

# Diyabet Hastalığının Verisel Ön Teşhisi/Tanısı

# Q Konu Başlıkları

1

Projenin Amacı

3

Grafikler Ve Modeller

2

Veri Seti

4

Eğitimler

# Projenin Amacı

Hastaların belirli özelliklerini kullanarak (yaş, vücut kitle indeksi, kan şekeri seviyeleri vb.), hastaların diyabet olup olmadığını belirlemektir. Diyabet, dünya genelinde yaygın bir sağlık sorunudur ve erken teşhis edilmesi ve tedavi edilmesi önemlidir. Bu nedenle, bu projenin amacı, hastaların diyabet riskini değerlendirmek için veri analitiği ve makine öğrenimi tekniklerini kullanmaktır.



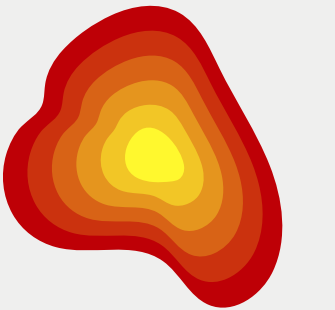
# Veri Seti

Veri setimiz, hastaların diyabet durumunu belirlemek için kullanılan 9 farklı özelliği içermektedir. Bu özellikler hamilelik sayısı, glikoz seviyeleri, tansiyon, cilt kalınlığı, insülin seviyeleri, vücut kitle indeksi (BMI), diyabet soyağacı fonksiyonu, yaş ve sonuç (diyabet olma durumu) özelliklerini kapsamaktadır. Veri seti, toplamda 768 örnekten oluşmaktadır.



# Isı Dağılım Haritası

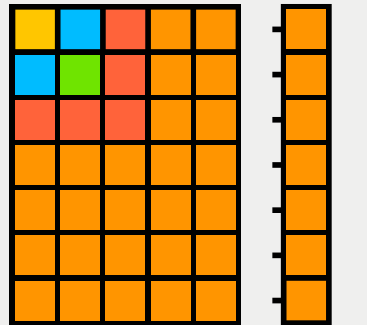
Veri setinin özellikleri arasındaki ilişkileri anlamak için korelasyon analizi yapılmış ve bu analizin sonuçları bir ısı haritası şeklinde görselleştirilmiştir. Aşağıda, özellikler arasındaki korelasyon katsayıları verilmiştir:



