T.C

KIRIKKALE ÜNİVESİTESİ

MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

MAKİNE ÖĞRENMESİ DERSİ FİNAL ÖDEVİ

Hazırlayan Öğrenci

130255065

NECDET YAPICI

Ders Sorumlusu

Dr. Arş. Gör. Enes Ayan

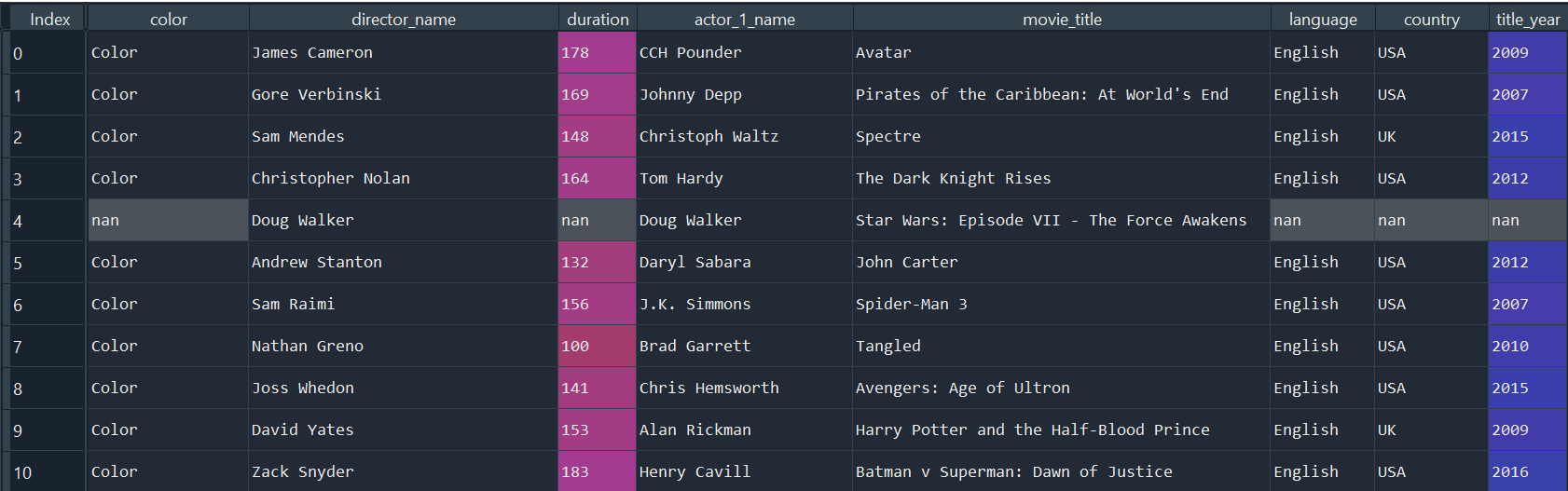
Kırıkkale 2020

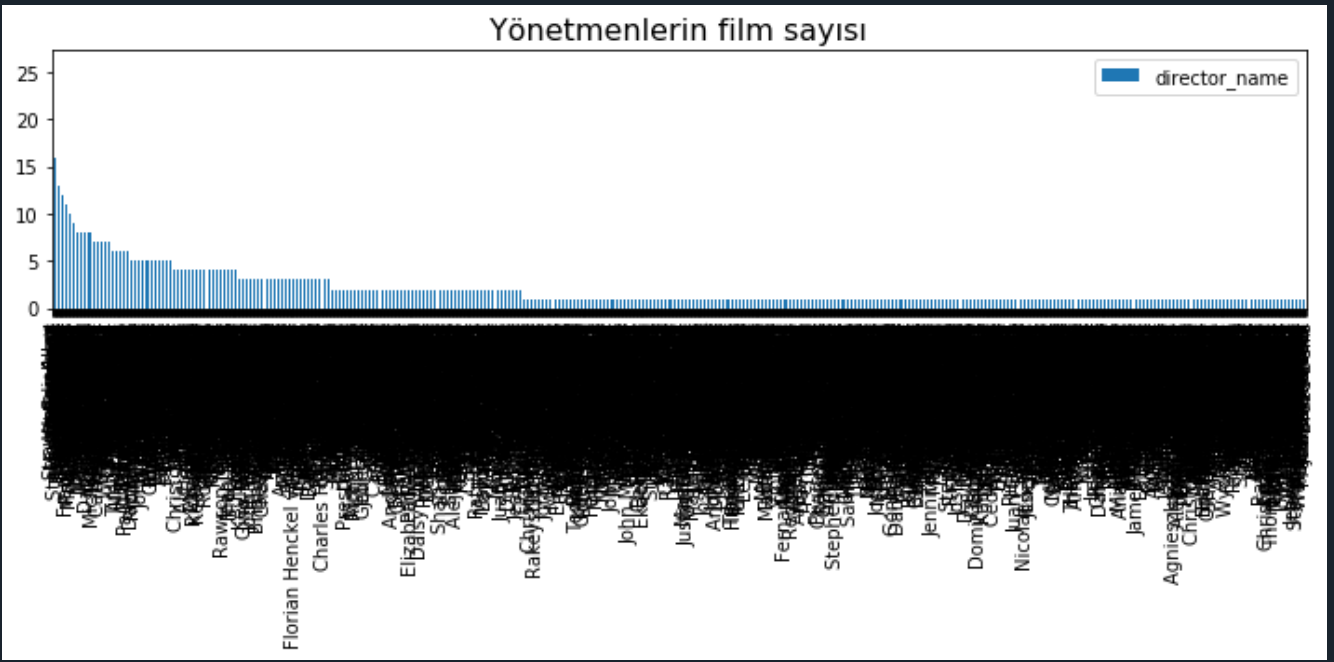
2019-2020 Bahar Dönemi Makine Öğrenmesi Dersi Final Proje Ödevi

