Banka müşterisi kayıp analizi

Abdullah Jamous

Siber Güvenlik ve Kriptografi Bölümü, Yıldız Teknik Üniversitesi, Türkiye

Özet İçinde bulunduğumuz çağda, veri analizi ve makine öğrenmesi tüm kuruluşlar için en önemli önceliktir. Nesnelerin İnterneti (IoT), büyük veri ve bulut bilgi işlem gibi İnternet teknolojilerinin hızla gelişmesiyle birlikte, bireyler, hükümet yetkilileri ve ordu, veri analizinde sorunlarla karşı karşıyadır. Üstel veri büyüme hızı göz önüne alındığında, araştırmacılar için zorlu bir görev, akıllı şehirler tasarlarken büyük miktarda veriyi etkin bir şekilde nasıl yönetecekleridir. Önerilen sonuçlar, %60 ile %80 arasında değişen sonuç doğruluk değerleri ile yüksek düzeyde bir belirsizlik elde etti.

Anahtar Kelimeler Makine Öğrenmesi, Naive Bayes, k-en yakın komşuluk, karar ağaçları.

I. Giriş

Kimin kalacağını ve gelecekte kimleri kaybedebileceğimizi tahmin edebilmek için müşteri kaybı analizi yapmak çok önemlidir. Makine öğrenimi kullanılarak yapılan veri analizi, geri çekilmeye maruz kalan ve katkıda bulunmaya çalışan müşterilerin örüntüsünün tahmin edilmesine katkıda bulunmuştur. Bu müşterilerle ilgilenmek ve karşılaştıkları sorunları bilmek ve Bu sorunlardan kurtulmak için bankalar, bu analizleri yaptıktan sonra, mümkün olan en fazla sayıda müşterinin kendilerinde kalmasını sağlamak için performans verimliliğini ve iş kalitesini artırabilir.

II. Veri Kümesi

Veri setini kaggle.com'dan aldık, veri setinde 10.000 satır ve 12 sütun var, sütunlar arasında sütunun kendisi ile ilgili verileri içerdiği için sildiğimiz bir sütun var, o da creadt_id ve ayrıca iki sütunumuz var. kategorik veri içerenleri x algoritmasını kullanarak sayısal verilere çevireceğiz, ayrıca eksik veri var mı kontrol edeceğiz, eksik veri olan satırları sileceğiz, gerekirse veriyi indireceğiz

III. Sınıflandırma Modelleri

Bu deneyde kullanılan yöntemler kısaca bahsetmek istiyorum:

I. k-en yakın komşuluk

KNN algoritması sınıflandırılmak istenen bir veriyi daha önceki verilerle olan yakınlık ilişkisine göre sınıflandıran bir algoritmadır. Örneğin; k = 3 olarak alırsak, yeni gelen verinin eski verilere olan uzaklıkları ölçülür ve en yakın 3 tanesi belirlenir. K sayısı genelde 3 ile 7 arasında seçilir. K sayısı tek sayı olması gerekiyor. K=5 seçilirse k=3 socundan daha başarılı çıkıyor.

II. Naive Bayes

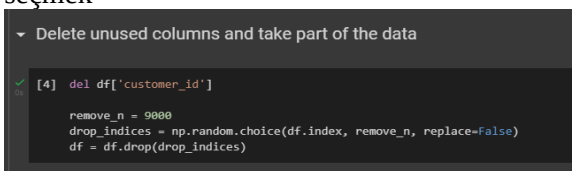
Naïve Bayes sınıflandırması olasılık ilkelerine göre tanımlanmış bir dizi hesaplama ile, sisteme sunulan verilerin sınıfını yani kategorisini tespit etmeyi amaçlar. Bir eleman için her durumun olasılığını hesaplar ve olasılık değeri en yüksek olana göre sınıflandırır. Az bir eğitim verisiyle çok başarılı işler çıkartabilir.

III. karar ağaçları (decision trees)

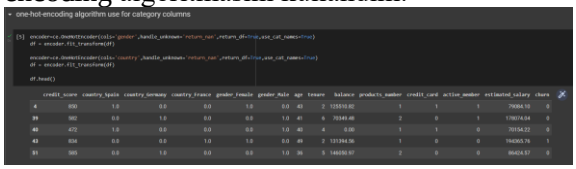
Amaç veri öğeleri basit kurallar kuralları bu kuralları öğrenerek bir değişkenin değerini tahmin eden modeli oluşturmaktır. Algoritma eksik değerleri desteklemez makineyi eğitmeden önce eksik değerleri (eksik değer) hesaplamamız gerekir. Rastgele Orman görünümü karar ağacı (Karar ağacı) gibi hem sınıflandırma (Sınıflandırma) hem de regresyon (Regresyon) problemlerinde kullanılabilir. Çalışma mantığı birden fazla karar ağacı oluşturur. Bir sonuç üreteceği zaman bu karar ağaçlarındaki ortalama değer alınır ve sonuç üretilir.

