

KÜTAHYA SAĞLIK BİLİMLERİ ÜNİVERSİTESİ
Bilgisayar Mühendisliği



Yapay Zeka Dersi Proje Tasarım Raporu

Celal ALTIN

March 14, 2024

Bu raporumda yapay zeka dersinde yapmış olduğum proje anlatılmaktadır.Raporun akışı aşağıdaki gibidir:

1. Giriş
2. Literatür Araştırması
3. Metodoloji
4. Kullanılacak Veriler
5. Beklenen Sonuçlar
6. Kaynakça

1 Giriş

2019 yılında başlayan COVID-19 salgını dünya çapında ciddi sağlık, ekonomik ve sosyal etkilere yol açmıştır.(Mart 2022 de yayınlanan bir habere göre [1] dünyada ölüm sayısının 18 milyonu aştığı hesaplanıyor, bu sayı ülkemizde ise Sağlık bakanlığının verilerine göre 31.05.2022 tarihine kadar 98.965 kişinin hayatını kaybetti yönünde.)Bugün bile bu etkilerin sonuçları geçmiş değildir.

Figure 1: Dünya Covit Haritasi [2]



Dünya çapındaki bu problemi daha iyi anlayabilmek ve ileride olası hastalıklarda daha etkili mücadele edebilmek için bilgisayar(Makine öğrenmesi kulanılarak yapılan bir araştırmaya göre sosyal medyanın da etkisiyle aşıya olan olumlu bakış artmıştır [3]) ve bilgisayar özelinde yapay zeka teknolojisinden faydalanmak en akılcı yöntemlerden birisidir.Bende bu projemde çeşitli kaynaklardan bulduğum Covit-19 verilerini kullanarak olası bir salgın hastalık dumunda gerçekleşebilecek seneryoyu gün yüzüne çıkartmayı amaçlıyorum.

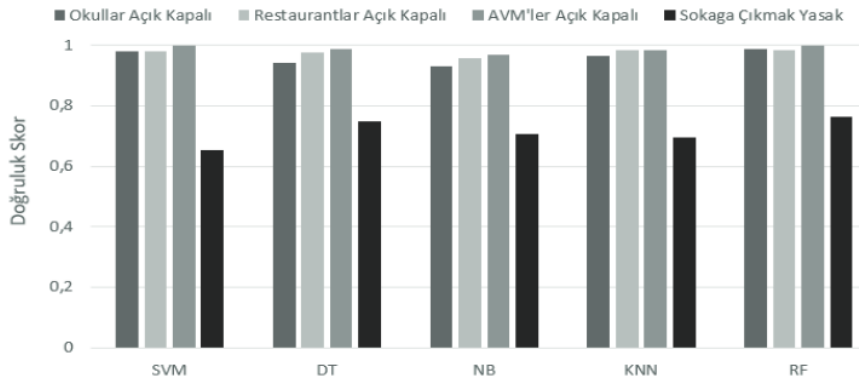
2 Literatür Araştırması

Koronavirüs, 2019 yılının Aralık ayında ilk olarak Çin'in Wuhan kentinde ortaya çıkmış ve 11 Mart 2020'de Dünya Sağlık Örgütü tarafından pandemi olarak ilan edilmiştir. Vaka sayılarını kontrol altına almak için pek çok ülke karantina, sokağa çıkma yasağı ve sosyal alanların bir süreliğine kapatılması

gibi çeşitli önlemler almıştır. Doğrulanmış vaka tahminlemesi pandemide olası planlamalar için büyük önem taşımaktadır. Gelecek verilerinin gerçeğe en yakın bir şekilde tahminlenmesi; pandemi döneminde lojistik, tedarik, hastane personel ve malzeme planlaması için kullanılabileceği gibi aşılama senaryolarında da girdi olarak kullanılabilir. Literatürde doğrulanmış vaka tahmininde makine öğrenmesi, bölme model, zaman serisi analizi gibi pek çok yöntem kullanarak tahminleme yapılan çalışmalar vardır. Bu çalışmada, Amerika Birleşik Devletleri'ndeki doğrulanmış vaka sayılarını kullanarak gelecek günlerdeki vaka tahminlerini çeşitli makine öğrenmesi modelleri yapılmıştır. Python ve R programlama dili kullanılarak yapılan tahminlemeler Prophet, Polinom Regresyon, ARIMA, Doğrusal Regresyon ve Random Forest modelleri ile yapılmıştır. Test verisiyle tahmin edilen verilerin performansları ortalama mutlak yüzde hatası (MAPE), ortalama karekök sapması (RMSE) ve ortalama mutlak hata (MAE) kullanılarak değerlendirilmiştir [4].