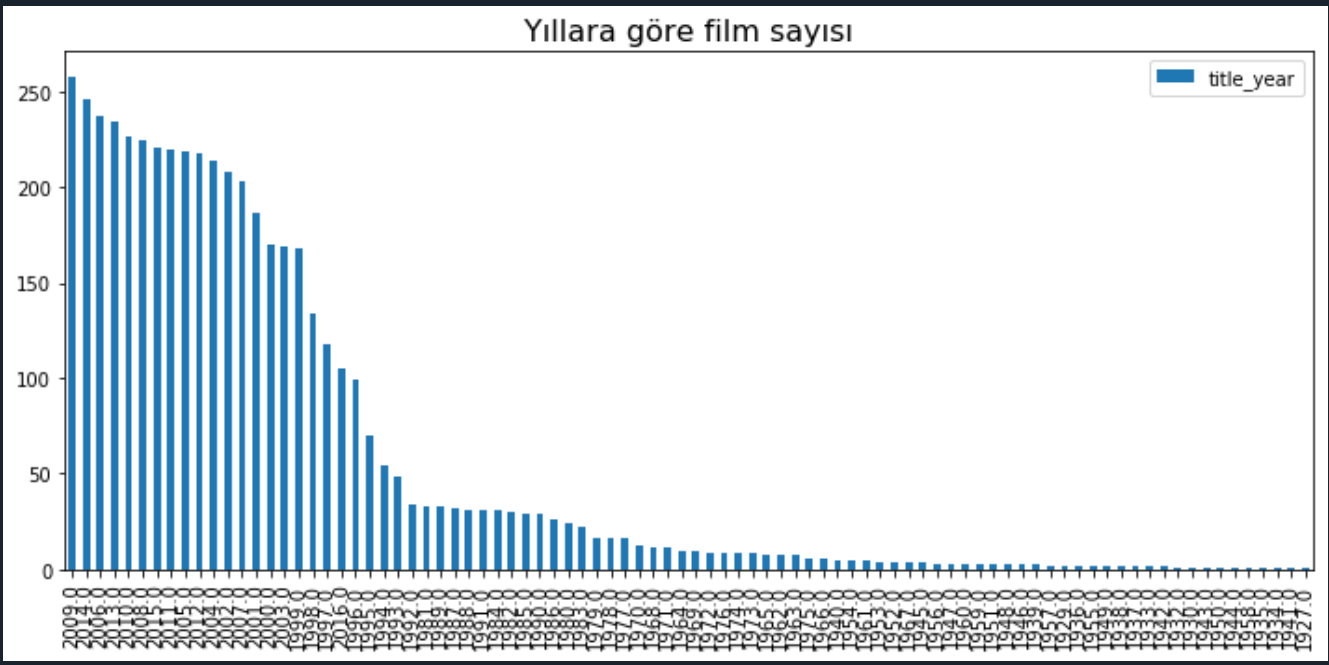
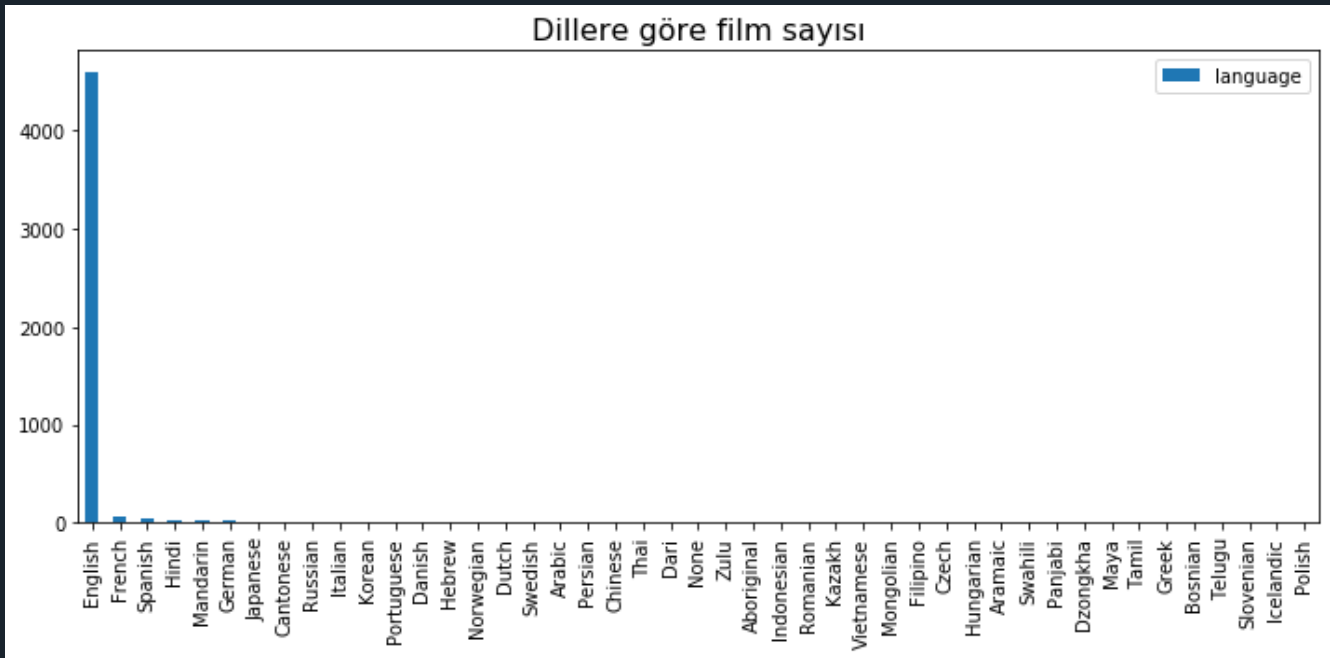
1- Analiz için kullanacağınız veri setini tanıtınız, problemin türünü girdilerini ve çıktısını açıkça ifade ediniz. Grafikler ile görsel ekleyebilirsiniz. Ayrıca kullandığınız veri setinin linkini de veriniz.

