MÜCAHİT TAHA AZ

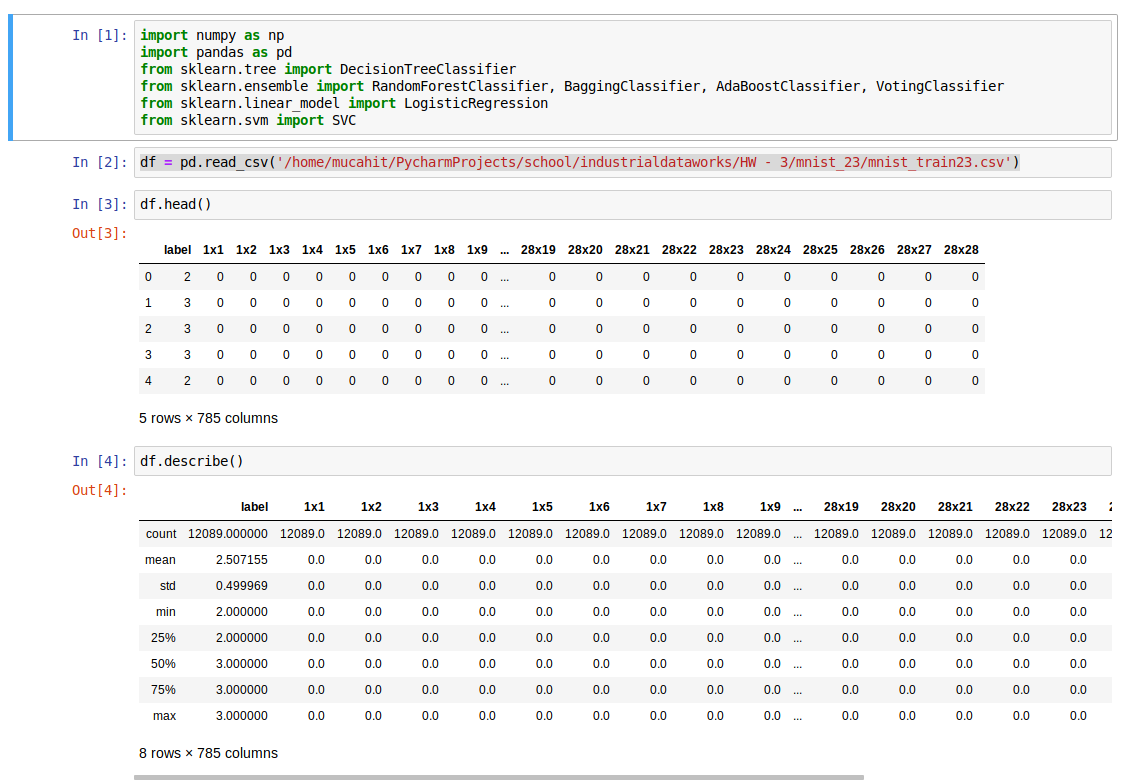
ENM - 593

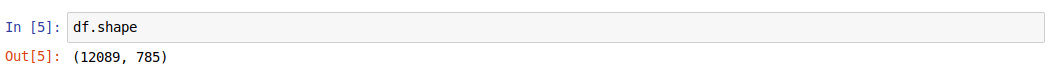
Endüstriyel Veri Madenciliği Uygulamaları

