



## GEBELİKTE ANNE SAĞLIĞI RİSK GRUPLARININ TAHMİNİNE YÖNELİK MAKİNE ÖĞRENMESİ TABANLI BİR KARAR DESTEK SİSTEM TASARIMI

İrem ŞENYER YAPICI<sup>1\*</sup>, Rukiye UZUN ARSLAN<sup>2</sup>

<sup>1</sup>Zonguldak Bülent Ecevit University, Faculty of Engineering, Department of Computer Engineering, 67100, Zonguldak, Türkiye

<sup>2</sup>Zonguldak Bülent Ecevit University, Faculty of Engineering, Department of Electrical and Electronics, 67100, Zonguldak, Türkiye

**Özet:** Gebelik döneminde anne sağlığı risklerinin erken tespiti ve uygun müdahalelerin yapılması, anne ve bebek sağlığı açısından hayati bir önem taşımaktadır. Bu süreçte, büyük veri kümelerinden elde edilen karmaşık ilişkileri ve desenleri otomatik olarak analiz edebilen makine öğrenme (MÖ) algoritmalarının kullanımı son derece kritik bir rol oynamaktadır. MÖ algoritmaları, büyük veri setlerindeki gizli bilgileri açığa çıkararak, gebelikle ilişkili risk faktörlerini daha doğru bir şekilde belirleme imkanı sunmaktadır. Bu bağlamda gerçekleştirilen bu çalışmada, gebelik sürecinde anne sağlığı risk seviyelerinin özellikle yüksek riskli hamileliklerin tahmininde başarının artırılmasına odaklanılmıştır. Bunun için öncelikle başarıyı artıracak önemli (kritik) özellikler belirlenmiş ve altı farklı makine öğrenme algoritması kullanılarak etkili bilgisayar temelli karar destek sistemi tasarlanmaya çalışılmıştır. Ki-Kare testi SelectKBest yöntemiyle birlikte uygulanarak, veri setindeki en kritik özelliklerin yaş, sistolik kan basıncı ve diyastolik kan basıncı olduğu tespit edilmiştir. Yanı sıra veri setindeki dengesizliği gidermek için sentetik azınlık aşırı örnekleme tekniğinden (SMOTE) yararlanılmıştır. Önerilen modelde kullanılan MÖ algoritmalarının başarımları hold-out performans değerlendirme yöntemiyle analiz edilmiştir. Elde edilen bulgular ışığında, SMOTE tekniğinin kullanılmasının gebelikteki risk seviyelerinin tahmininde model başarımlarını artırmada olumlu bir etkiye sahip olduğu belirlenmiştir. Önerilen modelde her bir sınıflandırma algoritması için en yüksek sınıflandırma başarımları yüksek risk sınıfı için elde edilmiştir. Kullanılan algoritmalar arasında, %97 başarı oranıyla en üstün performansa sahip olanın ekstrem gradyan artırma algoritması olduğu tespit edilmiştir. Genel olarak elde edilen sonuçlar, önerilen modelin yüksek risk taşıyan gebeliklerin tespitinde son derece etkili olduğunu doğrulamaktadır. Bu bulgu, önerilen MÖ temelli karar destek sisteminin uzman hekimlere gebelik sürecinde daha doğru teşhisler koyma ve gerekli müdahaleleri daha hızlı bir şekilde gerçekleştirme konusunda önemli bir destek sağlama potansiyeline sahip olduğunu göstermektedir.

**Anahtar kelimeler:** Gebelik risk sınıfları, Anne sağlığı, Makine öğrenmesi, Özellik seçimi, SMOTE


### A Machine Learning-Based Decision Support System Design to Predict Maternal Health Risk Groups during Pregnancy


**Abstract:** Early detection of maternal health risks during pregnancy and appropriate interventions are vital for maternal and infant health. In this process, the use of machine learning (ML) algorithms that can automatically analyze complex relationships and patterns from large datasets plays a critical role. By revealing hidden information in large data sets, ML algorithms offer the opportunity to more accurately identify pregnancy-related risk factors. In this context, this study focuses on increasing the success of predicting maternal health risk levels during pregnancy, especially in high-risk pregnancies. For this purpose, firstly, important (critical) features that will increase the success are identified and the most effective computer-based decision support system is designed by using six different machine learning algorithms. By applying the Chi-Square test in combination with the SelectKBest method, it was determined that the most critical characteristics in the dataset were age, systolic blood pressure and diastolic blood pressure. In addition, synthetic minority oversampling technique (SMOTE) was utilized to address the imbalance in the dataset. The performance of the ML algorithms used in the proposed model is analyzed by hold-out performance evaluation method. In the light of the findings, it was determined that the use of the SMOTE technique has a positive effect on improving model performance in predicting risk levels in pregnancy. In the proposed model, the highest classification performance for each classification algorithm was obtained for the high risk class. Among the algorithms used, the extreme gradient boosting algorithm was found to have the best performance with a 97% success rate. Overall, the results confirm that the proposed model is highly effective in detecting high-risk pregnancies. This finding shows that the proposed ML-based decision support system has the potential to provide significant support to specialist physicians in making more accurate diagnoses and performing the necessary interventions more quickly during pregnancy.

**Keywords:** Pregnancy risk classes, Maternal health, Machine learning, Feature selection, SMOTE

\*Sorumlu yazar (Corresponding author): Zonguldak Bülent Ecevit University, Faculty of Engineering, Department of Computer Engineering, 67100, Zonguldak, Türkiye

E mail: senyerirem@gmail.com (İ. ŞENYER YAPICI)

İrem ŞENYER YAPICI  <https://orcid.org/0000-0003-0655-340X>

Rukiye UZUN ARSLAN  <https://orcid.org/0000-0002-2082-8695>

Gönderi: 19 Mart 2024

Kabul: 29 Nisan 2024

Yayınlanma: 15 Mayıs 2024

Received: March 19, 2024

Accepted: April 29, 2024

Published: May 15, 2024

Cite as: Şenyer Yapıcı İ, Uzun Arslan R. 2024. A machine learning-based decision support system design to predict maternal health risk groups during pregnancy. BSJ Eng Sci, 7(3): 509-520.



