Višeklasna klasifikacija bez benignih podataka

Current relation

Relation: ALL_CICFLOW_TRAIN_MULTI_NO_BEN Attributes: 79

Instances: 21996 Sum of weights: 21996

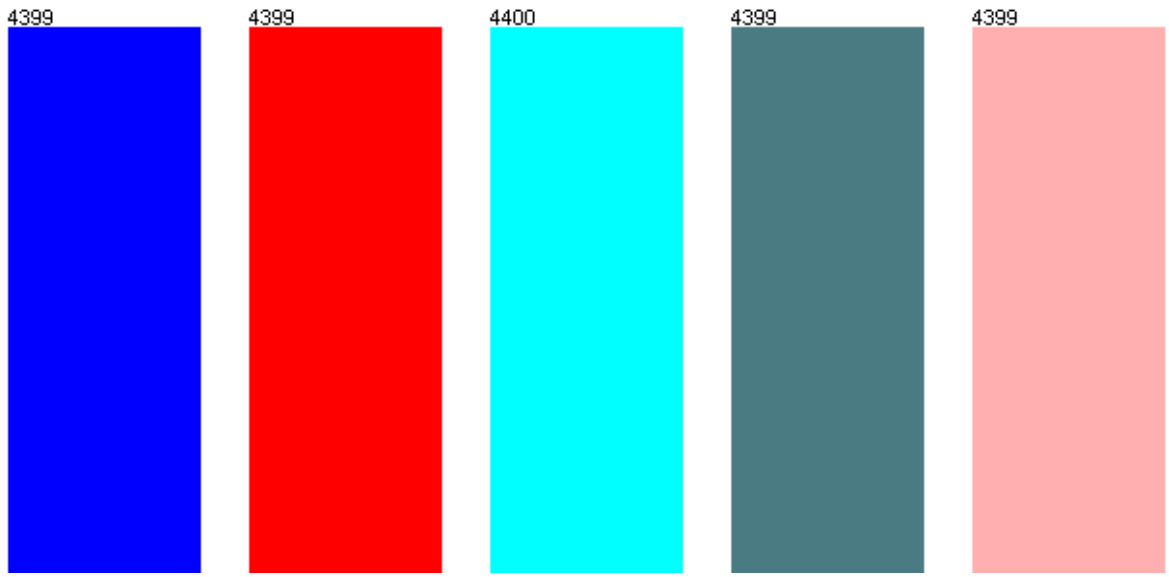
Selected attribute

Name: classification Type: Nominal

Missing: 0 (0%) Distinct: 5 Unique: 0 (0%)

No.	Label	Count	Weight
1	DoS Hulk	4399	4399
2	FTP-Patator	4399	4399
3	DoS Slowhttptest	4400	4400
4	SSH-Patator	4399	4399
5	PortScan	4399	4399

Class: classification (Nom) Visualize All



Kod klasifikacije ove vrste podataka problem je sličnost paketa po parametrima. Klasa zavisi od gotovo svih atributa i njihove međusobne povezanosti.

Naive Bayes

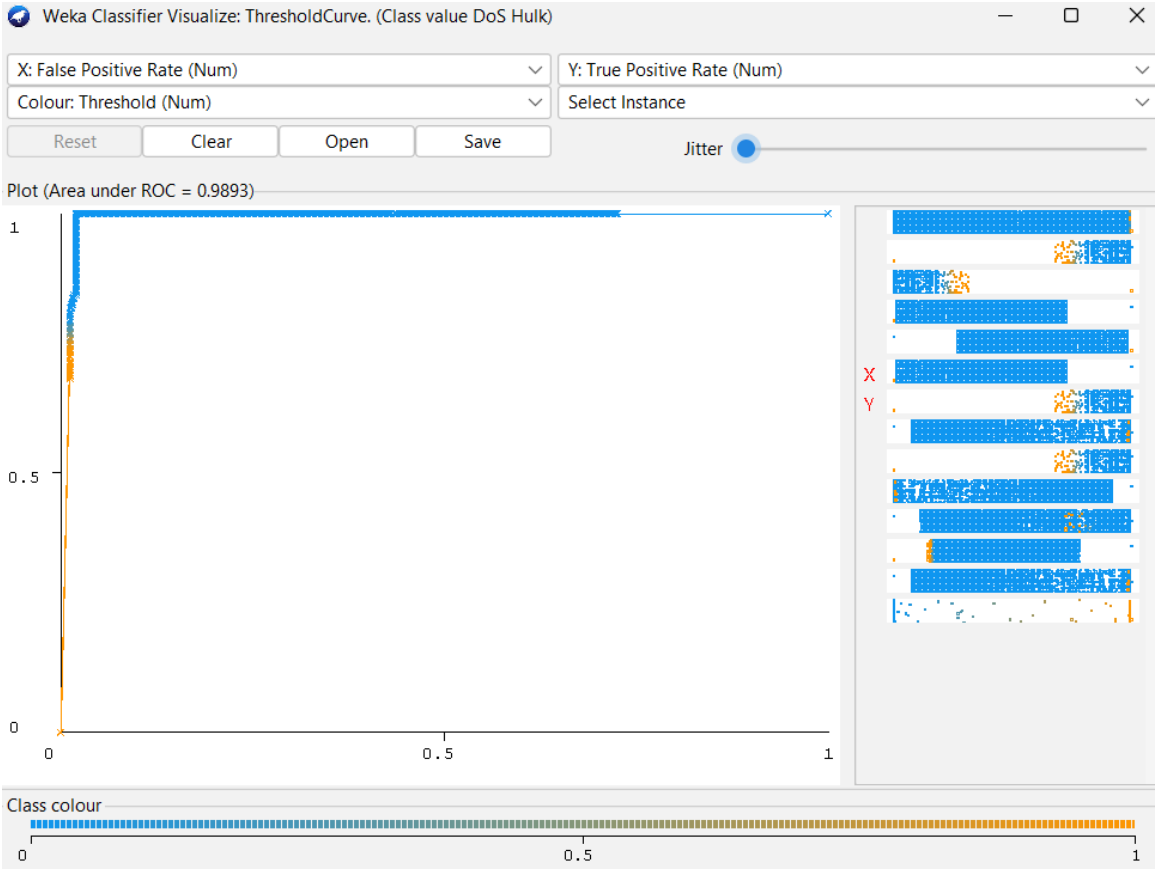
Bazira se na primeni Bayes-ove teoreme. Jednostavna statistička tehnika koja određuje verovatnoću da nepoznati slog pripada određenoj klasi. Naïve Bayes klasifikator uvodi pretpostavku da je efekat vrednosti jednog atributa na pripadnost klase NEZAVISTAN u odnosu na vrednosti ostalih atributa. Zbog brzine može se koristiti za real-time klasifikaciju podataka.