**ÖDEV - 3**

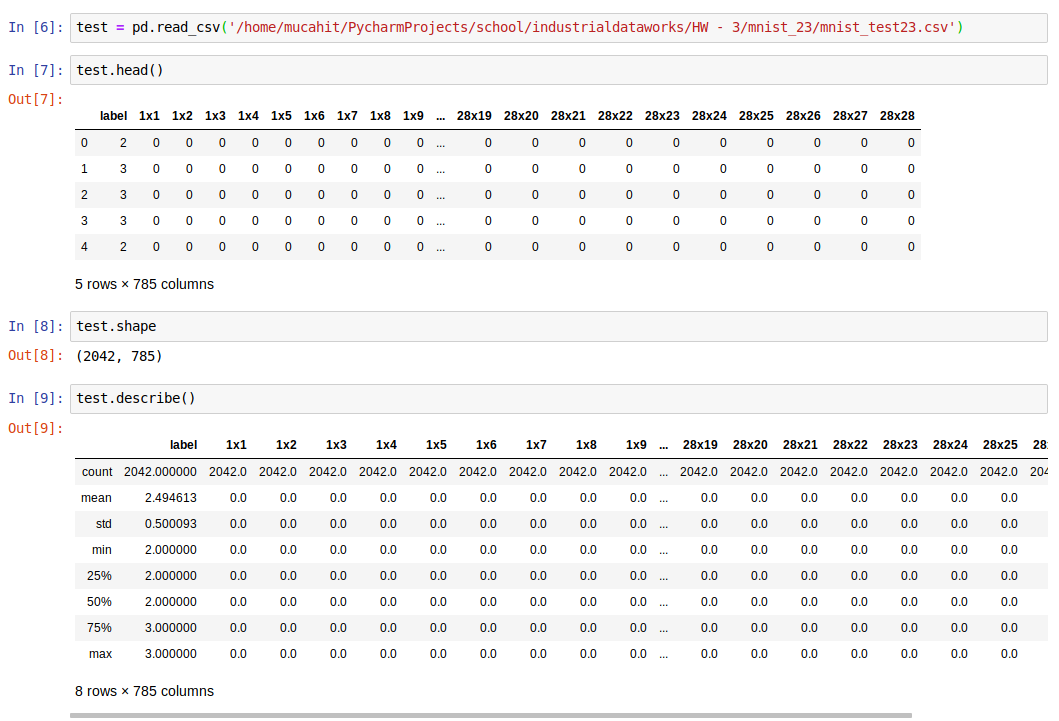
**Soru - 1**

Gerekli kütüphaneler yüklenir , veriseti yüklenir , verisetiyle ilgili temel bilgiler incelenir.



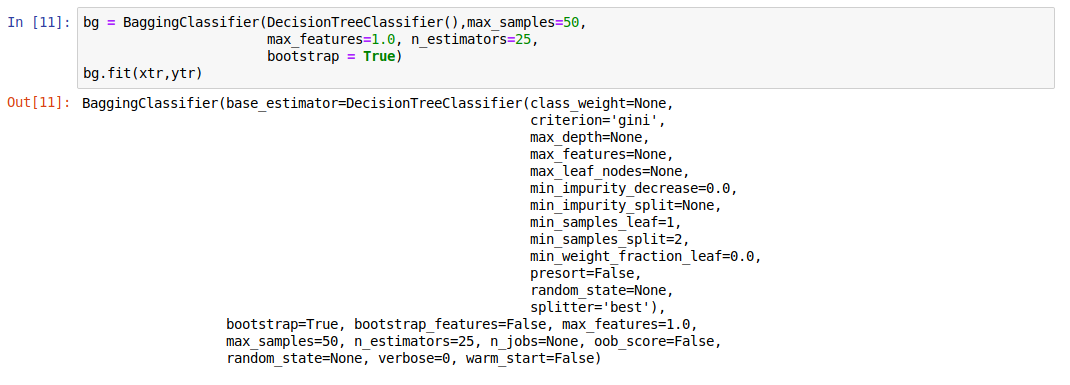


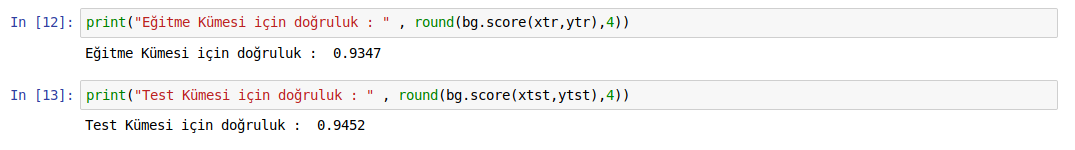
Test verisi yüklenir ve onunda temel bilgileri incelenir.



Burada label sutünu ayrılarak test ve train verilerinin eğitilmeye uygun hale getirilmesi sağlanır.



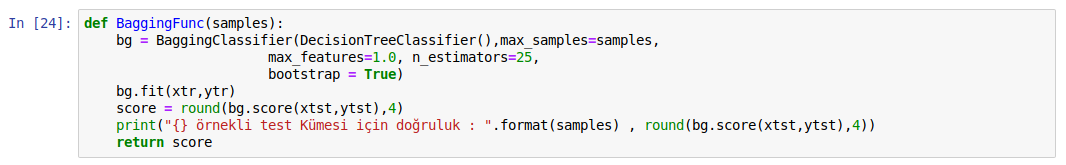
50 örneklemli bir bootstrap torbalama sınıflandırıcı eğitilir. Karar ağacı sınıflandırıcısı kullanılır.

Test ve Train datasetleri için sonuçlar yazdırılır.

