**KARABÜK ÜNİVERSİTESİ**

**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**BİYOMEDİKAL MÜHENDİSLİĞİ**



**BİYOMEDİKAL MÜHENDİSLİĞİNDE YAPAY SİNİR AĞI UYGULAMALARI**

**LİNEAR REGRESYON, RANDOM FOREST REGRESYON VE DECİSİON TREE REGRESYON UYGULAMALARI VİZE RAPORU**

**ADI: ESRA**

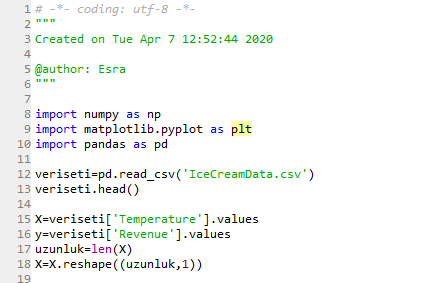
**SOYADI: ALGIM**

**NO: 1928142005**

Bir veri setinin doğrusal regresyon, rasgele orman regresyon ve karar ağacı regresyon ile modellenerek, modele ait denklemin grafiğinin çizilmesi ve performansının Python kütüphaneleri ile elde edilmesi amaçlanmıştır.

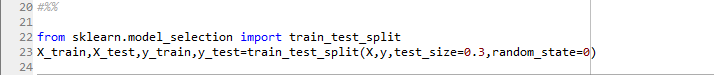
Öncelikle **veriseti** isimli DataFrame nesnesi pandas kütüphanesi yardımıyla oluşturulmuştur. IceCreamData.csv isimli dosya 2 adet özniteliğe sahip ve 500 satırdan oluşmaktadır. Bir dondurmacı hava sıcaklığına bağlı günlük gelirini tahmin edebilecek bir model oluşturmak istiyor. Bu örnekte kurulması planlanan regresyon modelinin sıcaklığa bağlı gelir tahmini oluşturması beklenmektedir. Bağımsız değişken sıcaklık, bağımlı değişkenimiz ise gelirdir.

Liste 1



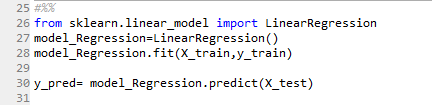
Liste 1’de öncelikle veriseti adlı dataframe içerisinden bağımsız değişken (X) ve bağımlı değişken (y) öznitelikleri, IceCreamData.csv dosyasındaki isimleri kullanılarak oluşturulmuştur.

Liste 2



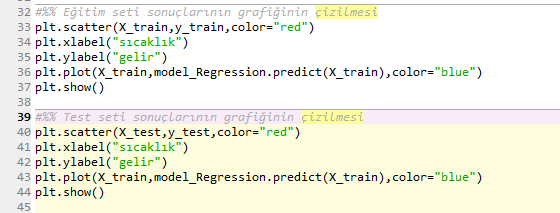
Liste 2’de veri seti eğitim ve test kümelerine bölünmüştür. Bunun için train\_test\_split() metodu kullanılmıştır. Bu metodun test\_size özelliği 0.0 ve 1.0 arasında bir değer alabilir. test\_size=0.3 verildiğinde %30 test seti ve %70 eğitim seti olarak ayrılması sağlanmıştır. random\_state özelliğine bir değer verilirse, train\_test\_split her çalıştığında aynı eğitim ve test kümesi elde edilir.

