Yusuf SEYİTOĞLU

1030516739

TOPIC

FORENSICS INVESTIGATION FROM FINGERPRINT MICROBES

ıNTRODUCTION TO PATTERN RECOGNITION

FINAL PROJECT REPORT

**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 değişkenlikler yapılarak, bu değişikliklerin başarı oranındaki etkisi gözlemlenmiştir. Bulunan sonuçları karşılaştırarak buna uygun algoritma seçilmiştir

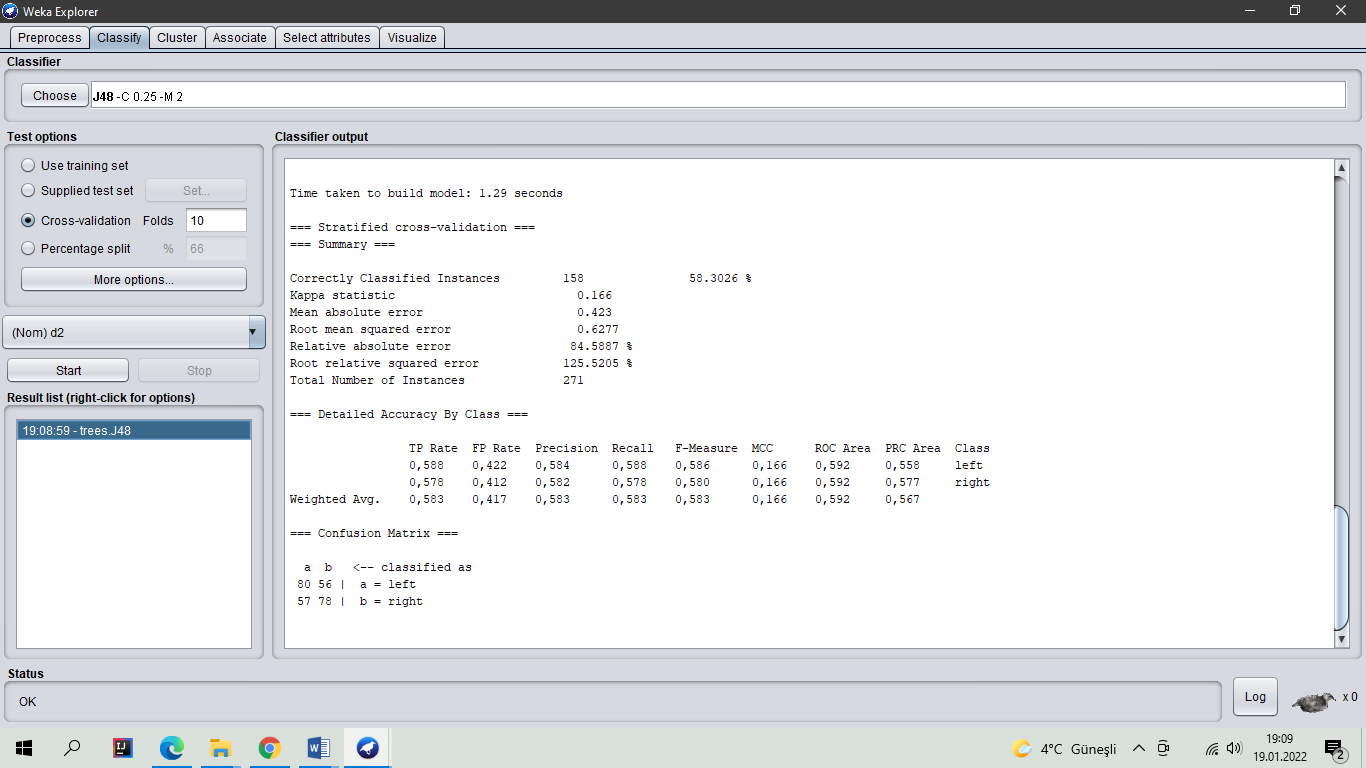
* DECİSİON TREE (J48)
* K-NEAREST NEİGHBORS (lBk)
* NAİVE BAYES
* VOTED PERCEPTRON

