

**T.C.
YALOVA ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
YAPAY SİNİR AĞLARI DERSİ**

**AMAZON ÇALIŞANLARININ
ŞİRKET İÇERİSİNDE BULUNAN KAYNAKLARA ERİŞİM İZİNLERİ**

HAZIRLAYANLAR

ATAKAN CEVAT COŞKUNDERE

150101070

MUHAMMED FURKAN BALTACI

170101065

ARALIK 2019

GİRİŞ

Projenin Tanımı

Şirket çalışanları'nın,şirketteki kariyerlerine göre her kaynağa erişim sağlamaları tehlikeli olabilir.Bilindiğine göre şirket kaynakları farklı yöneticilerin erişiminde ve sorumluluğundadır.

Veriler 2010 ve 2011'den toplanan gerçek verilerden oluşmaktadır. Amazon çalışanlarının zaman içinde kaynaklara elle erişmesine izin veriliyor veya reddediliyor. Görünmeyen bir çalışan grubu için onay / reddi öngörmek için bu tarihsel verilerden öğrenebilecek bir algoritma oluşturarak tahmin çalışması yaptık.(Dataset'i Kaggle.com dan edindik.)

Dataset Tanımı

Column Name	Description
Action	erişim onaylandıysa 1, onaylanmadıysa 0
Resource	Personel ID (Ancak kullanılmadı)
Mgr_Id	Yönetim Idleri
Role_Rollup_1	Şirket rolü 1 (bkz:Mühendis)
Role_Rollup_2	Şirket rolü 2 (bkz:Perakende Satış)
Role_Deptname	Rol açıklaması (bkz:Satışçı)
Role_Title	İşletme Ünvanı(bkz:Kıdemli Mühendis)
Role_Family_Desc	Rol Tanımı (bkz: Yazılım Mühendisi)
Role_Family	Yöneticilik Yetkisi(bkz: Satış Müdürü)
Role_Code	Şirketteki Ünvan Kodları (bkz:Yönetici)

Action sütunu harici diğer sütun verileri giriş (input) verileri olarak kullanıldı.Action sütunu ise çıkış(output) verisi olarak kullanıldı.

Projenin Faydası

Şirket içerisinde işleyişin daha adil bir şekilde yapılmasını sağlar.Eğer proje doğru şekilde kullanılırsa, erişim hakkına sahip bir çalışan erişiminden mahrum kalmayıp gerekli yetkiyi alırsa,işini daha verimli yapar ve bu şekilde iş veren bu verimi kar'a çevirmiş olur.

Dataset'in Kullanıldığı Başka Projeler

Shijian Tang, Jiang Han and Yue Zhang adlı kişiler Amazon Employee Access Control System çalışmalarında kullanmışlardır.Çalışmalarında tarih belirtilmemiştir.(Department of Electrical Engineering)

Email: sjtang@stanford.edu

ANALİZ

Auc-Roc Eğrisi Nedir ?

Auc-Roc Eğrisi,eşik ayarlarında sınıflandırma problemleri için performans ölçümünde kullanılan bir yöntemdir.Auc'un yüksek olması(tabloda sol üste yani 1'e yakınlığı) model tahmininin dağa sağlıklı yapıldığını gösterir.(1'e 1, 0'a) True Positive ve False Positive olarak iki parameter çizer.

Confusion Matrix nedir ?

Çıktının iki veya daha fazla olduğu yerlerde makine öğrenmesi sınıflandırması problemleri performansını ölçmek için vardır . Tahmin edilen ve gerçek değerlerle 4 farklı kombinasyonlu bir tablodur .True Positive,False Pozitif,False Negative,False Pozitif

Kullandıklarımız

Anaconda Jupyter-Notebook üzerinde çalıştık.Çalışmada Danışmanlı Öğrenme algoritması kullanılmıştır.Random olarak belirlediğimiz bias ve ağırlık değerlerimizi oluşturan bir fonksiyon oluşturuldu.