Liste 3

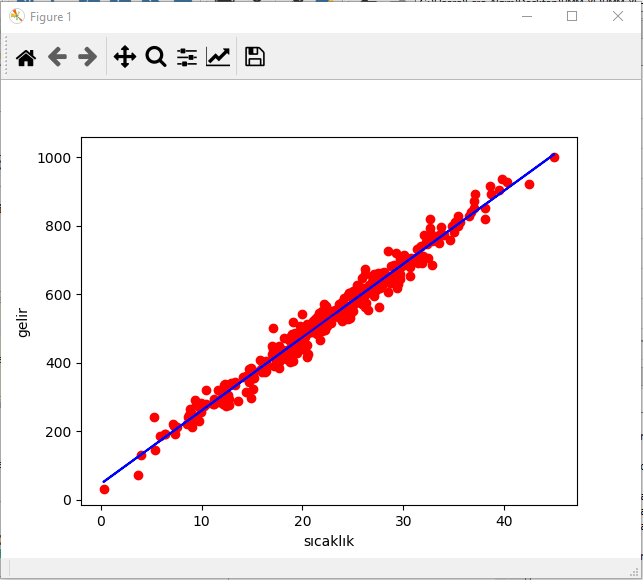


Liste 3’te linear\_model modülüne ait LinearRegression sınıfı kullanılarak doğrusal regresyon modeli oluşturulmuştur. Burada model\_regresyon isimli nesne oluşturulmuş ve daha sonra bu nesneye fit() metodu uygulanıp, modelin eğitim veri seti ile eğitilmesi gerçekleştirilmiştir. Modelin ürettiği tahminler y\_pred dizisinde toplanmıştır.

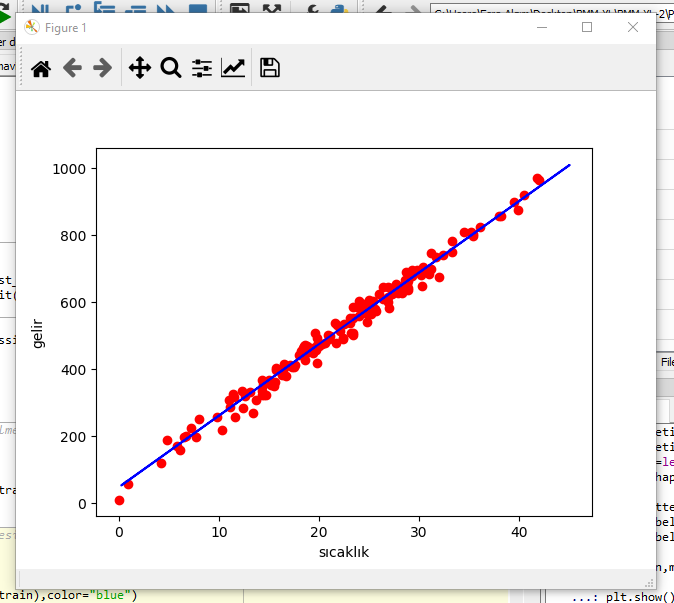
Liste 4



Liste 4’te doğrusal regresyon modeli için çizdirilen eğitim ve test seti grafikleri Şekil1 ve Şekil 2’ de gösterilmiştir.