## 1. Giriş

Günümüzde tıp ve sağlık alanında sağlık hizmetlerinin hastalık önleme konusunda kişi merkezli bir yaklaşım benimsemesi gerektiği kabul edilmektedir. Bu bağlamda, hastalık risklerinin daha hassas bir şekilde değerlendirilmesine olanak tanıyan makine öğrenimi (MÖ) tekniklerinin tıp ve sağlık alanındaki kullanımı giderek artmaktadır. MÖ'nin temel amacı, insan müdahalesinin minimum düzeyde olduğu ya da hiç olmadığı mevcut verilerden öğrenmeyi otomatik olarak gerçekleştiren öğrenme algoritmalarının geliştirilmesidir (Diamantoulaki ve ark., 2022). Literatürde, tıbbi tanı süreçlerinde uzman hekimlere yardımcı olmak amacıyla geliştirilen çok sayıda MÖ temelli karar destek sistemleri bulunmaktadır. Anne sağlığının izlenmesi ve geliştirilmesinde karar destek sistemlerinin kullanımı; hamilelik, doğum ve doğum sonrası dönemlerde annenin ve bebeğin sağlığını korumak adına oldukça önem arz etmektedir. Anne sağlığı, hamilelik, doğum ve doğum sonrası dönemde annenin fiziksel ve psikolojik sağlığını kapsayan bir kavramken; anne sağlığı riski ise bu süreçlerde anne ve bebeğin sağlığını olumsuz etkileyebilecek potansiyel riskleri ifade etmektedir. Bu riskler, anne adayının veya bebeğin sağlığını olumsuz yönde etkileyebilecek çeşitli durumları içermekte olup, gebelik sürecinde veya doğum sonrası dönemde ortaya çıkabilmektedir. Örneğin, yüksek tansiyon, diyabet, enfeksiyonlar, gebelik zehirlenmesi gibi sağlık sorunları anne sağlığı riski olarak değerlendirilmektedir. Ayrıca, gebelik komplikasyonları, doğum sonrası depresyon, doğum sonrası kanamalar gibi durumlar da anne sağlığı riski olarak kabul edilmektedir. Bu bağlamda gebelik sürecinde yüksek riskli gebeliklerin saptanması anne ve fetus sağlığının korunması açısından hayati bir rol oynamaktadır. Bu riskler öngörüldüğü veya teşhis edilebildiği takdirde gerekli önlemler alınabilmekte ve uygun tedavi yöntemleri planlanabilmektedir. Böylelikle anne ve fetus ölümleri kısmen engellenebilmektedir. Gebelikteki risk faktörlerine göre gebelik süreci düşük, orta ve yüksek risk düzeyi olmak üzere üç sınıfa ayrılmaktadır. MÖ algoritmaları sayesinde, geniş veri setleri üzerinde analizler yaparak gebelik sürecinde etkili olan risk faktörleri tanımlanabilmekte ve bu bilgiler kullanarak da gebeliğin risk düzeyleri sınıflandırılabilir. Böylelikle, uzmanlara gebelik sürecini daha yakından takip ederek olası riskleri daha erken aşamada teşhis etmelerinde ve uygun tedavi süreçlerini planlamalarında destek olunmaktadır. Bu bağlamda literatürdeki çalışmalar incelendiğinde, MÖ algoritmalarının gebelikteki risk faktörlerinin belirlenmesinde ve risk düzeylerinin sınıflandırılmasında oldukça başarılı olduğu görülmektedir (Al-Hindi ve ark., 2020; Finlayson ve ark., 2020; Macrohon ve ark., 2022; Togunwa ve Babatunde, 2023; Tokmak, 2023). Örneğin, Rai ve arkadaşları (2018), gebelik komplikasyonları sonucu artan anne ve bebek ölüm risk seviyelerinin parametrelerini değerlendirmek amacıyla anketler aracılığıyla uzmanlardan alınan bilgilere dayanarak yeni

bir model geliştirmişlerdir. Geliştirilen modelin başarımı 117 hamile kadından alınan verilerden 14 özellik seçilerek, yapay sinir ağı (YSA) ve Naive Bayes (NB) algoritmaları kullanılarak analiz edilmiştir. Analizler sonucunda YSA ile %80'lik bir başarımla elde edilirken, NB ile %70 doğruluk oranına ulaşılmıştır. Ahmed ve ark. (2020) tarafından anne sağlık risk faktörlerini tespit etmek için bulut tabanlı bir anne sağlık sistemi önerilmiştir. Önerilen sistemde risk faktörlerinin analizinde, riskin yoğunluğuna göre kategorize etme ve sınıflandırma yaklaşımları kullanılmıştır. Farklı MÖ algoritmaları arasında yapılan karşılaştırmalar sonucunda, gebelik risk seviyesinin sınıflandırılması ve tahmin edilmesi durumunda modifiye karar ağacı (KA) algoritmasının en yüksek doğruluğu verdiği tespit edilmiştir. Ahmed ve Kasem (2020) çalışmalarında risk faktörlerine dayalı olarak gebelikteki risk düzeyini keşfetmek için MÖ algoritmalarından yararlanmışlardır. Çalışmada Pima-Indian-diyabet veri seti kullanılarak risk faktörleri analiz edilmiş ve çeşitli MÖ algoritmaları karşılaştırılmıştır. Yapılan analizler sonucunda gebelik risk seviyesini sınıflandırılmasında ve tahmininde lojistik model ağacı algoritmasının en yüksek doğruluğu verdiği tespit edilmiştir. Buna ek olarak, seçilen birkaç gebe kadının verileri IoT (nesnelerin interneti) özellikli cihazlar aracılığıyla toplanarak, aynı işlemler bu veri kümesine de uygulanmıştır. Karşılaştırmalar sonucunda, risk tahmininin mevcut Pima-Hint diyabet veri kümesi ve gerçek veri kümesi için aynı olduğu gösterilmiştir. 2022 yılında yapılan bir çalışmada, hamilelik sırasında anne sağlığında ortaya çıkabilecek riskleri erken evrede teşhis etmek için rasgele orman (RO) algoritması kullanılmıştır. KA, k-en yakın komşu (KNN), NB gibi algoritmaların da karşılaştırıldığı makalede, RO algoritması %73,37'lük doğruluk oranı ile en yüksek başarımla verdiği tespit edilmiştir. Ayrıca çalışmada modelin genetik algoritmayla optimize edilmesi durumunda başarımın daha da yükseldiği gösterilmiştir (Ramdhani ve ark., 2022). 2022 yılında yapılan başka bir çalışmada özyinelemeli özellik seçimi yapılarak hamilelik dönemindeki risk faktörlerinin sınıflandırılmasında lojistik regresyon (LR), NB ve RO algoritmalarının aynı başarımla verdiği gösterilmiştir (Edayath, 2022). Umoren ve ark. (2022) anne ölümlerine dair risk tahmin modellerini inceleyerek, risk tahminlerini uygulanabilir hale getirmek için KA algoritmasına dayalı bir model geliştirmişlerdir. Geliştirdikleri modelin başarımı, destek vektör makineleri (DVM) dayalı modelle kıyaslanmış ve KA yaklaşımının daha yüksek bir başarımla sahip olduğu ortaya koyulmuştur. Pawar ve ark. (2022) çalışmalarında anne sağlığı risk tahmini için sekiz farklı MÖ algoritması kullanmışlardır. Özellik seçiminin Gini indeksine göre yapıldığı çalışmada en yüksek başarımla RO (%70,21) algoritmasıyla elde edilmiştir. Mutlu ve ark. (2023) IoT tabanlı risk izleme sistemleri aracılığıyla toplanan bir veri seti üzerinde anne sağlığı riskini belirlemek için altı farklı MÖ algoritması kullanmışlardır. Yaş, sistolik kan basıncı, diyastolik kan basıncı, solunum hızı ve kalp atışı

