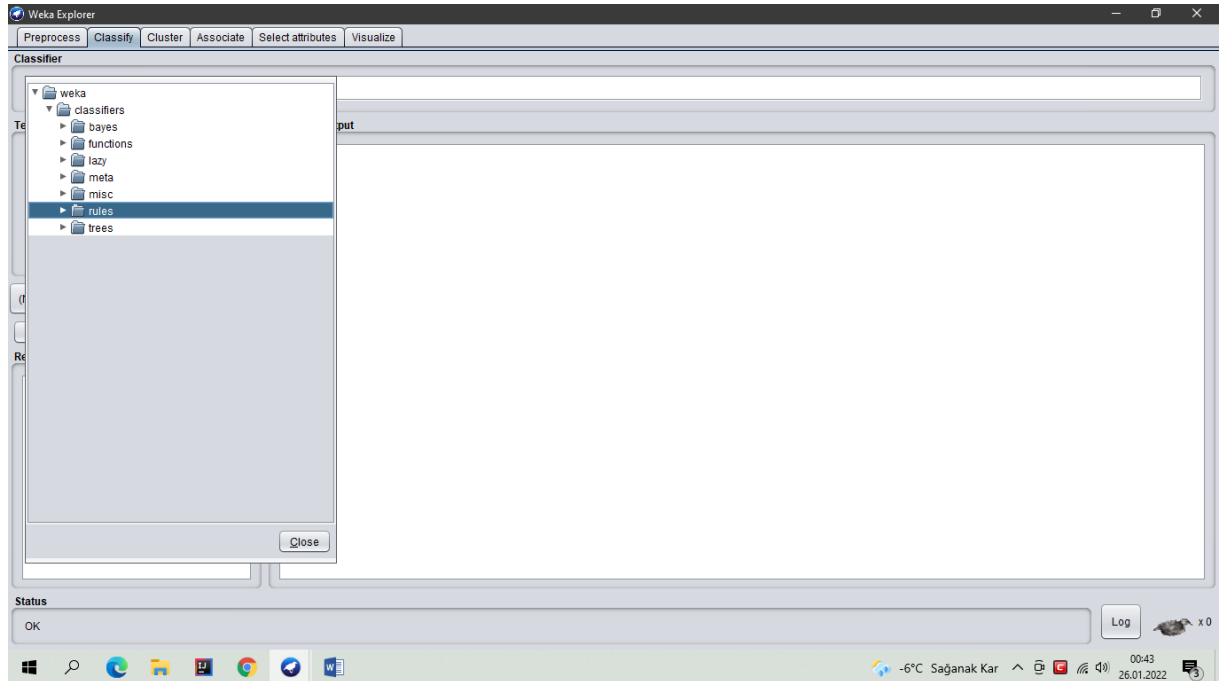
Projede veri analizi için WEKA yazılımı kullanılmıştır. Birkaç farklı algoritma ile veri setinin başarı oranları test edilmiştir. Algoritmaların başarı oranlarına etki eden parametrelerde değişiklikler yapılarak, bu değişikliklerin başarı oranındaki etkisi gözlemlenmiştir. Bulunan sonuçları karşılaştırarak buna uygun algoritma seçilmiştir

- ***STOCHASTIC GRADIENT DESCENT(SGD)***
- ***LOCALLY WEIGHTED LEARNING(LWL).***
- ***K* ALGORITHM (KStar)***
- ***NAIVE BAYES***
- ***RANDOM FOREST***
- ***SUPPORT VECTOR MACHINES(SMO)***
- ***VOTED PERCEPTRON***
- ***RANDOM TREE***
- ***DECİSİON TREE (J48)***
- ***K-NEAREST NEİGHBORS (1Bk)***
- ***LOGİCTIC REGRESSION (Logictic)***

Yapılacak olan başarı oranı karşılaştırmaları, yukarıda belirtilen algoritmalar ile yapılacaktır.



İlk önce indirdiğimiz veri setinin transpozmesini alıp yeni bir dosya oluşturuyoruz. Bu dosyayı “Weka” yazılımına yükledikten sonra “Sample” verisini tutan sütunu remove ediyoruz.



Ardından Classify butonuna tıklayıp buradan “Weka”nın içerisinde hazır bulunan algoritmalarından gereken algoritmaları seçiyoruz.

ALGORİTMALARIN DOĞRULUK KARŞILAŞTIRILMASI

DECİSİON TREE (J48)

Başarı Oranı: %58.3026

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
- ☒ Cross-validation Folds **10**
- ☐ Percentage split % **66**

More options...

(Nom) d2

Start Stop

Result list (right-click for options)

- 00:50:26 - trees.J48
- 00:50:50 - trees.J48

Classifier output

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

Correctly Classified Instances      158           58.3026 %
Kappa statistic                    0.166
K&B Relative Info Score           15.6069 %
K&B Information Score              42.3028 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          73217.1172 bits
Complexity improvement (Sf)        -72946.0648 bits
Mean absolute error                 0.423
Root mean squared error             0.6277
Relative absolute error             84.5887 %
Root relative squared error         125.5205 %
Total Number of Instances          271

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0,588    0,422    0,584     0,588    0,586     0,166    0,592    0,558    left
          0,578    0,412    0,582     0,578    0,580     0,166    0,592    0,577    right
Weighted Avg.    0,583    0,417    0,583     0,583    0,583     0,166    0,592    0,567

=== Confusion Matrix ===

  a  b  <-- classified as
80 56 | a = left
57 78 | b = right
```

Status

OK Log x 0

SENSITIVITY: $\frac{80}{136}$

SPECIFICITY: $\frac{78}{135}$

NAIVE BAYES

Başarı Oranı: %?????

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section on the left shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section on the right displays the following results:

Correctly Classified Instances 0 0 %
Kappa statistic 1 0 %
K&S Relative Info Score NaN %
K&S Information Score 0 bits NaN bits/instance
Class complexity | order 0 0 bits 0 bits/instance
Class complexity | scheme 0 bits NaN bits/instance
Complexity improvement (Sf) 0 bits NaN bits/instance
Mean absolute error NaN
Root mean squared error NaN
Relative absolute error NaN %
Root relative squared error NaN %
Unclassified Instances 271 100 %
Total Number of Instances 271

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
?	?	?	?	?	?	?	0,500	0,502	left
?	?	?	?	?	?	?	0,500	0,498	right
Weighted Avg.	?	?	?	?	?	?	?	?	

=== Confusion Matrix ===

a b <-- classified as
0 0 | a = left
0 0 | b = right

The 'Result list' on the left shows three entries: '00:50:26 - trees.J48', '00:50:50 - trees.J48', and '00:54:00 - bayes.NaiveBayes', with the last one selected. The 'Status' bar at the bottom shows 'OK'.

K-NEAREST NEIGHBOURS (IBk)

Başarı Oranı: %58.6716

Weka Explorer

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

Classifier

Choose **IBk -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A "weka.core.EuclideanDistance -R first-last"**

Test options

☐ Use training set
☐ Supplied test set **Set...**
☒ Cross-validation **Folds** 10
☐ Percentage split % 66
More options...

(Nom) d2

Start **Stop**

Result list (right-click for options)

- 00:50:26 - trees.J48
- 00:50:50 - trees.J48
- 00:54:00 - bayes.NaiveBayes
- 00:55:42 - lazy.IBk**

Classifier output