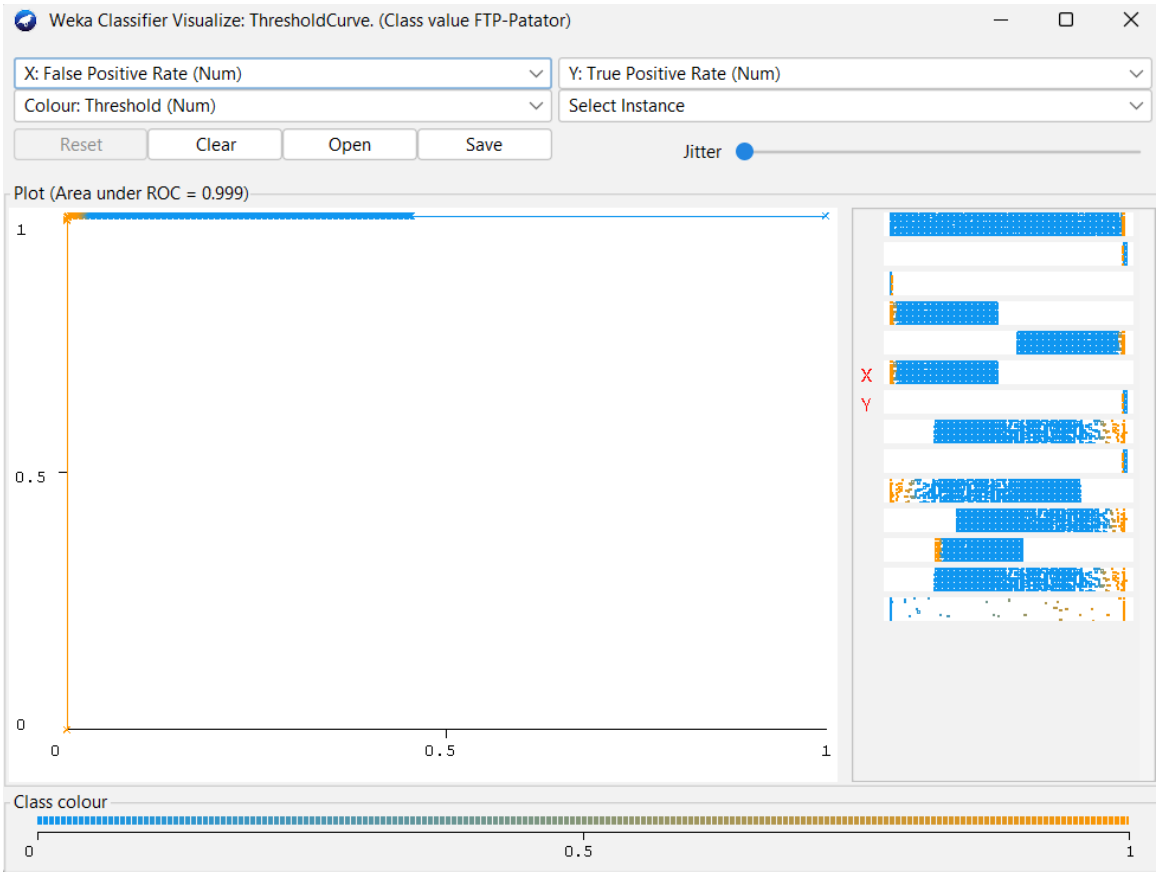
Cross-validation

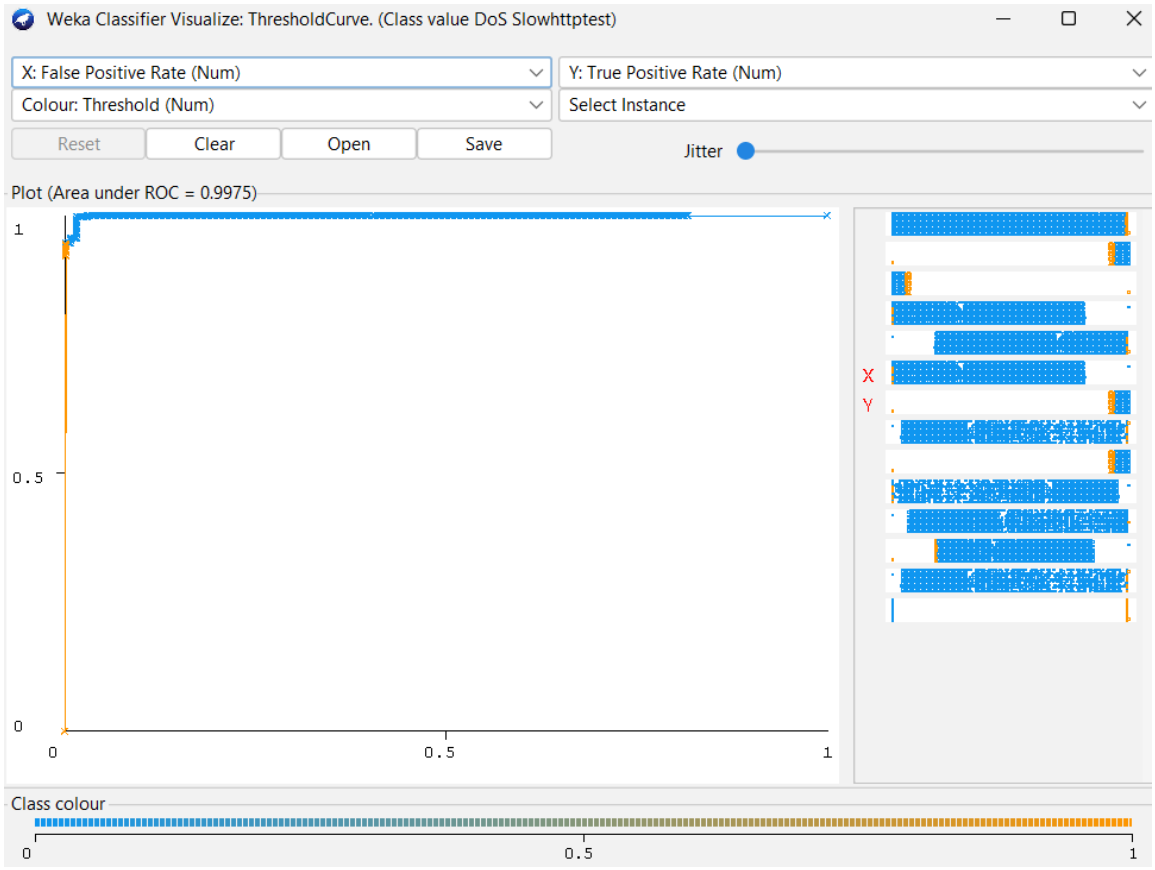
Classifier output

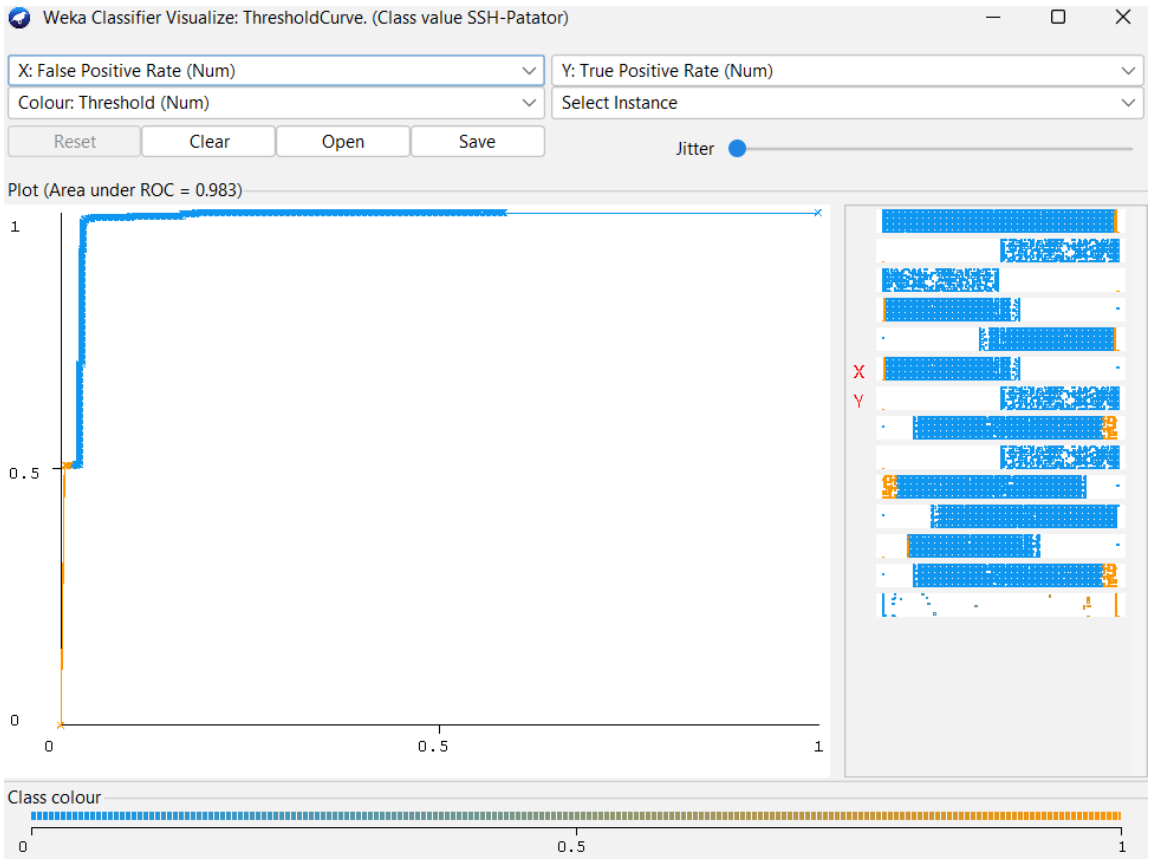
```
_Idle_Min
  mean                              1.1463          -0.4577           0.224          -0.4577          -0.4556
  std. dev.                         1.3068           0.0005          1.1392           0.0005           0.0664
  weight sum                          4399             4399            4400             4399             4399
  precision                         0.0031           0.0031          0.0031           0.0031           0.0031



Time taken to build model: 0.18 seconds

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

Correctly Classified Instances       18467               83.9562 %
Incorrectly Classified Instances      3529               16.0438 %
Kappa statistic                          0.7995
Mean absolute error                      0.0637
Root mean squared error                  0.2508
Relative absolute error                 19.9176 %
Root relative squared error             62.6945 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.759    0.013    0.935      0.759    0.838      0.809    0.989     0.927     DoS Hulk
                 0.997    0.022    0.917      0.997    0.956      0.945    0.999     0.995     FTP-Patator
                 0.947    0.004    0.985      0.947    0.966      0.958    0.998     0.984     DoS Slowhttptest
                 0.506    0.014    0.898      0.506    0.648      0.622    0.983     0.905     SSH-Patator
                 0.988    0.147    0.627      0.988    0.767      0.724    0.982     0.891     PortScan
Weighted Avg.    0.840    0.040    0.873      0.840    0.835      0.812    0.990     0.940

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 3341  370    8  244  436 |    a = DoS Hulk
    0 4387    9    2    1 |    b = FTP-Patator
  230    0 4166    2    2 |    c = DoS Slowhttptest
    0   14    7 2228 2150 |    d = SSH-Patator
    1   11   38    4 4345 |    e = PortScan
```