hızı gibi özellikler göz önünde tutularak geliştirdikleri modelin başarımını farklı performans metriklerine göre değerlendirmişlerdir. Yapılan analizler sonucunda en yüksek başarıma KA algoritmasıyla (%89,16) ulaşılmıştır. Şahin ve ark. (2023) anne sağlığı riski tahmininde sekiz farklı MÖ algoritmasının başarımı temel bileşenler analizi (PCA) ve lineer diskriminant analizi (LDA) uygulanması durumları için analiz edilmiştir. Yapılan analizler sonucunda LDA'nın sınıflandırıcı başarımları pozitif yönde etkilediği gösterilmiştir.

Yukarıda bahsedilen çalışmalar, hamilelik sırasında anne sağlığı risklerinin tahmin edilmesinde MÖ ve yapay zekâ tekniklerinin etkin bir şekilde uygulanabileceğini göstermektedir. Ancak, hangi tekniğin daha etkili olduğu konusunda kesin bir cevap bulunmamaktadır. Çünkü her bir tekniğin avantajları ve dezavantajları bulunmakla birlikte, sonuçlar birçok faktöre bağlı olarak değişebilmektedir. Örneğin, bazı durumlarda bir tekniğin başarısı veri setinin özelliklerine, boyutuna ve kalitesine bağlıken, diğer durumlarda kullanılan öznelik seçimi, model hiperparametrelerinin ayarlanması veya eğitim sürecindeki diğer faktörler belirleyici olabilmektedir. Dolayısıyla, en etkili tekniği belirlemek için daha fazla araştırma yapılması ve farklı yöntemlerin kapsamlı bir şekilde karşılaştırılması gerekmektedir. Bu bağlamda gerçekleştirilen bu çalışmada altı farklı MÖ algoritmalarını kullanarak gebelik sırasında oluşabilecek riskler sınıflandırılmıştır. Çalışmanın temel

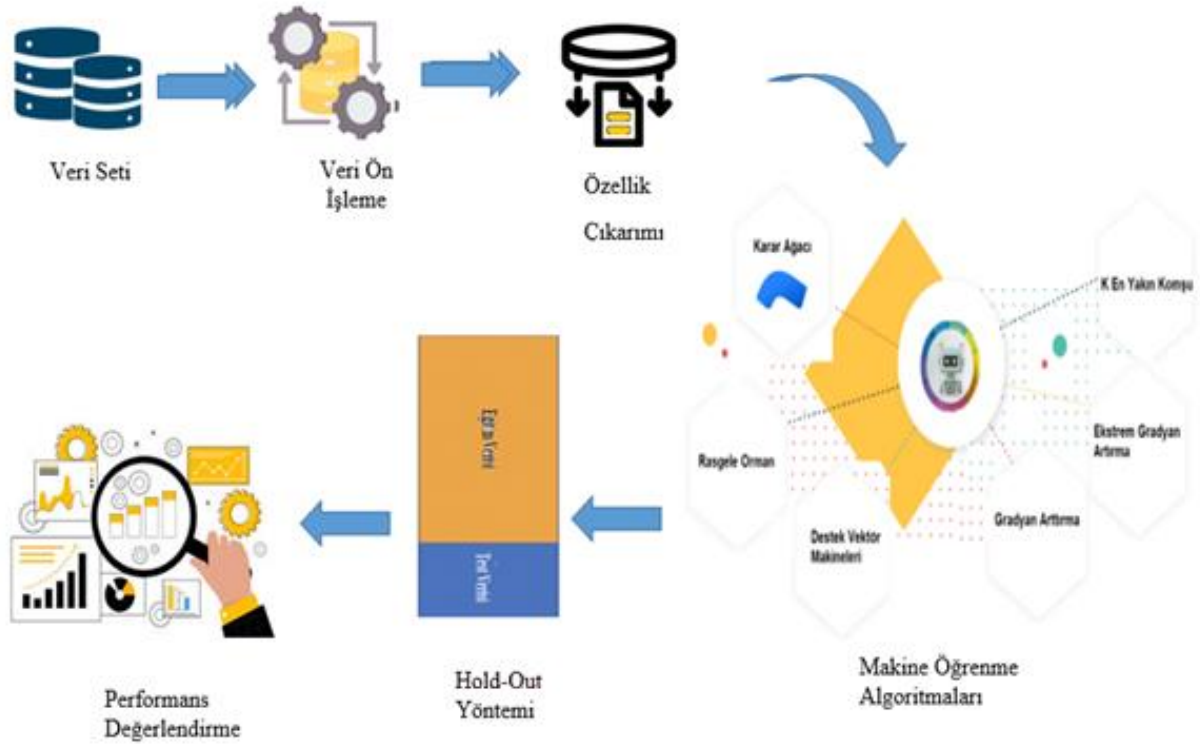
hedefi, yüksek riskli gebeliklerin tespitinde MÖ algoritmalarının ne kadar etkili olduğunu değerlendirmek ve en uygun modeli belirlemektir. Bu çalışmanın, anne sağlığı risklerinin özellikle yüksek gebelik risk seviyesinin erken teşhisinde ve uygun önlemlerin alınmasında uzman hekimlere ve anne adaylarına önemli bir katkı sağlayacağı düşünülmektedir.

## 2. Materyal ve Yöntem

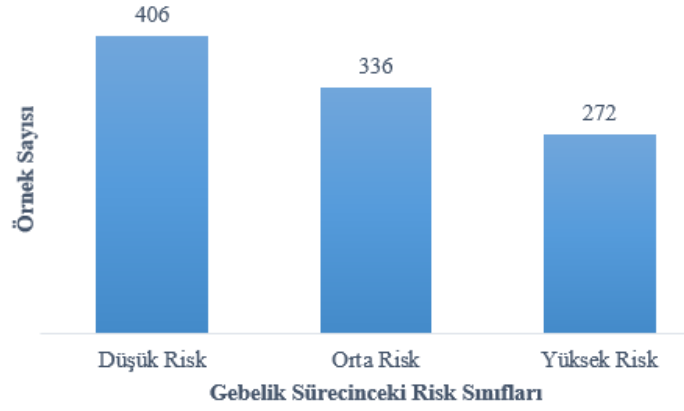
Bu bölümde, gebelik sürecinde anne sağlığı risk seviyesinin tahmine yönelik önerilen bilgisayara destekli karar destek sisteminde kullanılan veri seti, ön işleme, özellik çıkarma yöntemleri ve kullanılan MÖ algoritmaları detaylı bir şekilde açıklanmıştır. Önerilen sistemin genel blok diyagramı Şekil 1'de gösterilmiştir.

### 2.1. Veri Seti

Çalışmada, açık erişimli UCI Machine Learning Repository veri tabanında yer alan "Maternal Health Risk" başlıklı veri seti kullanılmıştır (Ahmed, 2023). Bangladeş'in farklı bölgelerindeki sağlık kuruluşlarından toplanan veri seti, IoT tabanlı bir risk izleme sistemi aracılığıyla elde edilmiş olup, 1014 örnek ve 7 özellik içermektedir. Tablo 1'de veri setindeki özellikler ve bunların açıklamaları detaylarıyla sunulmuştur. Veri seti, Şekil 2'de gösterildiği gibi 272 yüksek risk (%26,82), 336 orta risk (%33,14) ve 406 düşük risk (%40,04) sınıflarına ait örnekler içermektedir. Tablo 2' de veri setine ait örnek bir görünüm verilmiştir.



Şekil 1. Önerilen modelin blok diyagramı.



Şekil 2. Kullanılan veri setindeki örneklerin risk sınıflarına göre dağılımı.

Tablo 1. Veri setinde yer alan özelliklerinin tanımı ve ölçü birimleri

Özellik	Tanımı	Ölçü Birimi
Yaş	Anne hamilelik yaşı	Nümerik (Yıl)
Sistolik Kan Basıncı	Kalbin kasılması sırasında arterlerdeki maksimum basınç	Nümerik (mmHg)
Diastolik Kan Basıncı	Kalbin gevşemesi sırasında arterlerdeki minimum basınç	Nümerik (mmHg)
Kan Şekeri	Kandaki glukoz miktarı	Nümerik (mmol/L)
Vücut Sıcaklığı	Annenin vücut sıcaklığı	Nümerik (°F)
Kalp Atış Hızı	Normal dinlenme durumunda dakikadaki kalp atış hızı	Nümerik (BPM)
Risk Seviyesi	Hamilelik risk düzeyi	Kategorik

Tablo 2. Kullanılan veri setindeki gebelik döneminde anne sağlığı risk sınıflarına ait örnek

Yaş	Sistolik Kan Basıncı	Diastolik Kan Basıncı	Kan Şekeri	Vücut Sıcaklığı	Kalp Atış Hızı	Risk Seviyesi
48	120	80	11	98	88	Yüksek Risk
17	90	63	6.9	101	70	Orta Risk
33	120	75	10	98	70	Yüksek Risk
30	120	80	7.8	98	70	Düşük Risk
40	140	100	13	101	66	Yüksek Risk
20	120	75	7.5	98	70	Düşük Risk
17	110	75	12	101	76	Yüksek Risk
34	120	75	8	98	70	Düşük Risk
25	140	100	7.01	98	80	Yüksek Risk

## 2.2. Özellik Seçimi

Öngörücü bir modelin geliştirilmesinde özellik seçimi, veri kümesinden tahmin değişkenine veya çıktısına en çok katkı sağlayan özelliklerin belirlenmesi sürecidir. Veri setindeki özelliklerin doğru seçimi modelin daha hızlı ve daha doğru sonuçlar üretmesine yardımcı olurken, gereksiz veya gürültülü özelliklerin çıkarılmasıyla modelin genelleme yeteneğini arttırmaktadır. Böylelikle daha etkili ve güvenilir bir öğrenme süreci sağlanmaktadır (Selvakuberan ve ark., 2008; Jahan ve ark., 2021). Sınıflandırma tekniklerinde uygulanan farklı özellik seçim yöntemleri bulunmaktadır. Bunlardan biri olan Ki-kare yöntemi, kategorik değişkenler arasındaki ilişkiyi istatistiksel açıdan değerlendirmektedir. Bu yöntemde özelliklerin hedef değişkenle bağımsızlığını incelenerek, değişkenler arasındaki ilişkinin anlamlı olup olmadığını belirlemektedir. Uygulamalarda genellikle, Ki-kare,

SelectKBest veya benzeri bir yöntemle birlikte kullanılmaktadır (Kırlı ve ark., 2023). Bu bağlamda gerçekleştirilen bu çalışmada Ki-Kare SelectKBest yöntemiyle birlikte kullanılarak uygulanmıştır. SelectKBest, bir sınıflandırıcı için veri setindeki en anlamlı ve etkili olan özellikleri seçmek için kullanılmaktadır. Yöntem veri setindeki her bir özelliğin sınıflandırıcıdaki önemini belirlemek için Ki-kare testini kullanarak, en yüksek skora sahip olan özellikleri seçmektedir. Böylelikle, verilerin sınıflandırılmasında gereksiz veya az katkı sağlayan özellikler elimine edilerek sınıflandırma doğruluğu artırılmaktadır.

## 2.3. Sentetik Azınlık Aşırı Örneklemeye Tekniği (SMOTE)

Sentetik Azınlık Aşırı Örneklemeye Tekniği (SMOTE), dengesiz veri setlerinde sınıf dengesini sağlamak için kullanılan bir yöntemdir. Bu teknik, azınlık sınıfına ait örneklerin sentetik olarak oluşturulmasını sağlayarak,

veri setindeki sınıflar arasındaki dengesizliği azaltmaktadır. SMOTE, azınlık sınıfındaki örnekler arasında benzerlik ölçüsüne dayalı olarak yeni sentetik örnekler oluşturmaktadır. Bu şekilde, azınlık sınıfı daha dengeli bir şekilde temsil edilerek, sınıflandırma performansları artırılmaktadır (Abacı ve Yıldız, 2023).

#### 2.4. Sınıflandırma Algoritmaları

Sınıflandırma; veri noktalarının belirli özelliklerine göre analiz edilerek, bunları önceden tanımlanmış sınıflara veya kategorilere atama sürecidir. Bu süreç, MÖ algoritmaları kullanılarak gerçekleştirilmekte ve genellikle bir modelin eğitilmesiyle sağlanmaktadır. Çalışma kapsamında hamilelik sürecinde anne sağlığı risklerinin sınıflandırılmasında altı farklı MÖ algoritması kullanılmıştır (Şekil 3).



Şekil 3. Kullanılan makine öğrenme algoritmaları.

$k$  en yakın komşu (KNN) algoritması, sınıflandırma ve regresyon problemleri için kullanılan popüler bir denetimli MÖ algoritmasıdır. Özellikle küçük boyutlu veri kümelerinde ve sınıflar arasındaki ayrımın belirgin olduğu durumlarda etkili bir şekilde çalışmaktadır. Bu algortmada yeni bir veri noktası sınıflandırılırken,  $k$  en yakın komşusu dikkate alınmaktadır. Bunun için genellikle Öklid, Manhattan veya Minkowski gibi uzaklık ölçümleri kullanılarak yeni veri noktası ile en yakın  $k$  komşuları arasındaki mesafeler hesaplanmaktadır. Hesaplama işleminden sonra  $k$  komşunun etiketlerine göre test örneğinin sınıfını belirlemektedir (Hacıbeyoğlu ve ark., 2023).

Karar ağacı (KA), bir veri kümesindeki özelliklerin değerlerine göre sınıflandırma veya regresyon yapmak için kullanılan bir algortmadır. Algortmada veri kümesini en iyi şekilde bölmek için bir dizi karar kuralı oluşturulmakta ve bu kurallar kullanılarak veri daha homojen alt kümelere ayırmaktadır. Böylelikle, basit kararlarla karmaşık ilişkileri açıklama ve yorumlama kolaylığı sağlanmaktadır (Yakut ve Bolat, 2020).

Rasgele Orman (RO), birçok karar ağacının bir araya gelerek oluşturduğu bir ensemble (topluluk) öğrenme algoritmasıdır. Algortmada her bir karar ağacı, rastgele özellikler kullanılarak eğitilmekte ve ardından bunların sonuçlarını bir araya getirilerek hem daha kararlı hem de

genelleştirilebilir bir tahmin yapılmaktadır (Yıldız ve ark., 2023).

Destek vektör makineleri (DVM), veri noktalarını sınıflandırmak veya regresyon yapmak için kullanılan parametrik olmayan güçlü bir denetimli MÖ algoritmasıdır. Algortma; veri noktalarını ayıran en iyi hiperdüzlemi bulmaya çalışarak, sınıflandırma problemlerinde optimum ayrımı sağlamaktadır (Elen ve ark., 2022).

Gradyan arttırma (Gradient Boosting, GB) zayıf tahminicileri (genellikle karar ağaçları) bir araya getirerek güçlü bir tahminci oluşturan MÖ algoritmasıdır. Bu algortma, tahmin edilen ve gerçek değer arasındaki hatayı en aza indirmek için ardışık öğrenme adımları kullanmaktadır. Her adımda, önceki modelin hatalarına odaklanarak yeni bir tahminci ekleyerek modelin performansını arttırmaktadır. Algortmanın temel prensibi, modelin doğru tahmin edemediği veya tahminde zorlanılan örnekler üzerine odaklanarak performansı iyileştirmektir (Friedman, 2002).

Ekstremler gradyan arttırma (Extreme Gradient Boosting, XGBoost), GB algoritmasının optimize edilmiş ve performansı artırılmış bir versiyonudur. Büyük veri kümeleriyle etkili bir şekilde çalışabilen bu algortma, paralel hesaplama yetenekleri ve düşük bellek kullanımıyla öne çıkmaktadır. Ayrıca özelleştirilebilir kayıp fonksiyonları ve düzenleme teknikleri gibi gelişmiş özelliklere sahip olması, modelin performansını artırırken aşırı uyumu azaltmaktadır (Gündoğdu, 2023).

#### 2.5. Performans Metrikleri

Karmaşıklık matrisi (confusion matrix), bir sınıflandırma modelinin performansının değerlendirilmesinde sıklıkla kullanılan bir yöntemdir. Karmaşıklık matrisi, bir sınıflandırıcının bazı test verilerine göre sınıflandırma performansını tanımlayan  $N \times N$  boyutlu bir matristir. Bu matris gerçek ve tahmin edilen sınıflar arasındaki ilişkiyi göstermektedir. Bu matris modelin tahminlerinin gerçek sınıflarla karşılaştırılmasını vermekte olup, genellikle dört ana kategoriye içermektedir: Doğru pozitif (DP), Doğru negatif (DN), Yanlış pozitif (YP), Yanlış negatif (YN). Bu kategoriler; modelin doğruluğunu, hassasiyetini, geri çağırma, F1 puanını ve diğer performans metriklerini hesaplamak için kullanılmaktadır. Çalışmada gebelik sürecinde anne sağlığı risk seviyesinin sınıflandırılması amacıyla önerilen MÖ temelli modelin performansı aşağıda verilen denklemlere göre hesaplanan doğruluk, kesinlik, geri çağırma ve F1 skor metriklerine göre değerlendirilmiştir (Eşitlik 1-4).

$$\text{Doğruluk (Accuracy, ACC)} = \frac{DP + DN}{DP + DN + YP + YN} \quad (1)$$

$$\text{Kesinlik (Precision, Prec)} = \frac{DP}{DP + YP} \quad (2)$$

$$\text{Geri çağırma (Recall)} = \frac{DP}{DP + YN} \quad (3)$$

$$F1 - \text{Skor} = 2 * \frac{\text{Recall} * \text{Prec}}{\text{Recall} + \text{Prec}} \quad (4)$$

### 3. Bulgular ve Tartışma

Yapılan bu çalışmada, MÖ algoritmalarının gebelik sürecinde anne sağlığı risk durumlarının (özellikle yüksek riskli hamilelik geçiren annelerin) tespitindeki başarımları ele alınmıştır. Bunun için 406 "düşük risk", 336 "orta risk" ve 272 "yüksek risk" olmak üzere toplam 1014 örnek içeren bir veri seti kullanılarak altı farklı MÖ algoritmasının başarımları araştırılmıştır. Bu doğrultuda çalışmada ilk olarak, algoritmaların sınıflandırma başarımlarını artırmak amacıyla gebelik sürecinde etkili olan majör risk faktörleri tespit edilmiştir. Bu amaç doğrultusunda Ki-Kare testi SelectKBest yöntemiyle uygulanarak, veri setindeki gereksiz bilgi içeren (etkisiz, anlamsız) özellikler çıkartılmıştır. Analizler sonucunda gebelik sürecinde anne sağlığı risk durumlarının tespitinde yaş, sistolik kan basıncı ve diyastolik kan basıncının daha belirleyici (etkili) olduğu tespit edilmiştir. Daha sonra bu üç özellik kullanılarak altı farklı MÖ algoritmasının sınıflandırma başarımları hold-out performans değerlendirme yöntemiyle analiz edilmiştir. Bununla birlikte çalışma kapsamında MÖ algoritmalarının başarımları, hold-out yönteminin veri setindeki dengesizlikleri elimine etmek amacıyla SMOTE tabanlı örnekleme yöntemiyle birlikte uygulanması durumu için de değerlendirilmiş ve sonuçlar karşılaştırmalı olarak verilmiştir. Hold-out yönteminde

veri seti %80'i eğitim %20'si test verisi olacak şekilde bölünmüştür. Çalışmada kullanılan MÖ algoritmaları, literatürde gebelik sürecinde anne sağlığı risk durumlarının tespitine yönelik yapılan çalışmalar incelenerek seçilmiştir. Bu doğrultuda, K-en yakın komşu (KNN), karar ağacı (KA), rasgele orman (RO) , gradyan arttırma (GB), ekstrem gradyan arttırma (XGBoost) ve destek vektör makineleri (DVM) algoritmaları kullanılmıştır. Phyton programında yapılan analizlerin sonuçları karmaşıklık matrisiyle incelenerek, modellerin performanslarını karşılaştırmak adına karmaşıklık matrisi ile hesaplanan doğruluk, duyarlılık, özgüllük ve F-skoru değerleri birlikte sunulmuştur. Bu bağlamda çalışmada ilk olarak KNN algoritmasına ait performans metrikleri elde edilerek, Tablo 3' de verilmiştir.

Tablo 3 incelendiğinde; SMOTE uygulanmayan veri setinde en yüksek doğruluk oranı %92 ile "yüksek risk" seviyesinde, en düşük doğruluk oranı ise %70 ile "orta risk" seviyesinde elde edilmiştir. Veri setine SMOTE uygulanması durumunda da benzer şekilde en yüksek (%93) ve en düşük (%72) doğruluk oranları sırasıyla "yüksek risk" ve "orta risk" seviyelerinde elde edilmiştir. Elde edilen sonuçlar içerisinden en yüksek sınıflandırma oranının elde edildiği yani hold-out yönteminin SMOTE ile birlikte uygulanması durumu için KNN algoritmasına ait karmaşıklık matrisi Şekil 4'de verilmiştir.

**Tablo 3.** KNN algoritmasına ait performans metrikleri

	Risk Sınıfı	Risk Sınıfı			
		(%)	Düşük	Orta	Yüksek
Hold-out	ACC		73	70	92
	Prec		89	38	79
	Recall		61	69	84
	F1-skor		72	49	81
Hold-out & SMOTE	ACC		80	78	93
	Prec		77	72	78
	Recall		70	66	98
	F1-skor		73	69	87

Gerçek Sınıf	Tahmini Sınıf		
	Düşük risk	Orta risk	Yüksek risk
Düşük risk	67	20	
Orta risk	22	58	1
Yüksek risk	7	10	59

**Şekil 4.** KNN algoritmasına ait karmaşıklık matrisi.

Şekil 4'de verilen KNN algoritmasına ait karışıklık matrisi incelendiğinde, test verilerinin sınıflandırılmasında %75.41'lik bir doğruluk oranı elde edilmiştir. Yani KNN algoritmasıyla 244 adet test verisinin 184'ü doğru 60 tanesi yanlış sınıflandırılmış olup, en yüksek sınıflandırma başarımı "yüksek risk" sınıfı için elde edilmiştir. Tablo 4'de KA algoritmasına ait performans metrikleri verilmiştir.

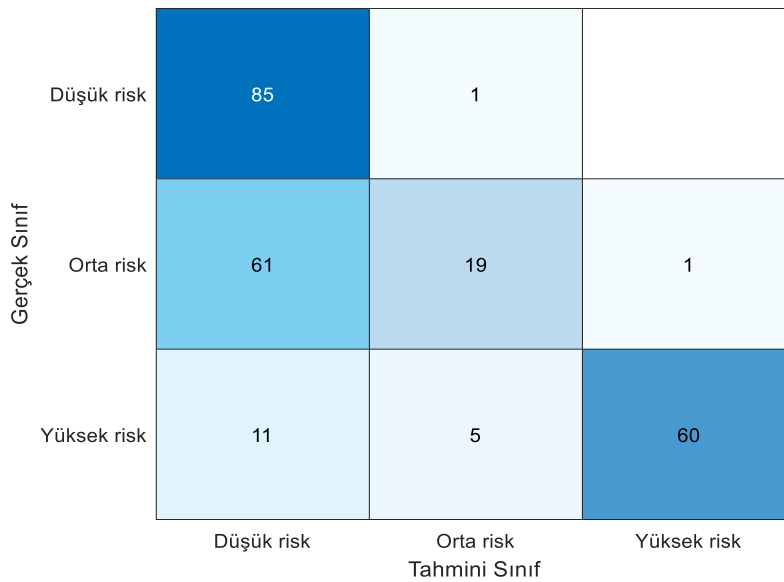
Tablo 4'de sunulan KA algoritmasına ait performans metriklerine göre; SMOTE uygulanmayan veri seti için en yüksek sınıflandırma başarım oranı %93 ile "yüksek risk" seviyesinde, en düşük sınıflandırma başarım oranı ise %67 ile "orta risk" seviyesinde tespit edilmiştir. Veri setine SMOTE uygulandığında ise; en yüksek sınıflandırma başarımına %93'lük bir oranla "yüksek risk" seviyesinde, en düşük sınıflandırma başarımına da %70'lik bir oranla "düşük risk" seviyesinde ulaşılmıştır. Her iki durum için de en yüksek sınıflandırma başarımlarının eşit olduğu ve bunların "yüksek risk" seviyesinde elde edildiği tespit edilmiştir. Buna dayanarak Şekil 5'de her iki durum için ayrı ayrı KA algoritmasına ait karmaşıklık matrisleri verilmiştir.

Şekil 5'den de açıkça görüldüğü üzere, veri setine SMOTE uygulanmadığında 244 tane test verisinin 173'ü doğru 71'i yanlış sınıflandırılırken; SMOTE uygulanmasını durumda ise 165 tanesi doğru ve 79 tanesi yanlış sınıflandırılmıştır. Veri setine SMOTE uygulanmaması durumunda %71'lik, SMOTE uygulanması durumunda ise %68'lik bir başarıma ulaşılmıştır. En yüksek sınıflandırma başarımları her iki durum içinde "yüksek risk" sınıfında edilmiştir. Tablo 5'de DVM algoritmasına ait performans metrikleri verilmiştir.

DVM algoritmasına ait sınıflandırma performans metrikleri incelendiğinde, veri setine SMOTE uygulanmaması halinde en yüksek doğruluk oranı %92 ile "yüksek risk" seviyesinde, en düşük doğruluk oranı ise %67 ile "orta risk" seviyesinde elde edilmiştir. Veri setine SMOTE uygulanması durumunda ise en yüksek doğruluk değeri %93 ile "yüksek risk" seviyesinde iken, en düşük doğruluk oranı %73 ile "düşük risk" seviyesindedir. En yüksek sınıflandırma oranının elde edildiği durum (yani hold-out yönteminin SMOTE ile birlikte uygulanması durumu) için algoritmaya ait karışıklık matrisi Şekil 6'da verilmiştir.

**Tablo 4.** KA algoritmasına ait performans metrikleri

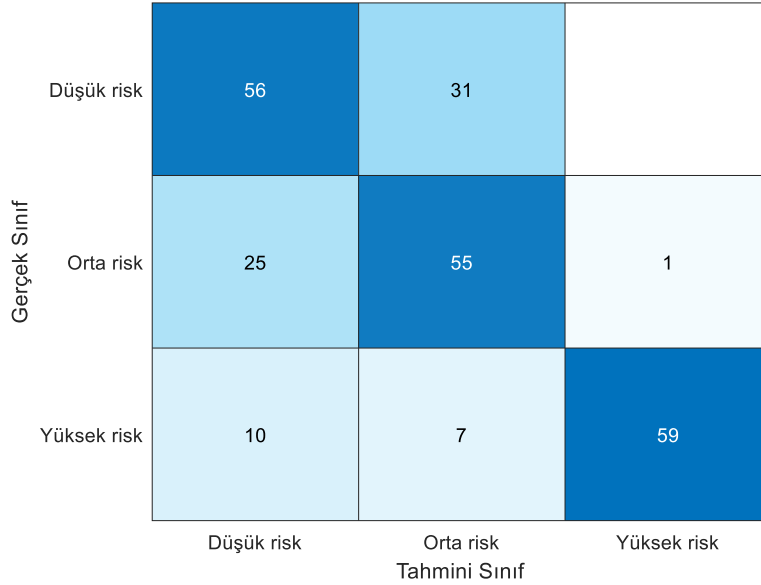
	Risk Sınıfı	Risk Sınıfı			
		(%)	Düşük	Orta	Yüksek
Hold-out	ACC		70	67	93
	Prec		50	70	83
	Recall		67	55	85
	F1-skor		67	55	85
Hold-out & SMOTE	ACC		70	72	93
	Prec		99,4	23	79
	Recall		54	76	98
	F1-skor		70	36	88



**Şekil 5.** KA ait karmaşıklık matrisi (a) Hold-Out, (b) Hold-out+SMOTE.

**Tablo 5.** DVM algoritmasına ait performans metrikleri

	Risk Sınıfı			
		(%)	Düşük	Orta
Hold-out	ACC	68	67	92
	Prec	95	21	79
	Recall	55	70	86
	F1-skor	70	32	82
Hold-out & SMOTE	ACC	73	74	93
	Prec	64	68	78
	Recall	62	59	98
	F1-skor	63	63	87



**Şekil 6.** DVM algoritmasına ait karmaşıklık matrisi.

Şekil 6’de gösterilen DVM algoritmasına ait karmaşıklık matrisi incelendiğinde test verilerinin sınıflandırılmasında %70’lik doğruluk oranı elde edilmiştir. Yani DVM algoritmasıyla 244 tane test verininin 170’i doğru, 74’ü yanlış tahmin edilmiş olup, en yüksek sınıflandırma başarımı “yüksek risk” sınıfı için elde edilmiştir. Tablo 6’da RO algoritmasına ait performans metrikleri sunulmuştur.

RO algoritmasına ait sınıflandırma performans metrikleri incelendiğinde, veri setine SMOTE uygulanması halinde en yüksek doğruluk oranı %93 ile “yüksek risk” seviyesinde, en düşük doğruluk oranı ise %83 ile “orta risk” seviyesinde elde edilmiştir. Veri setine SMOTE uygulanması durumunda ise en yüksek doğruluk değeri %95 ile “yüksek risk” seviyesinde iken, en düşük doğruluk oranı %87 ile “düşük risk” ve “orta risk” seviyelerindedir. Analizlerde en yüksek sınıflandırma oranının elde edildiği durum (yani hold-out+SMOTE durumu) için RO algoritmasına ait karmaşıklık matrisi Şekil 7’de verilmiştir.

Şekil 7’de verilen karmaşıklık matrisi incelendiğinde; RO algoritmasıyla test verilerinin sınıflandırılmasında %85’lik bir sınıflandırma başarımı elde edilmiştir. Yani test için ayrılan 244 verinin 207’si doğru tahmin

edilirken, 37’si yanlış tahmin edilmiştir. En yüksek sınıflandırma başarımına “yüksek risk” sınıfında ulaşılmıştır. Tablo 7’de GB algoritmasına ait performans metrikleri sunulmuştur.