IV. Deneysel Analiz

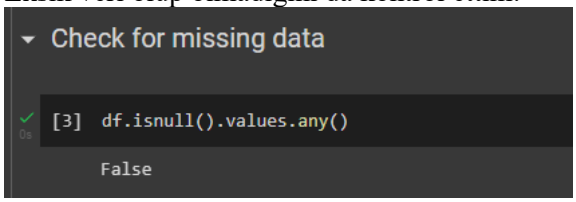
Veri kümesinde yapılan ön işlem : Eğitimde kullanılmayan sütunları silmek ve 1000 satır veri seçmek



Kategorik veri içeren iki sütunumuz var, bu veriyi sayısal veriye dönüştürmek için one-hot-encoding algoritmasını kullandım.

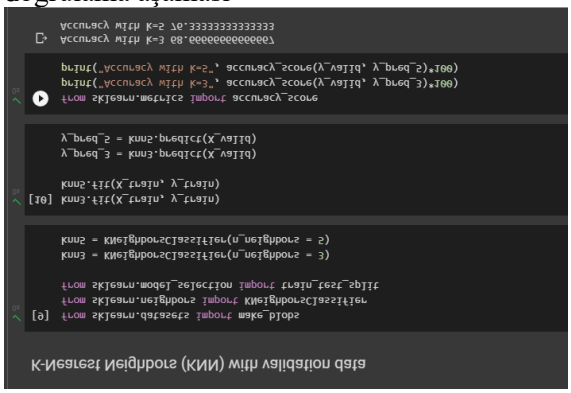


Eksik veri olup olmadığını da kontrol ettim.

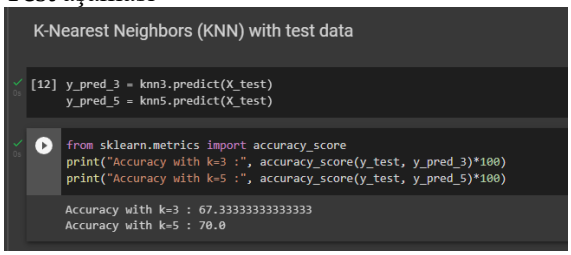


Ondan sonra datayı parçalara ayrıldım, 400 satır eğitmek için, 300 satır doğrulamak için, 300 satır test yapmak için

İkinci aşama doğrulama aşamasıdır:
Knn algoritması
doğrulama aşaması

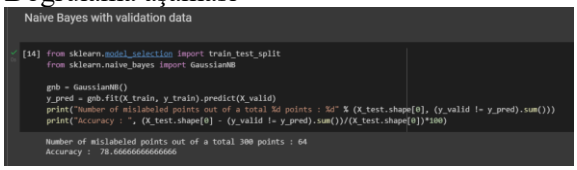


Test aşaması

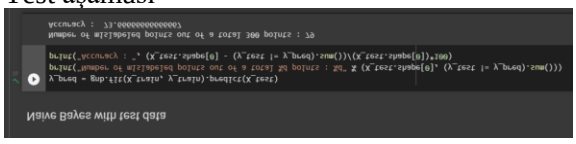


Önemli gördüğüm şey k değeri ne kadar büyük olsa o kadar iyi başarı oranı alabiliriz.

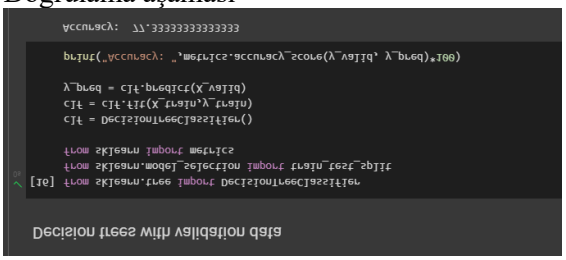
Naive Bayes algoritması:
Doğrulama aşaması



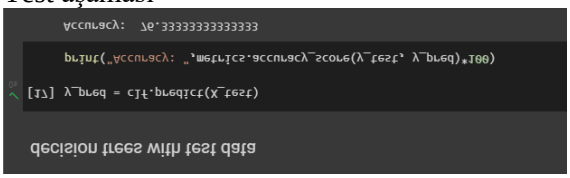
Test aşaması



Decision trees
Doğrulama aşaması



Test aşaması



Gördüğüm kadarıyla öğretmiş olduğum algoritmalardan en iyi sonuç veren Naive Bayes algoritmasıdır.

IV. Öğrenme modelinizin başarısın

Başarı oranı 60% lardan 80%lere kadar.