



MNİST VERİ SETİ

Sunum 3

Özet

Veri setine makine öğrenmesi uygulandı



1. GİRİŞ

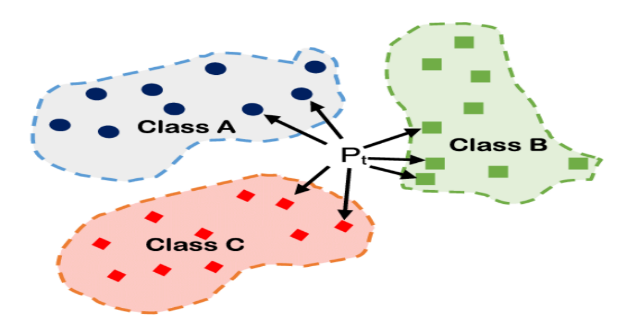
Mnist veri setine 10000 kirli veri eklendi. 4 adet makine öğrenme modeli kuruldu. Kurulan modeler eğitildi ve model tuning işlemi yapılarak en uygun parametreler alınarak en uygun modeller kuruldu.

1. Kulanılan Modeller

KNN , SCV RBF , Yapay Sinir Ağları (mlpc) ve CART modeleri kullanıldı

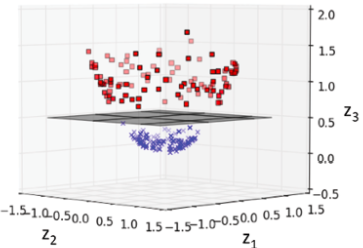
1. Yapılan Çalışmalar

K-En Yakın Komşu Modeli



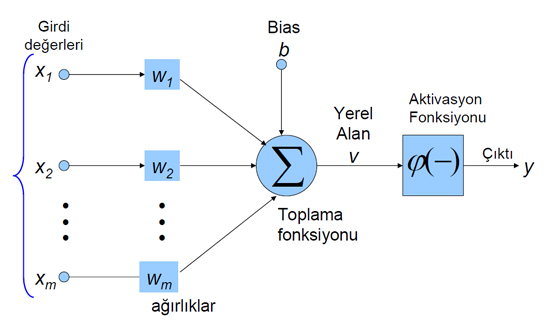
KNN (K-Nearest Neighbors) algoritması, en basit anlamıyla, tahmin edilecek değerin bağımsız değişkenlerinin oluşturduğu vektörün en yakın komşularının hangi sınıfa yoğun olduğu bilgisi üzerinden sınıfını tahmin etmeye dayanır. KNN algoritması iki temel değer üzerinden tahmin yapar

# Destek Vektör Makinesi



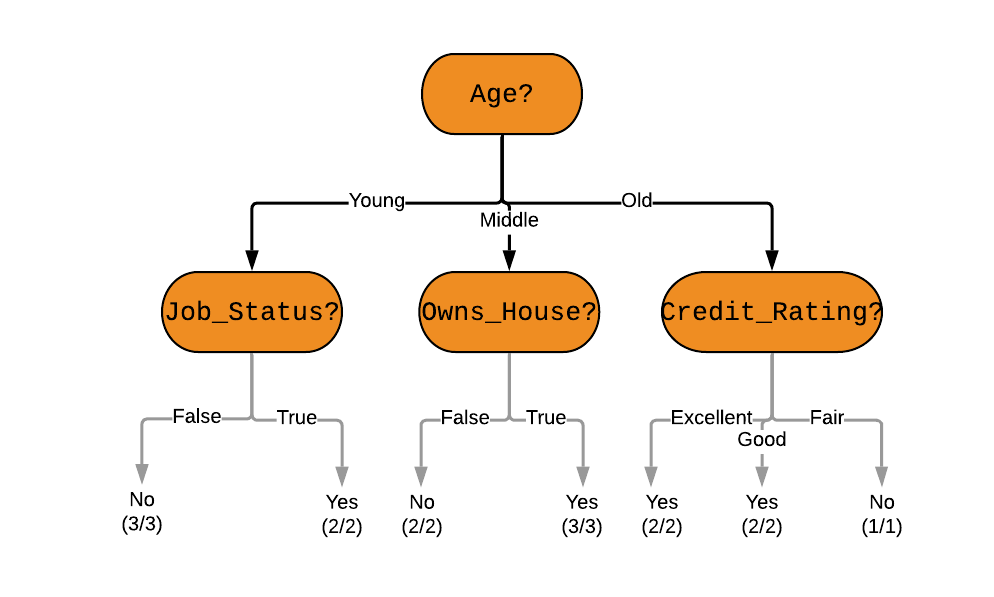
SVC RBF algoritması, veri noktalarını sınıflandırmak için bir karar sınırı oluşturur. Ancak, diğer SVC türlerinden farklı olarak, RBF çekirdek fonksiyonunu kullanır. Bu çekirdek fonksiyonu, veri noktalarını yüksek boyutlu uzaylara projelendirmek ve ardından bu uzayda sınıflandırma yapmak için kullanılır.