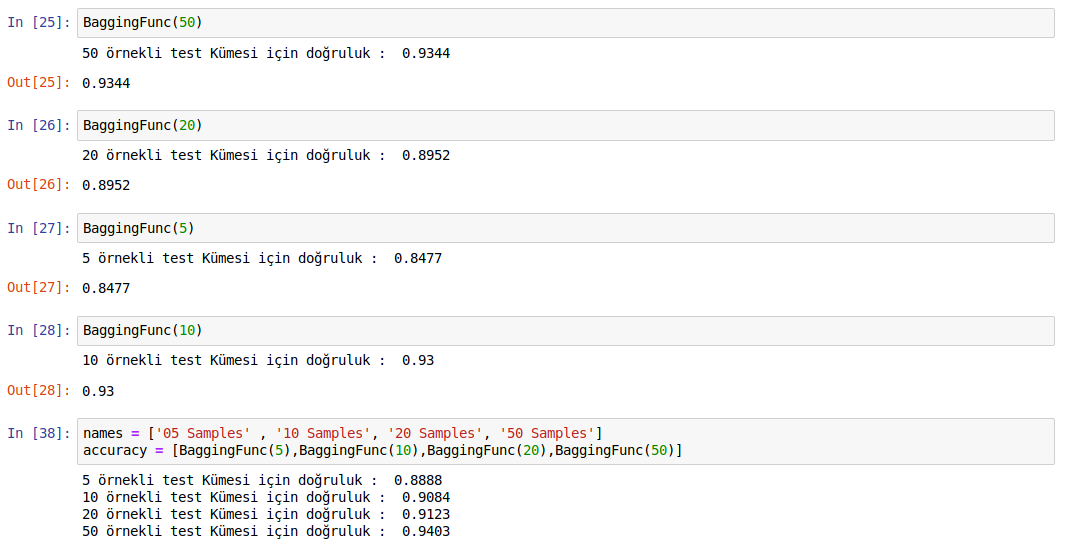
**Soru - 2**

Bir önceki sorudaki gibi en başta veriseti ve kütüphaneler hazırlanır.

Sonrasında ise fonksiyon tanımlanır.

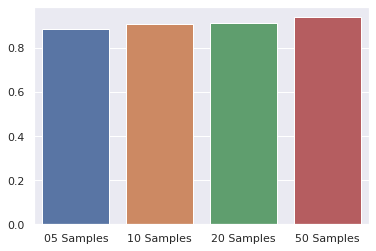


Soruda istenen örneklem miktarları için sırayla fonksiyon çalıştırılır.

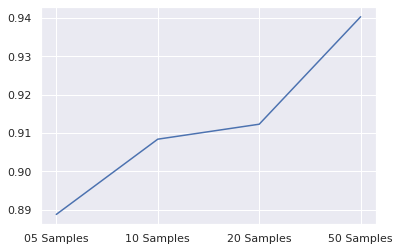


Grafik ile gösterim yapılır.