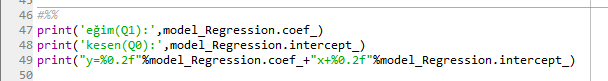


Şekil 1. Eğitim veri setine ait grafik



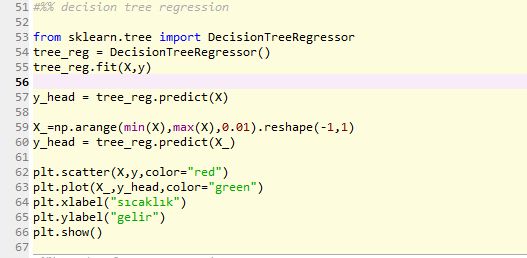
Şekil 2. Test veri setine ait grafik

Liste 5

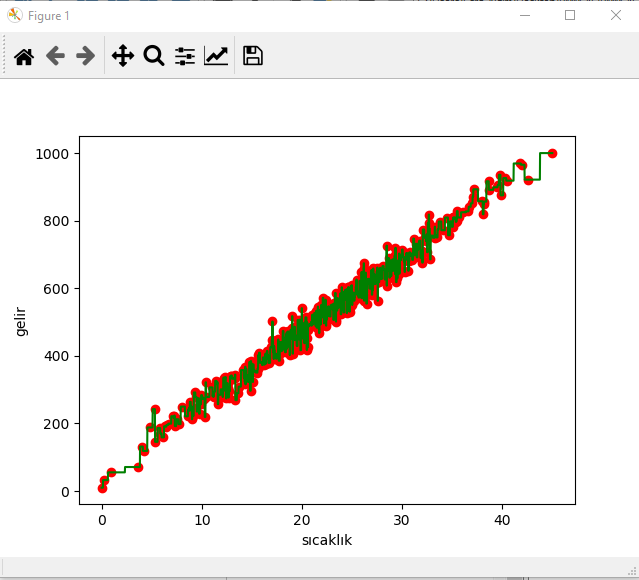


Liste 5’te kurulan modelin katsayıları hesaplanmıştır. model\_regresyon nesnesine ait coef\_ özelliği ile eğim, intercept\_ özelliği ile de kesen hesaplanır. Daha sonra bu özellikler ekrana yazdırılır.

Liste 6

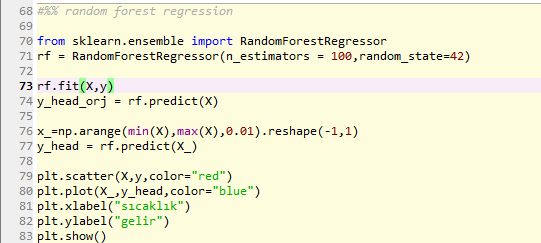


Liste 6’da tree\_reg modülüne ait DecisionTreeRegressor sınıfı kullanılarak karar ağacı regresyon modeli oluşturulmuştur. Burada tree\_reg isimli nesne oluşturulmuş ve daha sonra bu nesneye fit() metodu uygulanıp, modelin eğitim veri seti ile eğitilmesi gerçekleştirilmiştir. Modelin ürettiği tahminler y\_head dizisinde toplanmıştır. Merdiven yapısını görebilmek için 59 ve 60. kod bloğu oluşturulmuştur.

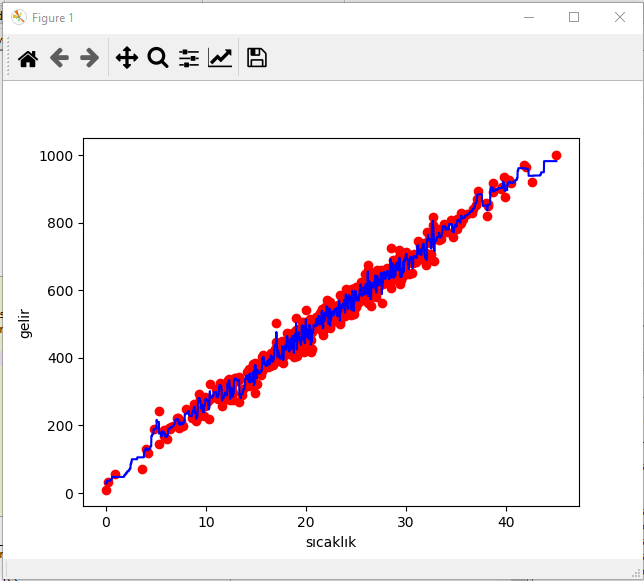


Şekil 3. Desion Tree Regresyon merdiven yapısı

Liste 7



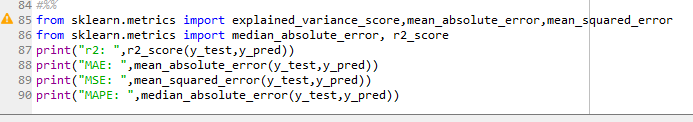
Liste 7’de RandomForestRegressor sınıfı kullanılarak rasgele orman regresyon modeli oluşturulmuştur. Burada rf isimli nesne oluşturulmuş ve daha sonra bu nesneye fit() metodu uygulanıp, modelin eğitim veri seti ile eğitilmesi gerçekleştirilmiştir. Modelin ürettiği tahminler y\_head dizisinde toplanmıştır. Merdiven yapısını görebilmek için 76 ve 77. kod bloğu oluşturulmuştur