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

**ALGORİTMALARIN DOĞRULUK KARŞILAŞTIRILMASI**

**DECİSİON TREE (J48)**

Başarı Oranı: %58.3026

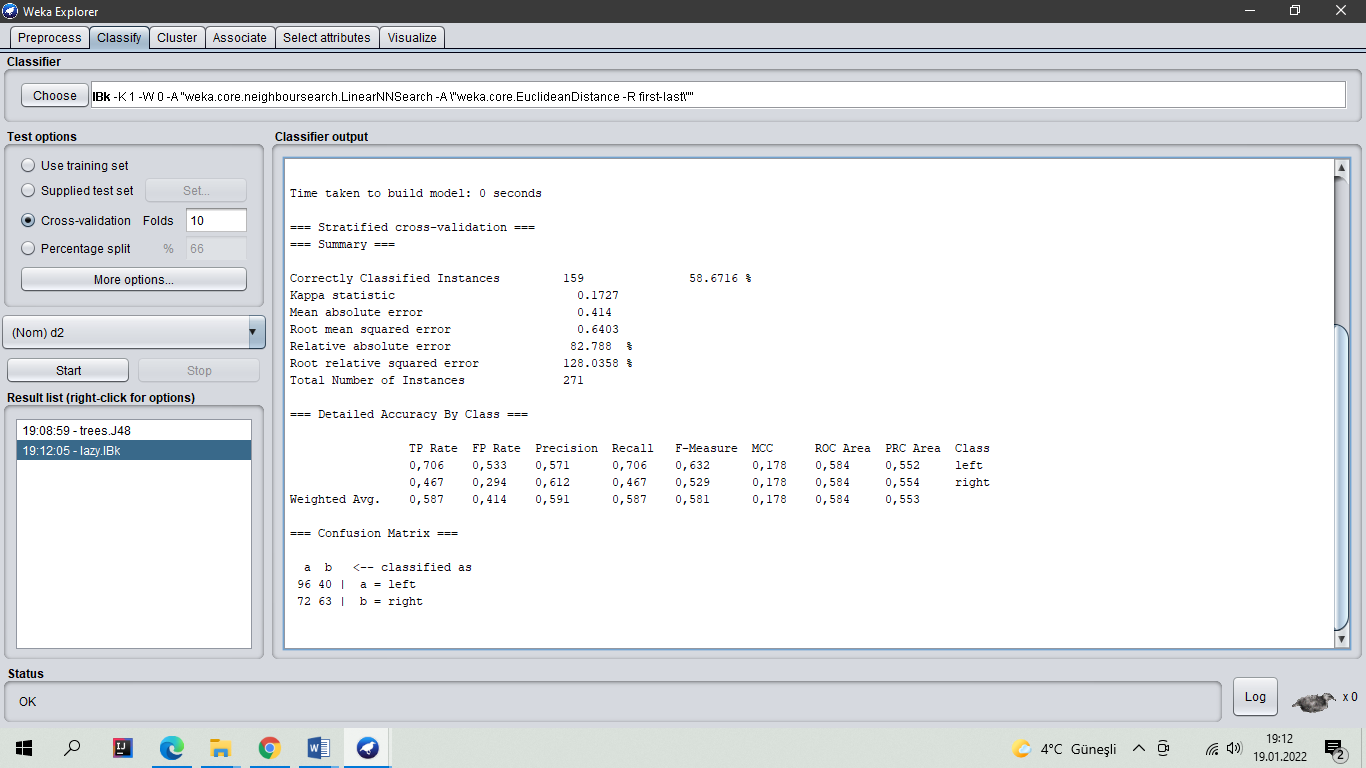


SENSİTİVİTY:

SPECİFİCİTY:

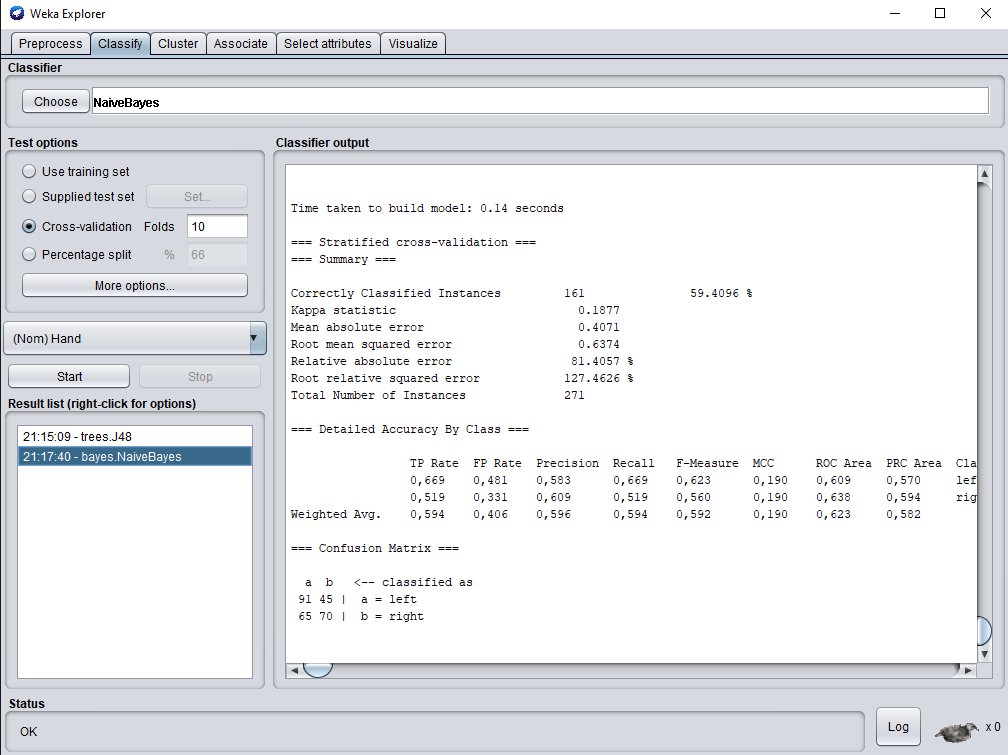
**K-NEAREST NEİGHBOURS (lBk)**

Başarı Oranı: %58.6716



SENSİTİVİTY:

SPECİFİCİTY:

**NAİVE BAYES**

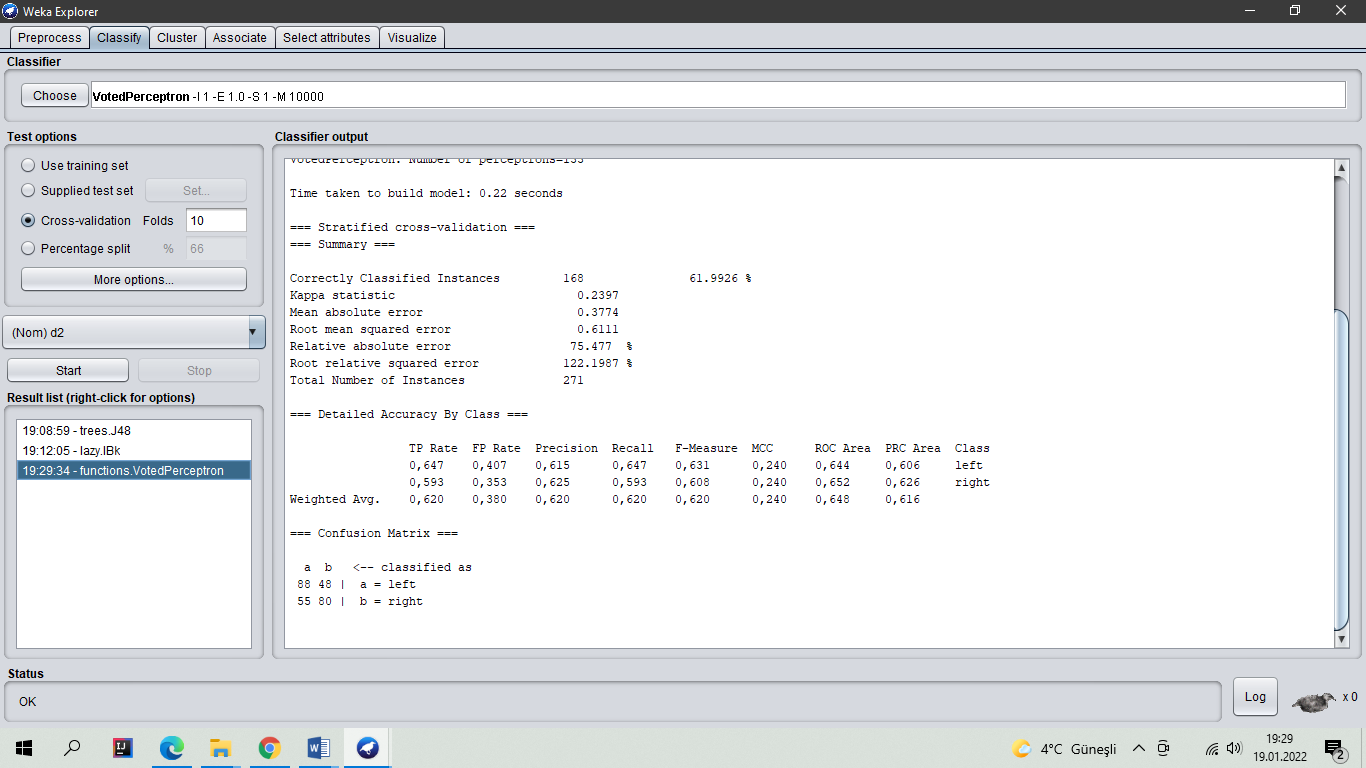
Başarı Oranı: %59.4096

SENSİTİVİTY:

SPECİFİCİTY:

**VOTED PERCEPTRON**

Başarı Oranı: %61.9926

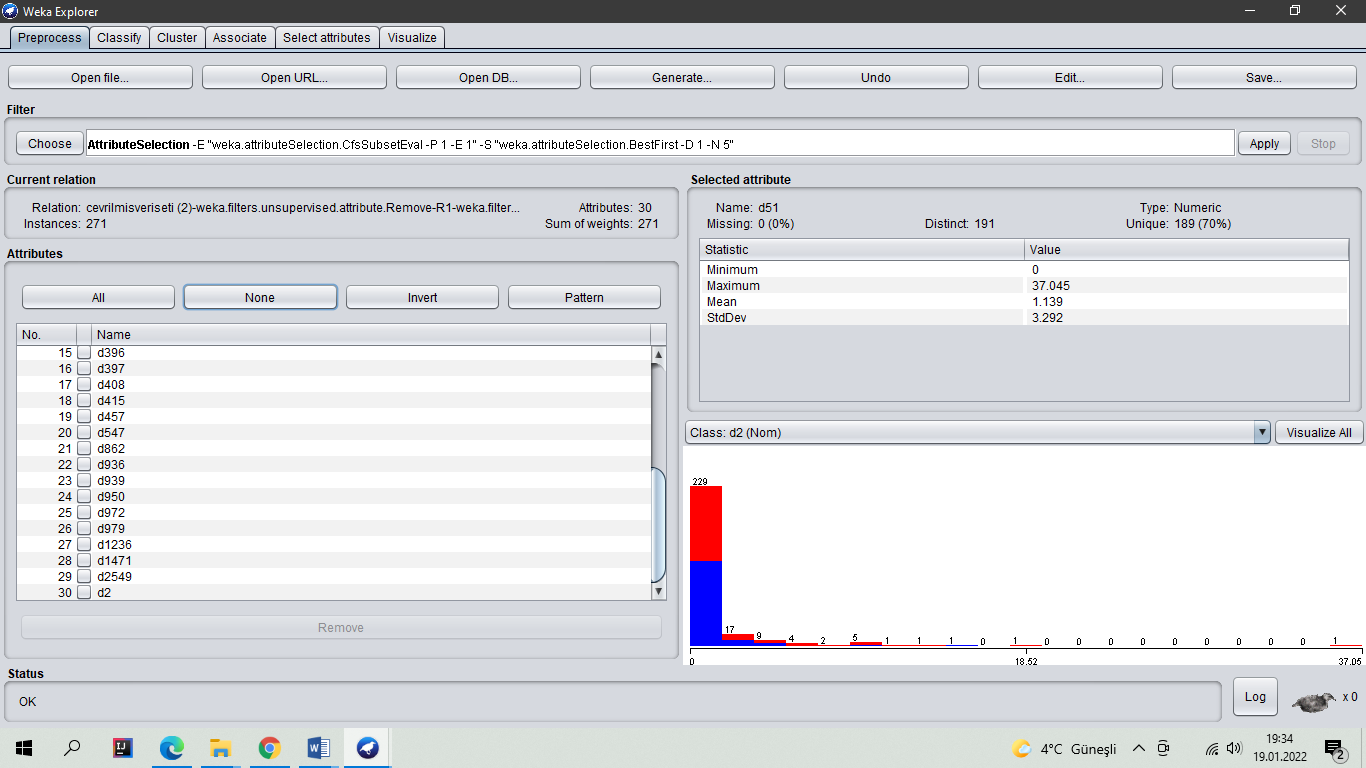


SENSİTİVİTY:

SPECİFİCİTY:

Hiçbir ön işleme tabi tutmadan ve parametre değiştirmeksizin yukarda belirtilen algoritmalar ile karşılaştırma yapıldığında en iyi sonucu “naive bayes” ve “voted perceptron” algoritmaları vermiştir. Şimdi ise verilere ön işlem uygulanarak analiz yapılacaktır.

**VERİ SETİNİN ÖN İŞLEME ALINMASI**

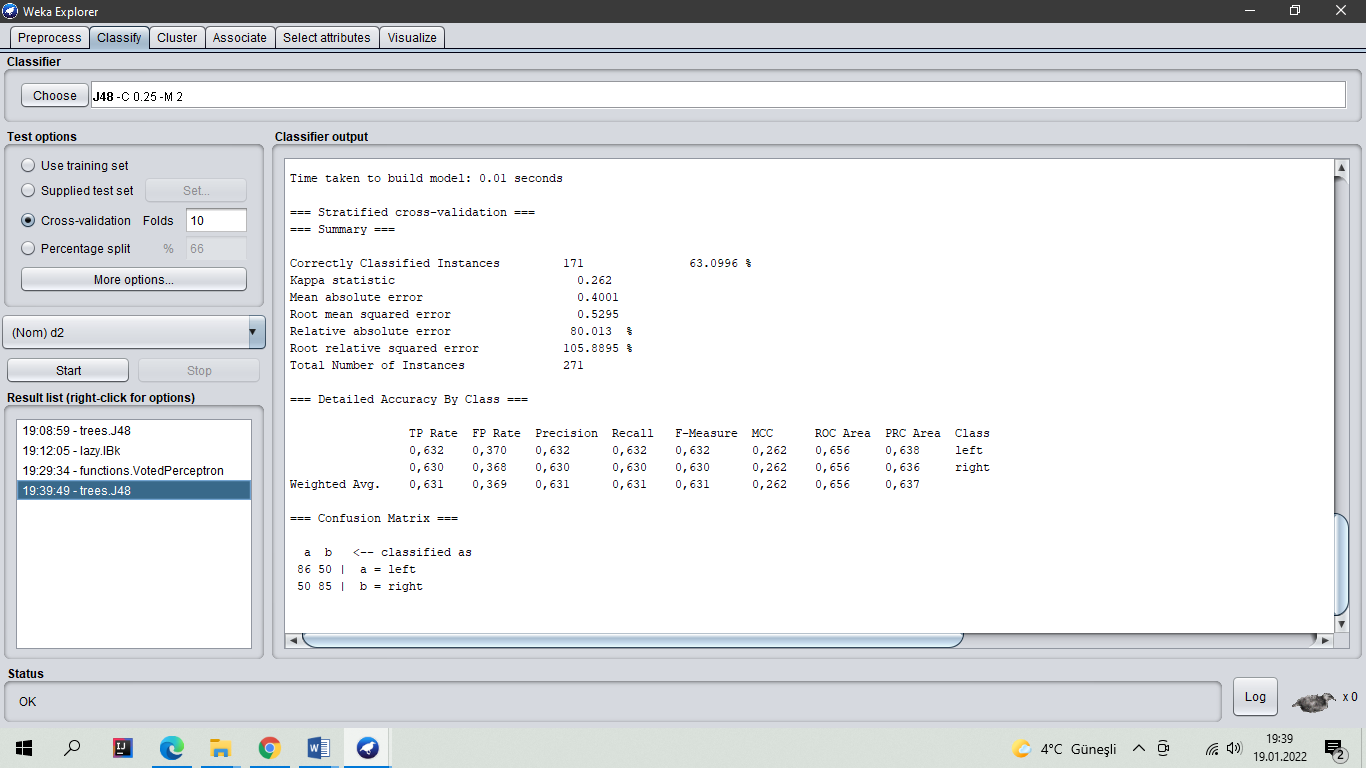


Yukarıdaki ekran görüntüsünde görüldüğü üzere “Best First” ile bir filtreleme işlemi yapılmıştır. “Best First” algoritması ile tekrarlanan verileri ayıklayıp veri setimizi 271’den 30 veriye düşürdük. Ön işleme alınmış bu veri seti ile yeniden başarı oranı karşılaştırması yapılacaktır.

