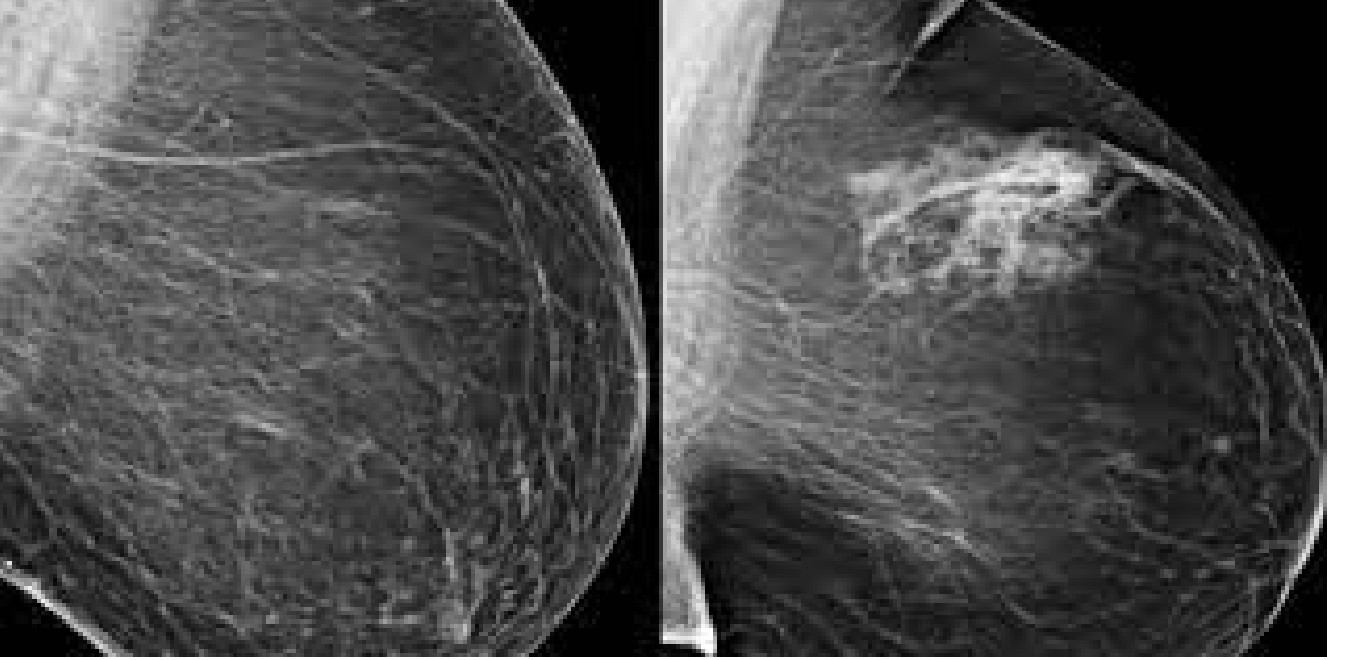


Göğüs Kanseri Sınıflandırması

Son Teslim : 18 Şubat 23:59



Classification

Sınıflandırma, istatistik ve makine öğrenimi alanlarında sıkça kullanılan bir analiz tekniğidir. Temel olarak, bir bağımlı değişkeni belirli sınıflara veya kategorilere atama amacını taşır. Bu ilişki, bağımsız değişkenlerin değerleri kullanılarak bağımlı değişkenin sınıflandırılması veya tahmin edilmesi olarak ifade edilir. Sınıflandırma analizi, genellikle kategorik veya sınıflandırılabilir verilerle çalışır ve bir model oluşturarak veri setindeki desenleri tanımlamayı ve yeni örneklerin hangi sınıfa ait olduğunu tahmin etmeyi amaçlar.

Sınıflandırma analizi, veriler arasındaki ilişkileri anlamak, desenleri tanımlamak ve sınıflandırılabilir örnekleri tahmin etmek için güçlü bir araçtır. Ancak, sınıflandırma modelleri her zaman doğru olmayabilir ve uygun model seçimi, veri ön işleme ve model değerlendirme süreçleri dikkatle yapılmalıdır.