Şekil 4. Random forest regresyonu merdiven yapısı

**Bu kısma kadar regresyon modeli elde edilmiştir. Modelin grafiği çizildi ve matematiksel olarak denklemi çıkarıldı.**

Liste 8



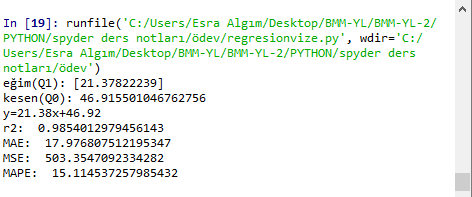
Elde edilen bu model mevcut veri setini ne ölçüde temsil ediyor?

Liste 8’de performans değerlendirmesinde kullanılan performans değerlendirme ölçütleri görülmektedir.

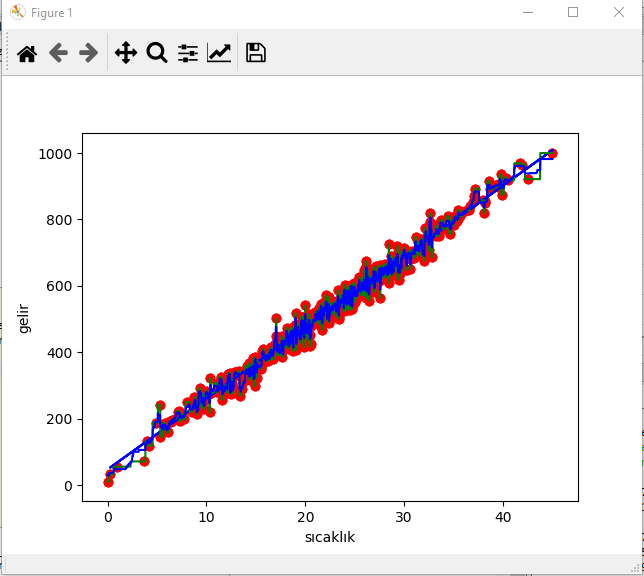
R2 değerinin 1’e yakın olması iyi bir model kurulduğunun göstergesidir.

MAE değerinin 0 ile sonsuz aralığında alacağı değer ne kadar düşük olursa modelden o kadar iyi performans elde edilir.

MSE değerinde aykırı değerler var ise o kadar yüksek değer hesaplanır.