**ÖN İŞLEME ALINMIŞ VERİ SETİ İLE KARŞILAŞTIRMA**

**DECİSİON TREE (J48)**

Başarı oranı: %63.0996

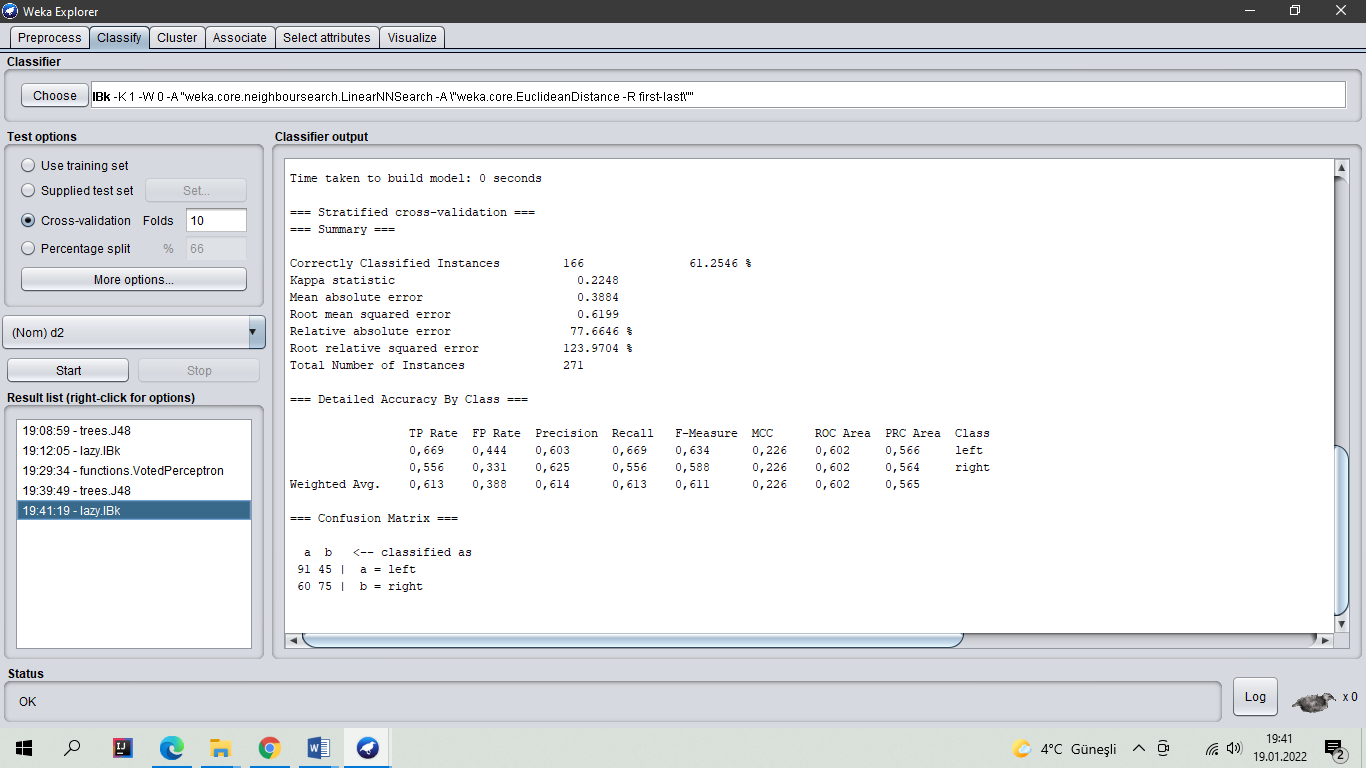


SENSİTİVİTY:

SPECİFİCİTY:

**K-NEAREST NEİGHBOURS (lBk)**

Başarı oranı: %61.2546

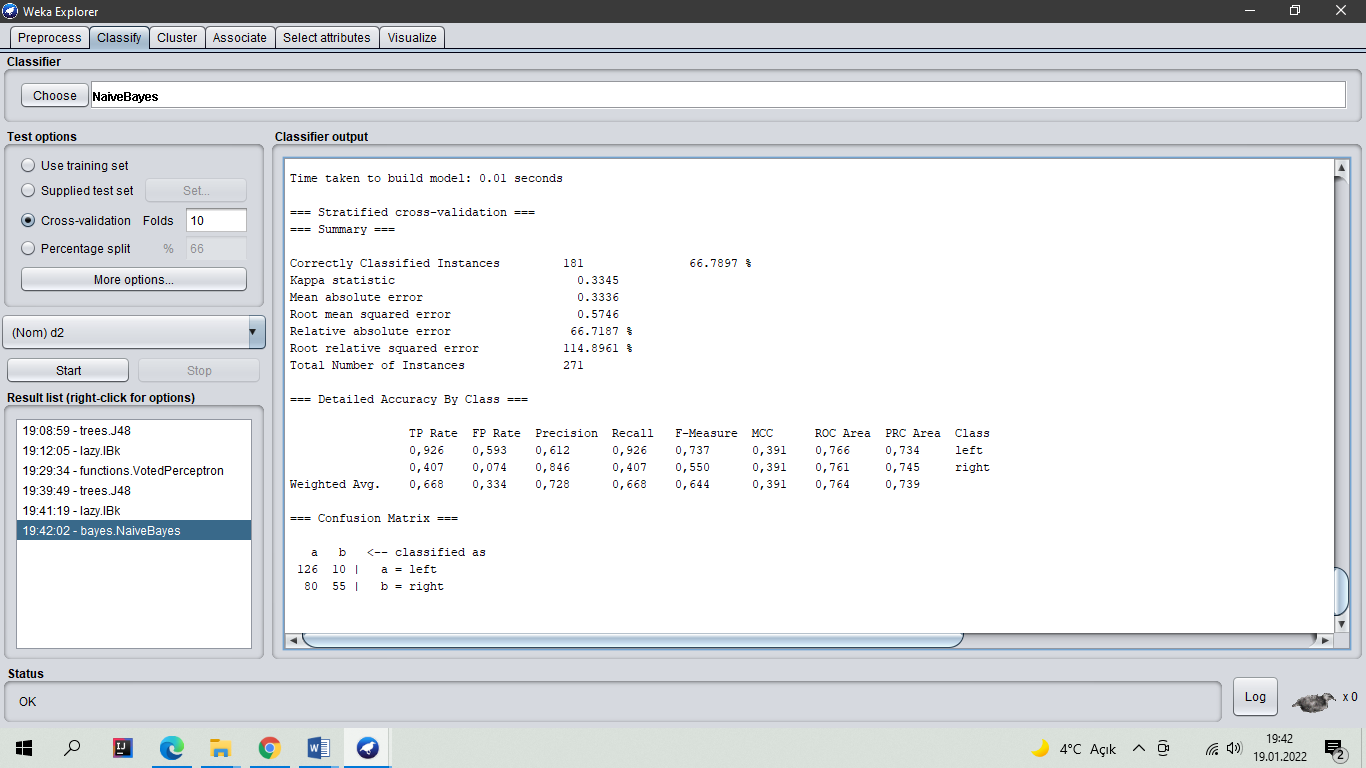


SENSİTİVİTY:

SPECİFİCİTY:

**NAİVE BAYES**

Başarı oranı: %66.7897

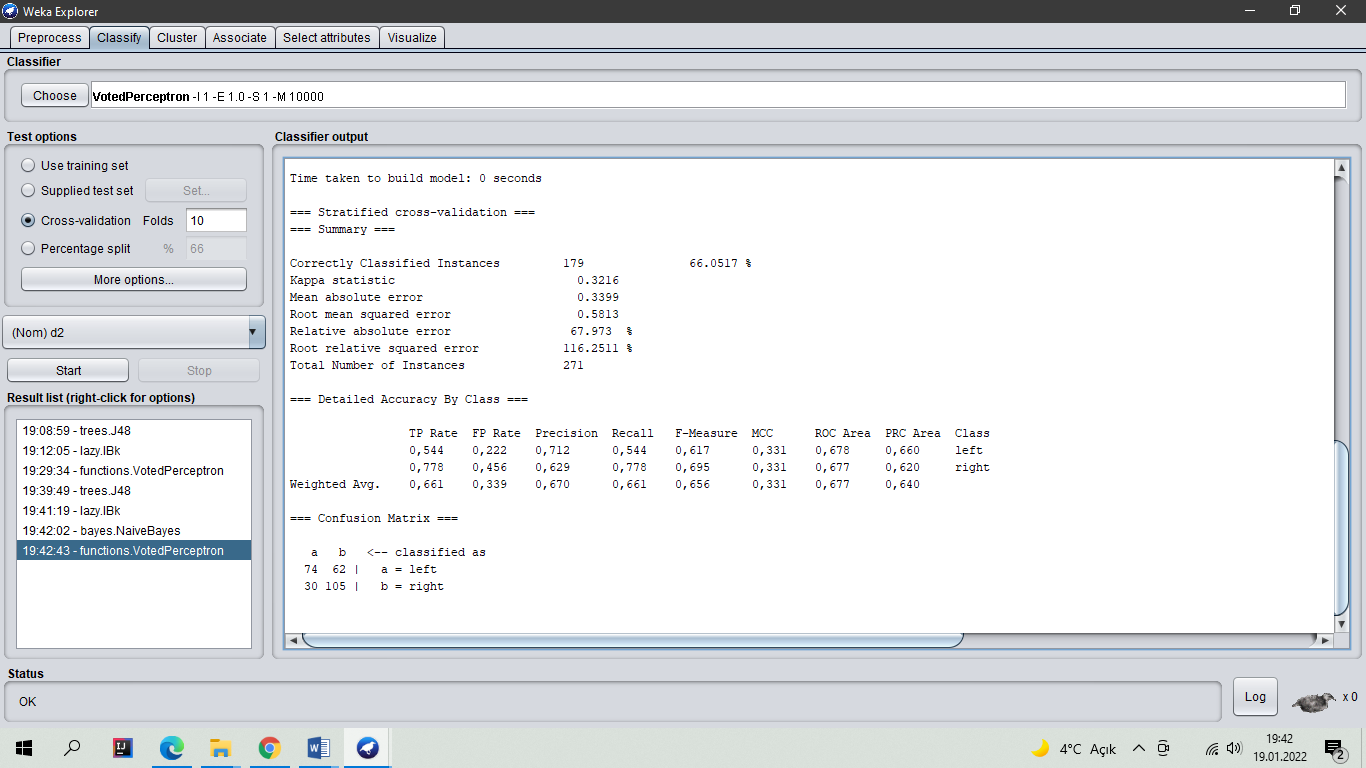


SENSİTİVİTY:

SPECİFİCİTY:

**VOTED PERCEPTRON**

Başarı Oranı: %66.0517



SENSİTİVİTY:

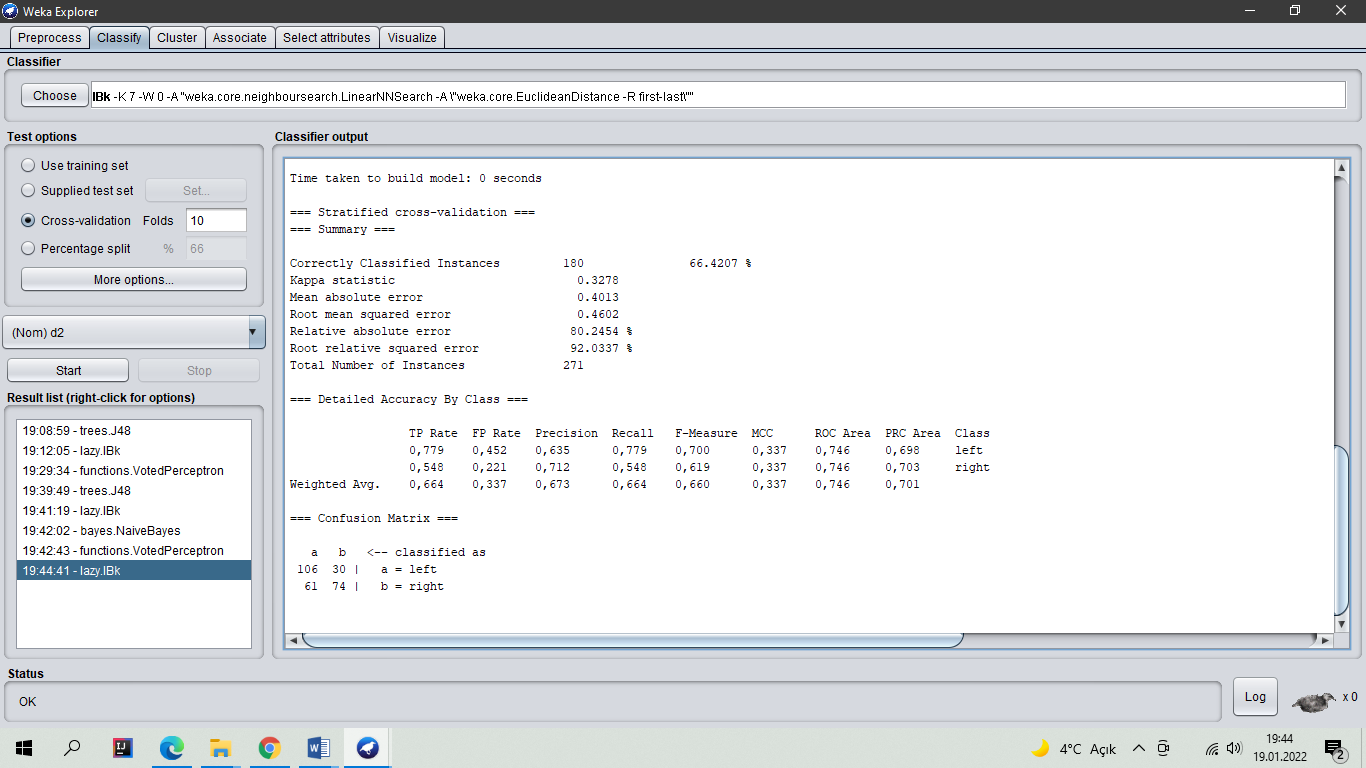
SPECİFİCİTY:

**PARAMETRE DEĞİŞİKLİĞİ İLE BAŞARIM KARŞILAŞTIRILMASI**

**K-NEAREST NEİGHBOURS (lBk)**

Başarı Oranı %66.4207

PARAMETRELER: K 7 -W 0



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

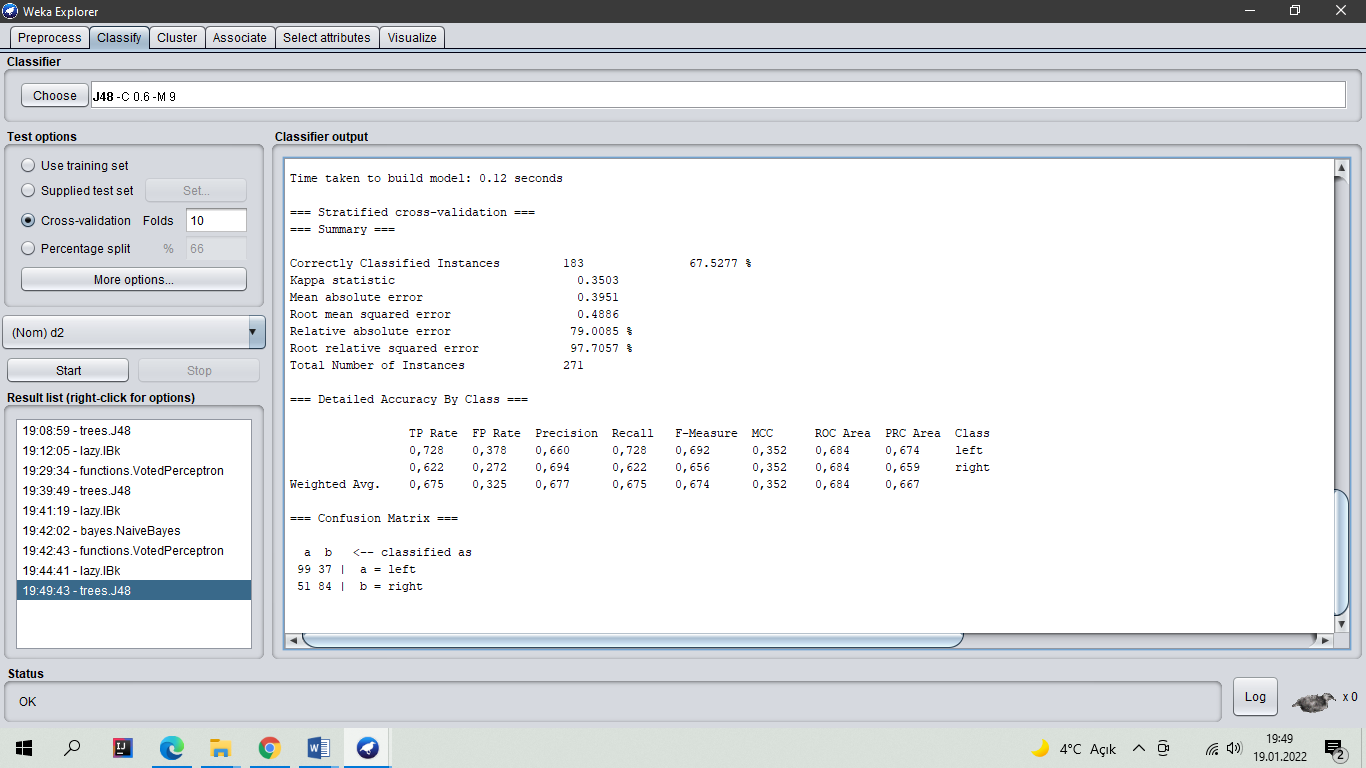
SENSİTİVİTY:

SPECİFİCİTY:

**DECİSİON TREE (J48)**

Başarı Oranı: %67.5277

PARAMETRELER: C 0.6 – M 9



Decision Tree algoritmasında “C 0.25” değeri “C 0.6” ile ve “M 2” değeri “M 9” ile değiştirilip başarı oranı arttırılmıştır. 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.

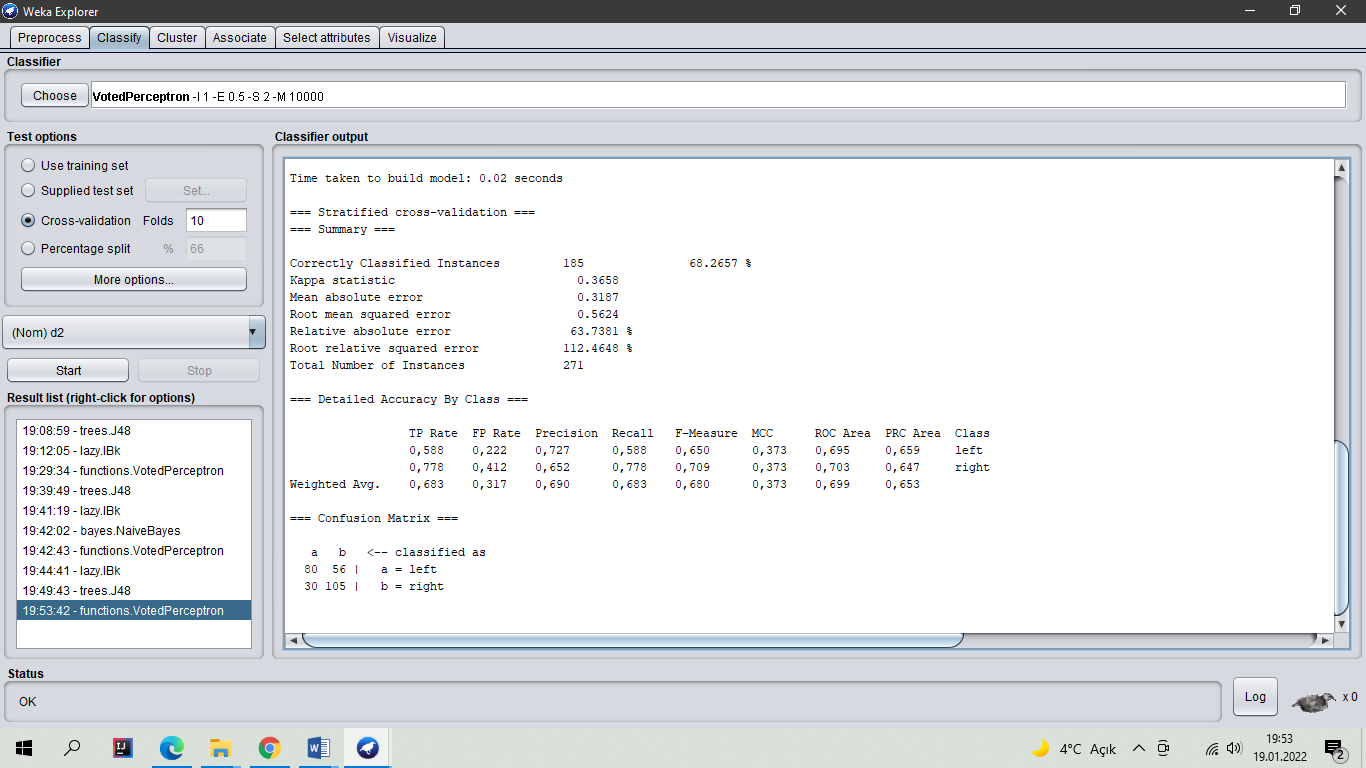
SENSİTİVİTY:

SPECİFİCİTY:

**VOTED PERCEPTRON**

Başarı Oranı: %68.2657

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



Denemeler sonucunda en iyi sonucu veren durum gözükmektedir. 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.

SENSİTİVİTY:

SPECİFİCİTY: