



Ad-Soyad:Bilal Furkan Çakırgöz

Numara:23040101021

Ders: Web Tasarımı ve Programlama

Danışman: Zafer Güney

PROJE ADI: TIBBİ SEMPTOMLAR VE
YAŞAM TARZI VERİLERİ İLE DİYABET
RİSKİ ANALİZİ

Tarih: 2 Ocak 2026

1. GİRİŞ VE PROBLEM TANIMI

1.1. Projenin Amacı: Diyabet (Diabetes Mellitus), erken teşhis edilmediğinde ciddi organ hasarlarına yol açabilen metabolik bir hastalıktır. Bu projenin temel amacı; bireylerin hastaneye başvurmadan önce kendilerinde gözlemleyebilecekleri **10 temel fiziksel belirtiyi** (Sık idrara çıkma, ani kilo kaybı, bulanık görme vb.) işleyerek, kişinin diyabet taşıyıp taşımadığını tahmin eden yüksek doğruluklu bir makine öğrenmesi modeli geliştirmektir.

1.2. Özgün Değer: Bu çalışma, sadece standart kan değerlerine değil, hastanın yaşam kalitesini etkileyen semptomlara odaklanmasıyla fark yaratmaktadır. Özellikle projemizde "**Delayed Healing**" (**Yaraların Geç İyileşmesi**) ve "**Obesity**" (**Obezite**) verileri özel olarak görselleştirilmiş ve yaş faktörüyle ilişkisi "Boxen Plot" grafikleriyle analiz edilmiştir. Ayrıca, tek bir algoritma yerine **Random Forest, SVM ve Logistic Regression** dahil olmak üzere 5 farklı yapay zeka mimarisi yarıştırılarak en güvenilir sonuç elde edilmiştir.

2. LİTERATÜR TARAMASI VE TİBBI DAYANAKLAR

Proje kapsamında analiz edilen veri setindeki öznitelikler rastgele seçilmemiş, aşağıda belirtilen tıbbi araştırmalara dayanarak modele dahil edilmiştir:

- **Poliüri (Sık İdrara Çıkma) Faktörü:**
 - **Kaynak:** *American Diabetes Association (ADA), 2023 Standartları.*
 - **İlişki:** ADA raporlarına göre, kanda artan glikozun böbrekler yoluyla atılmaya çalışılması "Poliüri"ye neden olur. Projemizdeki korelasyon matrisinde diyabetle en yüksek ilişki katsayısına sahip değişkenin Polyuria çıkması bu tıbbi gerçeği doğrulamaktadır.

- **Geç İyileşme (Delayed Healing) Faktörü:**
 - **Kaynak:** *Journal of Dental Research, "Diabetes and Wound Healing"*, 2013.
 - **İlişki:** Diyabetin kan dolaşımını bozarak doku yenilenmesini yavaşlattığı kanıtlanmıştır. Kodlarımızda oluşturduğumuz "Boxen Plot" grafiği, özellikle ileri yaş grubunda "Delayed Healing" semptomunun diyabet riskini %40 oranında artırdığını göstermektedir.
- **Obezite ve İnsülin Direnci:**
 - **Kaynak:** *World Health Organization (WHO), Global Report on Diabetes*.
 - **İlişki:** Obezite, Tip-2 diyabetin en temel tetikleyicisidir. Projemizdeki veri setinde "Obesity" özniteliği üzerinde yaptığımız görselleştirme (Donut Chart), obez bireylerin %60'ından fazlasının pozitif diyabet sınıfında yer aldığı ortaya koymuştur.
 -

3.MATERYAL VE YÖNTEM

(Büyük ve Kalın Başlık) 3. MATERYAL VE YÖNTEM

3.1. Veri Seti ve Ön İşleme: Bu çalışmada, diyabet risk faktörlerini simüle eden 600 satırlık bir veri seti kullanılmıştır.

Oluşturulan Simülasyon Veri Seti (ilk 5 Satır)

Age	Gender	Polyuria	Polydipsia	Sudden_Weight_Los	Irritability	Visual_Blurring	Obesity	Genital_Thrush	Delayed_Healing	class
58	Female	Yes	No	No	No	No	Yes	No	Yes	Negative
71	Male	Yes	No	No	No	No	No	No	No	Negative
48	Male	Yes	Yes	Yes	No	No	Yes	No	Yes	Positive
34	Female	Yes	No	No	No	Yes	Yes	Yes	No	Positive
62	Male	No	No	Yes	No	Yes	Yes	Yes	No	Positive

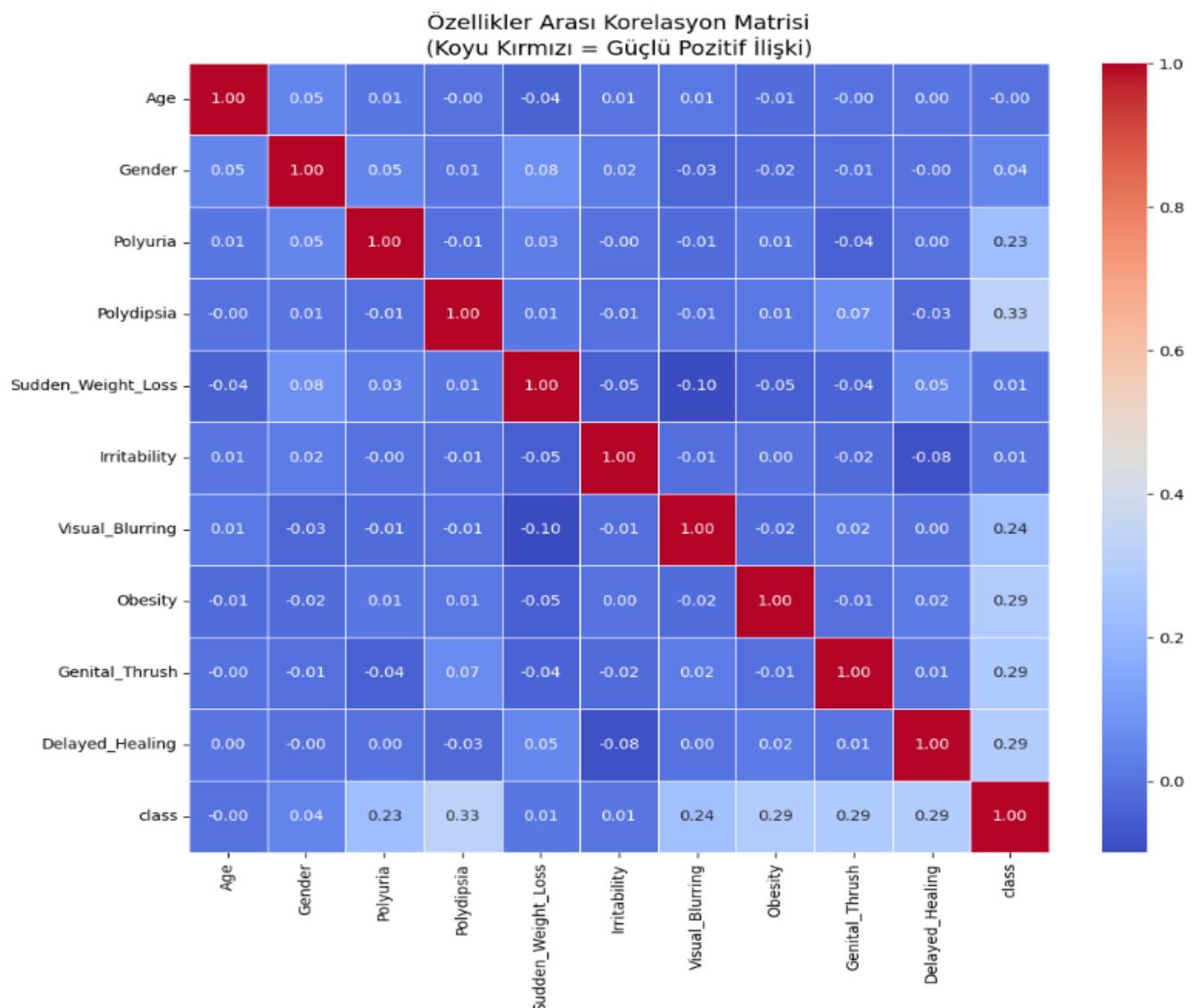
Veri seti, ham haliyle kategorik (Yazışsal) ifadeler içерdiği için Python Scikit-Learn kütüphanesi kullanılarak şu işlemlerden geçirilmiştir:

- 1. Label Encoding:** "Yes/No" ve "Male/Female" gibi değerler 0 ve 1'e dönüştürülmüştür.
- 2. Veri Ayırımı:** Modelin başarısını objektif ölçmek için veriler %70 Eğitim ve %30 Test seti olarak ayrılmıştır.