Console değerleri



Şekil 5. Program sonu çıktısı

KAYNAKLAR

https://www.kaggle.com/vinicius150987/ice-cream-revenue

**KARABÜK ÜNİVERSİTESİ**

**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**BİYOMEDİKAL MÜHENDİSLİĞİ**



**BİYOMEDİKAL MÜHENDİSLİĞİNDE YAPAY SİNİR AĞI UYGULAMALARI**

**SUPPORT VECTOR MACHİNES , RANDOM FOREST CLASSİFİER VE KNN CLASSİFİER UYGULAMALARI VİZE RAPORU**

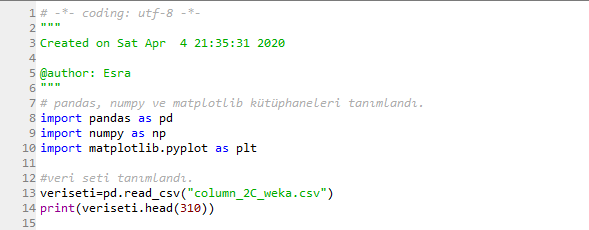
**ADI: ESRA**

**SOYADI: ALGIM**

**NO: 1928142005**

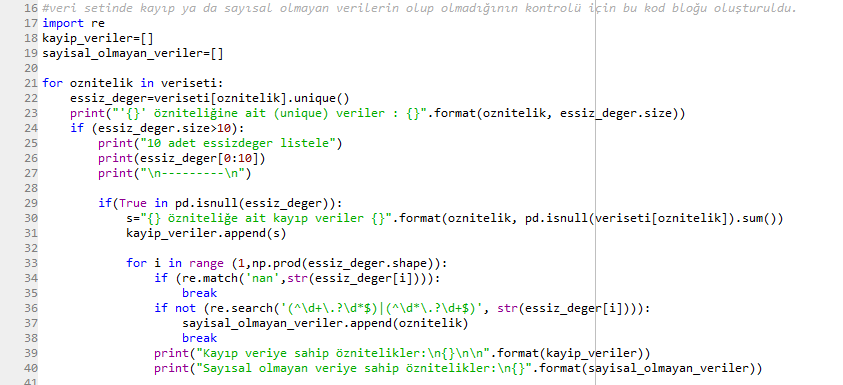
Bir veri setinin KNN yakın komşu algoritması, rasgele orman sınıflandırma ve destek vektör makineleri ile modellenerek, veri setinde olmayan bir hasta için  pelvis ve lomber omurganın şekli ve yöneliminden elde edilen altı biyomekanik özellik ile hastanın normal mi ya da anormal mi olduğu tahmin edilecektir.

Liste 1



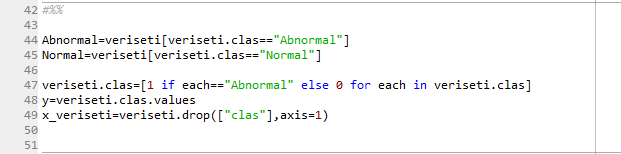
Öncelikle **veriseti** isimli DataFrame nesnesi pandas kütüphanesi yardımıyla oluşturulmuştur.

Liste 2



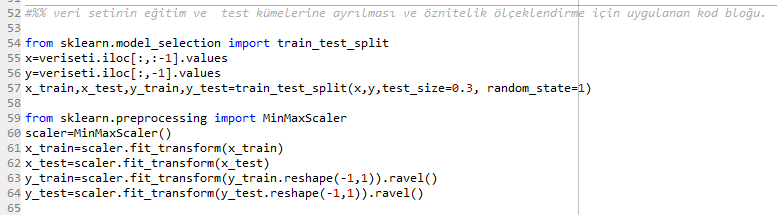
Veri setinde kayıp ya da sayısal olmayan verilerin olup olmadığının kontrolü için Liste 2’deki kod bloğu oluşturulmuştur. Kod bloğu veri setini daha iyi tanımak için yararlı olacaktır. Bu kod bloğunda, her özniteliğin içerdiği eşsiz(unique) değerler görülebilmektedir. Eşsiz değerin çok olması durumu dikkate alınarak eğer 10’dan fazla eşsiz değer var ise ekrana yazdırılmaktadır. Kodda kullanıla **re** (regular expression) kütüphanesi sayısal olmayan verilerin araştırılmasında kullanılmıştır. Ekran çıktısı incelendiğinde kayıp veri olmadığı görülmüştür. Veri ön işleme yapılmıştır.

Liste 3



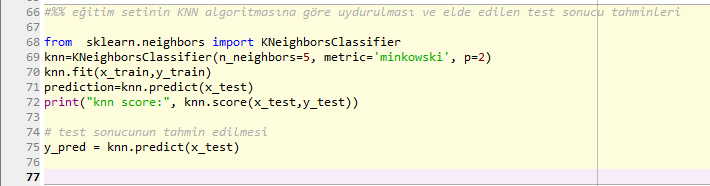
Veri setinde clas özniteliğinin sayısal olmayan ifade içerdiği ve bunların Abnormal ve Normal olarak iki özelliğe sahip olduğu görüldü. Abnormal ve Normal içeriğini içerenler Abnormal ve Normal dizilerine aktarıldı. X ekseni clas harici, y eksenimiz clas’ı içerecek şekilde eksenler oluşturuldu.