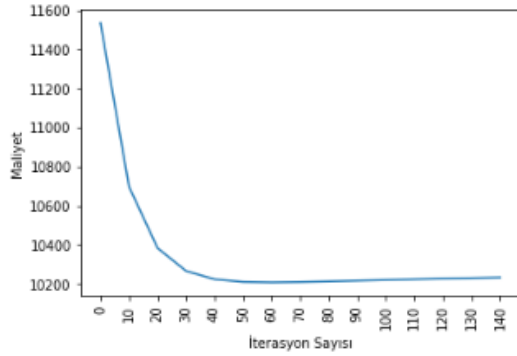
İleri Yayılım ,(forward_propagation) Metodunda bu ağırlıklar ve biaslar kullanılarak hem tahminler hemde loss değerleri elde edildi.Bu değerler toplanarak maliyet bulundu(computeCost).Geri Yayılımda (backward_propagation), amaç geriye doğru gidilip,ağırlıkların ve bias değerlerin türevleri elde edildi.Sonrasında Update metodunda parametreler güncellendi.Prediction metodunda ise ileri yayılımdan dönen değerler kullanılarak tahminler yapıldı.Son olarak “ikiKatmanlıYsa” metodunda modelimiz oluşturuldu.Not:İleri Yayılımda tanh(-1,1) ve sigmoid(0,1) aktivasyon fonksiyonları kullanılmıştır.Öğrenme Katsayısı(learning rate),model için fazlasıyla önemlidir.Ağırlıkların değişim miktarını belirlemektedir.Eğer büyük seçilirse ağı minimuma ulaşamayabilir çok küçük seçilirse de öğrenme zamanını arttıracaktır.

SONUÇLAR

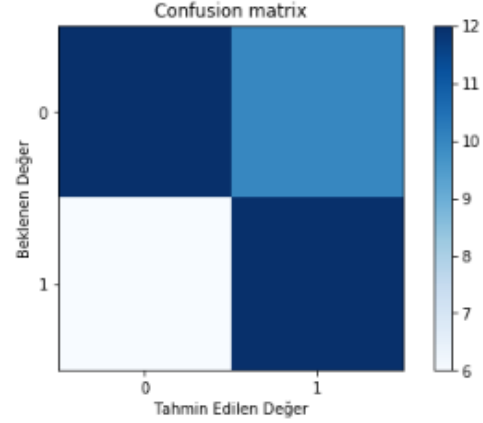
İterasyon Sayısı 150 sabit,Katman Sayısı ve Öğrenme Oranı değişimi

Katman Sayısı=13

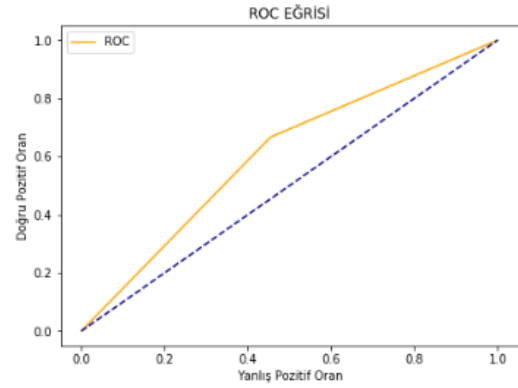
Öğrenme Oranı=0.03



train doğruluğu:55.34591194968554 %
test doğruluğu:60.0 %

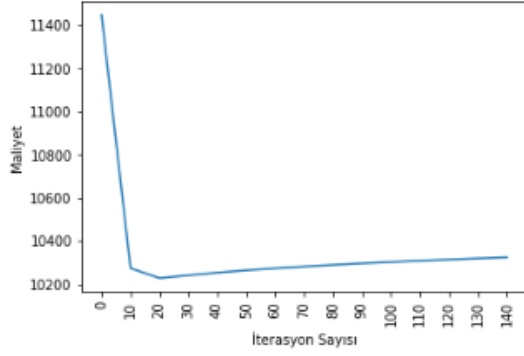


train doğruluğu:55.34591194968554 %
test doğruluğu:60.0 %

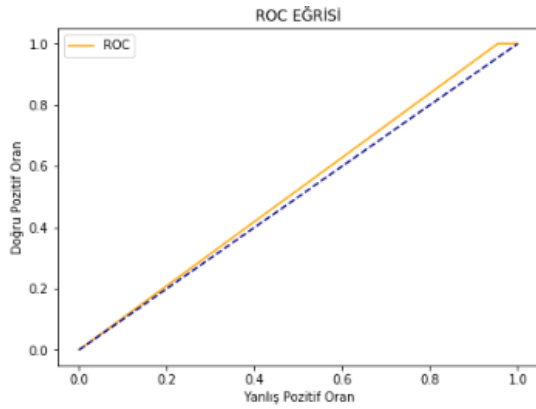
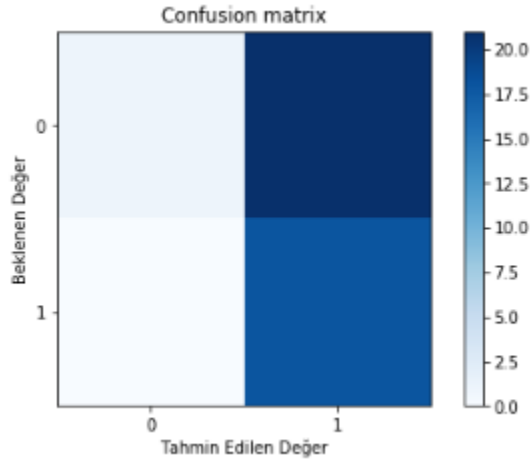


