**ASTIM VERİLERİNDEN SEMPTOM RİSK ANALİZİ**

Ömer Faruk Oruç-2118121032

**ÖZET**

Bu çalışmada ilk olarak astım verileri pandas,numpy,matplotlib gibi kütüphaneler kullanılarak yorumlanmıştır.Grafikler kullanılarak verilerin daha iyi anlaşılması amaçlanmıştır.Ardından bu veriler kullanılarak kullanıcıdan birtakım parametler girilmesi istenip bu girdilere göre semptom risk analizi yapan bir yapay zeka modeli eğitilmiştir.Eğitimle ilgili Loss,Accuracy ve F1 Score grafikleri çizdirilmiştir.Yapay zeka modeli kullanıcı arayüzüne bağlı hale getirilip kullanıma hazır hale getirilmiştir.

**Anahtar Kelimeler:**Pandas,Astım,Semptom,Risk

**GİRİŞ**

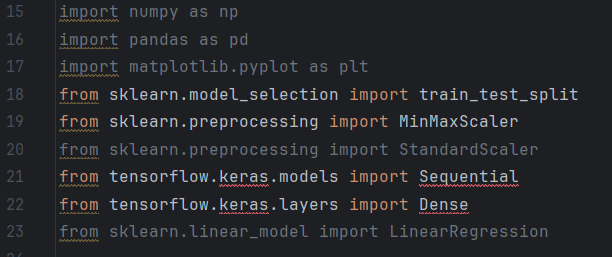
İlk önce verilerin işlenmesi ve tabloların çizilmesi için gerekli kodlar eklendi.Import os kodu işletim sistemi ile etkileşim için kullanılır.Pandas verileri manipüle etmek ve analiz etmek için kullanılır.Numpy bilimsel hesaplamalar için kullanılır.Matplob verileri görselleştirmeye yarar.Seaborn istatiksel veri görselleştirme ve scikit-learn ise makine öğrenimi için kullanılır.

Pd.read\_csv bir csv dosyasından veri kümesi yüklememizi sağlar.Verilerin alındığı yerde bu csv dosyasındaki yaş,cinsiyet vb. parametreler birbirlerine eşit olarak ayarlandığı için veriler üzerinde değişim sağlanmıştır.Semptomların Görülme Sıklığı kısmında bütün semptomların listeleme işlemi yapılır ve bu semptomların veri içerisindeki sayıları hesaplanıp ekrana yazdırılır.Yaş Gruplarına Göre Dağılım bölümünde benzer şekilde yaş grupları ve bu yaş gruplarından kaç adet olduğu ekrana yazdırılır.Aynı işlemler cinsiyet dağılımı ve hastalığın ağırlığının dağılımı bölümünde de yapılıp verideki cinsiyet sayıları ve hastalığın şiddetlerinin sayısı hesaplanıp ekrana yazdırılır.Daha sonra semptomların görülme sıklığı çubuk grafik kullanılarak görselleştirilir ve daha iyi anlaşılması sağlanır.Bu grafikten astım hastalarının hangi belirtilere daha çok sahip olduğu,hangi belirtilere daha az sahip olduğu net bir şekilde anlaşılabilir.Çıkan sonuca göre astım hastalarında en çok görülen 3 belirti sırayla :kuru öksürük,burun tıkanıklığı ve burun akıntısı’dır.En az görülen belirti ise boğaz ağrısıdır.Yaş dağılımları pasta grafiğinde görselleştirildiğinde ise bu hastalığın en çok görülen yaş aralıklarının 0-9 yaş ve 10-19 yaş aralığı olduğu görülür.Yaş dağılımı gibi cinsiyet dağılımı da pasta grafik üstünde gösterilir.Kullanılan veride belirtilen cinsiyetlerin sayısında neredeyse eşitlik söz konusudur(%50.1 erkek %49.9 kadın).Fakat veride cinsiyetleri belirtilmeyen hastalar da mevcuttur.Benzer şekilde hastalığın şiddeti de grafikleştirilmiştir.Bu grafikte de hastalığın şiddetleri açısından yakın sonuçlar görülmektedir.Semptomların şiddetinin bilinmediği ya da belirtilmediği hastalar vardır.Korelasyon analizi kısmında hastalığın semptomlarının kendi arasında korelasyon matrisi çizdirilmiştir.Bu matristen semptomların birbiriyle olan bağlantısı anlaşılır.İki semptomun kesiştiği yerdeki sayı ne kadar 1’e yakınsa ve dolayısıyla ne kadar kırmızı renge yakınsa bu iki semptom arasında bir ilişki vardır denilebilir ve ikisinin aynı anda görülme ihtimali fazla demektir.Onun hemen altında ise daha detaylı bir korelasyon matrisi bulunmaktadır.Bu matriste tüm belirtiler,tüm yaş grupları,her iki cinsiyet ve hastalığın şiddetinin kendi aralarında olan ilişkisi incelenmiştir[1][3].Tabloda bulunan hücrelerin rengi ne kadar kırmızıya yakın maviye uzaksa bu hücrede kesişen iki eksende bulunan verilerin birlikte bir hastada bulunma olasılığı o kadar fazladır.Hastalığın şiddeti ve semptomlar arasındaki ilişki kısmında ise hastalığın şiddeti ve bunların semptomlarla olan bağlantısı incelenmiştir.Tablodaki sonuçlar 1’e ne kadar yakınsa o semptom ve hastalığın şiddeti arasındaki ilişki o kadar kuvvetlidir.Semptomlar ve yaş arasındaki ilişki kısmında da yine benzer şekilde yaş gruplarıyla semptomlar arasındaki ilişki incelenmiştir.Tablodaki sonuç 1’e ne kadar yakınsa o yaş grubunda belirtilen semptomun görülme olasılığı o kadar fazladır.Cinsiyet ve semptomlar arasındaki ilişki kısmı ile yaş grupları ve hastalığın şiddeti arasındaki ilişki kısmında da tablodaki sonuçlarda 1’e yakınlık ile o iki veri arasındaki birlikte bulunma oranı o kadar yüksektir.



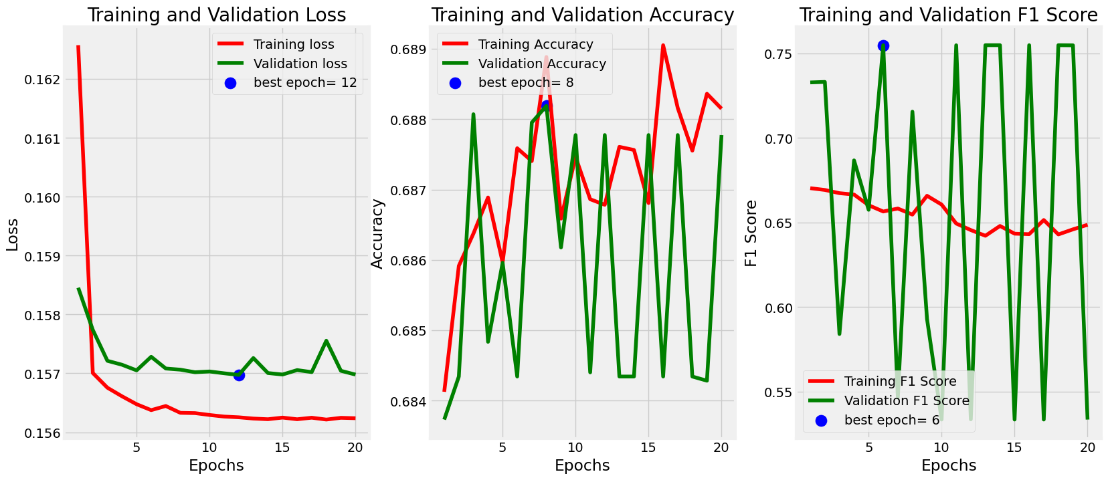
Kullanıcı arayüzü oluşturma kısmında ise Tkinter kütüphanesinden faydalanılır.Tkinter kütüphanesine ek olarak Numpy ve TensorFlow kütüphaneleri kullanılmıştır.Tensorflow makine öğrenimi için gerekli olan bir kütüphanedir.model = tf.keras.models.load\_model("model.keras") kodu ile TensorFlow modeli,model.keras adlı dosyadan yüklenir.Model giriş parametlerine göre semptom risk analizi yapmak için eğitilmiştir[4].window.geometry kodu ile açılacak arayüzün pencere boyutu ayarlanmıştır.window.title kısmında ise pencereye bir ad verilmiştir.Ardından label kullanılarak kullanıcıdan istenilecek girdiler açılan pencereye yazdırılmıştır.Label’a yazılacak yazı için renk eklenip arayüz görsel anlamda daha iyi hale getirilmeye çalışılmıştır[2].gender\_var ile cinsiyet seçimi için bir tamsayı değişkeni oluşturulmuştur ve ardından bu değişkene -1 değeri atanmıştır.Eğer radiobuttondan “Erkek” seçilirse bu değere 1,“Kadın” seçilirse 2 atanır.Yaş verisi için de bir değişken oluşturulur.Textbox kullanılarak kullanıcıdan yaşını girmesi sağlanır.Radiobutton kullanılarak kullanıcıdan istenen bilgilerin “Evet” veya “Hayır”seçilerek alınması sağlanır.Bunun için her semptom için bir tamsayı değişkeni tanımlanmıştır.Her değişken başlangıçta -1 değerindedir.Kullanıcının girdileri bu tamsayı değişkenlerinin aldığı değerlere göre değerlendirilir.

