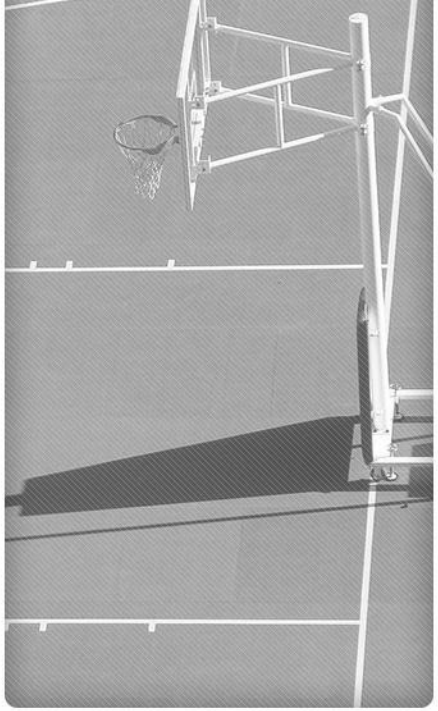





Business Intelligence

Prediction the best shooters in the NBA

Instructor: Prof. Dr. Adem Karahoca
Student Ahmet Hacıoğlu
Student ID: 041801123

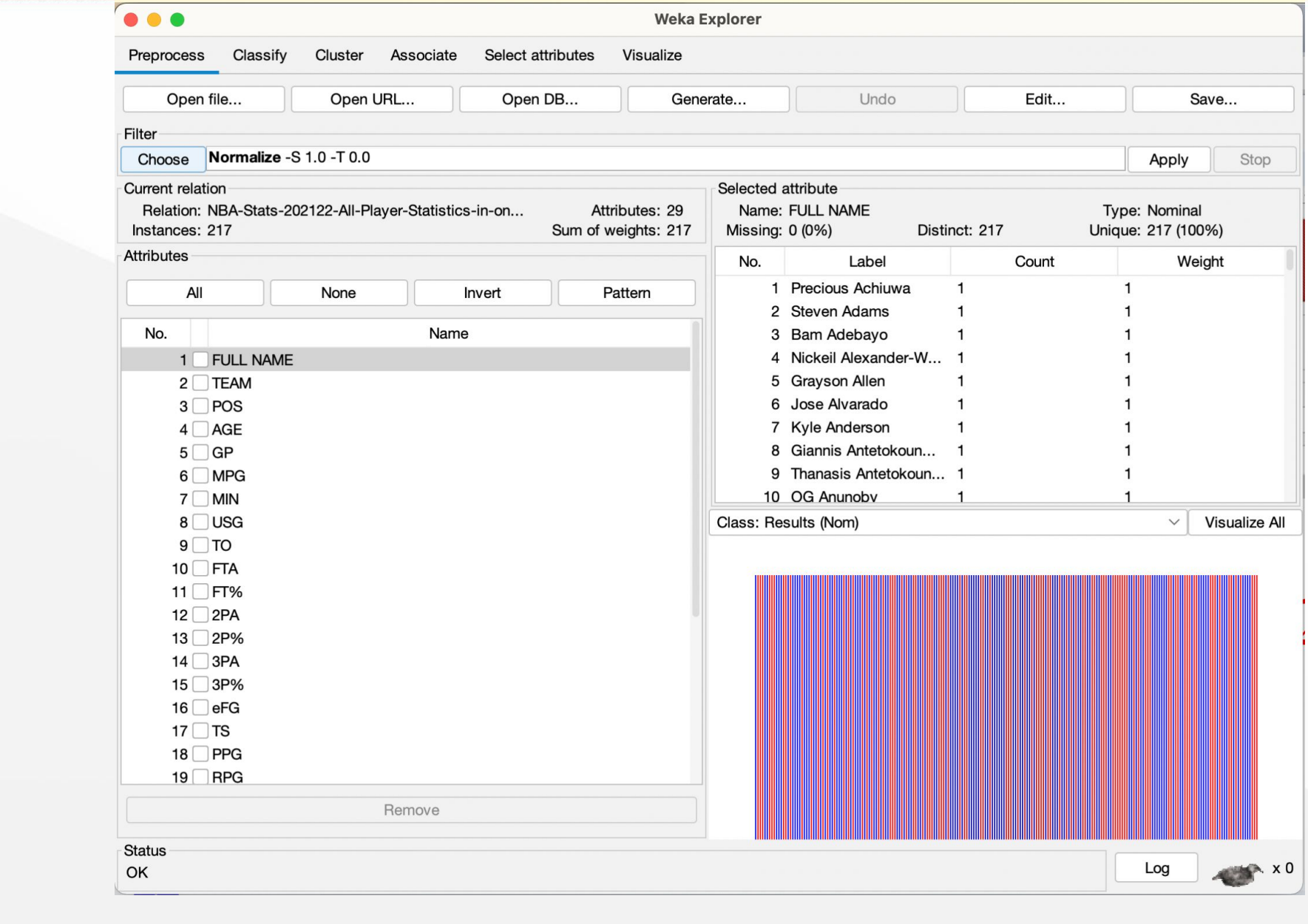
The main purpose of this project…

- *Predict the best shooters using key shooting statistics from basketball players.*

Why did I choose this topic?