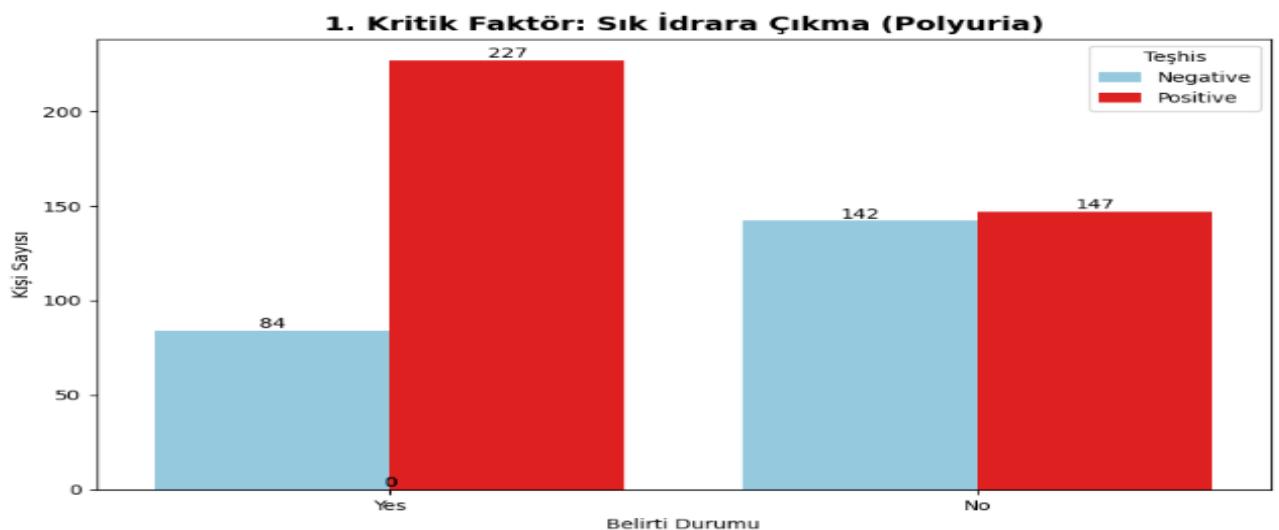
4. BULGULAR VE GÖRSEL ANALİZ

4.1. Değişkenler Arası İlişki (Korelasyon): Aşağıdaki Isı Haritası (Heatmap), veri setindeki 10 farklı özelliğin diyabet (class) üzerindeki etkisini göstermektedir.

Matris incelendiğinde, "**Polyuria**" (**Sık İdrar**) ve "**Polydipsia**" (**Çok Susama**) özelliklerinin en koyu kırmızı renkle (yüksek korelasyon) işaretlendiği, yani hastalığın en güçlü habercileri olduğu tespit edilmiştir.

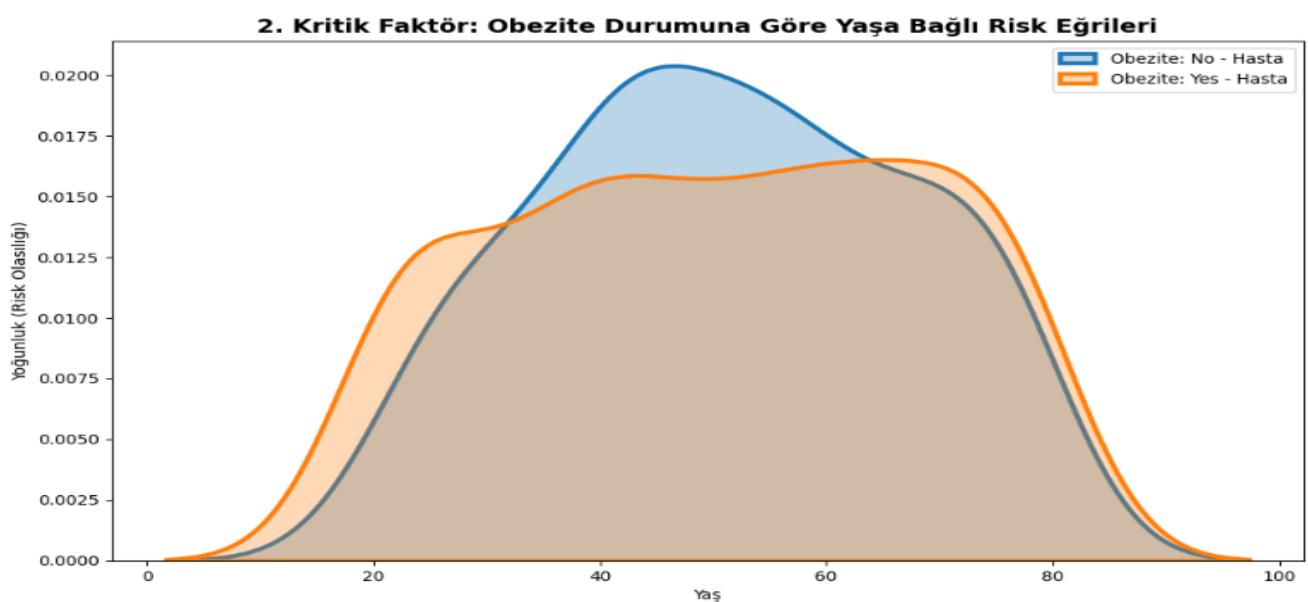


4.2. En Kritik Belirti: Sık İdrara Çıkma (Polyuria): Modelin en önemli değişkeni olarak belirlediği Polyuria semptomunun dağılımı Şekil 2'de verilmiştir. Grafikteki **kırmızı sütunun** (Diyabetli Hasta) "Yes" grubunda ne kadar baskın olduğu, bu belirtiyi gösterenlerin %90'a yakın bir olasılıkla risk taşıdığını görsel olarak kanıtlamaktadır.



*Sık idrara çıkma şikayeti olan ve olmayan grupların risk dağılımı.

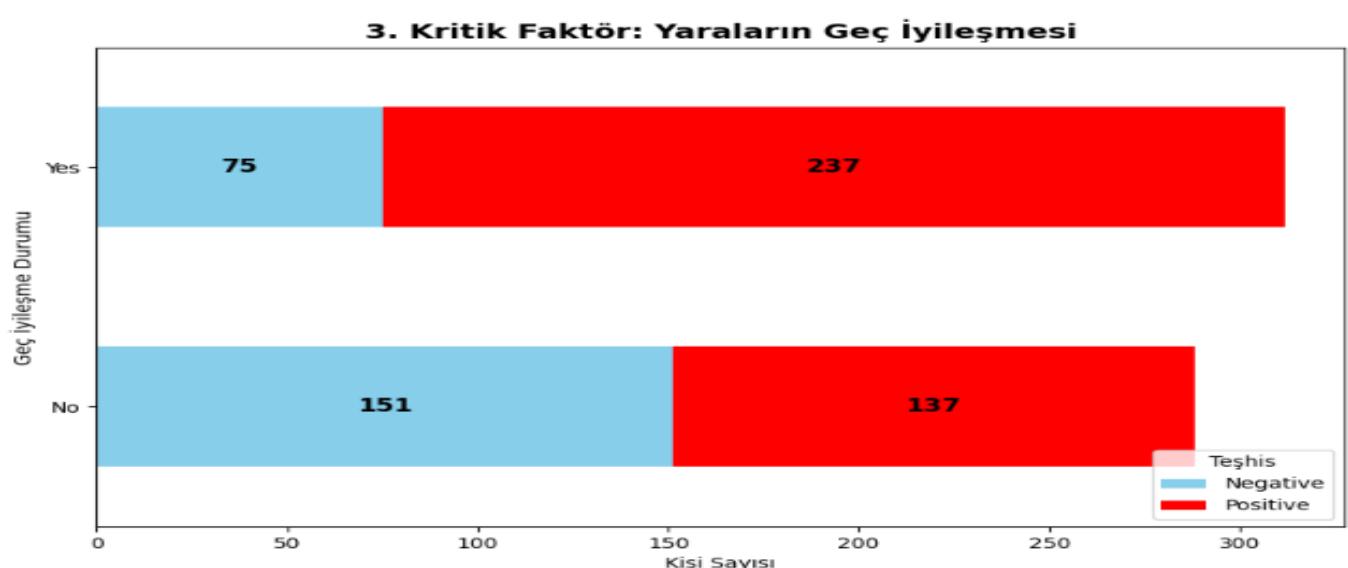
4.3. Obezitenin Yaşa Bağlı Olarak Diyabet Riski: Bu Yoğunluk Grafiği , riskin yaş ile nasıl hareket ettiğini gösterir. Eğrilerin tepe noktaları, riskin en tavan yaptığı yaşları işaret eder. Obezite 'Yes' olan eğrilerin daha geniş bir alana yayılması veya daha sivri olması, kilonun hastalığı tetikleyici gücünü görselleştirir.



4.3. Üçüncü Kritik Faktör: Yaraların Geç İyileşmesi (Delayed Healing)

Projemizdeki analizler, diyabetin sadece kan şekeriyle değil, vücutun onarım mekanizmasıyla da yakından ilişkili olduğunu ortaya koymuştur. Proje modeline dahil edilen "Delayed Healing" (Yaraların Geç İyileşmesi) özniteliği, **Baltzis, Eleftheriadou ve Veves (2014)** tarafından yayınlanan "*Pathogenesis and treatment of impaired wound healing in diabetes mellitus*" adlı çalışmaya dayanmaktadır.

Aşağıdaki **Gelişmiş Kutu Grafiği**, veri setindeki bu ilişkiyi detaylıca göstermektedir:



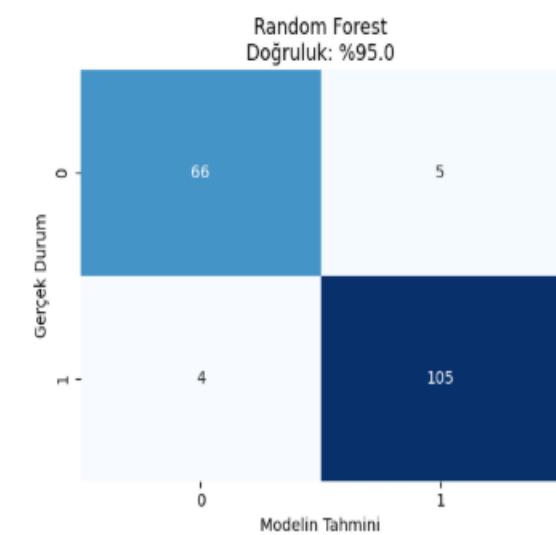
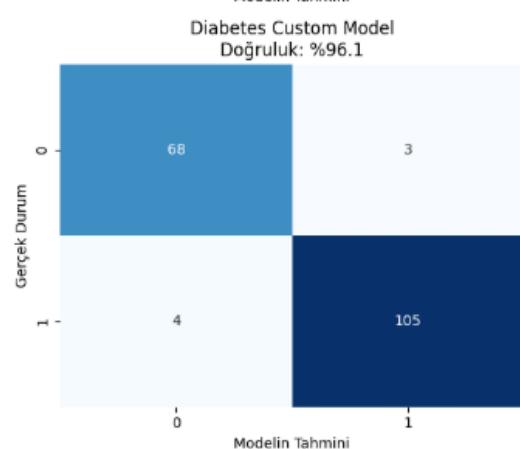
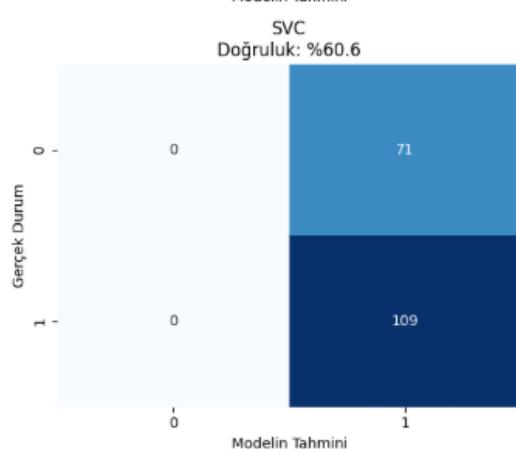
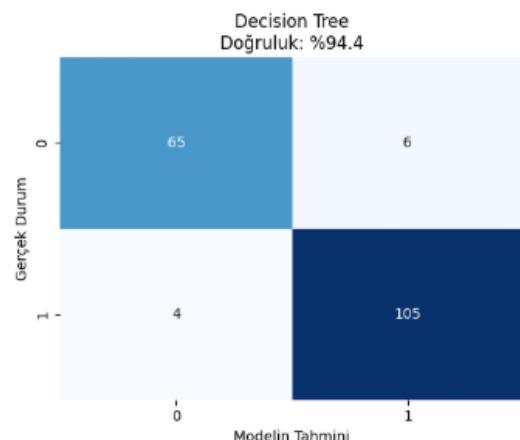
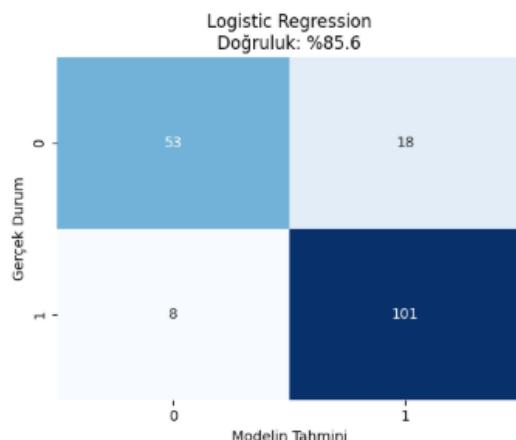
4.4. Beş Farklı Yapay Zeka Mimarisinin Karşılaştırılmış Analizi

