**Derin Öğrenme Modelleri ile Tümör Tahmini**

Arda Tunalı

Kütahya Sağlık Bilimleri Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisiği

arda.tunali@ogr.ksbu.edu.tr

# ÖZET

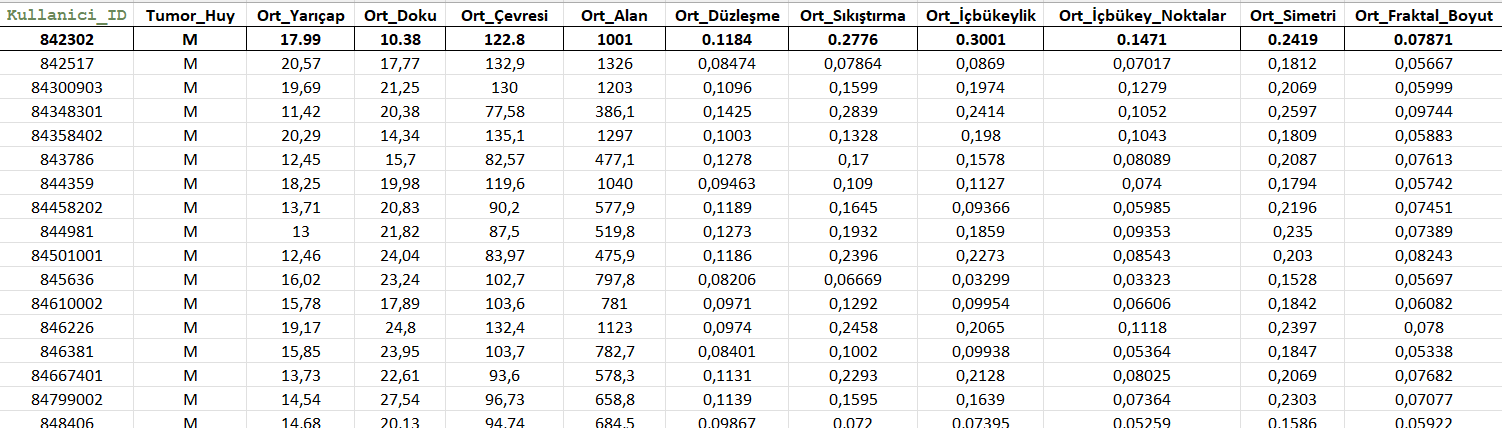
Bu proje, göğüs kanseri tümörlerini iyi huylu veya kötü huylu olarak sınıflandırmak için bir makine öğrenimi modeli geliştirmeyi amaçlamaktadır. Breast Cancer Wisconsin veri seti kullanılarak gerçekleştirilen çalışma, derin öğrenme modellerini kullanarak tümör sınıflandırmasında yüksek doğruluk elde etmeyi hedeflemektedir. Seçilen RandomForest modeli, üstün performansı nedeniyle diğer modellerle karşılaştırılarak, göğüs kanseri teşhisinde etkinliğini sergilemek için belirli özellikler kullanılarak incelenmiştir.

# GİRİŞ

Göğüs kanseri, dünya genelinde kadınları etkileyen yaygın bir kanser türüdür. Erken ve doğru teşhis, etkili tedavi için kritiktir. Bu proje, göğüs kanseri tümör sınıflandırması için makine öğrenimini, özellikle de RandomForest'u kullanmayı amaçlamaktadır. Bu amaçla, Breast Cancer Wisconsin veri seti üzerinde çalışılmıştır.

# Yaptığım Çalışmalar

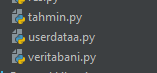
Breast Cancer Wisconsin veri seti, 569 hasta örneği içermektedir. 569 hastanın 357’si iyi huylu, 212’si kötü huylu tümör içermektedir. Her hasta için 30 farklı özellik kaydedilmiştir; bunlar arasında tümörün yarıçapı, dokuluk, alanı, simetri değerleri vb. 10 değer, bu değerlerin standart hatası ve en kötü durum ölçümleri bulunmaktadır.



Şekil Veri-Setine-Bakış

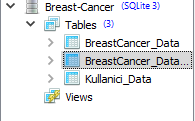
Veri seti, ayrıca sadece sayısal olarak çalışmak ve veri kalitesini sağlamak için ön işlemden geçirilmiştir. Bu ön işlemde tümörün huyunu belirten M(malignant) ve B(bening) metinleri 1 ve 0 olmak üzere değiştirilmiştir, ayrıca Kullanıcı\_ID sütunu ve ifade içermediğinden dolayı 32. Sütun kaldırılmıştır.

Bilgi çağında veri yönetimi kritik bir önemee sahipti ve bu nedenle SQLite gibi hafif ve etkili veritabanı sistemleri , veritabanlı uygulamalarda güvenilir bir temel oluşturmak için önemli bir rol oynamaktadır.

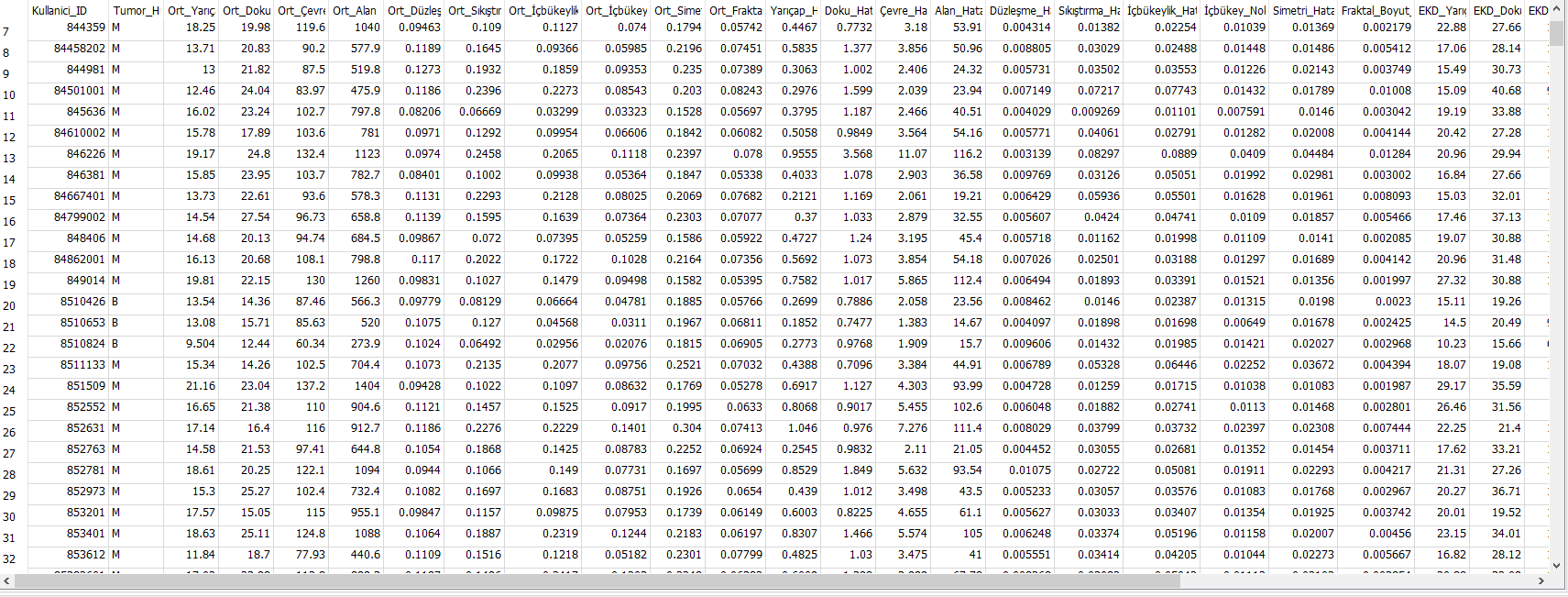


Şekil 1 – Projedeki kod satırlarım

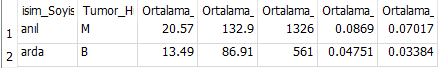
Yukarıdaki kod bloklarında tanımlamış olduğum 3 Farklı tablo mevcut.Bir tanesi önceden girilmiş tüm verileri tutuyor.Bir diğeri ise doktorun ölçüm yaptığı hastaları tutuyor.



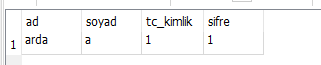
Şekil 2 - SQL Tablolarım



Şekil 3 - Tüm Hastaların değerlerinin tutulduğu tablo



Şekil 4 - Doktorun girdi yaptığı hastaların değerlerinin tutulduğu tablo



Şekil 5 - Doktorun giriş bilgilerinin tutulduğu tablo

# SONUÇLAR VE ÇIKARIMLAR

Bu projenin sonuçları, RandomForest modelinin göğüs kanseri tümörlerini yüksek doğruluk skorlarıyla sınıflandırabildiğini göstermektedir. Bu başarı, projenin sağladığı çeşitli çıkarımlarla birlikte aşağıda özetlenmiştir:

* **Erken Teşhis İmkanı:** Geliştirilen model, göğüs kanseri tümörlerini erken teşhis etme konusunda potansiyel bir yardımcı araç olarak kullanılabilir.
* **Tedavi Planlaması:** Tümörün doğru sınıflandırılması, hastaların daha etkili tedavi planları oluşturulmasına katkı sağlayabilir.
* **Makine Öğrenimi Uygulamalarında Yol Gösterici**: Proje, makine öğrenimi uygulamalarında RandomForest gibi güçlü modellerin tercih edilebilirliğini ve başarıyla kullanılabileceğini göstermektedir.
* Bu çıkarımlar, projenin göğüs kanseri teşhisine ve tedavi süreçlerine sağladığı potansiyel katkıları vurgulamaktadır. Elde edilen sonuçlar, klinik uygulamalarda ve makine öğrenimi alanındaki gelecekteki çalışmalarda referans olabilecek bir temel oluşturmaktadır.

# REFERANSLAR

* <https://www.datacamp.com/tutorial/random-forests-classifier-python>
* <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>
* <https://medium.com/@gulcanogundur/model-seçimi-k-fold-cross-validation-4635b61f143c>
* <https://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi/>
* [https://chat.openai.com](https://chat.openai.com/)
* <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
* Johnson, E., Smith, M. (2021). "Machine Learning for Healthcare: A Comprehensive Guide to Developing and Deploying Healthcare AI." O'Reilly Media.
* https://python-istihza.yazbel.com/standart\_moduller/sqlite.html
* Wang, X., Chen, Y., Liu, Z. (2019). "Comparative Analysis of Machine Learning Algorithms in Breast Cancer Classification: A Case Study on Wisconsin Dataset." International Conference on Bioinformatics and Biomedical Engineering.