- *Real-world application*
- *Interest-Relevance-Motivation*
- *A Huge Industry*
- *Valuable data analysis and modeling skills*

Attributes


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Normalize -S 1.0 -T 0.0
Apply
Stop
Current relation
Relation: NBA-Stats-202122-All-Player-Statistics-in-on...
Attributes: 29
Instances: 217
Sum of weights: 217
Selected attribute
Name: FULL NAME
Type: Nominal
Missing: 0 (0%)
Distinct: 217
Unique: 217 (100%)
Attributes
All
None
Invert
Pattern
No. Name
1 FULL NAME
2 TEAM
3 POS
4 AGE
5 GP
6 MPG
7 MIN
8 USG
9 TO
10 FTA
11 FT%
12 2PA
13 2P%
14 3PA
15 3P%
16 eFG
17 TS
18 PPG
19 RPG
Remove
No. Label Count Weight
1 Precious Achiuwa 1 1
2 Steven Adams 1 1
3 Bam Adebayo 1 1
4 Nickeil Alexander-W... 1 1
5 Grayson Allen 1 1
6 Jose Alvarado 1 1
7 Kyle Anderson 1 1
8 Giannis Antetokoun... 1 1
9 Thanasis Antetokoun... 1 1
10 OG Anunoby 1 1
Class: Results (Nom)
Visualize All
Status
OK
Log
x 0

The Important Attributes for Selecting Good Shooter

1. **Field Goal Percentage (FG%):** This measures the percentage of field goals a player makes compared to the number of attempts. This can give you a good sense of how accurate a player is when shooting.
2. **Three-Point Percentage (3P%):** This measures the percentage of three-point shots a player makes compared to the number of attempts. This can be particularly important if you're looking for players who are good at shooting from beyond the arc.
3. **Effective Field Goal Percentage (eFG%):** This is a more advanced metric that takes into account the fact that three-point shots are worth more than two-point shots. It adjusts a player's field goal percentage based on how many three-pointers they make.
4. **True Shooting Percentage (TS%):** This is another advanced metric that takes into account free throw shooting as well as two-point and three-point shooting. It gives you a sense of how efficient a player is at scoring.
5. **Usage Rate (USG%):** This measures how often a player is involved in their team's offensive plays. A player with a high usage rate may be more likely to take a lot of shots, but may also be more prone to turnovers.

DataSet

Viewer

Relation: NBA-Stats-202122-All-Player-Statistics-in-one-Page

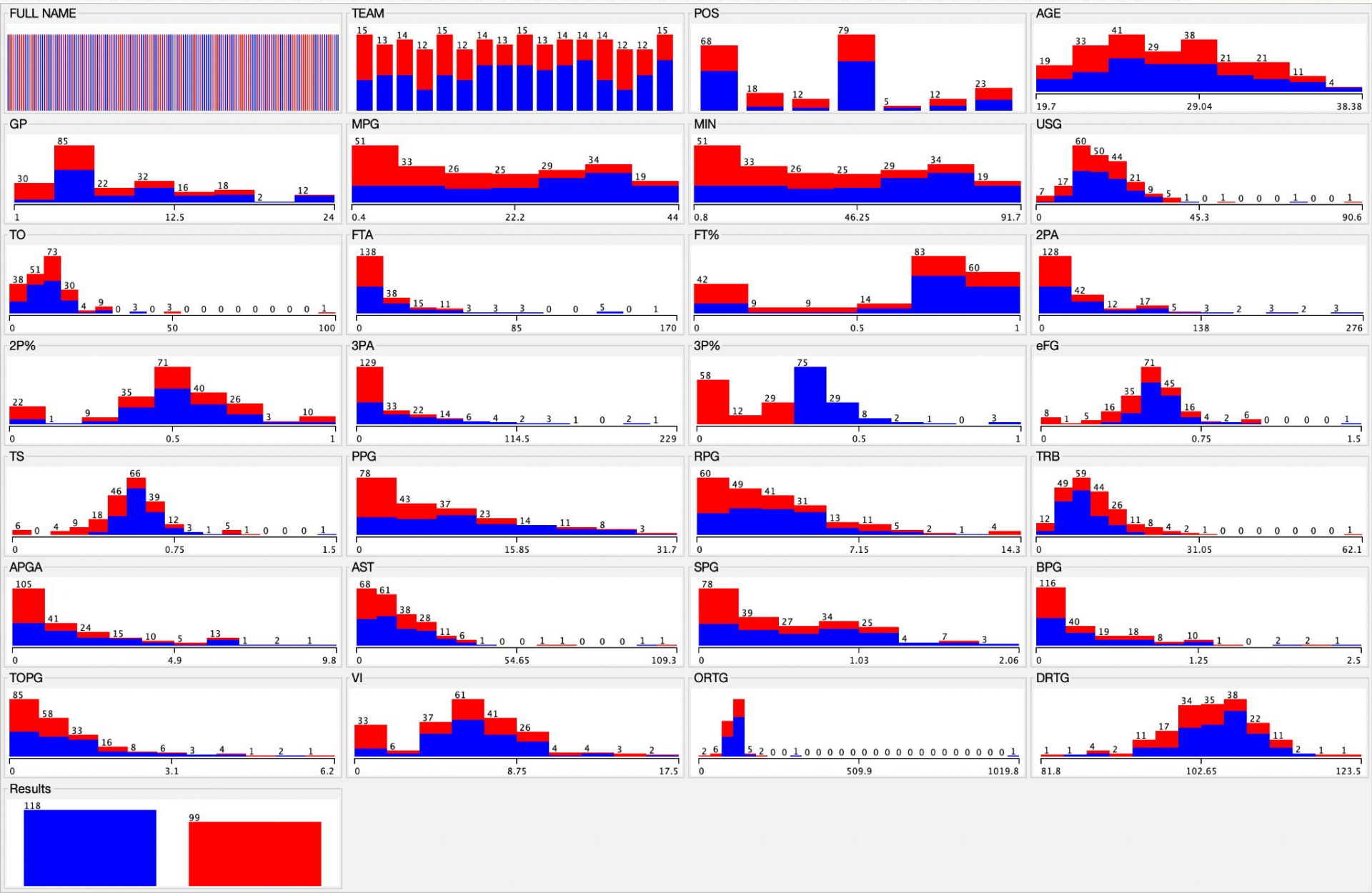
No.	1: FULL NAME Nominal	2: TEAM Nominal	3: POS Nominal	4: AGE Numeric	5: GP Numeric	6: MPG Numeric	7: MIN Numeric	8: USG Numeric	9: TO Numeric	10: FTA Numeric	11: FT% Numeric	12: 2PA Numeric	13: 2P% Numeric	14: 3PA Numeric	15: 3P% Numeric	16: eFG Numeric	17: TS Numeric	18: PPG Numeric	19: RPG Numeric	20: TRB Numeric	21: APGA Numeric	22: AST Numeric	23: SPG Numeric	24: BPG Numeric	2
1	Precious Ach...	Tor	F	22.74	6.0	27.8	57.8	18.2	13.8	10.0	0.6	36.0	0.556	16.0	0.313	0.529	0.541	10.2	4.8	10.7	1.0	5.7	0.17	0.83	
2	Steven Adams	Mem	C	28.91	7.0	16.2	33.8	10.8	13.4	11.0	0.545	21.0	0.429	0.0	0.0	0.429	0.464	3.4	6.4	20.9	2.1	17.6	0.14	0.14	
3	Bam Adebayo	Mia	C-F	24.92	18.0	34.1	71.0	17.9	15.4	76.0	0.763	174.0	0.598	1.0	0.0	0.594	0.638	14.8	8.0	13.8	2.7	12.6	1.0	0.72	
4	Nickeil Alexa...	Uta	G	23.79	1.0	4.7	9.8	24.4	0.0	1.0	1.0	2.0	1.0	0.0	0.0	1.0	1.025	5.0	1.0	12.4	1.0	70.1	1.0	0.0	
5	Grayson Allen	Mil	G	26.69	12.0	25.4	53.0	13.6	9.4	11.0	0.636	34.0	0.529	48.0	0.396	0.567	0.576	8.3	2.9	5.9	1.3	7.2	0.67	0.33	
6	Jose Alvarado	Nor	G	24.18	6.0	19.5	40.6	16.8	15.3	13.0	0.769	17.0	0.588	16.0	0.375	0.576	0.62	8.0	1.3	4.0	1.5	11.4	1.17	0.17	
7	Kyle Anderson	Mem	F-G	28.74	12.0	18.4	38.3	12.8	14.5	18.0	0.611	39.0	0.667	12.0	0.25	0.598	0.611	6.0	4.3	12.4	1.8	13.7	0.92	0.67	
8	Giannis Antet...	Mil	F	27.53	12.0	37.3	77.7	38.5	13.6	134.0	0.679	244.0	0.537	41.0	0.22	0.507	0.552	31.7	14.2	19.5	6.8	37.0	0.67	1.33	
9	Thanasis Ant...	Mil	F	29.91	8.0	2.6	5.3	9.2	0.0	3.0	0.333	3.0	0.667	0.0	0.0	0.667	0.579	0.6	0.5	10.0	0.1	7.0	0.13	0.0	
10	OG Anunoby	Tor	F	24.92	6.0	36.1	75.2	21.2	10.1	16.0	0.75	41.0	0.61	41.0	0.341	0.561	0.584	17.3	4.0	6.8	2.5	11.5	1.0	0.17	
11	Deandre Ayton	Pho	C	23.9	13.0	30.5	63.6	22.9	9.6	33.0	0.636	160.0	0.644	4.0	0.5	0.646	0.653	17.9	8.9	17.1	1.7	9.3	0.38	0.77	
12	Desmond Bane	Mem	G	23.98	12.0	35.6	74.2	18.3	8.4	35.0	0.857	71.0	0.465	88.0	0.489	0.613	0.645	18.8	3.7	5.4	2.3	9.7	1.0	0.67	
13	Dalano Banton	Tor	F	22.61	4.0	1.9	4.0	34.4	17.7	6.0	0.5	2.0	1.0	0.0	0.0	1.0	0.754	1.8	0.5	16.2	0.3	25.2	0.25	0.0	
14	Scottie Barnes	Tor	F	20.88	4.0	33.3	69.3	18.9	9.3	16.0	0.813	30.0	0.533	12.0	0.167	0.452	0.52	12.8	9.0	16.6	4.3	19.7	1.0	0.25	
15	Will Barton	Den	G	31.44	5.0	34.4	71.7	19.9	12.7	6.0	0.667	38.0	0.421	28.0	0.393	0.492	0.503	13.8	5.6	10.1	2.8	12.3	0.8	0.2	
16	Charles Bass...	Phi	C-F	21.63	3.0	4.2	8.7	9.2	0.0	1.0	0.0	2.0	0.5	0.0	0.0	0.5	0.41	0.7	1.7	24.5	0.3	11.9	0.0	0.33	
17	Malik Beasley	Min	G	25.55	6.0	19.9	41.4	19.3	13.0	6.0	0.833	19.0	0.579	25.0	0.32	0.523	0.547	8.5	3.3	9.3	0.7	5.6	0.33	0.17	
18	Davis Bertans	Dal	F	29.59	18.0	10.7	22.3	15.6	4.7	12.0	0.833	4.0	0.75	51.0	0.373	0.573	0.606	4.1	1.4	7.9	0.3	4.9	0.44	0.11	
19	Patrick Bever...	Min	G	33.93	6.0	32.3	67.2	15.6	17.0	22.0	0.682	23.0	0.522	26.0	0.346	0.52	0.562	11.0	3.2	5.5	4.8	23.1	1.17	1.17	
20	Khem Birch	Tor	C	29.72	6.0	10.5	21.9	13.9	15.8	0.0	0.0	12.0	0.5	4.0	0.5	0.563	0.563	3.0	1.5	8.7	0.5	7.2	0.17	0.17	
21	Bismack Biy...	Pho	C	29.8	9.0	9.5	19.9	13.7	23.4	6.0	0.5	17.0	0.647	0.0	0.0	0.647	0.636	2.8	2.1	12.9	0.6	8.0	0.11	0.22	
22	Nemanja Bjel...	Gol	F	34.1	15.0	10.0	20.8	12.9	15.9	7.0	0.571	26.0	0.577	8.0	0.375	0.574	0.58	2.9	2.1	11.6	1.1	15.3	0.27	0.13	
23	Bojan Bogda...	Uta	F	33.16	6.0	35.7	74.4	22.7	11.6	24.0	0.792	48.0	0.583	33.0	0.333	0.549	0.59	18.0	4.2	6.8	1.7	8.5	0.33	0.0	
24	Bogdan Bog...	Atl	G	29.83	4.0	26.7	55.7	24.7	7.0	10.0	0.8	23.0	0.478	26.0	0.346	0.5	0.534	14.3	4.8	10.7	3.0	21.2	0.25	0.25	
25	Isaac Bonga	Tor	G	22.61	1.0	2.8	5.8	14.6	0.0	2.0	1.0	0.0	0.0	0.0	0.0		1.136	2.0	1.0	21.9	0.0	0.0	0.0	0.0	
26	Devin Booker	Pho	G	25.63	10.0	36.6	76.2	28.9	13.9	53.0	0.887	110.0	0.464	65.0	0.431	0.531	0.587	23.3	4.8	7.7	4.4	18.7	0.5	0.4	
27	Chris Boucher	Tor	F-C	29.43	6.0	21.6	45.1	17.3	4.1	10.0	0.9	27.0	0.741	15.0	0.4	0.69	0.722	11.2	5.8	16.6	0.2	1.3	0.17	1.17	
28	Tony Bradley	Chi	C-F	24.44	2.0	3.9	8.1	28.1	0.0	0.0	0.0	5.0	1.0	0.0	0.0	1.0	1.0	5.0	2.0	25.7	0.5	109.3	0.0	0.0	
29	Mikal Bridges	Pho	F	25.8	13.0	38.5	80.2	14.9	6.8	30.0	0.933	105.0	0.505	33.0	0.394	0.525	0.572	13.3	4.7	7.1	2.8	10.2	1.08	1.0	
30	Armoni Brooks	Tor	G	24.03	4.0	1.9	4.0	24.4	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.5	16.2	0.0	0.0	0.0	0.0	
31	Dillon Brooks	Mem	G-F	26.4	11.0	30.5	63.5	25.1	10.7	25.0	0.64	100.0	0.35	72.0	0.347	0.422	0.44	14.6	2.7	4.7	2.7	13.8	1.0	0.27	
32	Troy Brown Jr.	Chi	G-F	22.89	3.0	12.2	25.4	23.9	15.0	0.0	0.0	6.0	0.5	11.0	0.182	0.353	0.353	4.0	2.7	11.0	1.0	13.2	0.67	0.0	
33	Bruce Brown	Bro	G-F	25.84	4.0	34.7	72.2	16.7	17.9	10.0	0.8	23.0	0.652	14.0	0.429	0.649	0.676	14.0	4.8	9.0	2.8	11.9	1.25	0.75	
34	Jaylen Brown	Bos	G-F	25.64	24.0	38.3	79.8	26.9	13.4	131.0	0.763	258.0	0.531	161.0	0.373	0.542	0.581	23.1	6.9	10.1	3.5	16.8	1.17	0.42	
35	Sterling Brown	Dal	G-F	27.35	9.0	2.8	5.9	27.9	13.3	7.0	0.714	6.0	0.5	4.0	0.0	0.3	0.42	1.2	0.9	18.2	0.3	18.4	0.44	0.22	
36	Jalen Brunson	Dal	G	25.79	18.0	35.0	72.8	27.8	5.1	90.0	0.8	236.0	0.504	75.0	0.347	0.508	0.553	21.6	4.6	7.7	3.7	20.0	0.78	0.06	
37	Reggie Bullock	Dal	G-F	31.26	18.0	39.3	81.9	11.2	4.8	18.0	0.889	20.0	0.45	131.0	0.397	0.576	0.598	10.6	4.6	6.7	1.7	6.3	1.22	0.06	
38	Trey Burke	Dal	G	29.59	10.0	3.7	7.7	32.8	3.9	6.0	1.0	12.0	0.583	10.0	0.4	0.591	0.649	3.2	0.3	4.7	0.4	23.5	0.1	0.0	
39	Jared Butler	Uta	G	21.81	1.0	4.7	9.8	20.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	12.4	0.0	0.0	0.0	0.0	