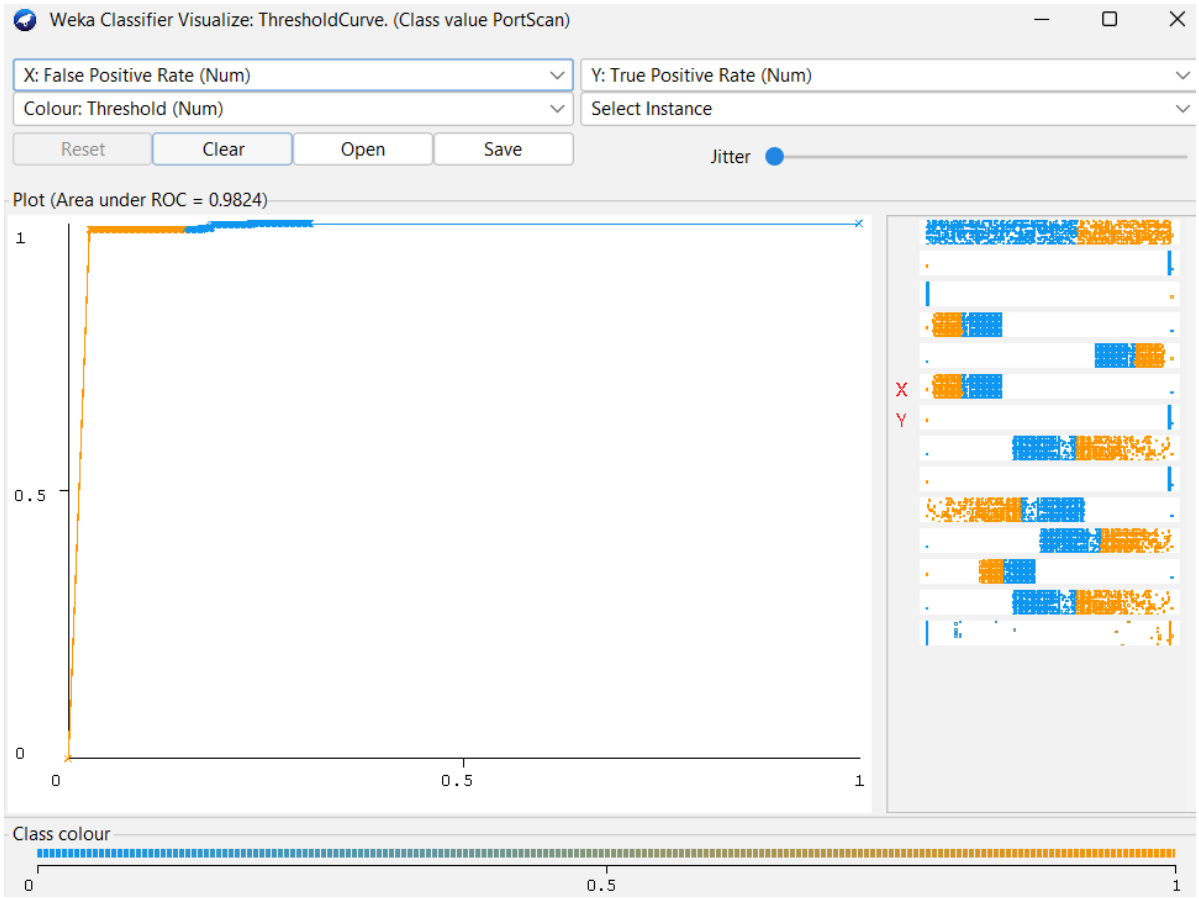
Algoritam je sa najvećom preciznošću klasifikovao DoS Slowhttptest napad, a sa najmanjom PortScan. Algoritam je dao ukupno 83.9562% tačnih klasifikacija kada je u pitanju cross-validation i 83.6152% kada je u pitanju testiranje nad test setom podataka koji je kreiran na isti način kao i trening set. ROC area je prilično zadovoljavajuća kada je u pitanju ova vrsta podataka za klasifikaciju.


Weka Classifier Visualize: ThresholdCurve. (Class value DoS Hulk)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.9893)
1
0.5
0
0
0.5
1
X
Y
Class colour
0
0.5
1


Weka Classifier Visualize: ThresholdCurve. (Class value FTP-Patator)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.999)
1
0.5
0
0
0.5
1
X
Y
Class colour
0
0.5
1


Weka Classifier Visualize: ThresholdCurve. (Class value DoS Slowhttptest)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.9975)
1
0.5
0
0
0.5
1
X
Y
Class colour
0
0.5
1


Weka Classifier Visualize: ThresholdCurve. (Class value SSH-Patator)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.983)
1
0.5
0
0
0.5
1
X
Y
Class colour
0
0.5
1


Weka Classifier Visualize: ThresholdCurve. (Class value PortScan)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.9824)
1
0.5
0
0
0.5
1
X
Y
Class colour
0
0.5
1

Test set

```
Correctly Classified Instances        4598               83.6152 %
Incorrectly Classified Instances       901               16.3848 %
Kappa statistic                          0.7952
Mean absolute error                      0.0656
Root mean squared error                  0.2546
Relative absolute error                 20.4902 %
Root relative squared error             63.6386 %
Total Number of Instances             5499

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.740    0.011    0.943      0.740    0.829      0.802    0.989     0.932     DoS Hulk
                 0.998    0.035    0.876      0.998    0.933      0.918    0.999     0.996     FTP-Patator
                 0.955    0.005    0.979      0.955    0.967      0.959    0.997     0.979     DoS Slowhttptest
                 0.505    0.020    0.866      0.505    0.638      0.605    0.983     0.909     SSH-Patator
                 0.982    0.134    0.647      0.982    0.780      0.738    0.989     0.943     PortScan
Weighted Avg.    0.836    0.041    0.862      0.836    0.830      0.804    0.991     0.952

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
  814  149    2   86   49 |    a = DoS Hulk
    0 1098    2    0    0 |    b = FTP-Patator
   48    0 1050    0    1 |    c = DoS Slowhttptest
    1    3    2  556  538 |    d = SSH-Patator
    0    3   17    0 1080 |    e = PortScan
```