```
=== Summary ===
Correctly Classified Instances      159           58.6716 %
Kappa statistic                    0.1727
K&S Relative Info Score            17.2554 %
K&S Information Score              46.7712 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          890.4318 bits
Complexity improvement (Sf)        -619.3794 bits
Mean absolute error                 0.414
Root mean squared error             0.6403
Relative absolute error             82.788 %
Root relative squared error         128.0358 %
Total Number of Instances          271

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0,587   0,414    0,591    0,587    0,581    0,178    0,584    0,553
left
Weighted Avg.   0,467   0,294    0,612    0,467    0,529    0,178    0,584    0,554
right

=== Confusion Matrix ===

  a  b  <-- classified as
96 40 | a = left
72 63 | b = right
```

Status

OK

Log

SENSITIVITY: $\frac{96}{136}$

SPECIFICITY: $\frac{63}{135}$

VOTED PERCEPTRON

Başarı Oranı: %61.9926

The screenshot shows the Weka Explorer interface with the Voted Perceptron classifier selected. The test options are set to cross-validation with 10 folds. The classifier output displays the following metrics:

Metric	Value
Correctly Classified Instances	168
Kappa statistic	0.2397
K&S Relative Info Score	24.5616 %
K&S Information Score	66.5748 bits
Class complexity order 0	271.0524 bits
Class complexity scheme	26138.6648 bits
Complexity improvement (Sf)	-25867.6124 bits
Mean absolute error	0.3774
Root mean squared error	0.6111
Relative absolute error	75.477 %
Root relative squared error	122.1987 %
Total Number of Instances	271

The detailed accuracy by class is as follows:

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,647	0,407	0,615	0,647	0,631	0,240	0,644	0,606	left
0,553	0,353	0,625	0,593	0,608	0,240	0,652	0,626	right
Weighted Avg.	0,620	0,380	0,620	0,620	0,240	0,648	0,616	

The confusion matrix is:

a \ b	left	right
88 48 a = left	88	48
55 80 b = right	55	80

LOGISTIC REGRESSION

Başarı Oranı: %55.3506

The screenshot shows the Weka Explorer interface with the Logistic Regression classifier selected. The test options are set to cross-validation with 10 folds. The classifier output displays the following metrics:

Metric	Value
Correctly Classified Instances	150
Kappa statistic	0.1069
K&S Relative Info Score	10.0839 %
K&S Information Score	27.3327 bits
Class complexity order 0	271.0524 bits
Class complexity scheme	24801.1986 bits
Complexity improvement (Sf)	-24530.1462 bits
Mean absolute error	0.4496
Root mean squared error	0.6667
Relative absolute error	89.9057 %
Root relative squared error	133.3224 %
Total Number of Instances	271

The detailed accuracy by class is as follows:

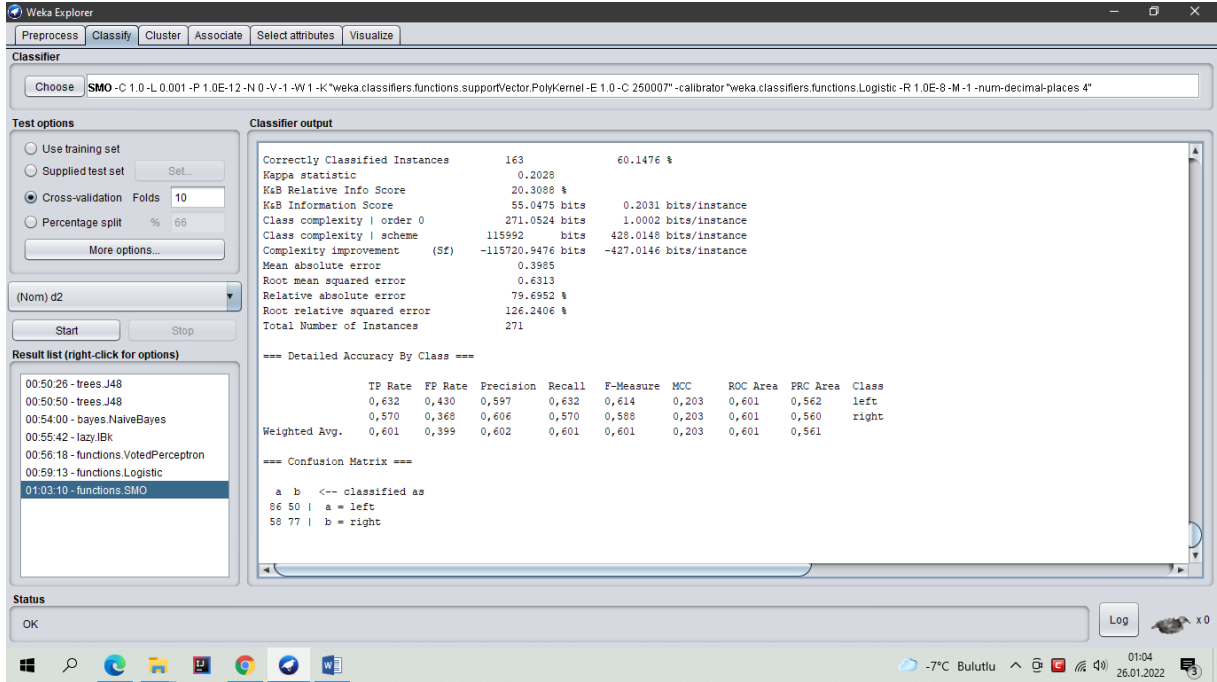
TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,574	0,467	0,553	0,574	0,563	0,107	0,564	0,536	left
0,533	0,426	0,554	0,533	0,543	0,107	0,549	0,549	right
Weighted Avg.	0,554	0,447	0,554	0,553	0,107	0,557	0,542	

The confusion matrix is:

a \ b	left	right
78 58 a = left	78	58
63 72 b = right	63	72

SUPPORT VECTOR MACHINES(SMO)

Başarı Oranı: %64.1476



The screenshot shows the Weka Explorer interface with the SMO classifier selected. The 'Test options' section is set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Metric	Value	Percentage
Correctly Classified Instances	163	60.1476 %
Kappa statistic	0.2028	
K&B Relative Info Score	20.3088	
K&B Information Score	55.0475 bits	0.2031 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	115992 bits	428.0148 bits/instance
Complexity improvement (Sf)	-115720.9476 bits	-427.0146 bits/instance
Mean absolute error	0.3985	
Root mean squared error	0.6313	
Relative absolute error	79.6952 %	
Root relative squared error	126.2406 %	
Total Number of Instances	271	

==== Detailed Accuracy By Class ====

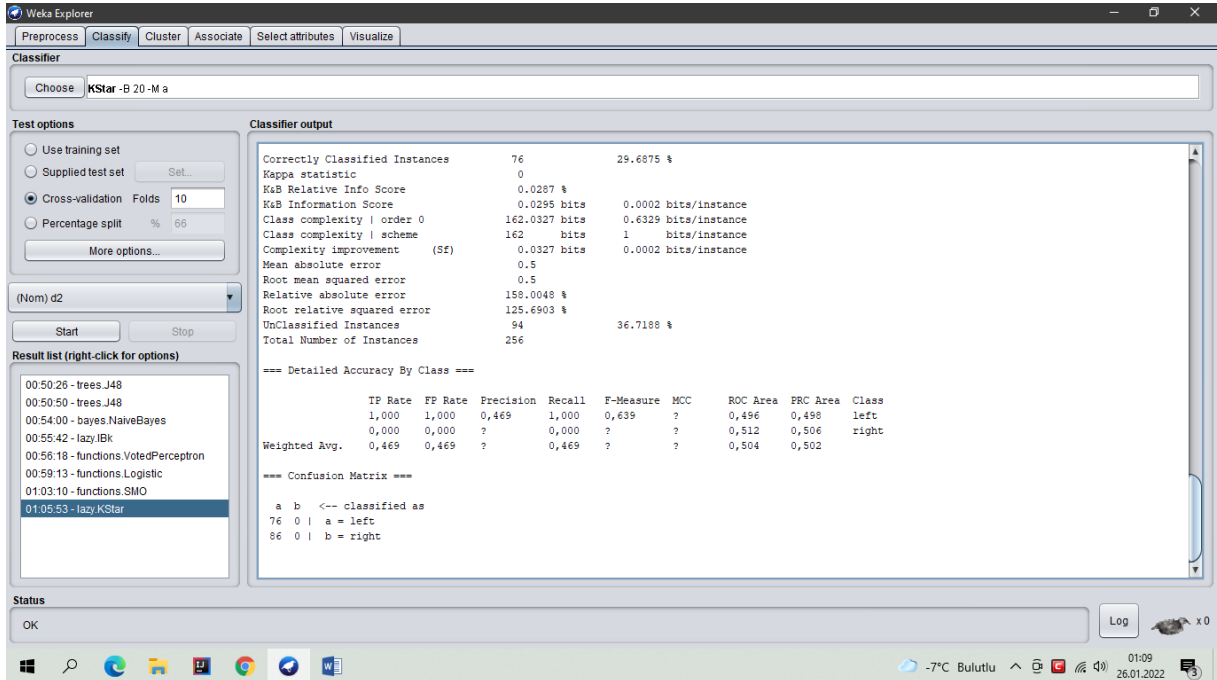
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,601	0,399	0,602	0,601	0,601	0,203	0,601	0,561	

==== Confusion Matrix ====

a \ b	classified as
86 50	a = left
58 77	b = right

K* ALGORITHM (KStar)

Başarı Oranı: %29.6875



The screenshot shows the Weka Explorer interface with the KStar classifier selected. The 'Test options' section is set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Metric	Value	Percentage
Correctly Classified Instances	76	29.6875 %
Kappa statistic	0	
K&B Relative Info Score	0.0287	
K&B Information Score	0.0295 bits	0.0002 bits/instance
Class complexity order 0	162.0327 bits	0.6329 bits/instance
Class complexity scheme	162 bits	1 bits/instance
Complexity improvement (Sf)	0.0327 bits	0.0002 bits/instance
Mean absolute error	0.5	
Root mean squared error	0.5	
Relative absolute error	158.0048 %	
Root relative squared error	125.6903 %	
Unclassified Instances	94	36.7188 %
Total Number of Instances	256	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,469	0,469	?	0,469	?	?	0,504	0,502	

==== Confusion Matrix ====

a \ b	classified as
76 0	a = left
86 0	b = right

RANDOM FOREST

Başarı Oranı: %64.5756

The screenshot shows the Weka Explorer interface with the Random Forest classifier selected. The classifier output window displays the following results:

```
=== Summary ===
Correctly Classified Instances      175          64.5756 %
Kappa statistic                    0.2912
K&S Relative Info Score           11.4489 %
K&S Information Score              31.0324 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          245.4996 bits
Complexity improvement (Sf)        25.5528 bits
Mean absolute error                 0.4539
Root mean squared error             0.4673
Relative absolute error             90.7788 %
Root relative squared error         93.4526 %
Total Number of Instances          271

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0,646   0,355   0,647    0,646   0,645     0,293   0,723    0,702
left
right

=== Confusion Matrix ===
  a b  <-- classified as
95 41 | a = left
55 80 | b = right
```

The result list on the left shows the following entries:

- 00:50:26 - trees.J48
- 00:50:50 - trees.J48
- 00:54:00 - bayes.NaiveBayes
- 00:55:42 - lazy.IBk
- 00:56:18 - functions.VotedPerceptron
- 00:59:13 - functions.Logistic
- 01:03:10 - functions.SMO
- 01:05:53 - lazy.KStar
- 01:10:57 - trees.RandomForest

RANDOM TREE

Başarı Oranı: %47.9705

The screenshot shows the Weka Explorer interface with the Random Tree classifier selected. The classifier output window displays the following results:

```
=== Summary ===
Correctly Classified Instances      130          47.9705 %
Kappa statistic                    -0.0406
K&S Relative Info Score           -4.0413 %
K&S Information Score             -10.9541 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          151434 bits
Complexity improvement (Sf)        -151162.9476 bits
Mean absolute error                 0.5203
Root mean squared error             0.7213
Relative absolute error            104.0466 %
Root relative squared error        144.2437 %
Total Number of Instances          271

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.   0,478   0,519   0,481    0,478   0,480     -0,041   0,480    0,492
left
right

=== Confusion Matrix ===
  a b  <-- classified as
65 71 | a = left
70 65 | b = right
```

The result list on the left shows the following entries:

- 00:50:26 - trees.J48
- 00:50:50 - trees.J48
- 00:54:00 - bayes.NaiveBayes
- 00:55:42 - lazy.IBk
- 00:56:18 - functions.VotedPerceptron
- 00:59:13 - functions.Logistic
- 01:03:10 - functions.SMO
- 01:05:53 - lazy.KStar
- 01:10:57 - trees.RandomForest
- 01:13:03 - trees.RandomTree

STOCHASTIC GRADIENT DESCENT (SGD)

Başarı Oranı: %63.0996

The screenshot shows the Weka Explorer interface with the SGD classifier selected. The classifier output pane displays the following results:

```
--- Summary ---
Correctly Classified Instances      171          63.0996 %
Kappa statistic                    0.2618
K&S Relative Info Score            26.2119 %
K&S Information Score              71.0479 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          107400 bits
Complexity improvement (Sf)        -107128.9476 bits
Mean absolute error                 0.369
Root mean squared error             0.6075
Relative absolute error             73.7919 %
Root relative squared error         121.4751 %
Total Number of Instances          271

--- Detailed Accuracy By Class ---

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.  0,631  0,369  0,632  0,631  0,630  0,262  0,631  0,583

--- Confusion Matrix ---

 a b  <-- classified as
91 45 | a = left
55 80 | b = right
```

LOCALLY WEIGHTED LEARNING (LWL)

Başarı Oranı: %55.3506

The screenshot shows the Weka Explorer interface with the LWL classifier selected. The classifier output pane displays the following results:

```
--- Summary ---
Correctly Classified Instances      150          55.3506 %
Kappa statistic                    0.107
K&S Relative Info Score            3.9885 %
K&S Information Score              10.811 bits
Class complexity | order 0         271.0524 bits
Class complexity | scheme          2419.319 bits
Complexity improvement (Sf)        -2148.2666 bits
Mean absolute error                 0.4842
Root mean squared error             0.5081
Relative absolute error             96.8256 %
Root relative squared error         101.5996 %
Total Number of Instances          271