# Isı Dağılım Haritası

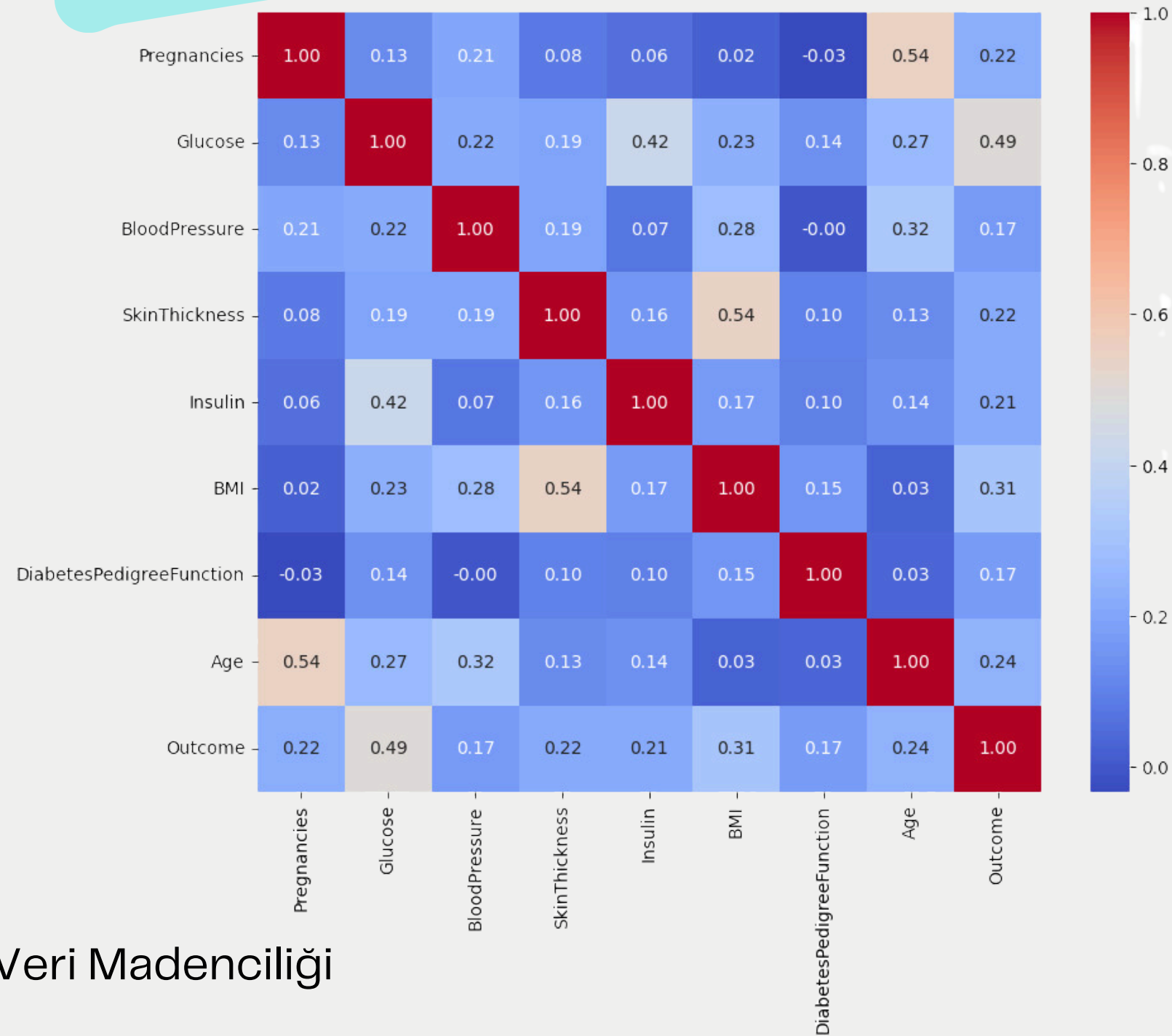
Veri setinin özellikleri arasındaki ilişkileri anlamak için korelasyon analizi yapılmış ve bu analizin sonuçları bir ısı haritası şeklinde görselleştirilmiştir. Aşağıda, özellikler arasındaki korelasyon katsayıları verilmiştir:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
Pregnancies	1.000000	0.127911	0.208522	0.082989	0.056027	0.021565	-0.033523	0.544341	0.221898
Glucose	0.127911	1.000000	0.218367	0.192991	0.420157	0.230941	0.137060	0.266534	0.492928
BloodPressure	0.208522	0.218367	1.000000	0.192816	0.072517	0.281268	-0.002763	0.324595	0.166074
SkinThickness	0.082989	0.192991	0.192816	1.000000	0.158139	0.542398	0.100966	0.127872	0.215299
Insulin	0.056027	0.420157	0.072517	0.158139	1.000000	0.166586	0.098634	0.136734	0.214411
BMI	0.021565	0.230941	0.281268	0.542398	0.166586	1.000000	0.153400	0.025519	0.311924
DiabetesPedigreeFunction	-0.033523	0.137060	-0.002763	0.100966	0.098634	0.153400	1.000000	0.033561	0.173844
Age	0.544341	0.266534	0.324595	0.127872	0.136734	0.025519	0.033561	1.000000	0.238356
Outcome	0.221898	0.492928	0.166074	0.215299	0.214411	0.311924	0.173844	0.238356	1.000000

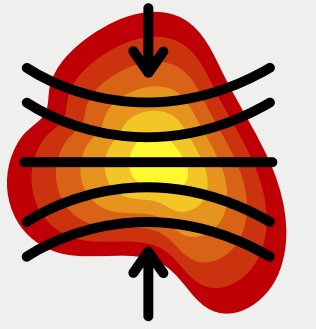




# Isı Dağılım Grafiği



Bu korelasyon analizi yapılmış olan tablonun ısı dağılım grafiği oluşturulmuş olup yanda görüntülmektedir.



# Çıkarımlar

Isı Haritasına göre diyabete en çok etki eden 3 faktör;

- Glikoz Değeri 0,49
- Vücut Kitle Endeksi 0,31
- Yaş 0,24

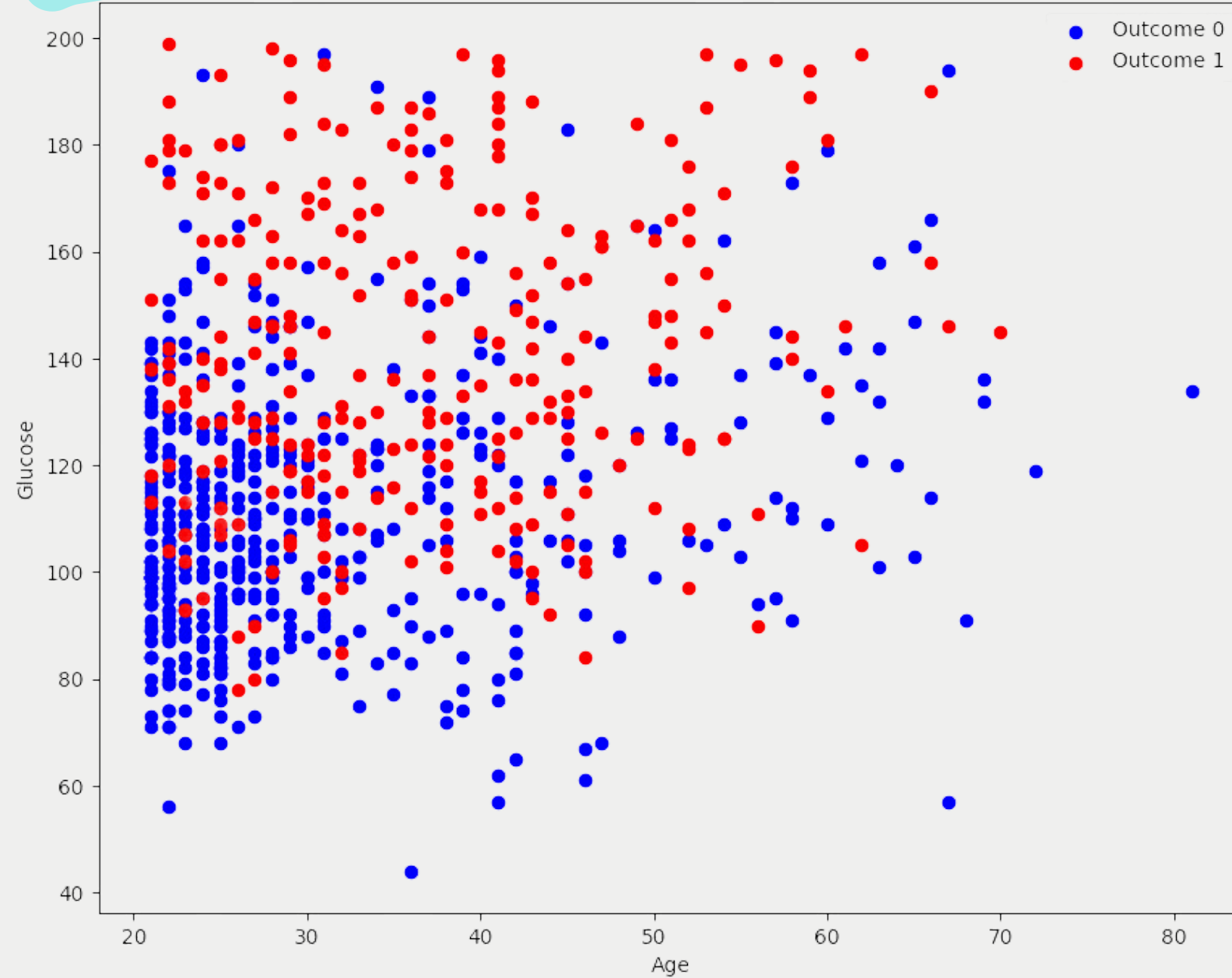
En az etki eden 3 faktör;

- Diyabetik Soy Ağacı 0,17
- Tansiyon 0,17
- Hamilelik Sayısı 0,22

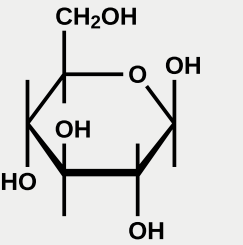




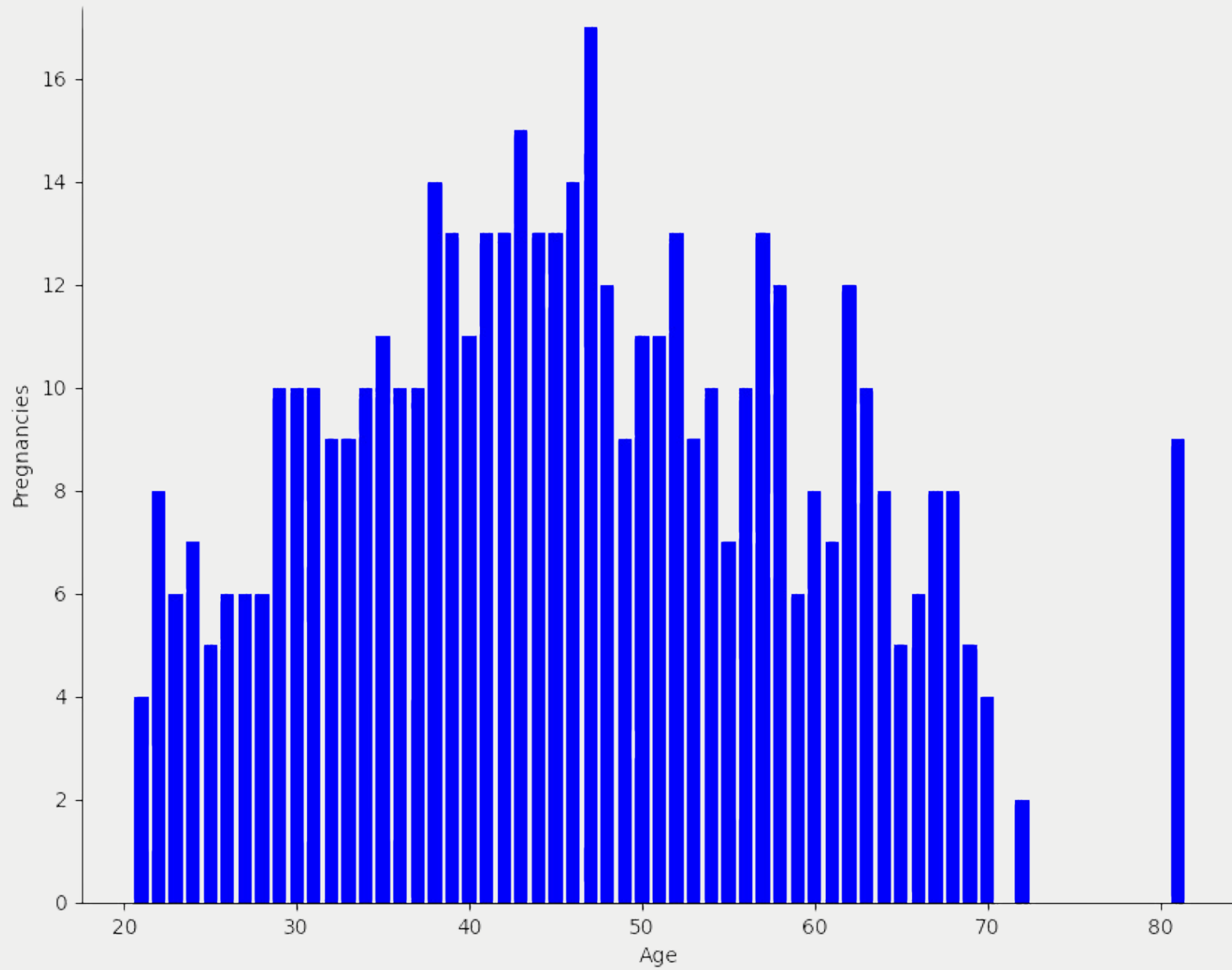
# Grafikler



Bu grafikte, veri setindeki her birey yaşına ve glikoz seviyesine dayalı olarak bir kabarcık olarak temsil edilmiştir. Kabarcıkların rengi, bireyin diyabet hastası olup olmadığını göstermektedir; kırmızı diyabet hastası olmayı, mavi ise diyabet hastası olmamayı simgeler.



# Grafikler



Bu grafik, her yaş grubundaki bireylerin maksimum hamilelik sayısını gösteren çubuklarla temsil edilmiştir.

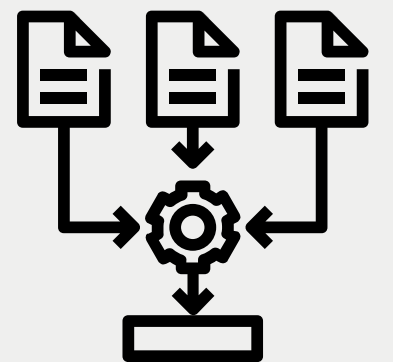


# Modeller

Bu projede

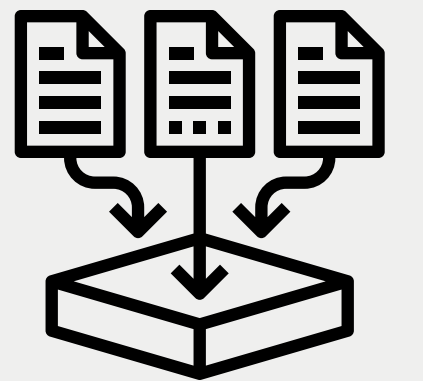
- Lojistik Regresyon
- Destek Vektör Makineleri
- Naive Bayes
- Karar Ağacı
- Yapay Sinir Ağları

olmak üzere toplam 5 model kullanılmıştır.



# Modellerin Eğitimi ve Test Grupları

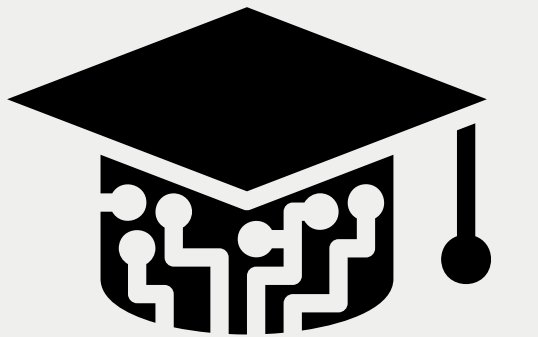
Veri seti, %66 eğitim ve %34 test olacak şekilde rastgele bölünmüştür. Eğitim setinin (%66) ve test setinin (%34) ayrı tutulmasıyla elde edilen sonuçlar, her bir model için ayrı ayrı değerlendirilmiştir. Bu değerlendirme sonuçları, her modelin eğitim ve test verisi üzerindeki performansını belirlemek için kullanılmıştır. Bu süreç, modelin doğruluğunu, hassasiyetini ve diğer performans metriklerini değerlendirmek için kullanılmıştır.



# Eğitimler

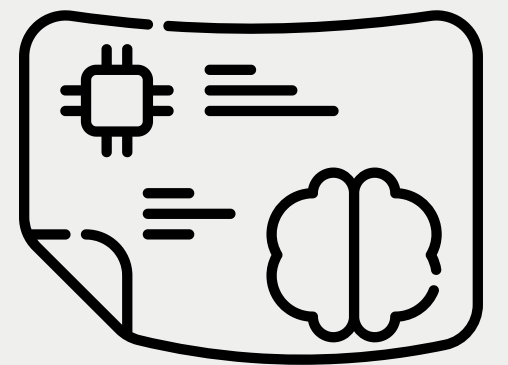
Eğitimler, test seti ve verinin tamamı olmak üzere iki defa gerçekleştirilmiştir. Her bir model için eğitim sonuçları elde edilmiş ve çeşitli metriklerle değerlendirilmiştir.

Her eğitim sonucunda modelin 0 ile 1 arasında bir doğruluk skoru vermesi beklenir.



# Eğitim Verisinin Bölünmeden Kullanıldığı Durum

- Lojistik Regresyon Doğruluk Skoru: 0.7669856459330144
- Destek Vektör Makineleri (SVM) Doğruluk Skoru: 0.7656527682843473
- Naive Bayes Doğruluk Skoru: 0.7734962406015038
- Karar Ağacı Doğruluk Skoru: 0.6717703349282296
- Yapay Sinir Ağı Doğruluk Skoru: 0.7318694463431306





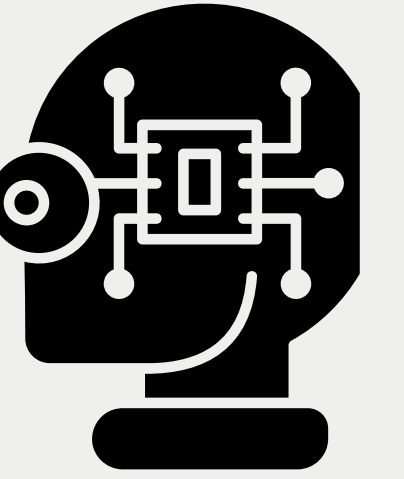
# Eğitim Verisinin Bölünerek Kullanıldığı Durum

- Lojistik Regresyon Doğruluk Skoru: 0.7823529411764707
- Destek Vektör Makineleri (SVM) Doğruluk Skoru: 0.786235294117647
- Naive Bayes Doğruluk Skoru: 0.786313725490196
- Karar Ağacı Doğruluk Skoru: 0.7173725490196078
- Yapay Sinir Ağı Doğruluk Skoru: 0.7034509803921569

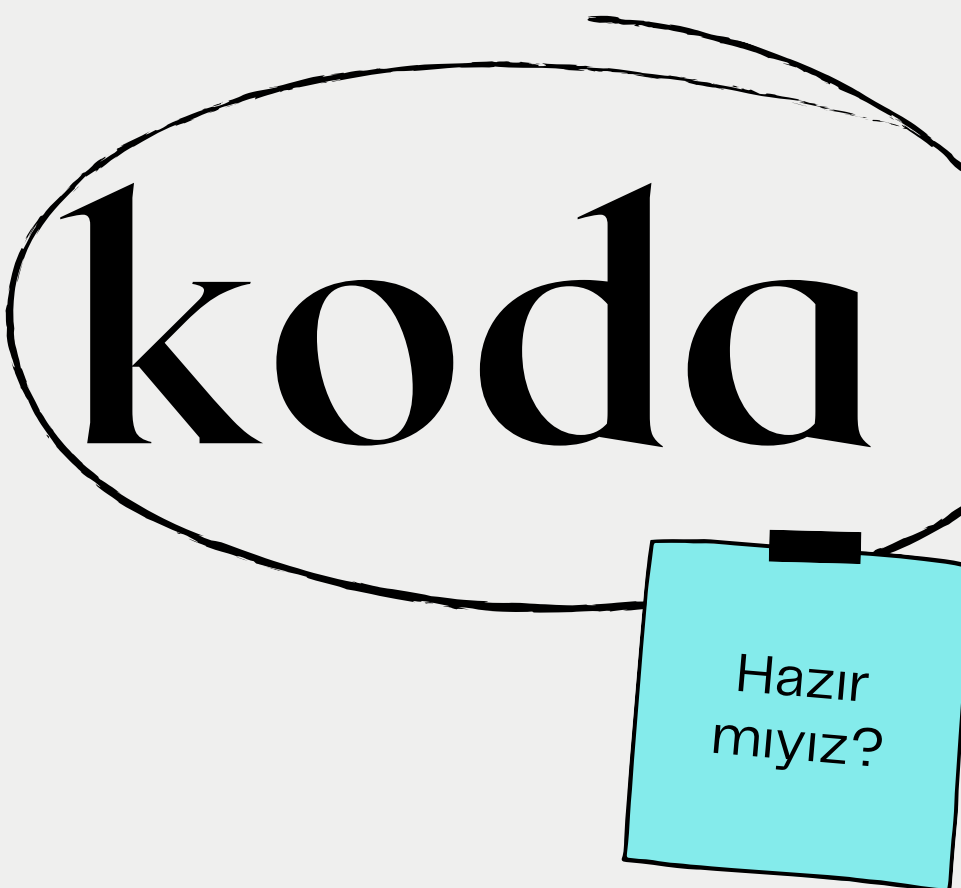


# Çıkarım

Tüm bu incelemelere ve hesaplamalara bakılarak en iyi modelin Bayes modeli olduğu çıkarımı yapılmış ve sonuç buna göre alınmıştır.



# Hadi koda geçelim



Hazır  
mıyız?

# Teşekkürler

Beni  
dinlediğiniz  
için  
teşekkürler