Random Forest

Random Forest (Slučajna šuma) je algoritam mašinskog učenja koji se zasniva na konceptu ansambl metoda, specifično ansambl stabala odlučivanja. Ansambl metode kombinuju više modela kako bi postigli bolje performanse u poređenju sa pojedinačnim modelima. Kod Random Forest-a, ovi modeli su stabla odlučivanja koja se grade na osnovu nasumično izabranih podskupova podataka (bagging) i nasumično izabranih atributa. Svako stablo radi nezavisno od ostalih, a konačna klasifikacija (ili regresija) se postiže kombinovanjem rezultata svih stabala putem glasanja ili proseka. Ovaj pristup ne samo da smanjuje varijansu modela, čime se smanjuje rizik od overfittinga, već takođe omogućava modelu da efikasno radi sa složenim podacima i velikim brojem atributa, čineći ga popularnim izborom za različite zadatke analize podataka.

Cross-validation

Classifier output

```
=== 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: 3.99 seconds

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

Correctly Classified Instances       21982               99.9364 %
Incorrectly Classified Instances        14                0.0636 %
Kappa statistic                          0.9992
Mean absolute error                      0.0007
Root mean squared error                  0.0152
Relative absolute error                  0.2202 %
Root relative squared error              3.7934 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    0.999      1.000    0.999      0.999    1.000     1.000     DoS Hulk
                 1.000    0.000    0.999      1.000    0.999      0.999    1.000     1.000     FTP-Patator
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     1.000     DoS Slowhttptest
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     1.000     SSH-Patator
                 0.999    0.000    1.000      0.999    1.000      1.000    1.000     0.999     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     1.000

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4397    0    2    0    0 |    a = DoS Hulk
    0 4397    1    1    0 |    b = FTP-Patator
    2    0 4396    2    0 |    c = DoS Slowhttptest
    0    3    0 4396    0 |    d = SSH-Patator
    3    0    0    0 4396 |    e = PortScan
```

Test set

Classifier output

```
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: 3.59 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.33 seconds

=== Summary ===

Correctly Classified Instances        5499               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0.0006
Root mean squared error                  0.011
Relative absolute error                  0.1934 %
Root relative squared error              2.743  %
Total Number of Instances             5499

=== 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     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     DoS Slowhttptest
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SSH-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PortScan
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1100    0    0    0    0 |    a = DoS Hulk
    0 1100    0    0    0 |    b = FTP-Patator
    0    0 1099    0    0 |    c = DoS Slowhttptest
    0    0    0 1100    0 |    d = SSH-Patator
    0    0    0    0 1100 |    e = PortScan
```

Ovaj algoritam je od svih imao najveću preciznost i kompletnost 100% nad test setom i 99% kada je u pitanju cross-validation. Nad test setom nema promašaja pri klasifikaciji, dok kod cross-validation imamo 0.0636% što predstavlja odlične rezultate. Rizik od overfiting-a je smanjen zbog prirode algoritma. Algoritam je ravnomerno imao 2 do 3 promašaja po klasi što je potpuno prihvatljivo za ovu vrstu podataka. Za razliku od prethodnog, ovaj je PortScan napade klasifikovao sa stopostotnom preciznošću.

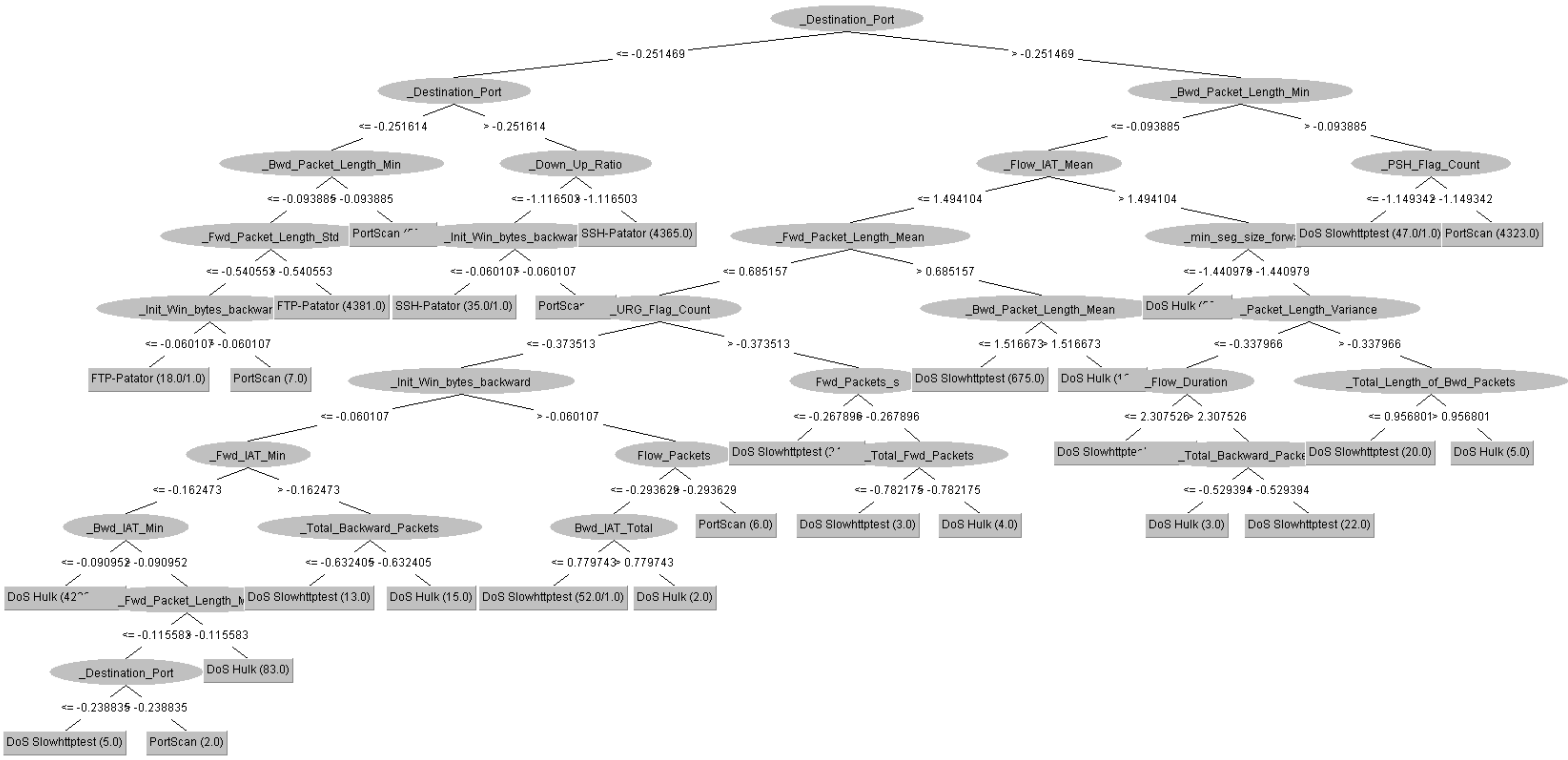
Tree J48

Tree J48 je algoritam za izgradnju stabla odlučivanja koji se temelji na C4.5 algoritmu. Ovaj algoritam koristi rekurzivni proces razdvajanja podataka na osnovu atributa kako bi se konstruisalo stablo odlučivanja. Počevši od korenskog čvora, algoritam bira najbolji atribut za razdvajanje podataka, koristeći metrike poput informacionog dobitka ili Gini indeksa. Svaki čvor se rekurzivno deli na manje podskupove, formirajući stablo dok se ne ispuni određeni kriterijum zaustavljanja. Nakon izgradnje, stablo se može orezati radi smanjenja složenosti i poboljšanja generalizacije. Tree J48 je popularan zbog svoje sposobnosti da radi sa različitim vrstama podataka i generiše interpretabilna stabla odlučivanja koja se često koriste u analizi podataka i mašinskom učenju.

Cross-validation

Classifier output

```
|   _Bwd_Packet_Length_Min > -0.093885
|   |   _PSH_Flag_Count <= -1.149342: DoS Slowhttptest (47.0/1.0)
|   |   _PSH_Flag_Count > -1.149342: PortScan (4323.0)

Number of Leaves  :     29

Size of the tree :      57


Time taken to build model: 1.07 seconds

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

Correctly Classified Instances       21977               99.9136 %
Incorrectly Classified Instances        19                0.0864 %
Kappa statistic                          0.9989
Mean absolute error                      0.0005
Root mean squared error                  0.0186
Relative absolute error                  0.161  %
Root relative squared error              4.6493 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.000    0.998      0.999    0.998      0.998    0.999     0.996     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 0.999    0.001    0.998      0.999    0.998      0.998    0.999     0.998     DoS Slowhttptest
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SSH-Patator
                 0.998    0.000    1.000      0.998    0.999      0.999    1.000     0.999     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4393    0    6    0    0 |    a = DoS Hulk
    0 4398    1    0    0 |    b = FTP-Patator
    5    0 4395    0    0 |    c = DoS Slowhttptest
    0    0    0 4399    0 |    d = SSH-Patator
    3    1    2    1 4392 |    e = PortScan
```



Izgled stabla odluke (zbog brojnosti atributa nije moglo stati celo na ekran).

Kada je u pitanju J48 algoritam, on daje nešto lošije rezultate u odnosu na Random Forest, a dosta bolje u odnosu na Naive Bayes algoritam. 99.9136% pogodaka pri cross-validation i 99.9273% pri testiranju na test setu podataka nam govori da je i ovo jedan od boljih algoritama za klasifikaciju ove vrste podataka. Najbolje klasifikovani napadi su FTP-Patator, SSH-Patator i PortScan sa 100% preciznošću. Kreiranje ovog modela trajalo je 3 puta kraće u odnosu na Random Forest zbog toga što se kreira jsamo jedno stablo odluke i nema semplovanja instanci.

Test set

Classifier output

```
Number of Leaves  :     29

Size of the tree :      57


Time taken to build model: 0.72 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.26 seconds

=== Summary ===

Correctly Classified Instances        5495               99.9273 %
Incorrectly Classified Instances         4                0.0727 %
Kappa statistic                          0.9991
Mean absolute error                      0.0004
Root mean squared error                  0.017
Relative absolute error                  0.1392 %
Root relative squared error              4.2403 %
Total Number of Instances             5499

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.996    0.000    1.000      0.996    0.998      0.998    0.999     0.999     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 1.000    0.001    0.996      1.000    0.998      0.998    1.000     0.998     DoS Slowhttptest
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SSH-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1096    0    4    0    0 |    a = DoS Hulk
    0 1100    0    0    0 |    b = FTP-Patator
    0    0 1099    0    0 |    c = DoS Slowhttptest
    0    0    0 1100    0 |    d = SSH-Patator
    0    0    0    0 1100 |    e = PortScan
```

Unpruned =TRUE

Pruning se obično koristi kako bi se smanjila složenost stabla i poboljšala generalizacija modela, sprečavajući overfitting. Nešto manji broj tačno klasifikovanih instanci to pokazuje.