Yapay Sinir Ağı

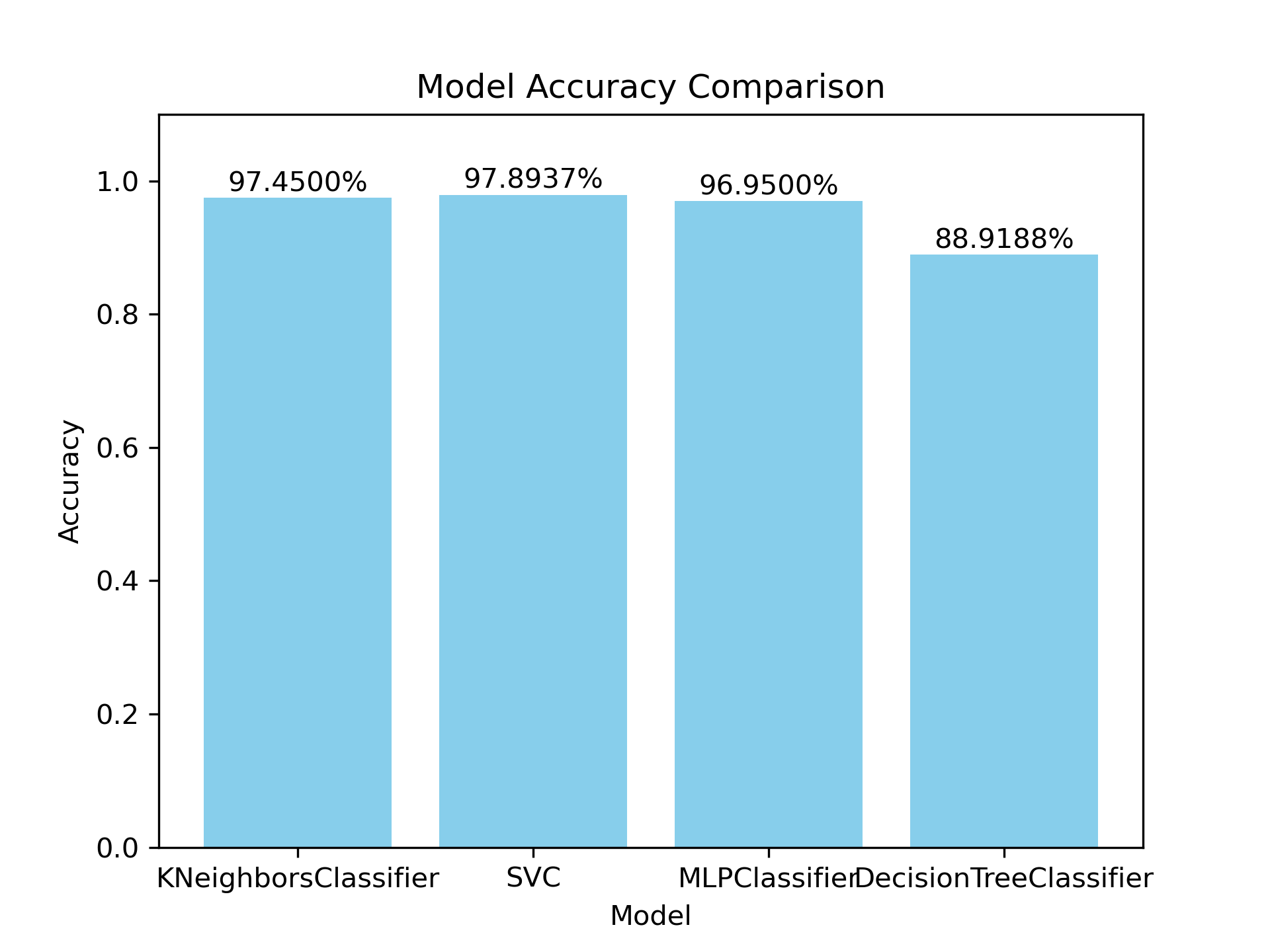


insan beyninin bilgi işleme şeklini referans alan sınıflandırma ve regresyon problemleri için kullanılabilen kuvvetli makine öğrenmesi algoritmalarındandan birisidir #### Bu algoritma, sınıflandırma ve regresyon problemleri için kullanılabilir. Sınıflandırma problemlerinde, veriler farklı kategorilere ayrılır ve bir veri noktasının hangi kategoriye ait olduğunu belirlemek için kullanılır. Regresyon problemlerinde ise, verilerin arasındaki ilişkiyi modellemek ve bir değişkenin değerini tahmin etmek için kullanılır.

Karar Ağaç Modeli



Bu algoritmanın temel amacı, veri setindeki karmaşık yapıları basit karar yapılarına dönüştürmektir. Yani, veri setindeki karmaşık ilişkileri anlamak ve bu ilişkileri daha anlaşılır ve yalın bir şekilde ifade etmek için kullanılır.¶ Özellikle heterojen veri setleri üzerinde etkili olan bu algoritma, belirlenmiş bir hedef değişkene göre homojen alt gruplara ayırma işlemi gerçekleştirir. Yani, veri setindeki farklı değişkenler arasındaki ilişkileri ve etkileşimleri belirleyerek, bu bilgiyi kullanarak veri setini daha küçük ve daha anlamlı parçalara böler.

Modellerin Karşılaştırılması

1. Sonuç

MNIST veri seti indirildikten sonra, veri setine 10.000 adet kirli veri eklendi. Bu kirli veriler, yanlış etiketlenmiş ya da bozulmuş görüntüler içerebilir. Bu kirli verilerin eklenmesi, modellerin performansını daha zorlu ve gerçekçi bir ortamda test etmek amacıyla yapılmıştır.

Destek Vektör Makinesi (SVC), diğer modellere kıyasla en yüksek kesinlik değerine ulaşarak MNIST veri seti üzerinde en iyi performansı göstermiştir. Bu sonuç, SVC'nin kirli verilerle bile yüksek doğrulukta sınıflandırma yapma yeteneğini ortaya koymaktadır.