Datasetimi, 1916-2016 yılları arasında yapılmış filmler ve bu filmlere ait etkileşimlerin olduğu bir set olarak seçtim. Datasetimim içinde 28 column bulunuyor ama benim kullanacağım sadece 8 column. Bu columnlar; color (object) , director\_name (object), duration (float64) , actor\_1\_name (object) , movie\_title (object) , language (object) , country (object) , title\_year (float64). Çıktı olarak yönetmenlere, dillere ve yıllara göre film sayısı olarak kullandım.

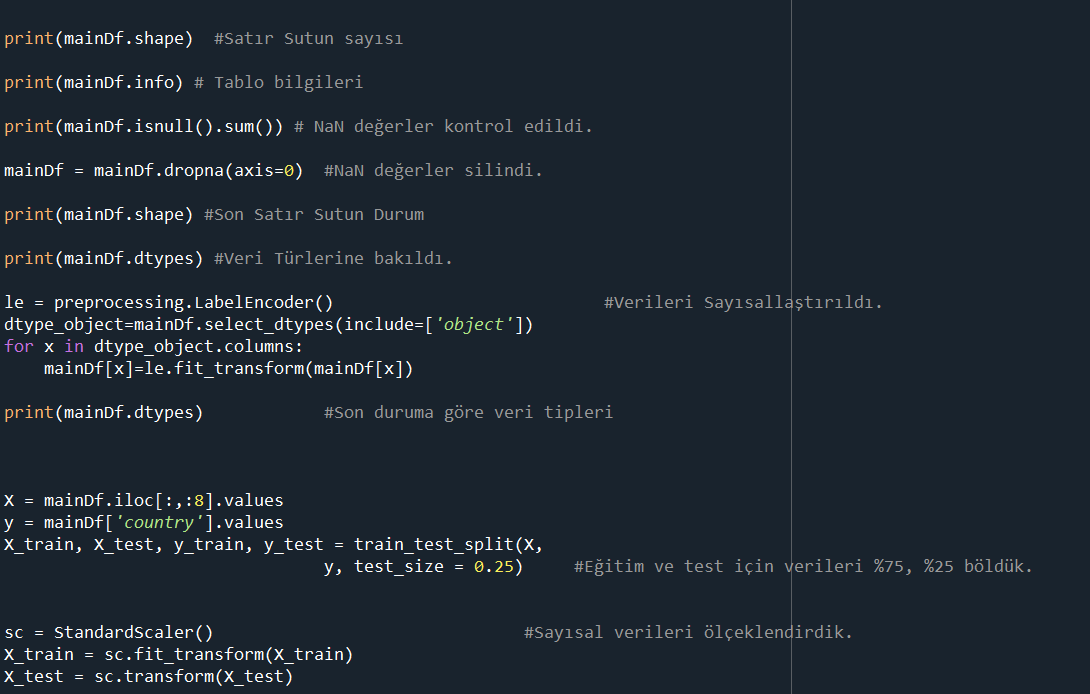
Kullanılan datasetin linki : <https://www.kaggle.com/karrrimba/movie-metadatacsv>







2- Kullanacağınız veri setini analiz ediniz, algoritmaya kadar ki ön işlem aşamalarını gösteriniz (boş veri silme, kategorik veriyi sayısala dönüştürme, ölçekleme, özellikleri ve çıktıyı ayırma, veriyi eğitim ve test olarak bölme gibi)



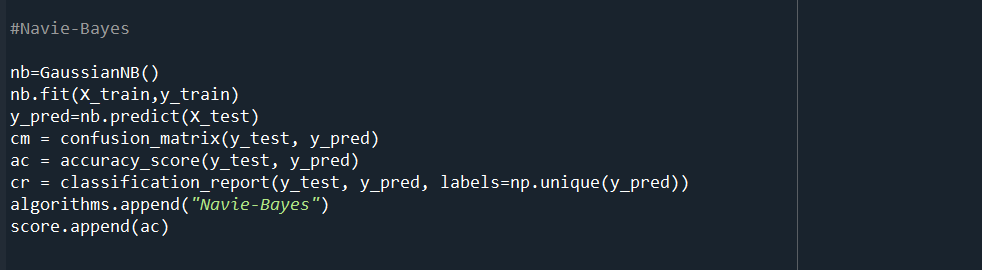
3- Seçeceğiniz yöntem veya yöntemler ile veriye dair problemin çözümünü gerçekleştiren aşamaları yazınız. (Probleminiz ne ise regresyon, sınıflandırma, kümeleme boyut küçültme gibi uygun algoritmayı seçiniz).