Classifier output

```
|   _Bwd_Packet_Length_Min > -0.093885
|   |   _PSH_Flag_Count <= -1.149342: DoS Slowhttptest (47.0/1.0)
|   |   _PSH_Flag_Count > -1.149342: PortScan (4323.0)

Number of Leaves  :     32

Size of the tree :      63


Time taken to build model: 1.02 seconds

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

Correctly Classified Instances       21975               99.9045 %
Incorrectly Classified Instances        21                0.0955 %
Kappa statistic                          0.9988
Mean absolute error                      0.0005
Root mean squared error                  0.0192
Relative absolute error                  0.1636 %
Root relative squared error              4.7999 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.998    0.000    0.998      0.998    0.998      0.998    0.999     0.997     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 0.999    0.001    0.998      0.999    0.998      0.998    0.999     0.998     DoS Slowhttptest
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SSH-Patator
                 0.998    0.000    1.000      0.998    0.999      0.999    1.000     0.999     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4391    0    8    0    0 |    a = DoS Hulk
    0 4398    1    0    0 |    b = FTP-Patator
    5    0 4395    0    0 |    c = DoS Slowhttptest
    0    0    0 4399    0 |    d = SSH-Patator
    3    1    2    1 4392 |    e = PortScan
```

Classifier output

```
Number of Leaves  :     32

Size of the tree :      63


Time taken to build model: 0.92 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.25 seconds

=== Summary ===

Correctly Classified Instances        5495               99.9273 %
Incorrectly Classified Instances         4                0.0727 %
Kappa statistic                          0.9991
Mean absolute error                      0.0004
Root mean squared error                  0.017
Relative absolute error                  0.128  %
Root relative squared error              4.24   %
Total Number of Instances             5499

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.996    0.000    1.000      0.996    0.998      0.998    0.999     0.999     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 1.000    0.001    0.996      1.000    0.998      0.998    1.000     0.998     DoS Slowhttptest
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     SSH-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1096    0    4    0    0 |    a = DoS Hulk
    0 1100    0    0    0 |    b = FTP-Patator
    0    0 1099    0    0 |    c = DoS Slowhttptest
    0    0    0 1100    0 |    d = SSH-Patator
    0    0    0    0 1100 |    e = PortScan
```

Ovde je preciznost svih klasa 100% osim DoS Slowhttptest-a koji je u 4 slučaja bio klasifikovan kao DoS Hulk napad kada je u pitanju testiranje na test setu. Kod cross-validation-a najviše ima upravo tih promašaja ali se i DoS Hulk negde meša sa PortScan napadom i FTP-Partatorom.

OneR algoritam

OneR je jednostavan klasifikator koji bira jedan atribut i za njega definiše pravila na osnovu kojih se radi klasifikacija.

Test set

Classifier output

```
Test mode:    user supplied test set:  size unknown (reading incrementally)

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

_Fwd_Header_Length:
        < -0.8190870005420706   -> PortScan
        < -0.7789051948693494   -> SSH-Patator
        < -0.7186324863602676   -> PortScan
        < -0.6483143264330056   -> DoS Hulk
        < -0.6181779721784646   -> FTP-Patator
        < -0.5779961665057434   -> SSH-Patator
        < -0.3670416867239571   -> DoS Hulk
        < -0.3469507838875965   -> DoS Slowhttptest
        < -0.296723526796695    -> DoS Hulk
        < -0.2665871725421541   -> DoS Slowhttptest
        < -0.1259508526876299   -> DoS Hulk
        < -0.0656781441785481   -> DoS Slowhttptest
        < 0.1151399813486972    -> DoS Hulk
        < 0.1352308841850578    -> DoS Slowhttptest
        < 0.3160490097123032    -> DoS Hulk
        < 0.3361399125486638    -> DoS Slowhttptest
        < 0.4265489753122865    -> DoS Hulk
        < 0.4566853295668274    -> DoS Slowhttptest
        < 0.5169580380759091    -> DoS Hulk
        < 0.5370489409122698    -> FTP-Patator
        < 1.1397760260030876    -> DoS Hulk
        < 1.9534575908756922    -> DoS Slowhttptest
        < 2.154366619239298     -> SSH-Patator
        < 2.2347302305847405    -> DoS Slowhttptest
        < 2.3150938419301825    -> SSH-Patator
        < 2.4155483561119855    -> DoS Slowhttptest
        >= 2.4155483561119855   -> SSH-Patator
(19249/21996 instances correct)
```

```
Correctly Classified Instances        4827               87.7796 %
Incorrectly Classified Instances       672               12.2204 %
Kappa statistic                          0.8472
Mean absolute error                      0.0489
Root mean squared error                  0.2211
Relative absolute error                 15.2755 %
Root relative squared error             55.273  %
Total Number of Instances             5499

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.681    0.013    0.932      0.681    0.787      0.757    0.834     0.698     DoS Hulk
                 0.996    0.101    0.711      0.996    0.830      0.797    0.947     0.709     FTP-Patator
                 0.736    0.003    0.982      0.736    0.841      0.821    0.866     0.775     DoS Slowhttptest
                 0.985    0.028    0.897      0.985    0.938      0.924    0.978     0.886     SSH-Patator
                 0.991    0.007    0.972      0.991    0.982      0.977    0.992     0.965     PortScan
Weighted Avg.    0.878    0.031    0.899      0.878    0.876      0.855    0.924     0.807

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
  749  310    9    1   31 |    a = DoS Hulk
    3 1096    0    1    0 |    b = FTP-Patator
   39  128  809  123    0 |    c = DoS Slowhttptest
    4    8    5 1083    0 |    d = SSH-Patator
    9    0    1    0 1090 |    e = PortScan
```

Kod ovog algoritma imamo nešto slabije performanse. Tačno klasifikovanih instanci je 87.7796% na test setu, a 87.4886% na cross-validaciji. To je bolji rezultat u odnosu na Naive Bayes, a dosta lošiji u odnosu na Tree algoritme. Sa najvećom preciznošću klasifikovani su DoS Slowhttptest napadi (0,982 i 0,976), a sa najmanjom FTP-Partator napadi (0,711 i 0,710). Najviše je bilo mešanja sa DoS Hulk i DoS Slowhttptest napadima. Dok je takođe i SSH-Partator značajan broj puta bio klasifikovan kao DoS Slowhttptest napad.