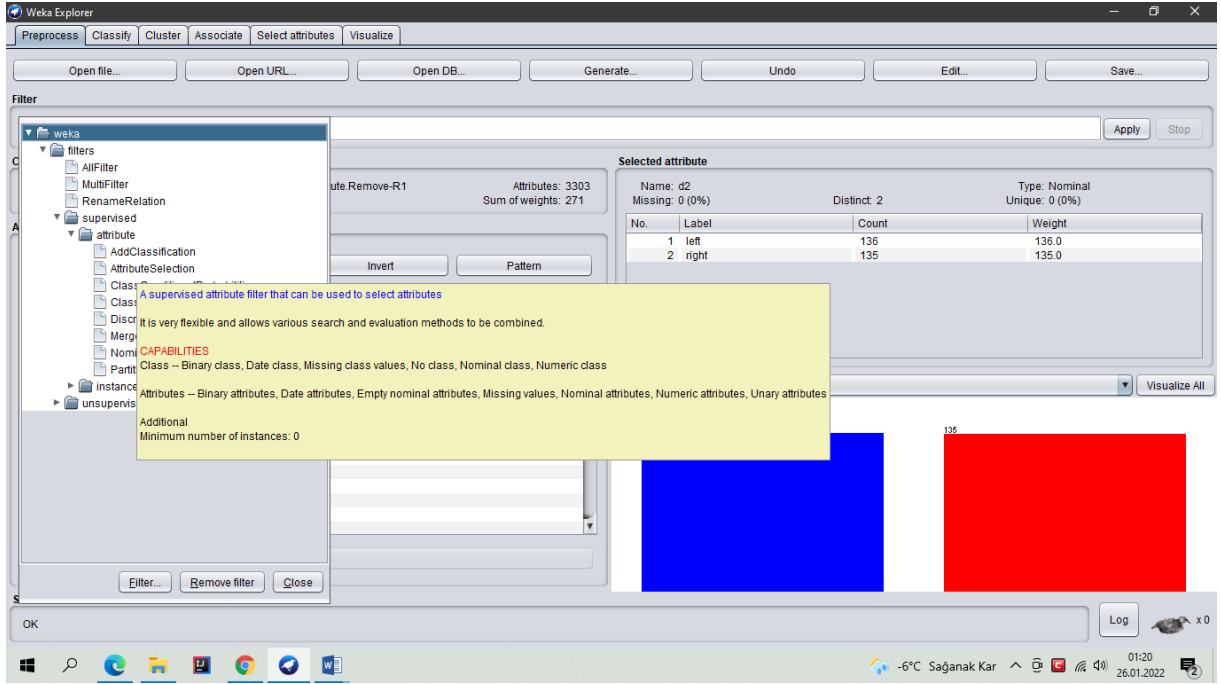
--- Detailed Accuracy By Class ---

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.  0,554  0,446  0,554  0,554  0,554  0,107  0,570  0,553

