



Breast Cancer Prediction

Hazırlayan:
Gökberk Kozak

Öğrenci No:
202210614057



Meme Kanseri: Tanı, Tedavi ve Farkındalık

Meme kanseri, kadın sağlığı açısından büyük bir tehdit oluşturan ve global düzeyde önemli bir halk sağlığı sorunu olarak öne çıkan bir kanser türüdür. Bu makalede, meme kanserinin kompleks yapısı, risk faktörleri, tanı süreci, çeşitli tedavi seçenekleri ve farkındalığın rolü üzerine detaylı bir bakış sunarak, bu hastalığın anlaşılmasını ve toplumda farkındalık yaratılmasını amaçlıyoruz.

Meme Kanseri: Derinlemesine Bir Anlayış:

Meme kanseri, meme dokusundaki hücrelerin kontrolden çıkarak anormal bir şekilde çoğalması sonucu oluşan bir kanser türüdür. Bu kanser, genellikle memede hissedilen kitleler, değişiklikler veya ağrılar şeklinde belirti verir. Genetik, hormonal ve çevresel faktörlerin bir kombinasyonu ile ortaya çıkabilir.

Risk Faktörleri ve Genetik Yatkınlık:

Meme kanserinin ortaya çıkmasında etkili olan birçok risk faktörü vardır. Genetik yatkınlık, önceki meme kanseri vakalarının aile geçmişi, hormonal değişiklikler, yaş, obezite ve genetik mutasyonlar bu faktörler arasında yer alır. Özellikle BRCA1 ve BRCA2 gen mutasyonları, genetik yatkınlığı artırarak meme kanseri riskini artırabilir.

Tanı Yöntemleri:

Meme kanseri genellikle düzenli mamografi taramaları, ultrason ve manyetik rezonans görüntüleme (MRG) gibi görüntüleme teknikleri kullanılarak teşhis edilir. Biyopsi gibi daha invaziv testlerle, kanserin tipi ve evresi belirlenir. Erken tanı, başarılı bir tedavi için kritik öneme sahiptir.

Tedavi Seçenekleri:

Meme kanseri tedavisi, hastanın genel sağlık durumu, kanserin evresi, tipi ve diğer bireysel faktörlere bağlı olarak belirlenir. Cerrahi müdahale, radyasyon terapisi, kemoterapi, hormon tedavisi ve hedefe yönelik tedaviler gibi bir dizi yöntem, bir tedavi planının parçalarını oluşturabilir. Bu tedaviler genellikle multidisipliner bir ekip tarafından yönetilir.

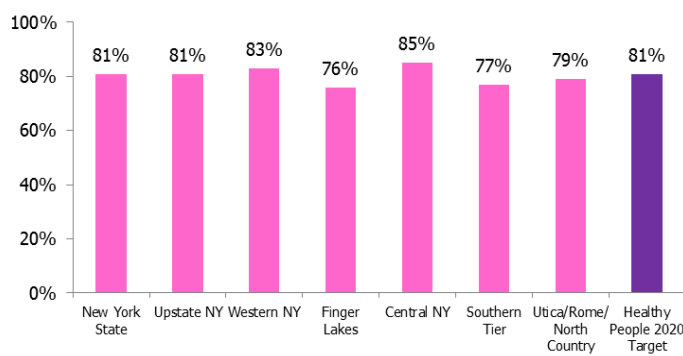
Farkındalık ve Önemli Rolü:

Meme kanseriyle mücadelede farkındalık yaratmak, toplumun bu konuda eğitilmesi ve düzenli tarama programlarına katılımın teşvik edilmesi, erken teşhis ve etkili tedavi için kritik önem taşır. Toplumda bilinçli bir farkındalık, genç kadınlar ve risk altındaki gruplar için özel destek programları ve eğitim kampanyaları ile birleştiğinde, meme kanseri ile mücadelede önemli bir adım atılabilir.

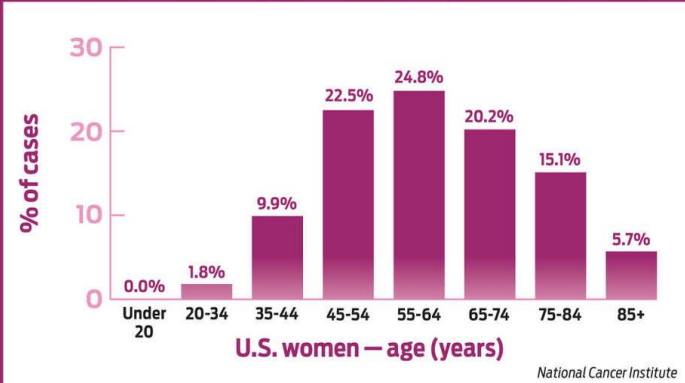
Sonuç:

Meme kanseri, kadın sağlığına yönelik ciddi bir tehdit oluşturmalarının yanı sıra, multidisipliner bir yaklaşımla yönetilebilir bir hastalıktır. Erken teşhis, etkili tedavi ve toplumda bilinçli bir farkındalık, bu zorlu yolculukta önemli kılavuzlardır. Bilimsel araştırmalar, tıp teknolojilerindeki gelişmeler ve toplumun eğitimi, meme kanseriyle mücadelede ilerlemeye katkıda bulunabilir ve umut vadeden bir gelecek için kapıları aralayabilir. Unutulmamalıdır ki, her bireyin ve toplumun bu konuda bilinçli bir rol oynaması, bu mücadelede önemli bir faktördür.

Percent of Women Aged 50-74 Years Receiving Breast Cancer Screening Based on Most Recent Guidelines



Breast cancer diagnoses by age — U.S. women



Makine Öğrenmesi Algoritmalarının Kullanımıyla Meme Kanseri Teşhisi: Geleceğin Tanı Yaklaşımı

Makine öğrenmesi, sağlık sektöründe önemli bir dönüşüm sağlayan ve tanı süreçlerinde etkili bir araç olarak öne çıkan bir teknolojidir. Bu yazıda, özellikle meme kanseri teşhisi konusunda makine öğrenmesi algoritmalarının kullanımının getirdiği yeniliklere odaklanacağız. Bu teknoloji, hastalığın erken teşhisi, doğru sınıflandırma ve bireyselleştirilmiş tedavi planlarının oluşturulması gibi önemli avantajlar sunmaktadır.

Makine Öğrenmesi ve Meme Kanseri:

Makine öğrenmesi algoritmaları, büyük veri setlerini analiz ederek desenleri tanıma ve karmaşık ilişkileri anlama yeteneği ile bilinir. Meme kanseri teşhisi için bu algoritmalar, hastaların klinik verilerini, genetik bilgilerini ve görüntüleme sonuçlarını kullanarak daha hassas ve hızlı bir teşhis süreci sağlamaktadır.

Veri Toplama ve Analiz:

Makine öğrenmesi algoritmaları, hastaların mamografi sonuçları, genetik profilleri, kişisel özellikleri gibi çeşitli veri setlerini kullanarak model oluşturur. Bu modeller, meme kanserinin belirli özelliklerini ve belirleyici faktörlerini belirleme yeteneği kazanır.

Destek Vektör Makineleri (SVM), Karar Ağaçları ve Derin Öğrenme:

Bu algoritmalar arasında destek vektör makineleri (SVM), karar ağaçları ve derin öğrenme modelleri, meme kanseri teşhisinde yaygın olarak kullanılan yöntemlerdir. SVM, veri setini iki sınıfa ayırarak, kanserli ve kansersiz vakaları doğru bir şekilde sınıflandırabilir. Karar ağaçları, hastaların özelliklerini ağaç benzeri bir yapıda analiz ederek teşhis sürecini açıklayabilir. Derin öğrenme ise karmaşık yapıdaki verileri anlama ve öğrenme yeteneği ile öne çıkar.

Erken Teşhis ve Tedavi Planlaması:

Makine öğrenmesi algoritmalarının en büyük avantajlarından biri, hastalığın erken evrelerinde dahi belirtileri tespit edebilme yetenekleridir. Bu erken teşhis, tedavi başarı şansını artırabilir ve hasta yaşam kalitesini iyileştirebilir. Ayrıca, bireyselleştirilmiş tedavi planları oluşturarak hastaya özel yaklaşımların geliştirilmesine olanak tanır.

Zorluklar ve Etik Konular:

Makine öğrenmesi uygulamalarıyla ilgili bazı zorluklar ve etik konular da mevcuttur. Veri güvenliği, algoritmaların şeffaflığı ve toplumun bu teknolojiye güveni gibi konular, makine öğrenmesi tabanlı teşhis uygulamalarının yaygınlaşmasını sınırlandıran faktörler arasında yer alabilir.

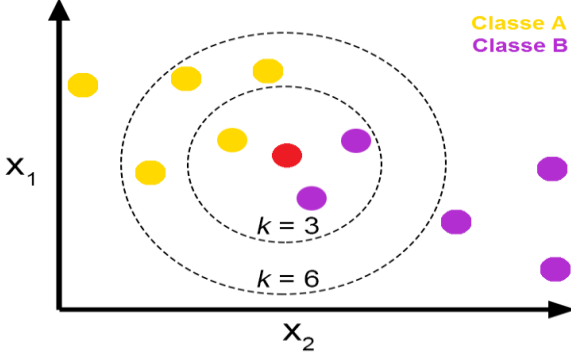
Sonuç:

Makine öğrenmesi algoritmalarının meme kanseri teşhisi alanında kullanımı, sağlık sektöründe önemli bir dönüşümü temsil eder. Bu algoritmaların geliştirilmesi ve kullanımının yaygınlaşmasıyla birlikte, meme kanseri teşhisi daha hassas, erken ve bireyselleştirilmiş bir şekilde gerçekleştirilebilir. Ancak, bu teknolojinin etik ve güvenlik konularının dikkate alınması, makine öğrenmesi uygulamalarının sağlık sistemlerine entegrasyonunda önemli bir rol oynayacaktır.

Kullanılan Algoritmalar

KNeighbors (K En Yakın Komşu):

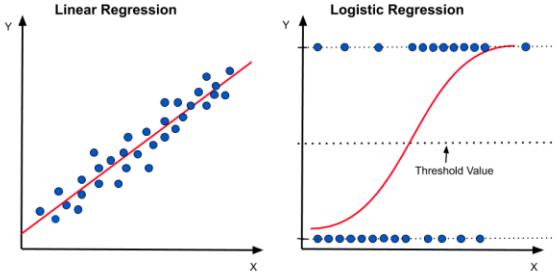
KNN algoritması, veri noktalarının benzer özelliklere sahip olanlarını gruplandırmasını sağlar. Algoritma, veri noktası için en yakın komşuları bulur ve bu komşuların sınıfını veri noktası için sınıflandırma yapar. Veri noktası için sınıflandırma, en çok tekrar eden komşu sınıflarına dayanır.



$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Logistic Regression (Lojistik Regresyon):

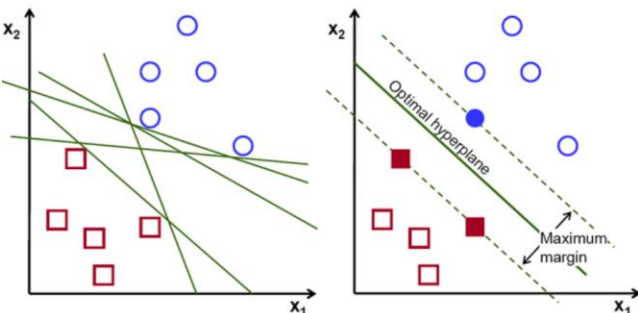
Lojistik regresyon, isminde “**regresyon**” geçmesine rağmen bir **sınıflandırma** algoritmasıdır. Yani görseldeki hayvanın kedi mi, köpek mi olduğu veya verilmiş olan bilgilerin bir erkeğe mi yoksa bir kadına mı ait olduğunu tahmin etme gibi iki sınıflı sınıflandırma problemlerinde sıkça kullanılır.



$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$
$$\Rightarrow P = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

SVM (Destek Vektör Makineleri):

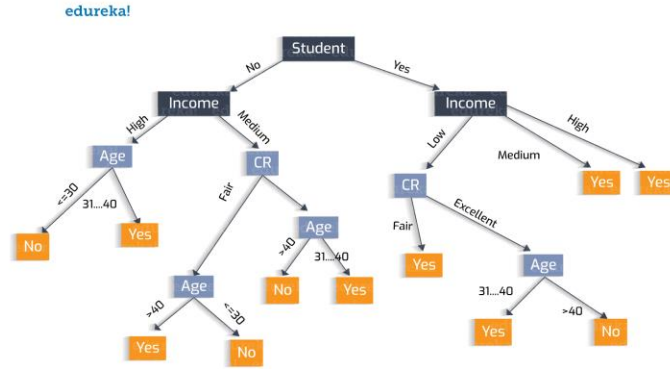
Destek Vektör Makineleri (Support Vector Machine) genellikle sınıflandırma problemlerinde kullanılan gözetimli öğrenme yöntemlerinden biridir. Bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer. Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar. Karmaşık ama küçük ve orta ölçekteki veri setleri için uygundur.



$$J(w, b, \alpha) = \sum_{i=1}^N \alpha_i + \frac{1}{2} w^T w - w^T \sum_{i=1}^N \alpha_i d_i x_i - b \underbrace{\sum_{i=1}^N \alpha_i d_i}_0$$
$$\Rightarrow Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j x_i^T x_j$$

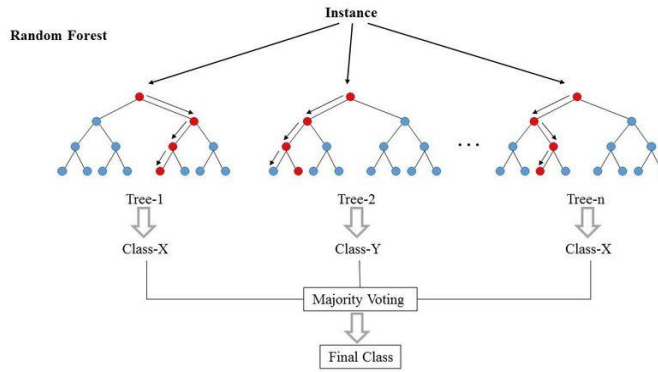
Decision Tree (Karar Ağacı):

Makine öğrenmesindeki temel algoritmalarından bir diğeri Karar Ağaçları (Decision Tree) Algoritmasıdır. En basit tanım ile, olası sonuç için oluşturulabilecek kararyollarının oluşturulmasıdır. Elimizdeki veri setindeki bir feature (kolon) seçip onun diğer alanlar üzerindeki dallanmalarını ilişkilendirip bir karar ağacı çıkartırız.



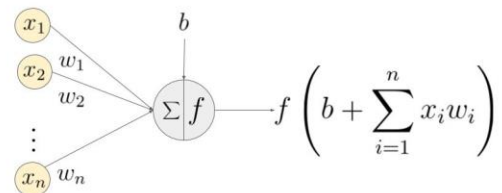
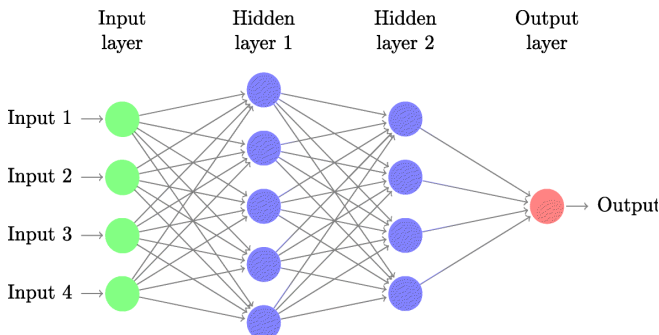
Random Forest:

Random Forest, çeşitli ağaçlar oluşturarak ve bu ağaçların sonuçlarını bir araya getirerek tahminler yapmak için kullanılan bir makine öğrenmesi algoritmasıdır. Her bir ağaç, belirli bir özellik alt kümesini kullanarak eğitilir ve birbirinden bağımsızdır. Bu nedenle, Random Forest, hem değişkenler arasındaki etkileşimleri hem de veri setindeki gürültüyü azaltarak, yüksek kaliteli tahminler yapabilir.



ANN (Artificial Neural Network - Yapay Sinir Ağı) & MLP :

Yapay sinir ağları, hayvan beyinlerindeki biyolojik sinir hücrelerinin (nöronların) işlevlerini taklit etme gayesiyle oluşturulmuş bir hesaplama modelidir. ANNs, öğrenme, genelleme ve tanıma yeteneklerine sahip olmaları için tasarlanmışlardır. Bu yetenekler, özellikle veri madenciliği, tanıma ve otomatik kontrol sistemleri gibi alanlarda oldukça değerlidir.



An example of a neuron showing the input ($x_1 - x_n$), their corresponding weights ($w_1 - w_n$), a bias (b) and the activation function f applied to the weighted sum of the inputs.

Breast Cancer Veri Seti

Meme Kanseri Veri Seti:

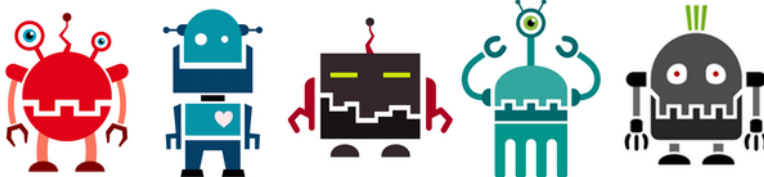
- Meme kanseri veri seti, genellikle meme dokusundaki hücresel özellikleri ve kanserli tümörlerin özelliklerini incelemek amacıyla oluşturulmuş bir veri kümesidir. Bu tür veri setleri genellikle klinik araştırmalarda ve makine öğrenimi modellerinin eğitiminde kullanılmaktadır. Bu veri setleri, meme kanserinin teşhisinde, sınıflandırılmasında veya prognozlanmasında yardımcı olabilecek çeşitli özellikleri içermektedir.
- Özellikle, meme kanseri veri setlerinde bulunan sütunlar, hücresel ve tümöral özellikleri temsil eden değerleri içerir. Bu değerler genellikle radyolojik görüntüler, histopatolojik incelemeler ve diğer klinik testler sonucu elde edilen ölçümleri içerir. Bu veri setleri üzerinde yapılan analizler, belirli özelliklerin veya kombinasyonlarının meme kanserinin teşhisi veya prognozu için belirleyici olup olmadığını anlamaya yönelik olabilir.
- Meme kanseri veri setleri, bilgisayar destekli teşhis sistemlerinin ve makine öğrenimi algoritmalarının geliştirilmesi için önemli bir kaynak olarak kullanılır. Bu tür sistemler, doktorlara ve sağlık profesyonellerine doğru teşhisler koyma ve tedavi planları oluşturma konusunda yardımcı olabilir. Ayrıca, bu veri setleri, kanser araştırmacılarına yeni bulgular ve ilişkiler keşfetme konusunda da önemli bir kaynak sağlar.
- Meme kanseri veri setleri genellikle geniş bir veri seti içerir, bu da çeşitli analizlerin ve modelleme çalışmalarının yapılabilmesine olanak tanır. Ancak, bu veri setlerinin kullanımı sırasında gizlilik ve etik konular dikkate alınmalıdır, çünkü bu veriler genellikle hassas sağlık bilgilerini içerir.

Kaggle Nedir:

Kaggle platformu, kullanıcılara geniş bir veri kümesi koleksiyonuna, öğrenme kaynaklarına ve çeşitli veri bilimi araçlarına erişim sağlar. İnsanlar, kendi projelerini başlatmak, veri setlerini keşfetmek, analizler yapmak ve çeşitli konularda yarışmalara katılarak yeteneklerini test etmek için Kaggle'ı kullanabilirler. Kaggle, aynı zamanda kullanıcıların birbirleriyle etkileşimde bulunabileceği bir topluluk platformu olarak da hizmet verir. Forumlar, yarışma tartışmaları, makale paylaşımları ve soru-cevap bölümleri gibi pek çok interaktif özellik sunar.

-Veri Setimizin bulunduğu siteden datasetimiz hakkında bilgi alabiliriz. Fakat daha fazla bilgi için analiz yapmalıyız.

kaggle



Veri Setindeki Sütunlar ve Açıklamaları

- **diagnosis:** Kanser teşhisi (malignant veya benign).
- **radius_mean:** Tümörün ortalaması alınan radyus (yarıçap) değeri.
- **texture_mean:** Gri tonlamalarının ortalaması alınan dokunun düzensizliği.
- **perimeter_mean:** Tümörün ortalaması alınan çevre (kenar) uzunluğu.
- **area_mean:** Tümörün ortalaması alınan alan.
- **smoothness_mean:** Gri tonlamalarının lokal değişkenliğinin ortalaması.
- **compactness_mean:** Tümörün kompaklığının ortalaması.
- **concavity_mean:** Tümörün oyukluğunun ortalaması.
- **concave points_mean:** Tümörün oyuk noktalarının ortalaması.
- **symmetry_mean:** Tümörün simetrisinin ortalaması.
- **fractal_dimension_mean:** Fraktal boyutunun ortalaması.
- **radius_se:** Tümörün radyusunun standart hatası.
- **texture_se:** Gri tonlamalarının dokunun düzensizliğinin standart hatası.
- **perimeter_se:** Tümörün çevresinin standart hatası.
- **area_se:** Tümörün alanının standart hatası.
- **smoothness_se:** Gri tonlamalarının lokal değişkenliğinin standart hatası.
- **compactness_se:** Tümörün kompaklığının standart hatası.
- **concavity_se:** Tümörün oyukluğunun standart hatası.
- **concave points_se:** Tümörün oyuk noktalarının standart hatası.
- **symmetry_se:** Tümörün simetrisinin standart hatası.
- **fractal_dimension_se:** Fraktal boyutunun standart hatası.
- **radius_worst:** Tümörün en kötü durumdaki radyusu.
- **texture_worst:** Gri tonlamalarının en kötü durumdaki dokunun düzensizliği.
- **perimeter_worst:** Tümörün en kötü durumdaki çevre uzunluğu.
- **area_worst:** Tümörün en kötü durumdaki alanı.
- **smoothness_worst:** Gri tonlamalarının en kötü durumdaki lokal değişkenliği.
- **compactness_worst:** Tümörün en kötü durumdaki kompaklığı.
- **concavity_worst:** Tümörün en kötü durumdaki oyukluğu.
- **concave points_worst:** Tümörün en kötü durumdaki oyuk noktaları.
- **symmetry_worst:** Tümörün en kötü durumdaki simetrisi.
- **fractal_dimension_worst:** En kötü durumdaki fraktal boyutu.

Veri Setinin Analizi

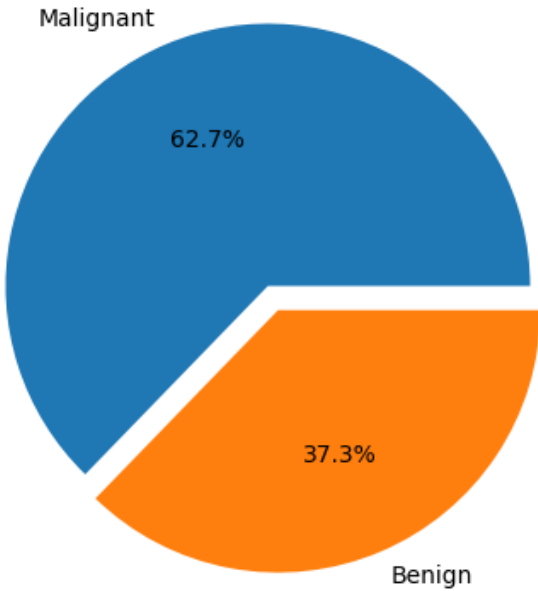
Python Pandas Kütüphanesi:

Pandas, Python programlama dilinde kullanılan güçlü bir veri analizi kütüphanesidir. Veri analizi süreçlerini hızlı ve etkili bir şekilde gerçekleştirmek için çeşitli araçlar sunar. Pandas'ın temel veri yapılarından biri DataFrame, tablo benzeri veri yapılarını temsil eder ve veri manipülasyonu için geniş bir fonksiyon seti sunar. Veri setlerini okuma, filtreleme, gruplama, birleştirme, temizleme ve istatistiksel analiz gibi birçok işlemi kolayca gerçekleştirebilirsiniz. Ayrıca, veri görselleştirmesi için diğer popüler kütüphanelerle (örneğin, Matplotlib ve Seaborn) entegre edilebilir. Pandas, veri analizi ve manipülasyonu süreçlerini hızlandırarak, araştırmacılara, veri bilimcilerine ve analistlere veri odaklı görevlerini daha etkili bir şekilde gerçekleştirme imkanı sağlar.

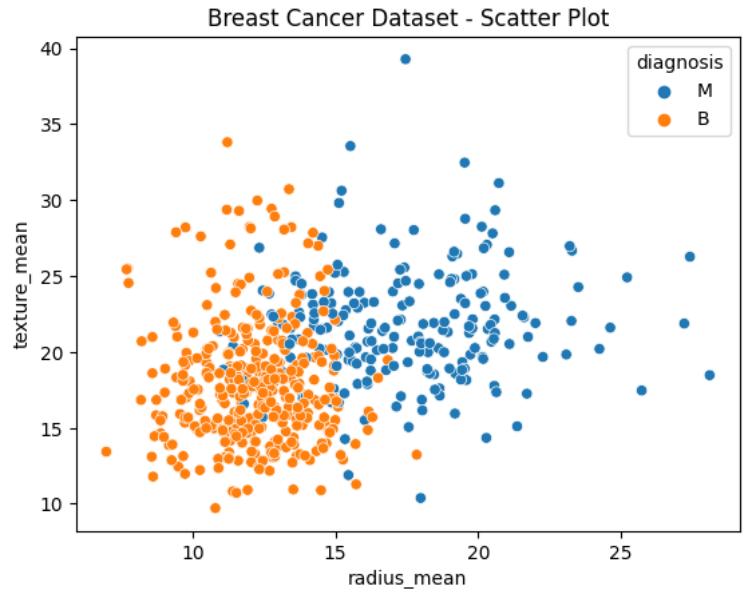
Python ile Veri Görselleştirme:

Python, zengin veri görselleştirme araçları ve kütüphaneleri sayesinde güçlü bir veri analizi ve görselleştirme platformu sunar. Matplotlib, Seaborn, Plotly gibi popüler kütüphaneler, çeşitli grafik türlerini oluşturmak ve veri setlerini anlamak için geniş bir yelpazede araçlar sağlar. Matplotlib temel bir çizim kütüphanesi olarak kullanılırken, Seaborn istatistiksel grafiği geliştirmek için optimize edilmiştir. Plotly interaktif grafikler oluşturmak için idealdir. Python, bu kütüphaneleri ve araçları kullanarak veri analistlerine, bilim insanlarına ve geliştiricilere, veri setlerini etkili bir şekilde keşfetme ve anlama yeteneği sunar. Veri görselleştirmesi, karmaşık veri setlerini daha anlamlı ve erişilebilir hale getirerek veri hikayelemesine katkıda bulunur.

Detayli Analiz Ve Veri Görselleştirme Kodda Bulunmakta Altındaki Görseller Veri Görselleştirmeden Bazı Örnekler.



Malignant (Kötü Huylu)
Benign (İyi Huylu)

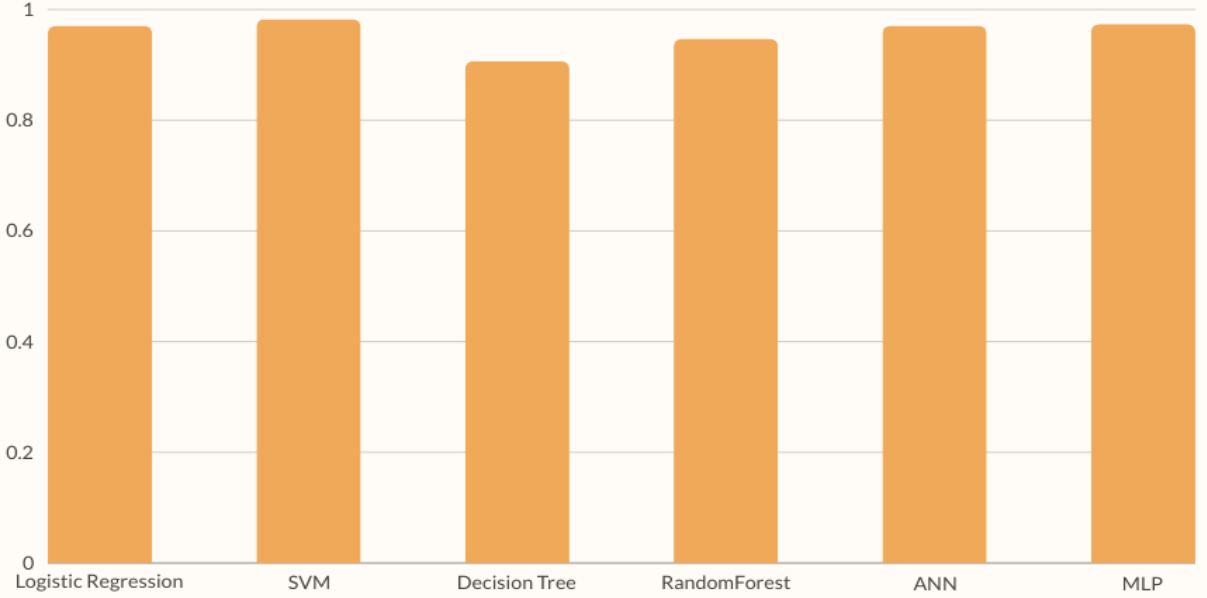


Algoritma Doğru Tahmin Oranları

Veri analizi sonuçları, makine öğrenmesi algoritmalarına veri sağlama sürecinde kritik bir rol oynar. Makine öğrenmesi algoritmalarının doğruluk oranları, veri setinin kalitesi, öznelik seçimi ve modelin yapılandırılmasına bağlı olarak değişebilir. İyi bir veri analizi, algoritmanın eğitildiği veri setini anlama, öznelikler arasındaki ilişkileri belirleme ve gereksiz bilgileri eleme yeteneği sağlar. Bu, modelin daha tutarlı ve genelleştirilebilir sonuçlar üretmesine yardımcı olur. Ayrıca, doğruluk oranlarını değerlendirirken, modelin aşırı uydurmayla mücadele etme ve yeni veri noktalarına genelleme yapabilme yeteneği de göz önünde bulundurulmalıdır. Veri analizi, makine öğrenmesi modellerinin geliştirilmesi ve optimize edilmesi sürecinde kritik bir aşama olarak, algoritmaların performansını artırmaya yönelik önemli bir rol oynar.

Görsel Canvada Oluşturuldu.

Algoritma Doğruluk Oranları



Doğruluk Oranları:

```
algorithms = ['KNeighbors', 'Logistic Regression', 'SVR', 'Decision Tree', 'RandomForest',  
'ANN', 'MLP']  
accuracy_scores = [0.9766081871345029, 0.9707602339181286, 0.9824561403508771,  
0.9064327485380117, 0.9473684210526315, 0.9707602339181286,  
0.9736842105263158]
```

Algoritma doğruluk sonuçlarına bakarak en yüksek SVM en düşük ise Decision Tree Algoritmaları karşımıza çıkıyor.

Kaynakça

- <https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15>
- <https://efecanxrd.medium.com/yapay-sinir-a%C4%9Flar%C4%B1-nedir-artificial-neural-networks-14c6bae75a44>
- <https://efecanxrd.medium.com/random-forest-karar-a%C4%9Fac%C4%B1-algoritmas%C4%B1-nedir-random-forest-algorithm-28a6172cdd50>
- <https://arslanev.medium.com/makine-%C3%B6%C4%9Frenmesi-karar-a%C4%9Fa%C3%A7lar%C4%B1-decision-tree-algoritmas%C4%B1-nedir-8349ecadff39>
- <https://medium.com/sfu-csmp/a-practical-guide-to-support-vector-machines-svm-ccd6a4d4dd04>
- <https://medium.com/deep-learning-turkiye/nedir-bu-destek-vekt%C3%B6r-makineleri-makine-%C3%B6%C4%9Frenmesi-serisi-2-94e576e4223e>
- <https://towardsdatascience.com/derivative-of-the-sigmoid-function-536880cf918e>
- <https://mfakca.medium.com/lojistik-regresyon-nedir-nas%C4%B1l-%C3%A7al%C4%B1%C5%9F%C4%B1r-4e1d2951c5c1>
- <https://efecanxrd.medium.com/knn-algoritmas%C4%B1-nedir-k-nearest-neighbor-8416db47a1c8>
- <https://arslanev.medium.com/makine-%C3%B6%C4%9Frenmesi-knn-k-nearest-neighbors-algoritmas%C4%B1-bdfb688d7c5f>
- <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>
- Google Colab
- **Slayt ve Kaynak Kodlar İçin:**
- Github.com/zegasega/Breast-Cancer