

PROJE TEKNİK ŞARTNAMESİ VE TANITIM RAPORU

GÖĞÜS KANSERİ SINIFLANDIRMA PROJESİ

Proje Geliştiricisi: OĞUZ ÖZAKIN

MALATYA,2021

İÇİNDEKİLER

SAYFA

BÖLÜM 1

GİRİŞ

1.1 PROJENİN AMAÇ VE KAPSAMI3

1.2 PROJE KONUSUNUN ANLAM VE ÖNEMİ3

BÖLÜM 2

2.1 PROJE ARAÇLARI4

2.2 PROJE AÇIKLAMASI VE KULLANILACAK OLAN TEKNOLOJİLER5

BÖLÜM 3

3.1 PROJE GELİŞİMİNİN HAFTALIK PLANLAMASI6

BÖLÜM 1

1.1

PROJENİN AMAÇ VE KAPSAMI

Göğüs Kanseri Sınıflandırma Projesi'nin temel amacı çeşitli algoritmalar yardımıyla eldeki verileri gruplandırarak kanser teşhisleri arasında iyi/kötü sınıflandırması yapmaktır.

1.2

PROJE KONUSUNUN ANLAM VE ÖNEMİ

Projenin temelde ilgilendiği konu bir kanser türü olan ‘göğüs kanseri’dir. Göğüs kanseri, göğüs dokusunu oluşturan hücre gruplarından birinin değişime uğraması ve kontrolsüz olarak çoğalması nedeniyle oluşan tümör sonucu ortaya çıkan bir hastalıktır. Küresel ölçekteki istatistiklere göre göğüs kanserinin görülme sıklığı 100000’de 40-50 civarlarında hesaplanmaktadır. Bu bilgiler doğrultusunda, kanser hücrelerinin sınıflandırılarak tedavi ve teşhis aşamalarında iyileştirilmeler yapılması büyük önem taşımaktadır.

BÖLÜM 2

2.1

PROJE ARAÇLARI

Proje, Python programlama dili ile gerçekleştirilecektir. Proje geliştirme ortamı olarak Anaconda üzerinde Spyder kullanılacaktır. Anaconda ücretsiz ve açık kaynaklı, Python ve R programlama dillerinin bilimsel hesaplama kullanımında paket yönetimini kolaylaştırmayı amaçlayan bir özgür ve açık kaynaklı dağıtımdır. Spyder ise Anaconda Navigator tarafından geliştirilmiş, Python dilinde kod yazmaya imkan sağlayan basit arayüzlü bir ortamdır.

Proje için kullanılan veriler bir CSV dosyası içinde tutulmaktadır. Kısaca açıklamak gerekirse CSV dosyaları içerisinde aralarına virgül koyularak birbirinden ayrılmış metin veya rakamlar yer almaktadır. Projedeki CSV dosyasında bulunan veri internet üzerinden hazır şekilde alınmış olup gerekli yerlerde üzerinde değişiklikler yapılarak geliştirme sağlanmaktadır.

PROJE AÇIKLAMASI VE KULLANILACAK TEKNOLOJİLER

Projede; Temel Bileşenler Analizi (PCA), K-En Yakın Komşu Algoritması (KNN) algoritma ve GridSearchCV yöntemlerinden faydalanılacaktır. Bunun yanında Python dilinin sağlamış olduğu bir matematik kütüphanesi olan Numpy ve bir veri işleme kütüphanesi olan Pandas kullanılacaktır. Ek olarak Matplotlib ve Seaborn kütüphaneleri de projede yer alacaktır. Matplotlib, veri görselleştirmesinde kullandığımız temel Python kütüphanesidir. 2 ve 3 boyutlu çizimler yapmamızı sağlar. Seaborn ise ilgi çekici ve bilgilendirici istatistiksel grafikler yapmak için kullanılan bir kütüphanedir.

PROJE GELİŞİMİNİN HAFTALIK OLARAK PLANLANMASI

- 17.03.2021** - PROJE SÖZLEŞME RAPORUNUN TESLİM EDİLMESİ
24.03.2021 - ÜZERİNDE ÇALIŞILACAK VERİLERİN İNCELENMESİ
31.03.2021 - VERİLER ÜZERİNDE (GEREKİYOR İSE) DEĞİŞİKLİKLERİN YAPILMASI VE KODLAMA KISMINA BAŞLANARAK GEREKLİ KÜTÜPHANELERİN TANITILMASI
07.04.2021 - VERİLERDEKİ TEŞHİS SAYILARIYLA SONUÇLARIN TESPİTİ VE GÖRSELLER İLE KARŞILAŞTIRILMASI
14.04.2021 - KEŞİFSEL VERİ ANALİZİ YÖNTEMİ KULLANILARAK VERİLER ARASINDAKİ İLİŞKİLERİN ORANTISAL OLARAK ÖLÇÜLMESİ
21.04.2021 - VERİ SETİNDEKİ AYKIRI VERİLERİN (OUTLIERS) TESPİTİ VE ÇIKARILMASI İŞLEMİ
28.04.2021 - BELİRLENEN AYKIRI VERİLERİN MATPLOTLIB KÜTÜPHANESİ YARDIMIYLA GÖRSELLEŞTİRİLMESİ VE SUNULMASI
05.05.2021 – VERİ SETİNİN EĞİTİM-TEST (TRAIN-TEST) OLARAK AYRILMASI
12.05.2021 - EĞİTİM-TEST AYRIMI SONUCUNDA BOXPLOT YARDIMIYLA VERİLERİN KARŞILAŞTIRILMASI VE YORUMLANMASI
19.05.2021 - KNN ALGORİTMASININ UYGULANMAYA BAŞLANMASI
26.05.2021 - KNN YARDIMIYLA EN İYİ PARAMETRELERİN BULUNMASI
02.06.2021 - KNN ALGORİTMASININ KODLANMASININ SONLANDIRILMASI VE SONUÇLARIN TESPİTİ
09.06.2021 – PROJENİN TAMAMLANMASI VE SUNUMUN YAPILMASI