Doğruluk 0.606060606060606

Öğrenme Oranı=0.1 oranına getirildiğinde;

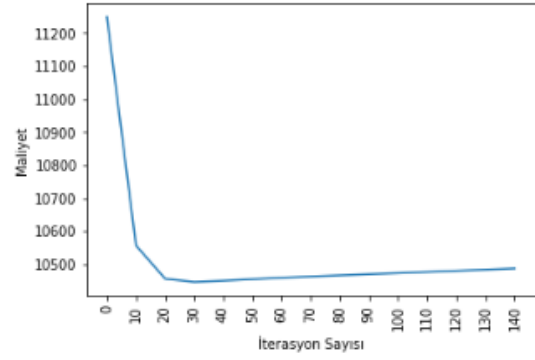


train doğruluğu:55.34591194968554 %
test doğruluğu:47.5 %

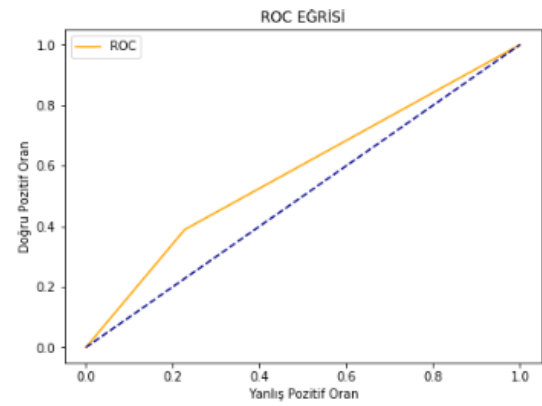
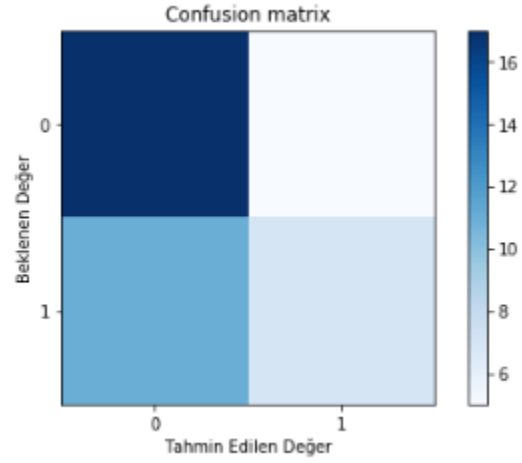


Doğruluk 0.5227272727272727

Katman Sayısı=8 yapıldığında;