Assignment 2

1-) Veriyi Araştırma

Bu ödev ikinci ödeviniz olduğundan, ilk ödevle kıyasla çok daha esnek bir tutum sergileyerek çoğu işi sizin yapmanızı bekliyoruz. Buna öncelikle datasetin nasıl toplandığı, güvenilir olup olmadığı (yıl , toplayan şirkete, diğer makalelerde kullanılma sayısı vb.) ve diğer makalelerde hangi amaçlar için kullanıldığını araştırmanızı ve güzel bir şekilde rapora dökmenizi istiyoruz. Bu konuda tamamen özgürsünüz ama bu ödevde bu kısma ağırlık vermenizi istiyoruz.

2-) EDA (Exploratory Data Analysis)

Bu kısımda veriyi daha iyi anlayıp, özellikler arasındaki ilişkiyi analiz ederek mantıklı bir şekilde veri ön işleme yapmanızı bekliyoruz. Bu kısımda da ilk kısımda olduğu gibi istediğiniz şekilde yapmakta özgürsünüz. Ancak sizden istediğimiz güzel bir şekilde model eğitip gerçekten de bir hastanede bu model kullanılacakmış ve insanların hayatını etkileyecekmiş gibi bu görevi simüle edip bu şekilde çalışmanız.

3-) Modellerin Kıyaslanması

Eşit bir kıyaslama yapabilmek açısından herkesten size verilen sitede de incelenmiş olan algoritmalar ile modellerinizi eğiterek kıyaslamanızı istiyoruz. Buna ek olarak modeller geliştirip daha iyi performans üretmeye çalışmak da aynı zamanda sizin göreviniz (En az bir ekstra model geliştirmenizi bekliyoruz). Unutmayın, yapacağınız bu iş gerçek hayat simülasyonu ve modeliniz gerçekten insanların hayatını etkileyebilir.

Model	Accuracy	Precision	Recall	ROC AUC Score
XGBoost Classification				
Logistic Regression				
Random Forest Classification				
Support Vector Classification				
Neural Network Classification				

Veri Seti

Veri setine şu adresten ulaşabilirsiniz :

<https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original>

Rapor ve Puanlama

Raporun içeriği sırasıyla:

1. Veri Araştırması: Verinin gerçek hayatta kullanılacakmış gibi araştırılması ve neden kullanılabilir yada kullanılamaz olduğunun kanıtlarla ifade edilmesi
2. EDA : Verinin anlaşılıp ön işlemeye tabi tutulması ve tüm aşamaların sebepleriyle açıklanması
3. Modellerin Kıyaslanması : Verilen modellerin ve en az ekstra bir modelin verilen metrikler ölçüsünde kıyaslanarak modellerin kullanılabilirlik sırasına sokulması ve bu sıranın yorumlanması

Örnek : RFC en yüksek precision ve accuracy değerlerine sahip olduğu için birinci sırada yer alıyor. Bundan sonraki model RFC'nin kullanılamayacağı durumlarda önerebileceğiniz ikinci model olmalı

Puanlama kısmı aşağıdaki gibi olacaktır:

1. Veri Araştırması : 25 Puan
2. EDA : 25 Puan
3. Modellerin Kıyaslanması ve Sonuç: 40 Puan
4. Ek Modelin Performansı: 10 Puan

Not : Sececeğiniz ek modelin sebebinin mantıklı olması ve sonuçlarının diğer modellerden düşük olmaması gerekmektedir.

Not : Raporları değerlendirirken ne kadar öğrenmiş olduğunuz dikkate alınmaya çalışılacaktır. Bu süreçte katılma sebebinizin ARGE biriminde rol alabilme şansının yanı sıra kendinizi geliştirmek için de olduğunu lütfen unutmayın. Sizden temennimiz amacınızın bir an önce ödevi teslim etmek değil, ödev içerisinde kullanmanızı istediğimiz modelleri güzel bir şekilde öğrenmiş olmanızdır. Unutmayın yapmış olduğunuz ödev sadece öğrendiğinizi değerlendirebilmemiz için bir araç, önemli olan bu süreç boyunca kendiniz geliştirmiş olmanız ve semester boyunca harcayacağınız vaktinizin olumlu ve verimli bir sonuç vermesidir. Sorularınızı whatsapp grubu üzerinden sorabilirsiniz. Şimdiden başarılar.

Ödev Teslim: Kodunuzu ipynb veya py uzantılı dosya olarak ve raporunuzu pdf formatı olarak **argeaiclub@gmail.com** mailine 18 Şubat Pazar günü 23:59'a kadar gönderebilirsiniz.