--- Confusion Matrix ---

 a b  <-- classified as
75 61 | a = left
60 75 | b = right
```

Hiçbir ön işleme tabi tutmadan ve parametre değiştirmeksizin yukarıda belirtilen algoritmalar ile karşılaştırma yapılmıştır. Şimdi ise verilere ön işlem uygulanarak analiz yapılacaktır.

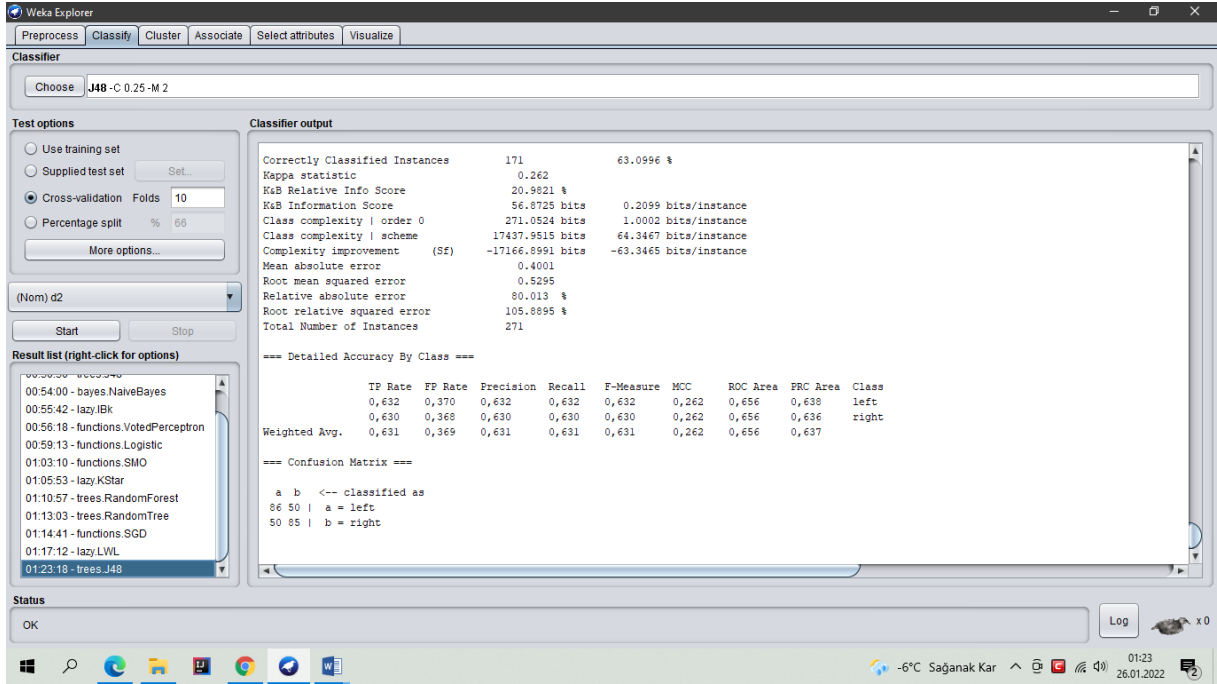


Preprocess sekmesinden görseldeki adımları izleyerek “Best First” algoritmasını kullanarak bir önışlem yapıyoruz. Bunun sonucunda ayıklayıp veri setimizi 271’den 30 veriye düşürüldü.

ÖN İŞLEME ALINMIŞ VERİ SETİ İLE KARŞILAŞTIRMA

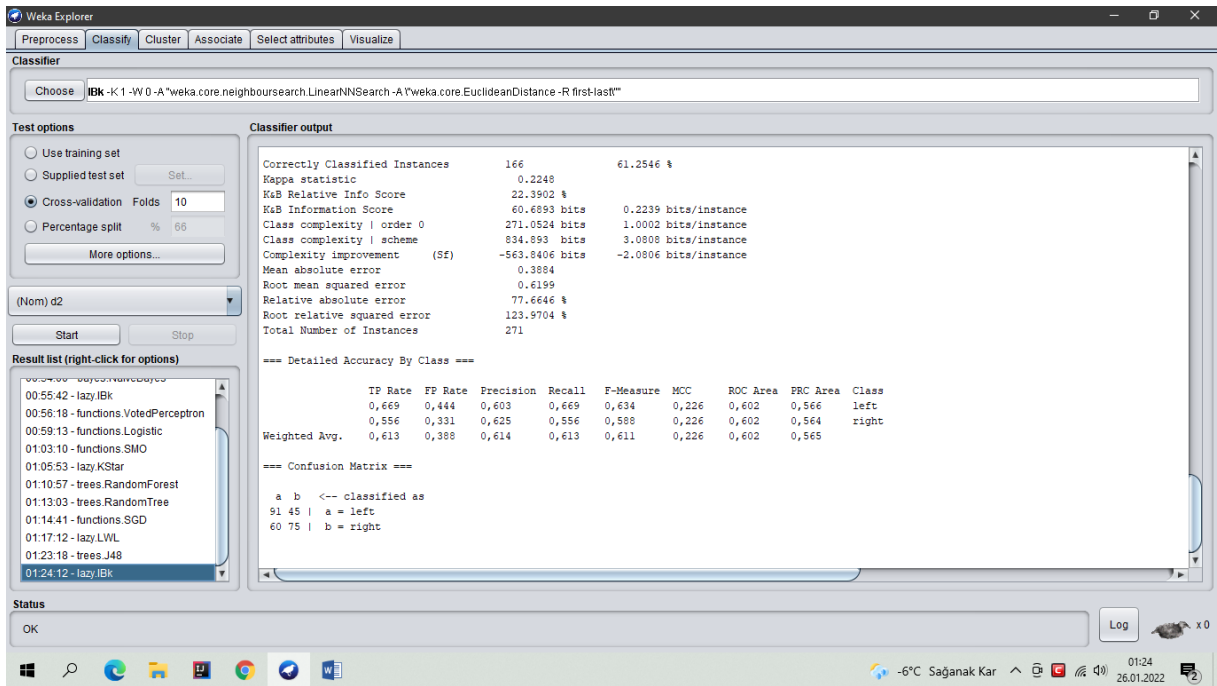
DECISION TREE (J48)

Başarı oranı: %63.0996



K-NEAREST NEIGHBOURS (IBk)

Başarı oranı: %61.2546



NAİVE BAYES

Başarı oranı: %66.7897

The Weka Explorer interface displays the results for the NaiveBayes classifier. The 'Classifier' tab is selected, and the 'NaiveBayes' classifier is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following metrics:

Metric	Value	Percentage
Correctly Classified Instances	181	66.7897 %
Kappa statistic	0.3345	
KsB Relative Info Score	33.291	%
KsB Information Score	90.2361 bits	0.333 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	4792.6 bits	17.6849 bits/instance
Complexity improvement (Sf)	-4521.5476 bits	-16.6847 bits/instance
Mean absolute error	0.3336	
Root mean squared error	0.5746	
Relative absolute error	66.7187 %	
Root relative squared error	114.8961 %	
Total Number of Instances	271	

The 'Result list (right-click for options)' shows a list of classifiers, with '01:24:33 - bayes.NaiveBayes' selected. The 'Status' bar shows 'OK'.

VOTED PERCEPTRON

Başarı Oranı: %66.0517

The Weka Explorer interface displays the results for the VotedPerceptron classifier. The 'Classifier' tab is selected, and the 'VotedPerceptron -1 -E 1.0 -S 1 -M 10000' classifier is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following metrics:

Metric	Value	Percentage
Correctly Classified Instances	178	66.0517 %
Kappa statistic	0.3216	
KsB Relative Info Score	32.0501	%
KsB Information Score	86.8727 bits	0.3206 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	60068.8079 bits	221.6561 bits/instance
Complexity improvement (Sf)	-58797.7555 bits	-220.6559 bits/instance
Mean absolute error	0.3399	
Root mean squared error	0.5813	
Relative absolute error	67.973 %	
Root relative squared error	116.2511 %	
Total Number of Instances	271	

The 'Result list (right-click for options)' shows a list of classifiers, with '01:24:53 - functions.VotedPerceptron' selected. The 'Status' bar shows 'OK'.

STOCHASTIC GRADIENT DESCENT(SGD)

Başarı Oranı: %70.1107

The screenshot shows the Weka Explorer interface with the SGD classifier selected. The 'Test options' section is set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Metric	Value	Percentage
Correctly Classified Instances	190	70.1107 %
Kappa statistic	0.4015	
K&S Relative Info Score	40.2318	
K&S Information Score	109.0492 bits	0.4024 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	86994 bits	321.0111 bits/instance
Complexity improvement (Sf)	-86722.9476 bits	-320.0109 bits/instance
Mean absolute error	0.2989	
Root mean squared error	0.5467	
Relative absolute error	59.7714 %	
Root relative squared error	109.3276 %	
Total Number of Instances	271	

Below the main results, there is a 'Detailed Accuracy By Class' table and a 'Confusion Matrix'.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,860	0,459	0,654	0,860	0,743	0,423	0,701	0,632	left
	0,541	0,140	0,793	0,541	0,643	0,423	0,701	0,658	right

The 'Confusion Matrix' shows:

a \ b	left	right
117 19 a = left	117	19
62 73 b = right	62	73

LOCALLY WEIGHTED LEARNING (LWL)

Başarı Oranı: %55.7196