Cross-validation

```
Correctly Classified Instances       19244               87.4886 %
Incorrectly Classified Instances      2752               12.5114 %
Kappa statistic                          0.8436
Mean absolute error                      0.05
Root mean squared error                  0.2237
Relative absolute error                 15.6392 %
Root relative squared error             55.9271 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.675    0.012    0.933      0.675    0.783      0.754    0.832     0.695     DoS Hulk
                 0.997    0.102    0.710      0.997    0.829      0.796    0.948     0.708     FTP-Patator
                 0.719    0.004    0.976      0.719    0.828      0.806    0.857     0.758     DoS Slowhttptest
                 0.990    0.031    0.887      0.990    0.936      0.921    0.979     0.880     SSH-Patator
                 0.994    0.006    0.975      0.994    0.984      0.980    0.994     0.970     PortScan
Weighted Avg.    0.875    0.031    0.896      0.875    0.872      0.852    0.922     0.802

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 2970 1251   64    3  111 |    a = DoS Hulk
    7 4386    0    6    0 |    b = FTP-Patator
  171  521 3164  544    0 |    c = DoS Slowhttptest
   12   22   10 4353    2 |    d = SSH-Patator
   23    1    4    0 4371 |    e = PortScan
```

K Nearest Neighbors (kNN) -IBk

Klasifikator koji se bazira na učenju po analogiji. Nepoznati slog se poredi sa trening slogovima koji su mu slični. Trening slogovi su opisani sa n atributa i svaki slog se može predstaviti kao tačka u n-dimenzionalnom prostoru. K je broj suseda iz trening seta.

Cross-validation KNN=1

```
Correctly Classified Instances       21976               99.9091 %
Incorrectly Classified Instances        20                0.0909 %
Kappa statistic                          0.9989
Mean absolute error                      0.0004
Root mean squared error                  0.0191
Relative absolute error                  0.1344 %
Root relative squared error              4.7673 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    0.998      1.000    0.999      0.999    1.000     0.998     DoS Hulk
                 0.999    0.000    1.000      0.999    0.999      0.999    0.999     0.999     FTP-Patator
                 0.999    0.000    1.000      0.999    0.999      0.999    0.999     0.999     DoS Slowhttptest
                 0.999    0.000    0.998      0.999    0.999      0.998    0.999     0.998     SSH-Patator
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    0.999     0.998

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4397    0    0    1    1 |    a = DoS Hulk
    0 4394    1    4    0 |    b = FTP-Patator
    1    0 4395    2    2 |    c = DoS Slowhttptest
    3    2    0 4394    0 |    d = SSH-Patator
    3    0    0    0 4396 |    e = PortScan
```

Ovaj algoritam daje rezultate ranga Tree algoritama. Broj pogodaka je veći kada je broj suseda koji se posmatraju veći jer je time smanjen uticaj šuma na klasifikaciju.

Cross-validation KNN=3

```
Correctly Classified Instances       21963               99.85    %
Incorrectly Classified Instances        33                0.15    %
Kappa statistic                          0.9981
Mean absolute error                      0.0006
Root mean squared error                  0.0208
Relative absolute error                  0.1899 %
Root relative squared error              5.1893 %
Total Number of Instances            21996

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.001    0.997      0.999    0.998      0.998    1.000     0.999     DoS Hulk
                 1.000    0.000    0.999      1.000    0.999      0.999    1.000     0.999     FTP-Patator
                 0.998    0.000    0.999      0.998    0.998      0.998    1.000     0.999     DoS Slowhttptest
                 0.997    0.000    0.999      0.997    0.998      0.997    0.999     0.998     SSH-Patator
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999     PortScan
Weighted Avg.    0.998    0.000    0.999      0.998    0.998      0.998    1.000     0.999

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4395    0    2    0    2 |    a = DoS Hulk
    0 4397    0    1    1 |    b = FTP-Patator
    2    1 4389    5    3 |    c = DoS Slowhttptest
    7    5    0 4387    0 |    d = SSH-Patator
    3    0    1    0 4395 |    e = PortScan
```

Test set KNN=3

```
Correctly Classified Instances        5498               99.9818 %
Incorrectly Classified Instances         1                0.0182 %
Kappa statistic                          0.9998
Mean absolute error                      0.0001
Root mean squared error                  0.0085
Relative absolute error                  0.0414 %
Root relative squared error              2.1321 %
Total Number of Instances             5499

=== 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     DoS Hulk
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     FTP-Patator
                 1.000    0.000    0.999      1.000    1.000      0.999    1.000     0.999     DoS Slowhttptest
                 0.999    0.000    1.000      0.999    1.000      0.999    1.000     0.999     SSH-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PortScan
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1100    0    0    0    0 |    a = DoS Hulk
    0 1100    0    0    0 |    b = FTP-Patator
    0    0 1099    0    0 |    c = DoS Slowhttptest
    0    0    1 1099    0 |    d = SSH-Patator
    0    0    0    0 1100 |    e = PortScan
```

Test set KNN=3

```
Correctly Classified Instances        5493               99.8909 %
Incorrectly Classified Instances         6                0.1091 %
Kappa statistic                          0.9986
Mean absolute error                      0.0004
Root mean squared error                  0.0168
Relative absolute error                  0.1198 %
Root relative squared error              4.1925 %
Total Number of Instances             5499

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     1.000     DoS Hulk
                 1.000    0.000    0.998      1.000    0.999      0.999    1.000     1.000     FTP-Patator
                 0.999    0.000    0.999      0.999    0.999      0.999    1.000     0.999     DoS Slowhttptest
                 0.996    0.000    0.998      0.996    0.997      0.997    1.000     0.999     SSH-Patator
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     PortScan
Weighted Avg.    0.999    0.000    0.999      0.999    0.999      0.999    1.000     1.000

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1099    0    0    1    0 |    a = DoS Hulk
    0 1100    0    0    0 |    b = FTP-Patator
    0    0 1098    1    0 |    c = DoS Slowhttptest
    1    2    1 1096    0 |    d = SSH-Patator
    0    0    0    0 1100 |    e = PortScan
```