**KNN**

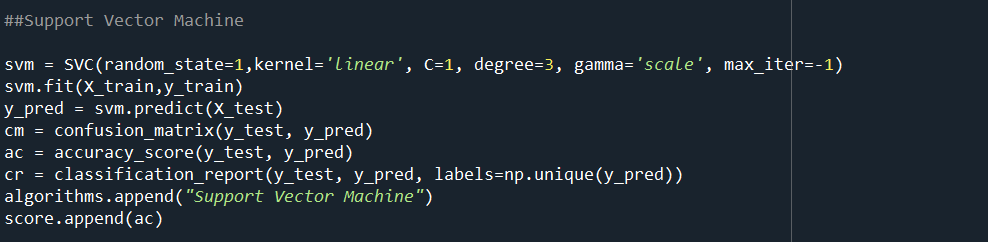
KNN yönteminde 2 tür denedim. KNN-2 de en mantıklı değer 9 olarak buldum ve 1. De bunu kullandım.



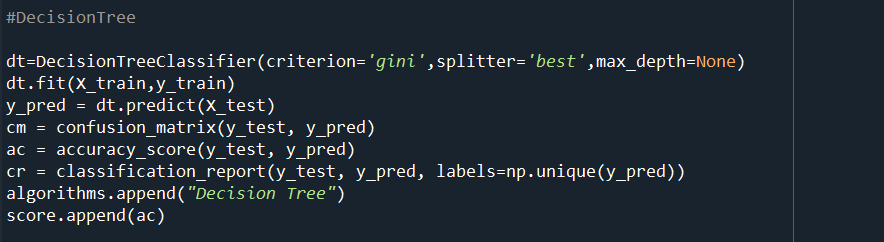
**Navie-Bayes**



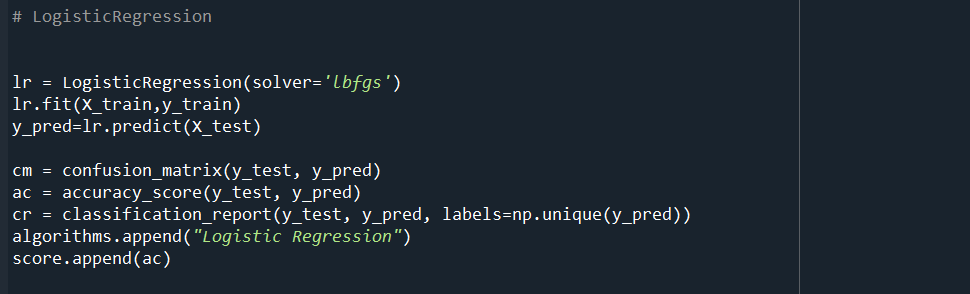
**SUPPORT VECTOR MACHİNE**



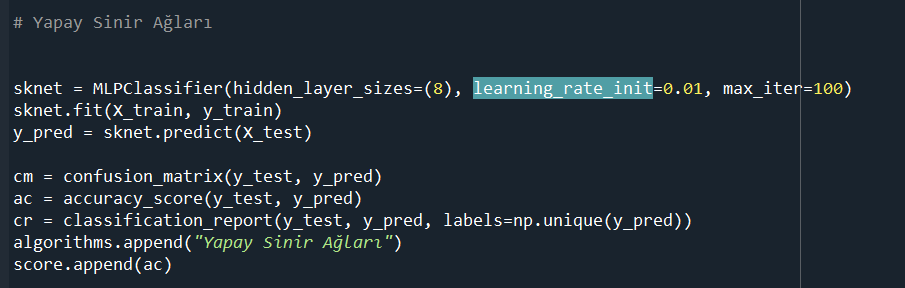
**DECİSİON TREE**



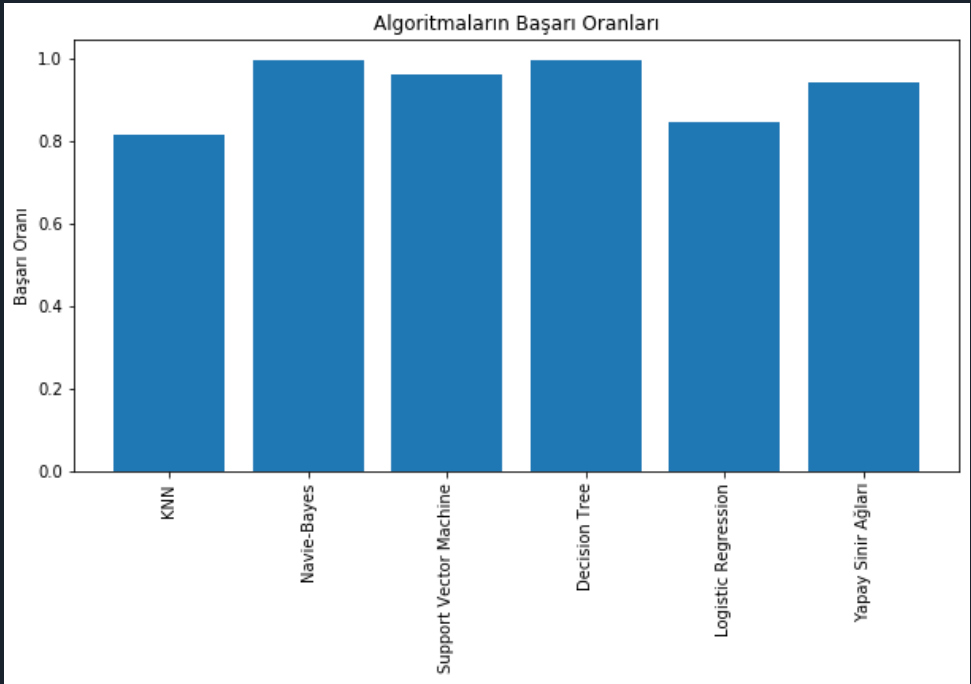
**LOGICTIS REGRESSION**

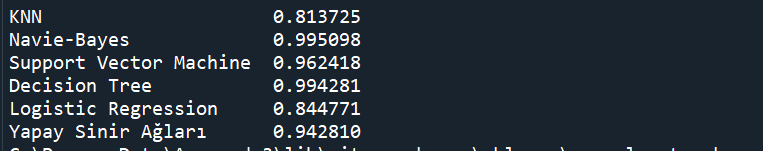


**YAPAY SİNİR AĞLARI**



4- Sonuçları bireysel olarak yorumlayınız sonuçlarınız hangi yöntemde iyi çıktı sizce sebepleri veya sonuçlarınız neden kötü çıktı bunun sebepleri nelerdir?





Bu problemim de sınıflandırma kararı aldım. Algoritmalardan aldığım sonuçlar üstte göründüğü gibi. Genel anlamda başarılı sonuçlar aldım.

Karar ağacında %99.4281 gibi bir sonuç almak yeterince başarılıydı. Bu sonuçlara göre en karmaşık olan KNN algoritması. Bu algoritmada karmaşıklığı giderebilir miyim diye for döngüsüyle analiz yapsam da maksimum alacağım oran %81.3725 oldu.

Sonuç olarak ;

En başarılı algoritma : Decision Tree

En başarısız algoritma : KNN

Not: Her bir maddenin altına çıktıları ve o çıktıyı elde ettiğiniz kod buluğunu yazınız. Haricinde tüm kodun yer aldığı tek bir Python dosyasını bu rapor ile birlikte gönderin. Gönderim şeklinde ise bu rapor ve tek Python dosyasını bir rar ya da zip dosyası hale getirin bu dosyayı isminiz ile isimlendirip [enesayan@yandex.com](mailto:enesayan@yandex.com) mail adresine final sınavından önce gönderin. Rapor üzerinde değişiklik yapmayın font ile oynamayın.