The screenshot shows the Weka Explorer interface with the LWL classifier selected. The 'Test options' section is set to 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Metric	Value	Percentage
Correctly Classified Instances	151	55.7196 %
Kappa statistic	0.1146	
K&S Relative Info Score	4.4397	
K&S Information Score	12.0339 bits	0.0444 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	1444.0998 bits	5.3288 bits/instance
Complexity improvement (Sf)	-1173.0474 bits	-4.3286 bits/instance
Mean absolute error	0.4826	
Root mean squared error	0.5093	
Relative absolute error	96.5161 %	
Root relative squared error	101.8402 %	
Total Number of Instances	271	

Below the main results, there is a 'Detailed Accuracy By Class' table and a 'Confusion Matrix'.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,522	0,407	0,563	0,522	0,542	0,115	0,588	0,547	left
	0,593	0,478	0,552	0,593	0,571	0,115	0,588	0,596	right

The 'Confusion Matrix' shows:

a \ b	left	right
71 65 a = left	71	65
55 80 b = right	55	80

K* ALGORITHM (KStar)

Başarı Oranı: %59.4096

The Weka Explorer interface displays the results for the KStar classifier. The 'Classifier' tab is selected, and the 'KStar-B 20-M a' model is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following metrics:

Metric	Value	Percentage
Correctly Classified Instances	161	59.4096 %
Kappa statistic	0.1872	
K&S Relative Info Score	18.0808	%
K&S Information Score	49.0084 bits	0.1808 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	737.8422 bits	2.7227 bits/instance
Complexity improvement (Sf)	-466.7898 bits	-1.7225 bits/instance
Mean absolute error	0.4099	
Root mean squared error	0.5775	
Relative absolute error	81.9781	%
Root relative squared error	115.4804	%
Total Number of Instances	271	

The 'Result list (right-click for options)' shows a list of models, with '01:29:45 - lazy KStar' selected. The 'Status' bar shows 'OK'.

RANDOM FOREST

Başarı Oranı: %70.1107

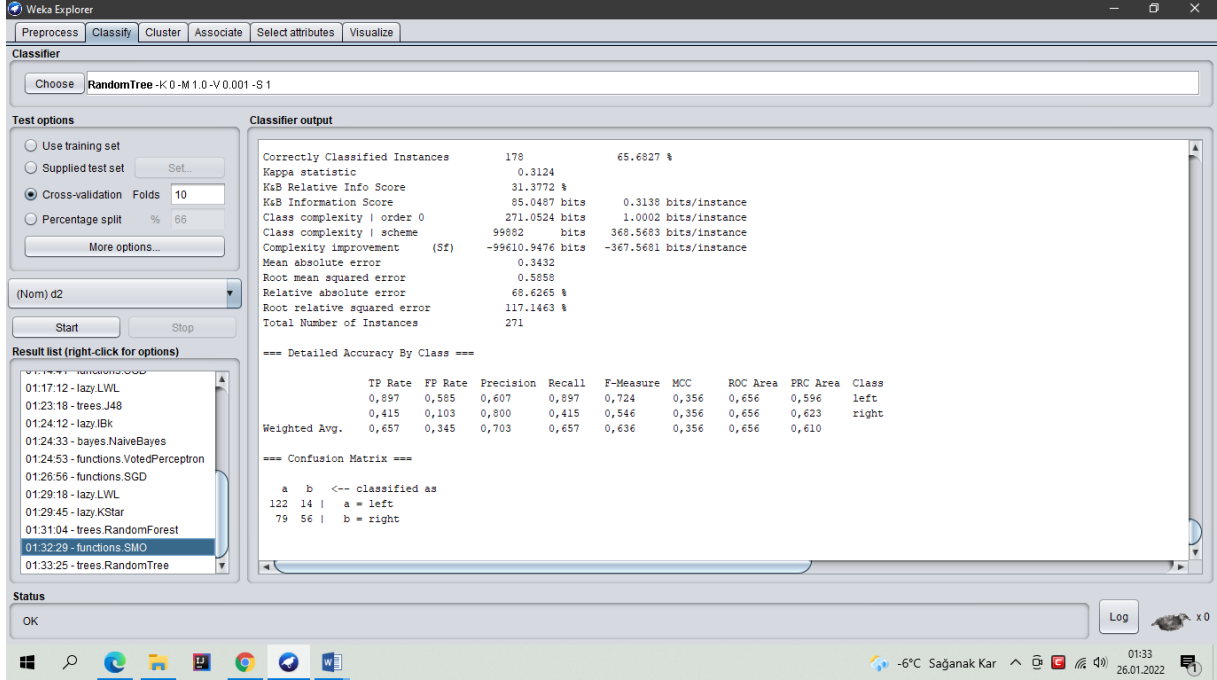
The Weka Explorer interface displays the results for the Random Forest classifier. The 'Classifier' tab is selected, and the 'RandomForest -P 100-I 100-num-slots 1-K 0-M 1.0-V 0.001-S 1' model is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section shows the following metrics:

Metric	Value	Percentage
Correctly Classified Instances	190	70.1107 %
Kappa statistic	0.4022	
K&S Relative Info Score	23.0978	%
K&S Information Score	62.607 bits	0.231 bits/instance
Class complexity order 0	271.0524 bits	1.0002 bits/instance
Class complexity scheme	233.0345 bits	0.8599 bits/instance
Complexity improvement (Sf)	38.0179 bits	0.1403 bits/instance
Mean absolute error	0.3987	
Root mean squared error	0.4488	
Relative absolute error	79.7248	%
Root relative squared error	89.7401	%
Total Number of Instances	271	

The 'Result list (right-click for options)' shows a list of models, with '01:31:04 - trees RandomForest' selected. The 'Status' bar shows 'OK'.

SUPPORT VECTOR MACHINES(SMO)

Başarı Oranı: %65.6827



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'RandomTree -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Correctly Classified Instances: 178 (65.6827 %)
Kappa statistic: 0.3124
K&S Relative Info Score: 31.3772 %
K&S Information Score: 85.0487 bits (0.3138 bits/instance)
Class complexity | order 0: 271.0524 bits (1.0002 bits/instance)
Class complexity | scheme: 99892 bits (368.5683 bits/instance)
Complexity improvement (Sf): -99610.9476 bits (-367.5681 bits/instance)
Mean absolute error: 0.3432
Root mean squared error: 0.5858
Relative absolute error: 68.6265 %
Root relative squared error: 117.1463 %
Total Number of Instances: 271

==== Detailed Accuracy By Class ===