PCA(Veri boyutunu azaltma yöntemleri sınıflandırma yapmak için harcanan zamanı ve bazı durumlarda sınıflandırma hatasını azaltmaya yardımcı olur. Zaman kritik uygulamalarda öznetelik elde etme evresinde harcanan zamanı azaltmak için, öznetelik seçme yöntemleri, tüm giriş değerlerinin ölçülmesini gerektiren boyut indirgeme yöntemlerine tercih edilir[5]) yöntemi kullanıldığında doğruluk değeri en yüksek RF algoritmasında, duyarlılık ve kesinlik değeri en yüksek SMV algoritmasında saptanmıştır. Bu yöntemin kullanılması sonucunda en düşük doğruluk NB algoritmasında, duyarlılık ve kesinlik değerleri en düşük NB ve DT algoritmalarında elde edilmiştir[6].

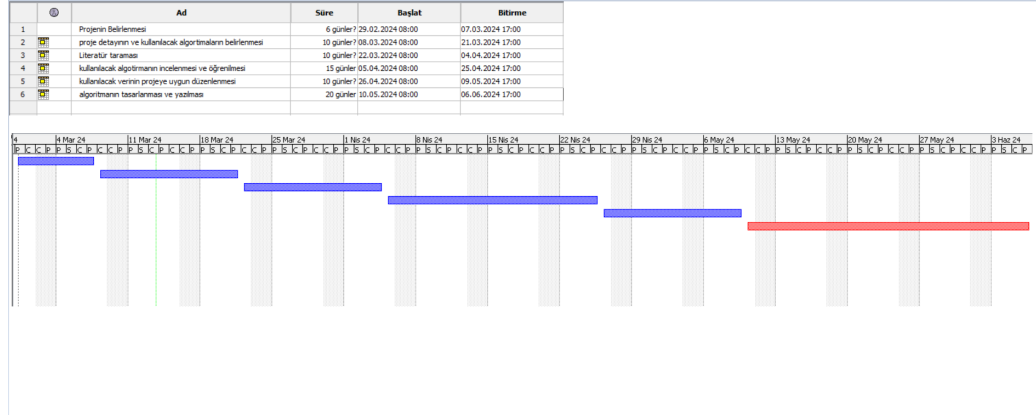
Figure 2: Dogruluk Tablosu



3 Metodoloji

Projeyi şematik planı aşağıdaki şekildedir.

Figure 3: GANTT CHART



Projede kullanacağımız veriler gün veya haftalık olarak pozitif hasta sayısını, vefat edenlerin sayısını, kullandığımız verilerdeki insanların yaşadığı şehrin yada ülkenin nüfusunu, iyileşen sayısını, toplam test sayısı gibi parametreleri içermelidir.

Veri işleme Normalizasyon işlemi- Araştırmalarda veri setlerinde verilerin bütünlüğünün sağlanması, veri tekrarının önlenmesi ve veri bütünlüğünün korunması ile performansının artırılması için normalizasyon yapılmaktadır. Daha sonra Çapraz doğrulama (Çapraz doğrulama, makine öğrenimi modellerinin başarı derecesini ortaya koymak için kullanılan yöntemdir. Çapraz doğrulama algoritma performansı hakkında bilgi verirken, verilerin daha verimli kullanılmasını sağlar.[6])-yöntemi kullanılacaktır. Daha sonra PCA yöntemi kullanılarak veri kümesini azalttıktan sonra RF(Random Forest) algoritmasına beslenecektir.

4 Kullanılacak veri

5 Beklenen Sonuçlar

Olası bir salgın hastalık durumunda oluşabilecek sonuçları tahmin etmede fikir vermesi ve projemin % 90 nın üzerinde doğrulukla sonuçlanması.

5 Kaynakçalar

- [1] BBC, “<https://www.bbc.com/turkce/haberler-dunya-60702679>,” 11 Mart 2022.
- [2] arcgis, “<https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6>,” 2024.
- [3] C. Çılğın Et Al., “Sentiment analysis of public sensitivity to covid-19 vaccines on twitter by majority voting classifier-based machine learning twitter’da covid-19 aşılarna karşı kamu duyarlılığının çoğunluk oylama sınıflandırıcısı temelli makine öğrenmesi ile duygu analizi,” *Journal of the Faculty of Engineering and Architecture of Gazi University*, pp. 1093–1104, 2023.
- [4] N. S. ÖZEN, S. SARAÇ, and M. KOYUNCU, “Covid-19 vakalarının makine Öğrenmesi algoritmaları ile tahmini: Amerika birleşik devletleri Örneği,” *Avrupa Bilim ve Teknoloji Dergisi*, no. 22, p. 134–139, 2021.
- [5] H. M. Genc, Z. Cataltepe, and T. Pearson, “A new pca/ica based feature selection method,” in *2007 IEEE 15th Signal Processing and Communications Applications*, pp. 1–4, IEEE, 2007.
- [6] E. SÜTCÜ and P. SHAMS, “Türkiye’de covid-19 günlük vaka sayısının makine öğrenmesi algoritmaları ile tahmin edilmesi,” *Anadolu Bil Meslek Yüksekokulu Dergisi*, vol. 16, no. 63, p. 197–213, 2021.