train doğruluğu:57.23270440251572 %
test doğruluğu:60.0 %

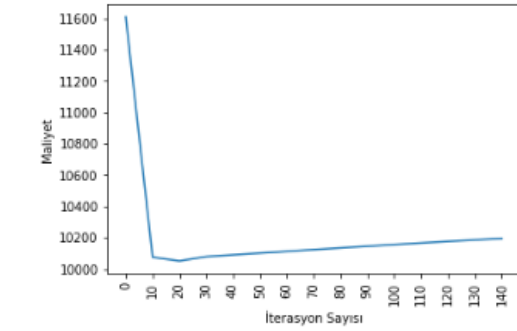


Doğruluk 0.5808080808080808

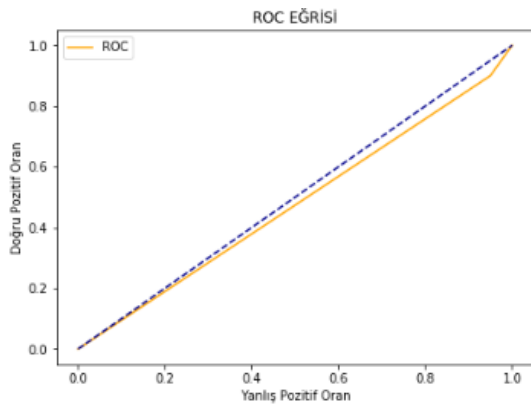
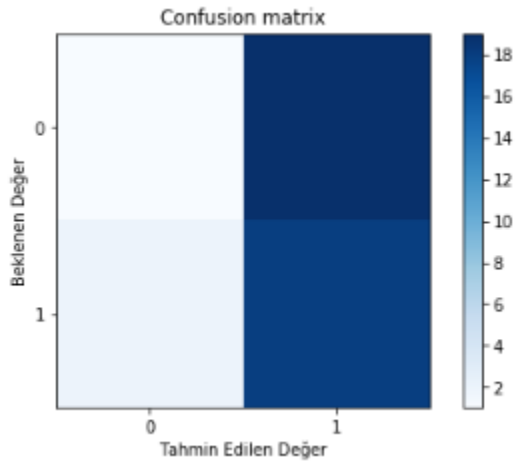
Öğrenme oranı'nın 0.03 den 0.1 e yükseltilmesi tahmin başarılarını düşürmüştür.Burdan anlıyoruz ki,parametreleleri güncellerken hedeften uzaklaşmıştır.

Modelimiz için;aynı durumdan katman sayısı 13 den 8 e indirildiğinde %13 lük bir artış olduğu gözleniyor.

Katman Sayısı=15 yapıldığında;



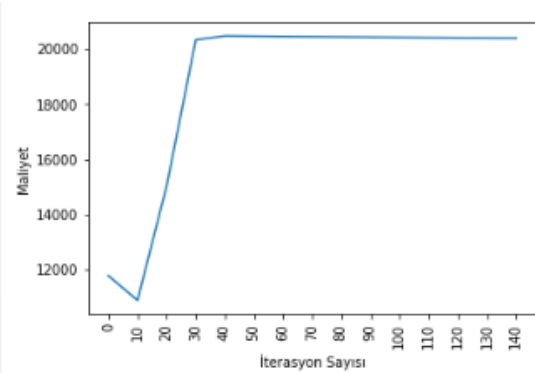
train doğruluğu:58.490566037735846 %
test doğruluğu:47.5 %



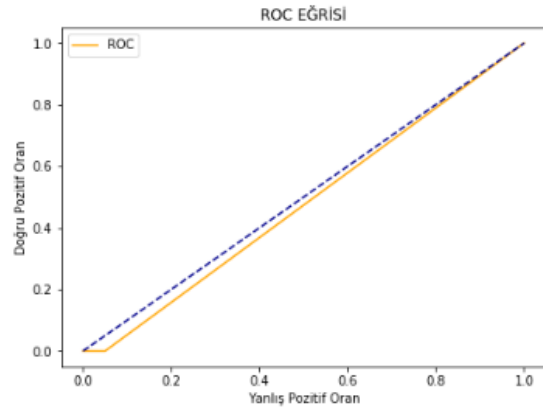
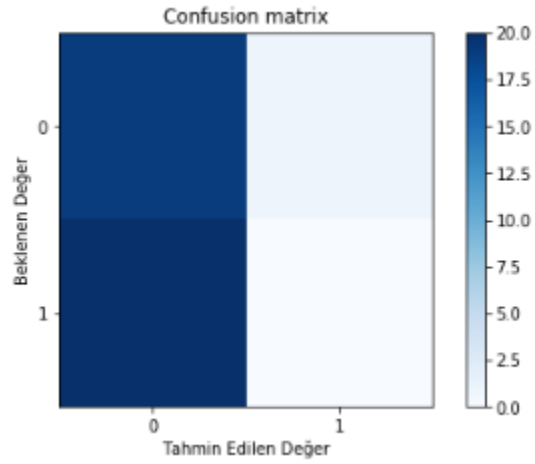
Doğruluk 0.47500000000000003

Katman sayısı 8 den 15'e çıkarıldığında performansda düşüş gözlemlendi.0 ların tahmininde problem olup,1 beklenen değerlerin daha doğru tahmin edildiği gözlemlendi.

Son olarak Öğrenme Oranı=0.7 oranına getirildiğinde;



train doğruluğu:45.911949685534594 %
test doğruluğu:47.5 %



Doğruluk 0.475

Burada öğrenme oranının fazla artırılması sonucunda maliyetteki artışı ve performansın azalışı görüntülendi.