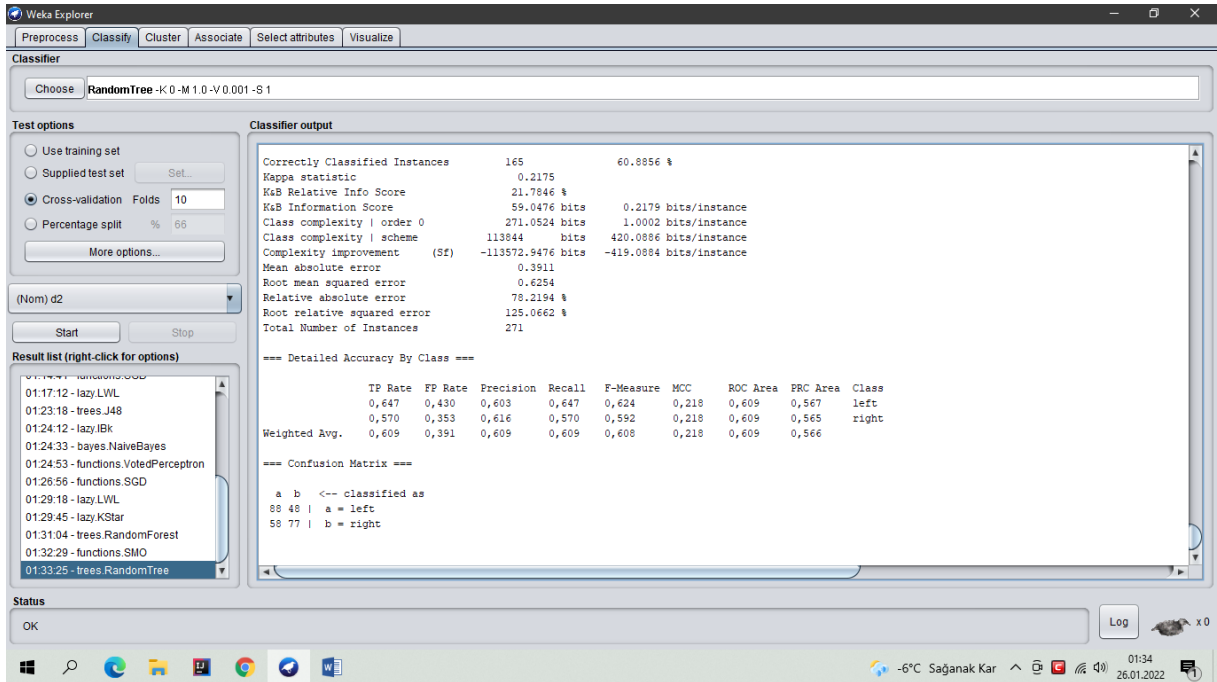
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,897	0,585	0,607	0,897	0,724	0,356	0,656	0,596	left
	0,415	0,103	0,800	0,415	0,546	0,356	0,656	0,623	right

==== Confusion Matrix ===

a \ b	left	right
122 14 a = left	122	14
79 56 b = right	79	56

RANDOM TREE

Başarı Oranı: %60.8856



The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'RandomTree -K 0 -M 1.0 -V 0.001 -S 1'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' section displays the following results:

Correctly Classified Instances: 165 (60.8856 %)
Kappa statistic: 0.2175
K&S Relative Info Score: 21.7846 %
K&S Information Score: 59.0476 bits (0.2179 bits/instance)
Class complexity | order 0: 271.0524 bits (1.0002 bits/instance)
Class complexity | scheme: 113844 bits (420.0886 bits/instance)
Complexity improvement (Sf): -113572.9476 bits (-419.0884 bits/instance)
Mean absolute error: 0.3911
Root mean squared error: 0.6254
Relative absolute error: 78.2194 %
Root relative squared error: 125.0662 %
Total Number of Instances: 271

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0,647	0,430	0,603	0,647	0,624	0,218	0,609	0,567	left
	0,570	0,353	0,616	0,570	0,592	0,218	0,609	0,565	right

==== Confusion Matrix ===

a \ b	left	right
88 48 a = left	88	48
58 77 b = right	58	77

LOGISTIC REGRESSION

Başarı Oranı: %70.8487

The screenshot displays the Weka Explorer interface with the 'Classifier' tab selected. The 'Logistic - R 1.0E-8 - M -1 - num-decimal-places 4' classifier is chosen. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

Correctly Classified Instances 192 70.8487 %
Kappa statistic 0.4165
K&S Relative Info Score 34.7904 %
K&S Information Score 94.3003 bits
Class complexity | order 0 271.0524 bits 1.0002 bits/instance
Class complexity | scheme 403.9609 bits 1.4906 bits/instance
Complexity improvement (Sf) -132.9085 bits -0.4904 bits/instance
Mean absolute error 0.3337
Root mean squared error 0.4428
Relative absolute error 66.7309 %
Root relative squared error 88.5486 %
Total Number of Instances 271

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,816	0,400	0,673	0,816	0,738	0,426	0,791	0,752	left
	0,600	0,184	0,764	0,600	0,672	0,426	0,790	0,815	right
Weighted Avg.	0,708	0,292	0,718	0,708	0,705	0,426	0,791	0,783	

==== Confusion Matrix ====

a	b	<-- classified as
111	25	a = left
54	81	b = right

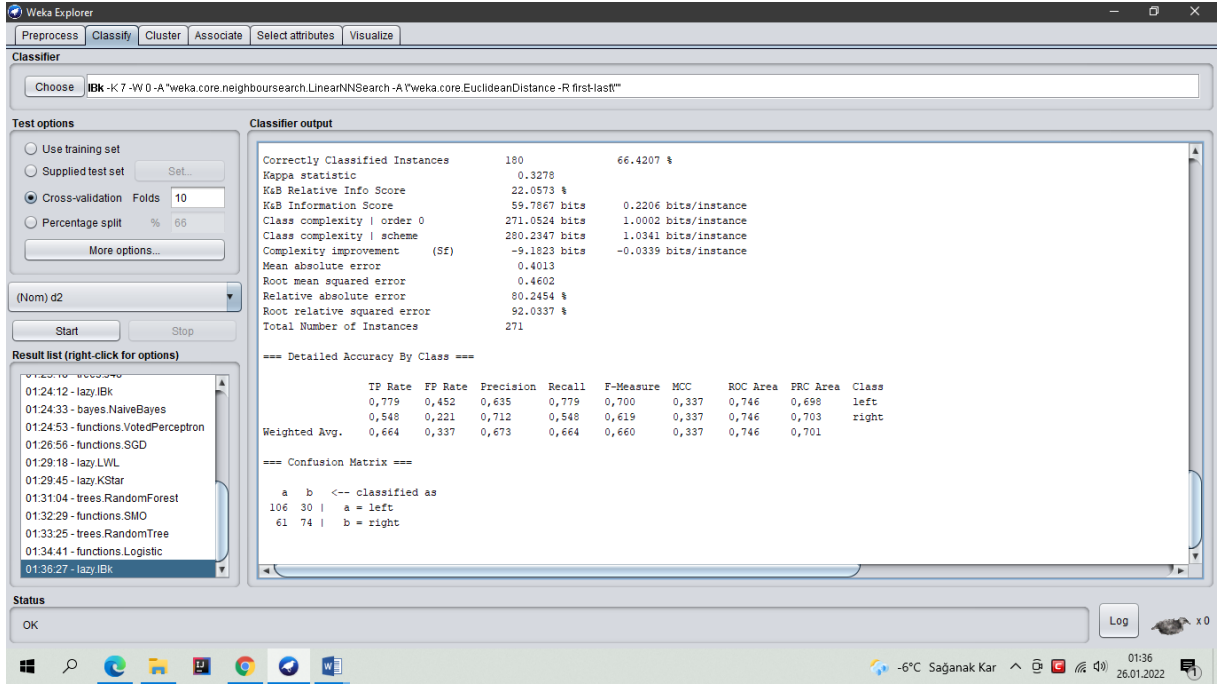
The 'Result list' on the left shows various classifiers, with 'functions.Logistic' selected at the bottom. The 'Status' bar at the bottom indicates 'OK'.

PARAMETRE DEĞİŞİKLİĞİ İLE BAŞARIM KARŞILAŞTIRILMASI

K-NEAREST NEIGHBOURS (IBk)

Başarı Oranı %66.4207

PARAMETRELER: K 7 -W 0



K - NEAREST NEIGHBOUR ALGORİTMASINDA “K1” değeri “K7” ile değıştirilip başarı oranı arttırılmıştır. Veri setinin ön işleme sokularak ve algoritmadaki parametre değeriinde değışim yapılarak başarı oranında artış gözlemlenmiştir. KNN algoritmasının en iyi sonucu bu parametreler ile bulunmuştur.

$$\text{SENSITIVITY: } \frac{106}{136}$$

$$\text{SPECIFICITY: } \frac{74}{135}$$

DECISION TREE (J48)

Başarı Oranı: %67.5277

PARAMETRELER: C 0.6 – M 9

Classify

Choose **J48-C 0.6-M 9**

Test options

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

More options...

(Nom) d2

Start Stop

Result list (right-click for options)

- 01:24:33 - bayes.NaiveBayes
- 01:24:53 - functions.VotedPerceptron
- 01:26:56 - functions.SGD
- 01:29:18 - lazy.LWL
- 01:29:45 - lazy.KStar
- 01:31:04 - trees.RandomForest
- 01:32:29 - functions.SMO
- 01:33:25 - trees.RandomTree
- 01:34:41 - functions.Logistic
- 01:36:27 - lazy.IBk
- 01:37:48 - trees.J48**

Classifier output