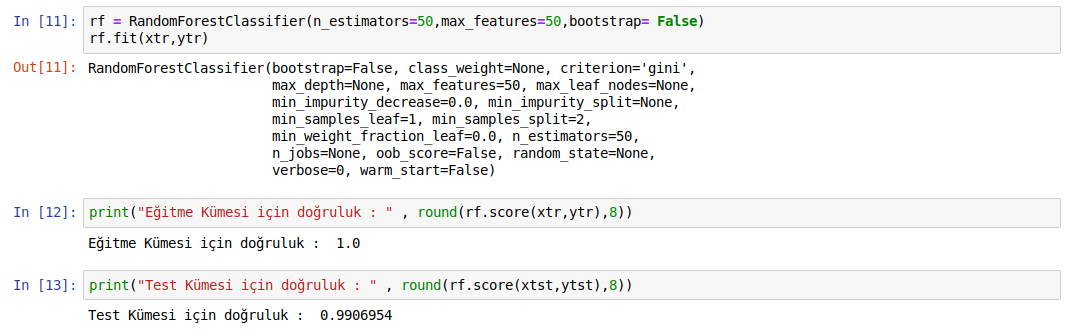






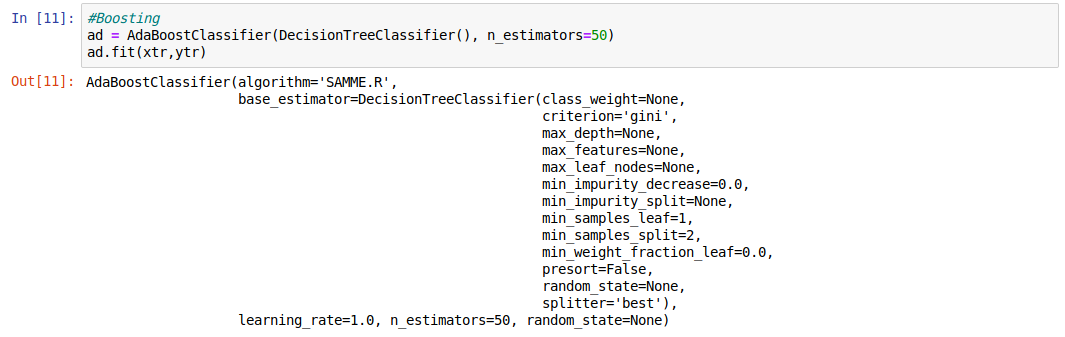
**Soru – 3**

Veri ve kütüphaneler yüklenir. Soruda rassal orman sınıflandırıcısı ile çözüm yapılması istenmiştir.

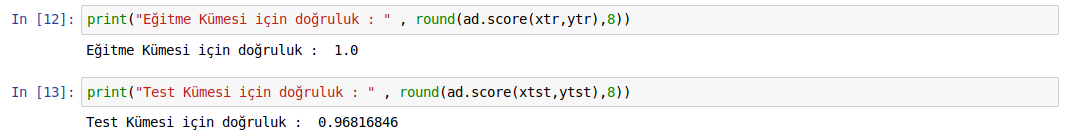
Eğitme kümesi kusursuz öğrenilmiş gözükmektedir ama test kümesinde kusursuzluk yoktur. Öğrenme seti overfitting (aşırı öğrenme) yapmış olabilir.

**Soru – 5**

Veriseti ve kütüphaneler yüklenir. Adaboost yardımı ile takviyeleme yapılır. Karar ağacı sınıflandırıcısı kullanılır.

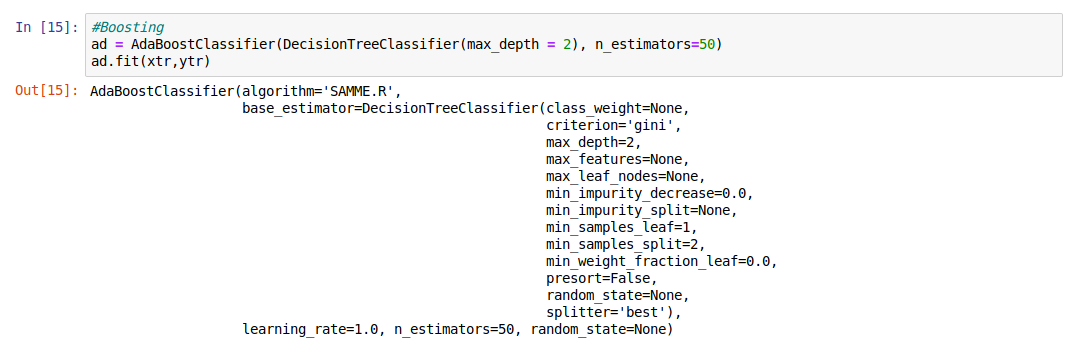


Soruda eğitim kümesi overfit olmuştur ve accuracy olarak 1 bulmuştur.

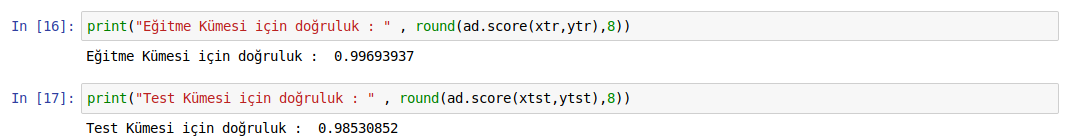


**Soru – 6**

Bir önceki soruda yaşanan overfittingin yeniden yaşanmaması için parametlerde düzenlemeler yapılır ve sınıflandırma yeniden yapılır.



Bu sefer overfitting yaşanmamıştır ve test kümesi için doğruluk artmıştır.



**Soru – 7**

a) Bagging yöntemi ile yapılan rassal orman sınıflandırıcı 0.99 ile en başarılıdır ama eğitim verisinde 1.0 doğruluk bulmasından dolayı overfitting yapmıştır. Bu durumda overfitting yapmadan en yüksek doğruluğu bulan adaboost ile yapılmış karar agacı sınıflandırması en başarılıdır.

b)Topluluktaki örneklem sayısı arttıkça doğruluk oranı artmaktadır.

c)Modelin karmaşıklığının artması modelin overfit etme ihtimalini arttırır. Bu da bagging ve boosting algoritmalarının performansını düşürür.