Add instance | Undo | OK | Cancel

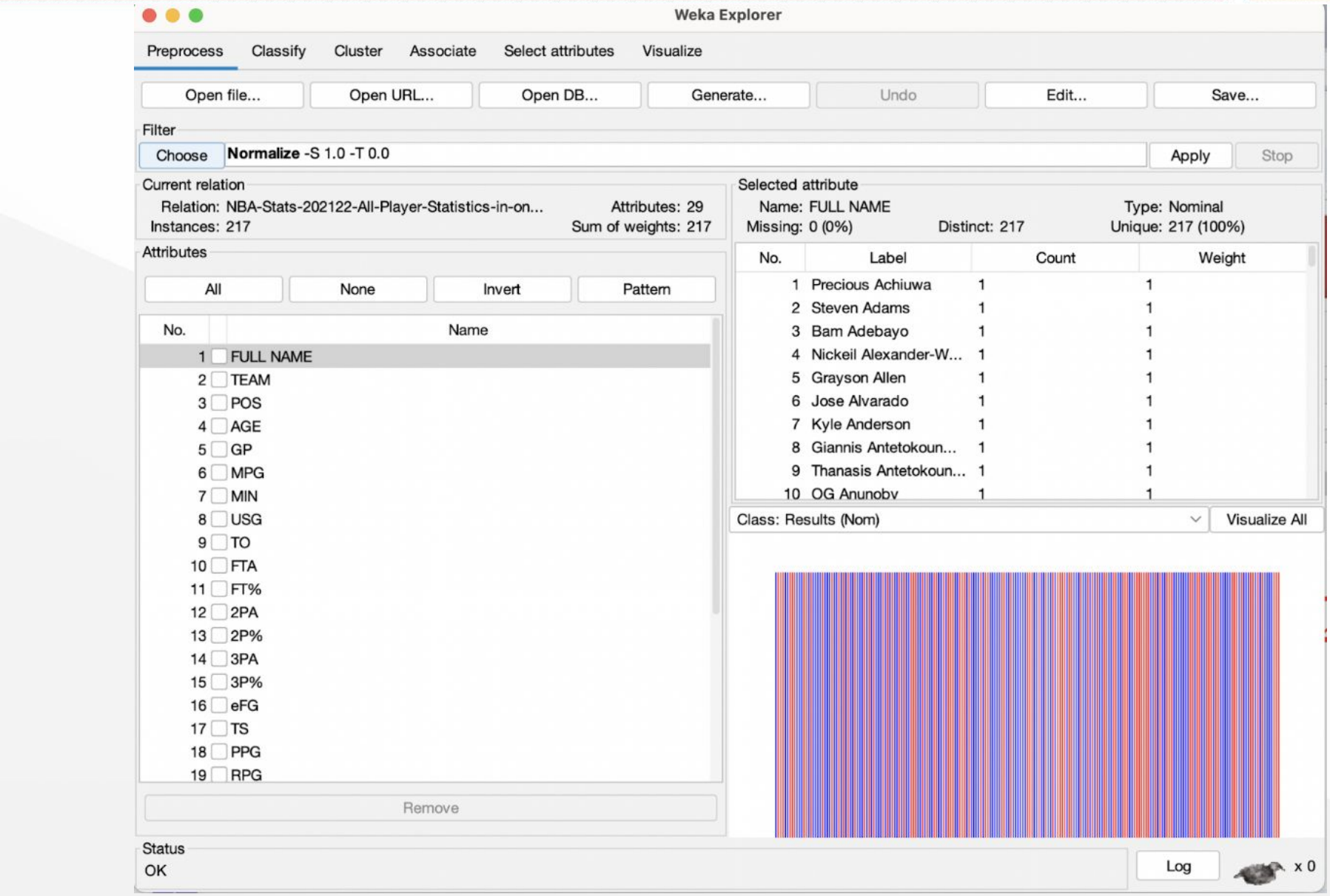
DataSet

- *The data set comes from 2021-2022 NBA play-offs.*
- *The data set have 29 different attributes and 217 instances.*

Visualize all attributes



Filter Choose



Classify - Which Algorithms?

1. *Naive Bayes*
2. *J48*
3. *Random Forest*
4. *SMO*
5. *K-nearest neighbor*
6. *Random Tree*

Naive Bayes

Weka Explorer

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

Classifier

Choose **NaiveBayes**

Test options

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

More options...

(Nom) Results

Start | Stop

Result list (right-click for options)

05:14:34 - bayes.NaiveBayes

Classifier output

```
  weight sum                    108        71
  precision                  6.5372    6.5372

DRTG
  mean                     104.7318  102.9194
  std. dev.                  5.3272     6.768
  weight sum                    108        72
  precision                  0.3446    0.3446



Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         174               80.1843 %
Incorrectly Classified Instances        43               19.8157 %
Kappa statistic                          0.6009
Mean absolute error                      0.2181
Root mean squared error                  0.4266
Relative absolute error                 43.9456 %
Root relative squared error             85.6421 %
Total Number of Instances              217

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.814    0.212    0.821      0.814    0.817      0.601    0.847     0.854     1
                 0.788    0.186    0.780      0.788    0.784      0.601    0.847     0.803     0
Weighted Avg.    0.802    0.200    0.802      0.802    0.802      0.601    0.847     0.831

=== Confusion Matrix ===

  a  b   <-- classified as
 96 22 |  a = 1
 21 78 |  b = 0
```

Status

OK

Log

x 0

J48

Weka Explorer

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

Classifier

Choose **J48** -C 0.25 -M 2

Test options

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

More options...

(Nom) Results

Start | Stop

Result list (right-click for options)

- 05:14:34 - bayes.NaiveBayes
- 05:15:08 - trees.J48

Classifier output

```
J48 pruned tree
------------------

3P% <= 0.3: 0 (99.0)
3P% > 0.3: 1 (118.0)

Number of Leaves  :     2

Size of the tree :      3


Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         217               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances              217

=== 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     1
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     0
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 118   0 |   a = 1
   0  99 |   b = 0
```

Status

OK

Log

x 0

RandomForest

Weka Explorer

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

Classifier

Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

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

More options...

(Nom) Results

Start | Stop

Result list (right-click for options)

Starts the classification

05:14:34 - bayes.NaiveBayes
05:15:08 - trees.J48
05:15:33 - trees.RandomForest

Classifier output

```
Test mode:    10-fold cross-validation

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.05 seconds

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

Correctly Classified Instances         216               99.5392 %
Incorrectly Classified Instances         1                0.4608 %
Kappa statistic                          0.9907
Mean absolute error                      0.2598
Root mean squared error                  0.2741
Relative absolute error                 52.3527 %
Root relative squared error             55.0236 %
Total Number of Instances              217

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.992     0.000     1.000      0.992     0.996      0.991    1.000     1.000     1
               1.000     0.008     0.990      1.000     0.995      0.991    1.000     1.000     0
Weighted Avg.  0.995     0.004     0.995      0.995     0.995      0.991    1.000     1.000

=== Confusion Matrix ===

   a   b   <-- classified as
 117   1 |   a = 1
   0  99 |   b = 0
```

Status

OK

Log

x 0

SMO

Weka Explorer

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

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functi

Test options

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

More options...