Liste 4



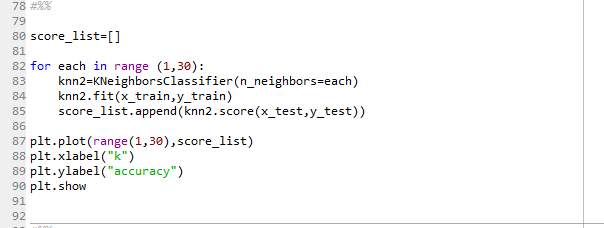
Liste 4’ te veri setinin eğitim ve test kümelerine ayrılması ve öznitelik ölçeklendirme için uygulanan kodlar görülmektedir. Veri setinin %70’i eğitim %30’u test için ayrılmıştır. Veri ön işleme adımlarından olan öznitelik ölçeklendirme işlemi veri setindeki tüm değerlerin 0 ile 1 arasında değer alan bir dönüşüme uğratmıştır.

Liste 5

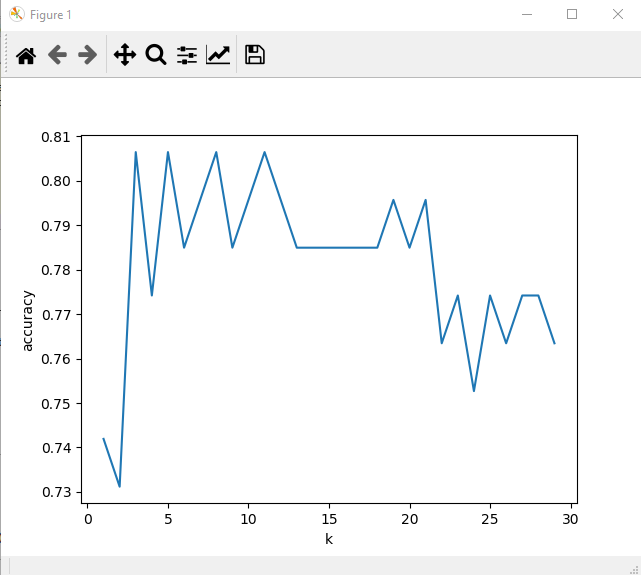


Liste 5’te eğitim setinin KNN algoritmasına göre uydurulması ve elde edilen test sonucu tahminleri görülmektedir. Scikit-learn kütüphanesinde varsayılan uzaklık ölçüm yöntemi olarak minkowski tanımlanmıştır. Öklid uzaklık yöntemi kullanmayı tercih ettiğim için p=2 ifadesi kullanıldı. n\_neighbors ifadesi ise algoritmadaki k değerini vermektedir. Modelin ürettiği tahminler y\_pred dizisinde toplanmıştır.

Liste 6



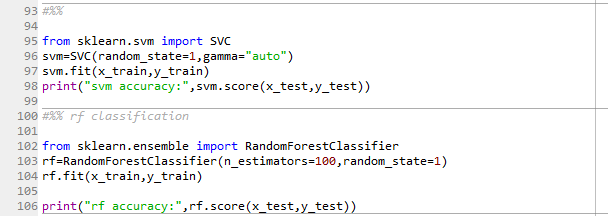
En iyi k değerini bulmak için score\_list oluşturuldu. Kodlar incelendiğinde k değeri için 1 ve 29 arasındaki tüm tamsayı değerleri bir döngü yardımıyla denenmiş ve elde edilen 29 farklı sonuç grafik ile gösterilmiştir.



Şekil 1. K değerlerine göre accuracy oranları grafiği

Şekil 1’de görüldüğü gibi en iyi k değerleri grafiğin pik verdiği noktalardır.

Liste 7



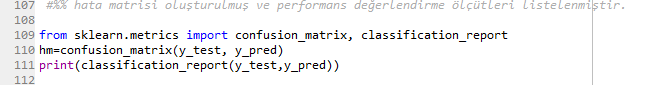
Liste 7’de Support vector machines yöntemi kullanmak için SVC sınıfı kullanarak sınıflandırma yapılmıştır. Svm isimli nesne oluşturulup, gama değeri kullanarak en iyi k değeri kullanılmıştır. fit() metodu uygulanıp, modelin eğitim veri seti ile eğitilmesi gerçekleştirilmiştir.