Dataset kullanılmak üzere koda eklenmiştir[5]. Eğitim kısmında ise aşağıdaki kütüphaneler tanımlanır.



Bu kodda bulunan Keras,TensorFlow üzerinde çalışan bir sinir ağı kütüphanesidir.Eğitimin hedef semptomu olarak nefes almada güçlük seçilmiştir.X ve y değişkenlerinde hedef çıkarılıp diğer parametreler tutulur.MinMaxScaler fonksiyonu kullanılarak verilerin ölçeklendirilmesi sağlanmıştır.Veri setinin %20’lik kısmı test seti olarak seçilmiştir.Random şekilde 42 parça bölünmesi sağlanmıştır.Mean Absolute Error ile hata hesaplamaları yapılmıştır.Ardından Sequential kullanılarak bir Keras modeli oluşturulmuştur.Bu modele Dense katmanları eklenip bu katmanlardaki nöron sayıları belirlenmiştir.İlk katman aynı zamanda giriş verisinin boyutunu belirtmektedir.Son katman ise çıkış katmanıdır ve sadece 1 nöron içerir.Summary kısmında modelin özet bilgisinin gösterilmesi sağlanır.Eğitim sürecinin geçmişi history adlı değişkene kaydedilmiştir.Epoch değeri modelin veriler üzerinde kaç kere eğitim yapacağını gösterir.Bu modelde 20 epoch değeri kullanılmıştır.Doğrulama verilerinin oranı ise %20 olarak seçilmiştir.Bu işlemlerin ardından modelin performansı değerlendirilir ve model.keras adıyla diske kaydedilir.Program çalıştırıldığında her seferinde tekrar model eğitilmez.Bu eğitilmiş model kullanılarak kullanıcıya risk ihtimali haber verilir.

**SONUÇ**

Yapılan bu çalışma sonucunda program kullanıcıdan çeşitli bilgileri aldıktan sonra eğer kullanıcı astım hastalığına sahipse nefes darlığı sıkıntısı çekme ihtimalini hesaplar.

Validation Loss tablosunda modelin eğitim ve doğrulamadaki kayıp durumları gösterilmektedir.20 epoch bulunan eğitimde en iyi değer 12. epochta alınmıştır.Kayıp 0’a ne kadar yakınsa eğitimin o derece iyi olduğu söylenebilir.Validation Accuracy tablosunda da modelin doğru tahmin ettiği örneklerdeki başarısını gösteren grafiktir.Burada en iyi değer 8. epochta tespit edilmiştir.Değerler 1’e yaklaştıkça doğruluk oranının arttığını anlayabiliriz.F1 Score kısmında ise modelin hem doğruluğu hem de hassasiyeti dikkate alınır.Accuracy grafiğinde olduğu gibi burada da değerler 1’e ne kadar yaklaşırsa o kadar doğruluğunun ve hassasiyetinin arttığı anlaşılır.

**REFERANSLAR**

1. Hsingc. (2023, September 11). *Asthma disease*. Kaggle. <https://www.kaggle.com/code/hsingc/asthma-disease> (Erişim tarihi 07.01.2024)
2. K-Gokulappadurai. (n.d.). *K-Gokulappadurai/asthma-risk-predicition-using-machine-learning-and-internet-of-things: An asthma risk predictor using machine learning by the help of Dataset*. GitHub. <https://github.com/K-GOKULAPPADURAI/Asthma-Risk-Predicition-Using-Machine-Learning-and-Internet-of-Things> (Erişim tarihi 07.01.2024)
3. Haque, R., Ho, S.-B., Chai, I., & Abdullah, A. (2021, September 10). *Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers*. F1000Research. <https://f1000research.com/articles/10-911/v1#referee-response-95805> (Erişim tarihi 07.01.2024)
4. Team, K. (n.d.). <https://keras.io/> (Erişim tarihi 07.01.2024)
5. dschettler8845. (2022, July 15). *How to load the dataset*. Kaggle. <https://www.kaggle.com/code/dschettler8845/how-to-load-the-dataset> (Erişim tarihi 11.01.2024)