(Nom) Results

Start | Stop

Result list (right-click for options)

- 05:14:34 - bayes.NaiveBayes
- 05:15:08 - trees.J48
- 05:15:33 - trees.RandomForest
- 05:16:26 - functions.SMO

Classifier output

```
 +       0.463  * (normalized) TOPG
 +       0.4043 * (normalized) VI
 +      -0.1876 * (normalized) ORTG
 +       0.0757 * (normalized) DRTG
 +       1.69

Number of kernel evaluations: 19065 (94.207% cached)



Time taken to build model: 0.08 seconds

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

Correctly Classified Instances         191               88.0184 %
Incorrectly Classified Instances        26               11.9816 %
Kappa statistic                          0.7561
Mean absolute error                      0.1198
Root mean squared error                  0.3461
Relative absolute error                 24.1448 %
Root relative squared error             69.4909 %
Total Number of Instances              217

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.941    0.192    0.854      0.941    0.895      0.761    0.874     0.835     1
                 0.808    0.059    0.920      0.808    0.860      0.761    0.874     0.831     0
Weighted Avg.    0.880    0.131    0.884      0.880    0.879      0.761    0.874     0.833

=== Confusion Matrix ===

   a   b   <-- classified as
 111   7 |   a = 1
  19  80 |   b = 0
```

Status

OK

Log

x 0

K-Nearest Neighbors

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) Results

Start | Stop

Result list (right-click for options)

- 05:14:34 - bayes.NaiveBayes
- 05:15:08 - trees.J48
- 05:15:33 - trees.RandomForest
- 05:16:26 - functions.SMO
- 05:16:55 - lazy.IBk

Classifier output

```
               DRTG
               Results
Test mode:    10-fold cross-validation

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

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0 seconds

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

Correctly Classified Instances         139               64.0553 %
Incorrectly Classified Instances        78               35.9447 %
Kappa statistic                          0.2648
Mean absolute error                      0.3609
Root mean squared error                  0.5965
Relative absolute error                 72.7214 %
Root relative squared error            119.7544 %
Total Number of Instances              217

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.746    0.485    0.647      0.746    0.693      0.269    0.628     0.622     1
                 0.515    0.254    0.630      0.515    0.567      0.269    0.628     0.554     0
Weighted Avg.    0.641    0.380    0.639      0.641    0.635      0.269    0.628     0.591

=== Confusion Matrix ===

  a  b   <-- classified as
 88 30 |  a = 1
 48 51 |  b = 0
```

Status: OK

Log

x 0

Random Tree

Weka Explorer

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

Classifier

Choose **RandomTree** -K 0 -M 1.0 -V 0.001 -S 1

Test options

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

More options...

(Nom) Results

Start | Stop

Result list (right-click for options)

- 05:14:34 - bayes.NaiveBayes
- 05:15:08 - trees.J48
- 05:15:33 - trees.RandomForest
- 05:16:26 - functions.SMO
- 05:16:55 - lazy.IBk
- 05:17:15 - trees.RandomTree

Classifier output

```
FULL NAME = Robert Williams III : 0 (1/0)
FULL NAME = Grant Williams : 1 (1/0)
FULL NAME = Patrick Williams : 1 (1/0)
FULL NAME = Ziaire Williams : 1 (1/0)
FULL NAME = Delon Wright : 1 (1/0)
FULL NAME = Thaddeus Young : 0 (1/0)
FULL NAME = Trae Young : 0 (1/0)
FULL NAME = Omer Yurtseven : 0 (1/0)

Size of the tree : 218

Time taken to build model: 0 seconds

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

Correctly Classified Instances         176               81.106  %
Incorrectly Classified Instances        41               18.894  %
Kappa statistic                          0.6113
Mean absolute error                      0.2166
Root mean squared error                  0.3856
Relative absolute error                 43.6539 %
Root relative squared error             77.4035 %
Total Number of Instances              217

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.932    0.333    0.769      0.932    0.843      0.629    0.865     0.854     1
                 0.667    0.068    0.892      0.667    0.763      0.629    0.864     0.813     0
Weighted Avg.    0.811    0.212    0.825      0.811    0.806      0.629    0.865     0.835

=== Confusion Matrix ===

   a   b   <-- classified as
 110   8 |   a = 1
  33  66 |   b = 0
```

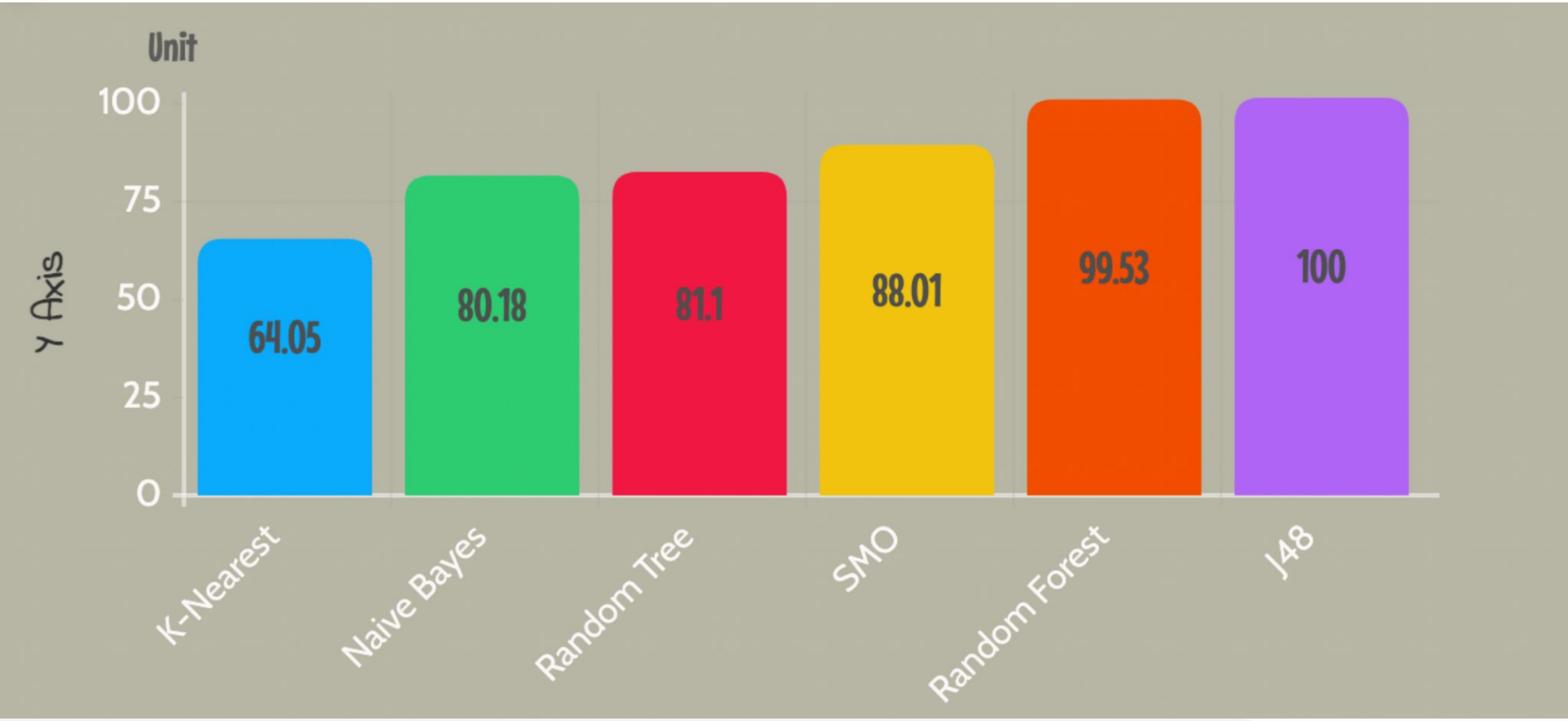
Status

OK

Log

x 0

Compare Classify Algorithms



- *J48 and Random Forest algorithms accuracy rates are perfect.*

Cluster - Which Algorithms?

1. *Simple k-means*
2. *Make Density Based Clusterer*
3. *Farthest First*
4. *Filtered Clusterer*

Simple K-means

Weka Explorer

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

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 5

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation
  (Nom) Results
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

05:29:15 - SimpleKMeans

Clusterer output

```
Missing values globally replaced with mean/mode

Final cluster centroids:
                                                Cluster#
Attribute                          Full Data           0           1
                                     (217.0)     (124.0)      (93.0)
=================================================================================
FULL NAME                  Precious Achiuwa Precious Achiuwa     Steven Adams
TEAM                                    Tor             Gol              Atl
POS                                       G               G                G
AGE                                 27.4306         27.8165          26.9161
GP                                   8.7143          10.379           6.4946
MPG                                 19.4304         25.1508          11.8032
MIN                                 40.4733         52.3839          24.5925
USG                                  18.935         19.2194          18.5559
TO                                   12.7868         12.2323          13.5262
FTA                                  18.1014         25.9516           7.6344
FT%                                   0.6232          0.7213           0.4925
2PA                                  39.4147         56.1371          17.1183
2P%                                    0.495          0.5059           0.4805
3PA                                  27.7281         41.7339           9.0538
3P%                                   0.2733          0.4007           0.1035
eFG                                   0.5212          0.5665           0.4609
TS                                    0.5538          0.6001           0.4921
PPG                                   8.4571         11.7339           4.0882
RPG                                   3.4037          3.9726           2.6452
TRB                                  10.7829          9.2194          12.8677
APGA                                  1.8286           2.529           0.8946
AST                                  14.0111         15.3694             12.2
SPG                                   0.5788          0.7297           0.3777
BPG                                   0.3585          0.4381           0.2524
TOPG                                  1.0796          1.4102           0.6388
VI                                    6.0396          6.8081           5.0151
ORTG                                117.5078        116.5578         118.7745
DRTG                                103.9956        104.6311         103.1481
Results                                    1               1                0
```

Status

OK

Log | x 0

Make Density Based Clusterer

Weka Explorer

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

Clusterer

Choose **MakeDensityBasedClusterer** -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation

(Nom) Results

[x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

05:29:15 - SimpleKMeans
05:30:15 - MakeDensityBasedClusterer

Clusterer output

```
Normal Distribution. Mean = 4.0882 StdDev = 4.0167
Attribute: RPG
Normal Distribution. Mean = 2.6452 StdDev = 2.7641
Attribute: TRB
Normal Distribution. Mean = 12.8677 StdDev = 8.6035
Attribute: APGA
Normal Distribution. Mean = 0.8946 StdDev = 1.1594
Attribute: AST
Normal Distribution. Mean = 12.2 StdDev = 15.608
Attribute: SPG
Normal Distribution. Mean = 0.3777 StdDev = 0.4117
Attribute: BPG
Normal Distribution. Mean = 0.2524 StdDev = 0.4323
Attribute: TOPG
Normal Distribution. Mean = 0.6388 StdDev = 0.8072
Attribute: VI
Normal Distribution. Mean = 5.0151 StdDev = 3.8952
Attribute: ORTG
Normal Distribution. Mean = 118.7745 StdDev = 98.1812
Attribute: DRTG
Normal Distribution. Mean = 103.1481 StdDev = 5.9386
Attribute: Results
Discrete Estimator. Counts =  4 91  (Total = 95)


Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      122 ( 56%)
1       95 ( 44%)


Log likelihood: -71.88017
```

Status

OK

Log | x 0

Farthest First

Weka Explorer

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

Clusterer

Choose **FarthestFirst** -N 2 -S 1

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation

(Nom) Results

[x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

- 05:29:15 - SimpleKMeans
- 05:30:15 - MakeDensityBasedClusterer
- 05:31:15 - FarthestFirst

Clusterer output

```
                TRB
                APGA
                AST
                SPG
                BPG
                TOPG
                VI
                ORTG
                DRTG
                Results
Test mode:    evaluate on training data


=== Clustering model (full training set) ===


FarthestFirst
==============

Cluster centroids:

Cluster 0
        Khris Middleton Mil F 30.85 2.0 35.8 74.5 22.0 30.3 3.0 1.0 10.0 0.4 14.0 0.429 0.5
Cluster 1
        Willy Hernangomez Nor C-F 28.06 1.0 1.9 4.0 90.6 0.0 0.0 0.0 3.0 0.333 1.0 0.0 0.25



Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      139 ( 64%)
1       78 ( 36%)
```

Status

OK

Log

x 0

Filtered Clusterer

Weka Explorer

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

Clusterer

Choose **FilteredClusterer** -F "weka.filters.AllFilter " -W weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation
  (Nom) Results
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

- 05:29:15 - SimpleKMeans
- 05:30:15 - MakeDensityBasedClusterer
- 05:31:15 - FarthestFirst
- 05:31:45 - FilteredClusterer

Clusterer output

```
Missing values globally replaced with mean/mode

Final cluster centroids:
                                                Cluster#
Attribute                          Full Data           0           1
                                     (217.0)     (124.0)      (93.0)
=============================================================================
FULL NAME                  Precious Achiuwa Precious Achiuwa Steven Adams
TEAM                                    Tor             Gol         Atl
POS                                       G               G           G
AGE                                 27.4306         27.8165     26.9161
GP                                   8.7143          10.379      6.4946
MPG                                 19.4304         25.1508     11.8032
MIN                                 40.4733         52.3839     24.5925
USG                                  18.935         19.2194     18.5559
TO                                  12.7868         12.2323     13.5262
FTA                                 18.1014         25.9516      7.6344
FT%                                  0.6232          0.7213      0.4925
2PA                                 39.4147         56.1371     17.1183
2P%                                   0.495          0.5059      0.4805
3PA                                 27.7281         41.7339      9.0538
3P%                                  0.2733          0.4007      0.1035
eFG                                  0.5212          0.5665      0.4609
TS                                   0.5538          0.6001      0.4921
PPG                                  8.4571         11.7339      4.0882
RPG                                  3.4037          3.9726      2.6452
TRB                                 10.7829          9.2194     12.8677
APGA                                 1.8286           2.529      0.8946
AST                                 14.0111         15.3694        12.2
SPG                                  0.5788          0.7297      0.3777
BPG                                  0.3585          0.4381      0.2524
TOPG                                 1.0796          1.4102      0.6388
VI                                   6.0396          6.8081      5.0151
ORTG                                117.5078        116.5578    118.7745
DRTG                                103.9956        104.6311    103.1481
Results                                    1               1           0
```

Status

OK

Log

x 0

Compare Clusters

- *The algorithms could not cluster correctly.*
- *Simple k-means clustering and Filtered Clusterer results were same.*
- *The most worst clustering algorithm is Farthest First.*
- *The best clustering algorithm was make density based clustering but this algorithm good with poor results.*

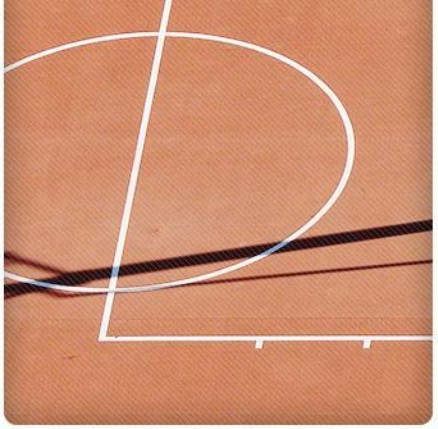
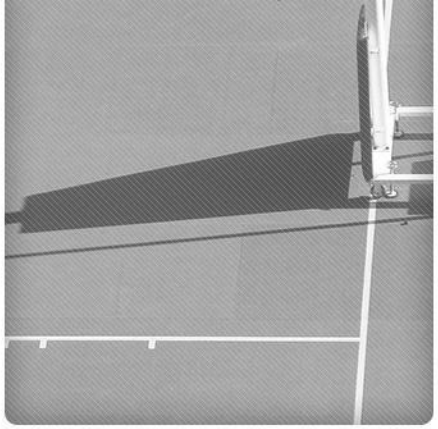
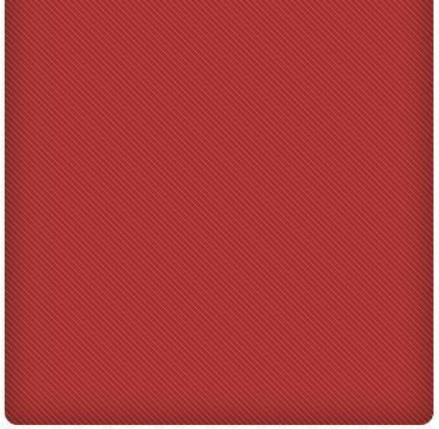
References

1. *Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2017). The WEKA Machine Learning Workbench. Online appendix for "Data Mining: Practical Machine Learning Tools and Techniques". Fifth Edition, Morgan Kaufmann, 2017. Available at http://www.cs.waikato.ac.nz/ml/weka/book.html.*

2. *Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter, 11(1), 10-18.*

3. *[BilgisayarKavramları]. (2015, May 10). WEKA Eğitim Serisi [Video]. Youtube.com. https://www.youtube.com/watch?v=5s8IgMfH698&list=PLh9ECzBB8tJP8vpIM91W1k_qoGI10ikh1*

4. *NBA (n.d.). NBA Advanced Stats. https://www.nba.com/stats*

THANK YOU