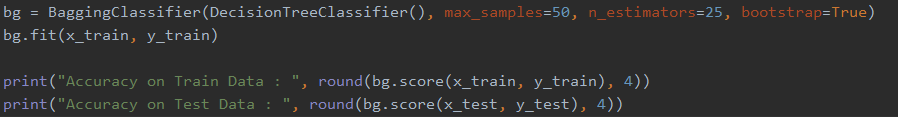
**ENDÜSTRİYEL VERİ MADENCİLİĞİ UYGULAMALARI**

**ÖDEV – 3 RAPOR**

**OĞUZHAN DİKİCİ**

**Soru 1:**

Gerekli importlar yapılıp,

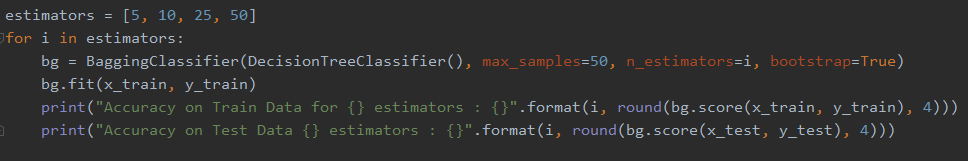


kodu ile

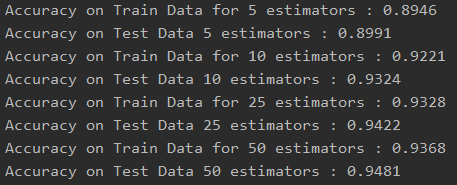


sonucuna ulaşıldı. Test başarısının daha yüksek olmasından iyi bir öğrenme olduğu yorumu yapılabilir.

**Soru 2:**

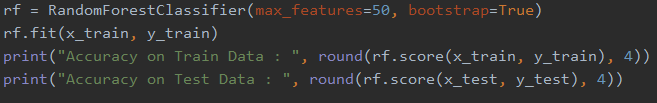
****

Basit bir for döngüsü ile 5, 10, 25, 50 ağaç öğrenildi.

****

Öğrenilen ağaç arttıkça sonucun yükseldiği görülüyor. Test-eğitim başarılarına da bakılırsa, öğrenilen ağaç miktarının artması iyiye yorumlanabilir.

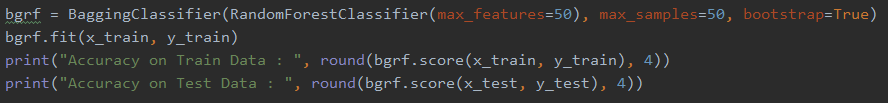
**Soru 3:**





Şu ana kadarki en başarılı algoritma.

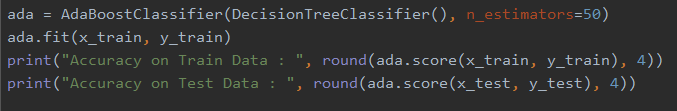
**Soru 4:**

****

****

Başarı bir öncekine kıyasla düşük ancak parametrelerle oynanırsa bir öncekinden daha da başarılı olabileceğini düşünüyorum.

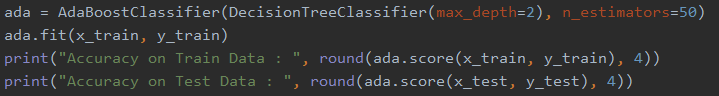
**Soru 5:**

****

****

Bariz bir overfitting varmış gibi gözükse de testteki başarısı bir overfite göre yüksek. Başarılı bir algoritma denebilir.

**Soru 6:**





Öğrenme sınırı koyulduğunda bariz bir şekilde overfitting önlenmiş ve testteki başarı yükselmiş oldu.

**Soru 7:**

1. Ödev kapsamında sonuçlara bakılınca ‘Random Forest’ ve ‘Adaboost’ aynı başarıda sayılabilir, ancak tüm topluluk yöntemleri arasında ‘Boosting’in daha başarılı olduğunu düşünüyorum.
2. Örneklem sayısı eğer elimizdeki toplam veri miktarı ise, bunun sonuca doğrudan pozitif bir etkisi olacaktır. Eğer eğitim için kullanılacak algoritmanın her bir iterasyonunda kullanılacak olan örneklem ise, buna pek yorum yapamıyorum. Elimizdeki veriye ve algoritmaya göre farklı sonuçlar çıkacağını düşünüyorum.
3. Her bir veri öğrenme durumunda, optimum bir model karmaşıklığı vardır. Bu nokta hata, bias ve varyansın en düşük olduğu noktadır. Bunun ötesinde ve gerisinde hata artar. Varyans ve bias ters ilişkide olur. Topluluk yöntemlerinde bias durumu yoktur ancak geri kalan durum tamamen geçerlidir.