```
Correctly Classified Instances      183      67.5277 %
Kappa statistic                    0.3503
K&B Relative Info Score           22.9502 %
K&B Information Score              62.2069 bits      0.2295 bits/instance
Class complexity | order 0        271.0524 bits      1.0002 bits/instance
Class complexity | scheme         4550.3619 bits      16.791 bits/instance
Complexity improvement (Sf)       -4279.3095 bits     -15.7908 bits/instance
Mean absolute error                0.3951
Root mean squared error            0.4086
Relative absolute error            79.0085 %
Root relative squared error        97.7057 %
Total Number of Instances         271

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      -----  -
      0,728    0,378    0,660    0,728    0,692    0,352    0,684    0,674    left
      0,622    0,272    0,694    0,622    0,656    0,352    0,684    0,659    right
Weighted Avg.   0,675    0,325    0,677    0,675    0,674    0,352    0,684    0,667

=== Confusion Matrix ===

  a  b  <-- classified as
99 37 | a = left
51 84 | b = right
```

Status

OK

Log

01:37 26.01.2022

Decision Tree algoritmasında “C 0.25” değeri “C 0.6” ile ve “M 2” değeri “M 9” ile değiştirilip başarı oranı artırılmıştır. Veri setinin ön işleme sokularak ve algoritmadaki parametre değerlerinde değişim yapılarak başarı oranında artış gözlemlenmiştir. Decision Tree algoritmasının en iyi sonucu bu parametreler ile bulunmuştur.

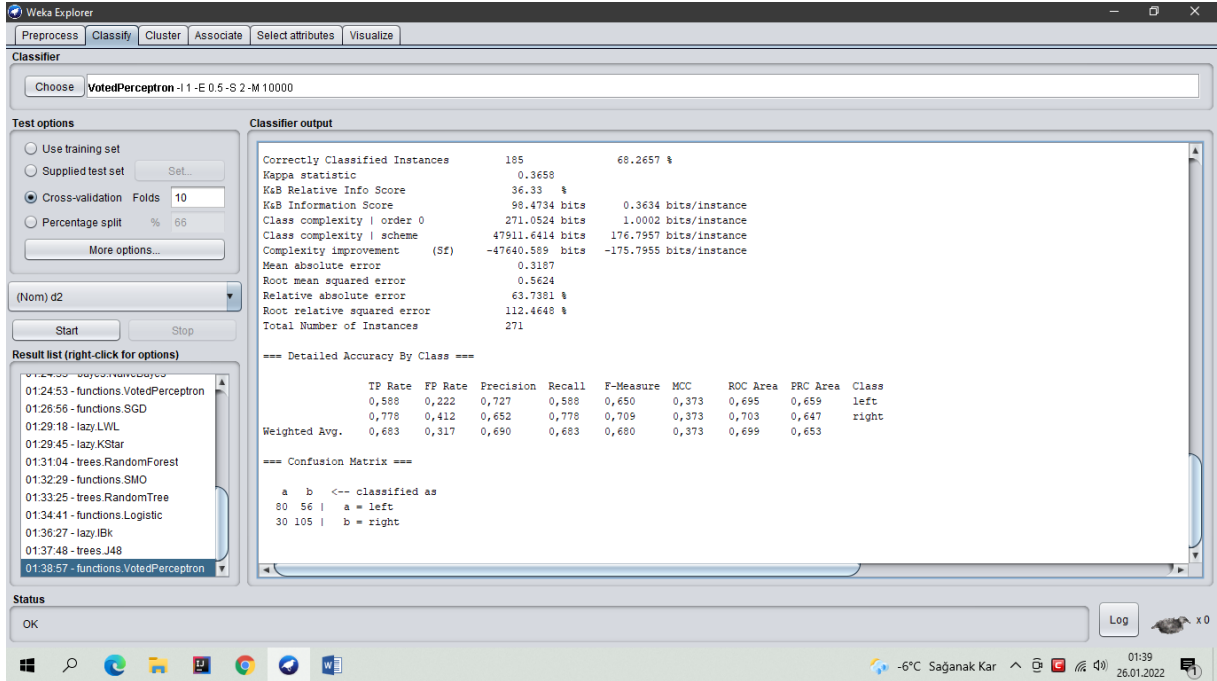
SENSITIVITY: $\frac{99}{136}$

SPECIFICITY: $\frac{84}{135}$

VOTED PERCEPTRON

Başarı Oranı: %68.2657

PARAMETRELER: I 1 – E 0.5 – S 2 – M 10000



Denemeler sonucunda en iyi sonucu veren durum gözükmemektedir. Voted perceptron algoritması kullanılarak verilerin parametreler ile alınan sonucu başarı oranı en iyi sonuçtur. Exponent değeri 1.0'dan 0.5'e ve Seed değeri 1'den 2'ye değiştirilerek bu sonuca varılmıştır.

$$\text{SENSITIVITY: } \frac{80}{136}$$

$$\text{SPECIFICITY: } \frac{105}{135}$$

SONUÇLAR

ALGORİTMALAR	MAE	RMSE	RAE(%)	RRSE(%)	DOĞRULUK ORANI (%)
(SGD)	0.369	0.6075	73.7919	121.4751	60.0996
(LWL)	0.4842	0.5081	96.8256	101.5996	55.3506
K*	0.5	0.5	158.0048	125.6903	29.6875
NAİVE BAYES	?	?	?	?	?
RANDOM FOREST	0.4539	0.4673	90.7788	93.4526	64.5756
SMO	0.3985	0.6313	79.6952	126.2406	64.1476
RANDOM TREE	0.5203	0.7213	104.0466	144.2437	47.9705
J48	0.423	0.6277	84.5887	125.5205	58.3026
IBK	0.414	0.6403	82.788	128.0358	58.6716
LOGİSTİC REGRESSION	0.4496	0.6667	89.9057	133.3324	55.3506
VOTED PERCEPTRON	0.3774	0.6111	75.477	122.1987	61.9926
WITH PREPROCESS					
(SGD)	0.2989	0.5467	59.7714	109.3276	70.1107
(LWL)	0.4826	0.5093	96.5161	101.8402	55.7196
K*	0.4099	0.5775	81.9781	115.4804	59.4096
NAİVE BAYES	0.3336	0.5746	66.7187	114.8961	66.7897
RANDOM FOREST	0.3987	0.4488	79.7248	89.7401	70.1107
SMO	0.3432	0.5858	68.6265	117.1463	65.6827
RANDOM TREE	0.3911	0.6254	78.2194	125.0662	60.8856
J48	0.4001	0.5295	80.013	105.8895	63.0996
IBK	0.3884	0.6199	77.6646	123.9704	61.2546
LOGİSTİC REGRESSION	0.3337	0.4428	66.7309	88.5486	70.8487
VOTED PERCEPTRON	0.3399	0.5813	67.973	116.2511	66.0517
CHANGE PARAMETERS					
J48	0.3951	0.4886	79.0085	97.7057	62.5277
IBK	0.4013	0.4602	80.2454	92.0337	66.4207
VOTED PERCEPTRON	0.3187	0.5624	63.7381	112.4648	68.2657

İLK OLARAK ALGORİTMALAR KULLANILDIĞINDA EN İYİ SONUCU **RANDOM FOREST** ALGORİTMASI VERMİŞTİR.

PREPROCESS UYGULADIKTAN SONRA EN İYİ SONUCU **LOGİSTİC REGRESSION** ALGORİTMASI VERMİŞTİR

İÇLERİNDEN BAZI ALGORİTMALARIN PARAMETRELERİ DEĞİŞTİRELEREK DENENDİĞİNDE EN İYİ SONUCU **VOTED PERCEPTRON** VERMİŞTİR