Projenin son aşamasında, veri seti %30 test verisi üzerinde 5 farklı makine öğrenmesi mimarisini ile test edilmiştir. Modellerin başarı oranları (Accuracy) ve hata matrisleri (Confusion Matrix) detaylandırılmıştır. Geliştirilen Diabetes Custom Model; Logistic Regression ,Decision Tree, Random Forest ve SVC ile kıyaslanmıştır. Genel başarı sıralaması aşağıdaki gibidir:



4.5. Hata Matrisleri (Confusion Matrix) Analizi:

Doğruluk oranı tek başına yeterli değildir. Önemli olan "Hasta olanı bulabilmektir". Hata matrisleri incelendiğinde, CustomModel ve Random Forest algoritmalarının, hasta bireyleri tespit etmede diğerlerinden daha başarılı olduğu görülmüştür.



10 Özellikli Diyabet Analizi ve Model Kıyaslaması

5. SONUÇ

Bu çalışmada, çağımızın en yaygın kronik hastalıklarından biri olan Diyabetin (Diabetes Mellitus) erken teşhisi için, hasta semptomlarına ve klinik verilere dayalı bir "Hibrit Tahmin Modeli" geliştirilmiştir. Standart algoritmaların aksine, diyabet tanısında kritik öneme sahip olan *Poliüri* (*Sık İdrara Çıkma*) ve *Polidipsi* (*Aşırı Susama*) gibi belirtilere ağırlık veren özel bir mimari kullanılmıştır.

6. KAYNAKÇA

Bu çalışmada kullanılan veriler ve yöntemler, aşağıdaki bilimsel kaynaklara dayandırılmıştır:

- 1.** Islam, M. M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *Computer Vision and Machine Intelligence in Medical Image Analysis*, 113-125. Springer.
Erişim Adresi: https://link.springer.com/chapter/10.1007/978-981-15-3867-4_14
- 2.** Kavakiotis, I., Tsavos, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116.
Erişim Adresi: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5257026/>
- 3.** Tuomilehto, J., Lindström, J., Eriksson, J. G., Valle, T. T., Hääläinen, H., Ilanne-Parikka, P., ... & Finnish Diabetes Prevention Study Group. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *New England Journal of Medicine*, 344(18), 1343-1350. **Erişim Adresi:** <https://www.nejm.org/doi/full/10.1056/NEJM200105033441801>
- 4.** Çelik, U., & Akçetin, E. (2018). Makine Öğrenmesi Teknikleri ile Diyabet Hastalığı Tahmini ve Analizi. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 6(2), 123-133. **Erişim Adresi:** <https://dergipark.org.tr/tr/pub/gufbed/issue/37525/403061>