RandomForestRegressor sınıfı kullanılarak rasgele orman sınıflandırma yapılmıştır. Burada rf isimli nesne oluşturulmuş ve daha sonra bu nesneye fit() metodu uygulanıp, modelin eğitim veri seti ile eğitilmesi gerçekleştirilmiştir.

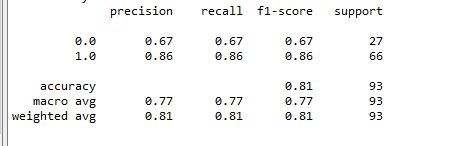


knn,svm,rf değerlerine göre knn ve rf’in daha yüksek kesinlikte olduğu, svm değerinin daha düşük kesinlikte olduğu görülmektedir.

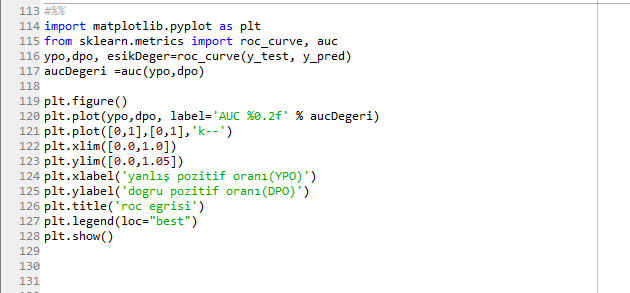
Liste 8

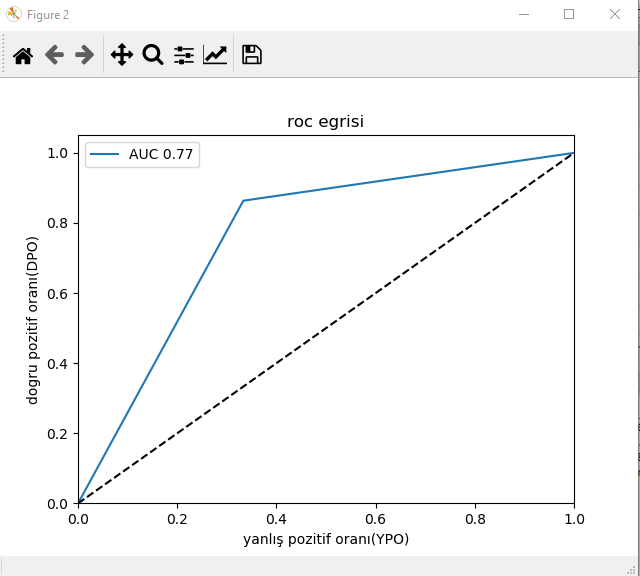


confusion\_matrix() fonksiyonu ile oluşturulan hata matrisi hm adlı listeye atanmıştır. Elde edilen hata matrisine ait performans değerlendirme ölçütleri kesinlik, duyarlılık ve f1 skorlama değerlerini içeren sonuçlar listelenmiştir.



Liste 8



Hata matrisi doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif olmak üzere dört kısımdan meydana gelir. Doğru pozitif ve doğru negatif örneklerin sayısının çok olması sınıflandırma performansı için olumlu yönde etki yapacaktır. Dengeli dağılıma sahip olmayan veri setleri için ROC eğrisi kullanılarak performans değerlendirmesi yapılabilir. Doğru pozitif oranı ve yanlış pozitif oranı ROC eğrisini oluşturan iki temel unsurdur. Doğru pozitif oranının yüksek, yanlış pozitif oranının düşük olması performans için olumlu olarak nitelendirilir. ROC eğrisinin altında kalan alanın 1’e yakın olması sınıflandırma performansı açısından hedeflenen bir sonuçtur. AUC değeri 0 ile 1 arsında değişir arzulanan değeri ise 1’e olabildiğince yakın olmasıdır.

Şekil 2. ROC eğrisi

KAYNAKLAR

https://www.kaggle.com/uciml/biomechanical-features-of-orthopedic-patients