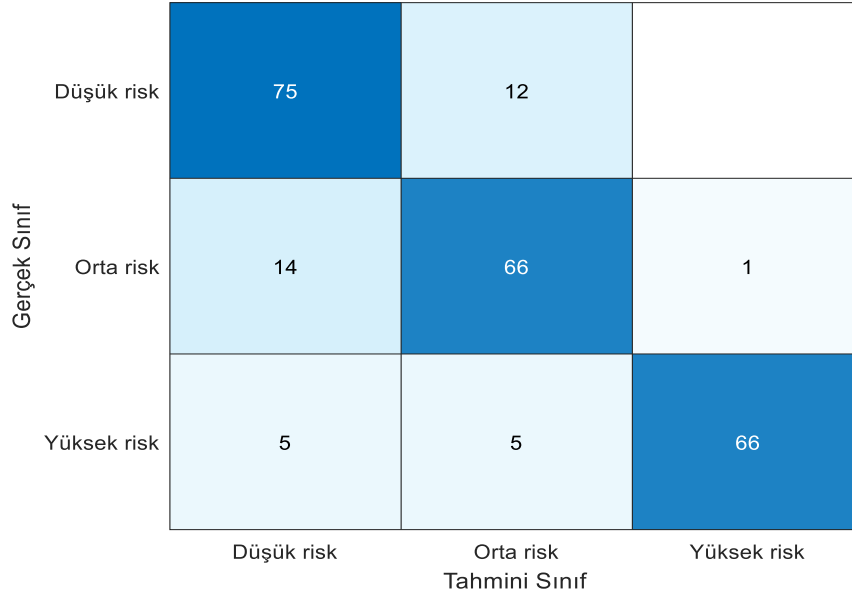
Tablo 7’de sunulan GB algoritmasına ait sınıflandırma performans metrikleri incelendiğinde, veri setine SMOTE uygulanmaması halinde en yüksek doğruluk oranı %93 ile “yüksek risk” seviyesinde, en düşük doğruluk oranı ise %78 ile “orta risk” seviyesinde olduğu görülmektedir. Veri setine SMOTE uygulanması durumunda en yüksek doğruluk değeri %95 ile “yüksek risk” seviyesinde, en düşük doğruluk değeri ise %82 ile “düşük risk” ve “orta risk” seviyelerindedir. En yüksek sınıflandırma oranının elde edildiği durum (yani hold-out+SMOTE durumu) için algoritmaya ait karmaşıklık matrisi Şekil 8’de verilmiştir.

Şekil 8’deki GB algoritmasına ait karmaşıklık matrisi incelendiğinde, test verilerinin sınıflandırılmasında %79’luk bir doğruluk oranı elde edilmiştir. Yani GB algoritmasıyla 244 tane test verisinin 193’ü doğru, 51’i yanlış sınıflandırılmış olup, en yüksek sınıflandırma başarımı “yüksek risk” sınıfı için elde edilmiştir.



**Tablo 6.** RO algoritmasına ait performans metrikleri

	Risk Sınıfı	Risk Sınıfı		
		(%)	Düşük	Orta
Hold-out	ACC	85	83	93
	Prec	85	76	79
	Recall	78	78	88
	F1-skor	81	77	83
Hold-out & SMOTE	ACC	87	87	95
	Prec	86	81	87
	Recall	80	80	99
	F1-skor	83	80	92



**Şekil 7.** RO algoritmasına ait karmaşıklık matrisi.

**Tablo 7.** GB algoritmasına ait performans metrikleri

	Risk Sınıfı	Risk Sınıfı		
		(%)	Düşük	Orta
Hold-out	ACC	80	78	93
	Prec	75	72	81
	Recall	74	70	88
	F1-skor	75	71	84
Hold-out & SMOTE	ACC	82	82	95
	Prec	75	77	87
	Recall	75	70	96
	F1-skor	75	73	91

**Tablo 8.** XGB algoritmasına ait performans metrikleri

	Risk Sınıfı	Risk Sınıfı		
		(%)	Düşük	Orta
Hold-out	ACC	88	85	95
	Prec	86	79	87
	Recall	83	81	89
	F1-skor	85	80	88
Hold-out & SMOTE	ACC	89	88	97
	Prec	85	85	91
	Recall	84	80	99
	F1-skor	84	83	95

Gerçek Sınıf	Düşük risk	65	22	
	Orta risk	16	62	3
	Yüksek risk	6	4	66
		Düşük risk	Orta risk	Yüksek risk
		Tahmini Sınıf		

Şekil 8. GB algoritmasına ait karmaşıklık matrisi.

Gerçek Sınıf	Düşük risk	74	13	
	Orta risk	11	69	1
	Yüksek risk	3	4	69
		Düşük risk	Orta risk	Yüksek risk
		Tahmini Sınıf		

Şekil 9. XGB algoritmasına ait karmaşıklık matrisi.

XGB algoritmasına ait performans metrikleri Tablo 8’de verilmiştir. Tablodan da açıkça görüldüğü üzere, veri setine SMOTE uygulanmaması halinde en yüksek doğruluk oranı %95 ile “yüksek risk” seviyelerinde, en düşük doğruluk oranı ise %85 ile “orta risk” seviyesinde elde edilmiştir. Veri setine SMOTE uygulanması durumunda ise en yüksek doğruluk değeri %97 ile “yüksek risk” seviyelerinde iken, en düşük doğruluk değeri %88 ile “orta risk” seviyesindedir. Analizler sonucunda en yüksek sınıflandırma oranının elde edildiği durum (yani hold-out+SMOTE) için algoritmaya ait karmaşıklık matrisi Şekil 9’da verilmiştir.

Şekil 9’deki XGB algoritmasına ait karmaşıklık matrisi incelendiğinde test verilerinin sınıflandırılmasında %87’lik doğruluk oranı elde edilmiştir. XGB algoritmasıyla test için ayrılan 244 verinin 212’si doğru, 32’si yanlış tahmin edilmiştir. En yüksek sınıflandırma

başarımı “yüksek risk” sınıfı için elde edilmiştir.

Elde edilen bulgulara dayanarak genel anlamda gebelik sürecinde anne sağlığı risk tahmininde SMOTE örnekleme tekniğinin kullanımı önerilen modelin başarımını önemli ölçüde arttırdığı belirlenmiştir. Bu bağlamda anne sağlığı risk veri setinden elde edilen üç farklı koşulun (düşük, orta ve yüksek) sınıflandırma sürecinde, kullanılan MÖ algoritmaları arasından en yüksek başarımları %87 ile XGB algoritmasıyla elde edilmiştir. Bu sınıflandırma başarımını sırasıyla %85 ile RO, %79 ile GB, %75 ile KNN, %70 ile DVM ve %68 ile KA sınıflandırıcısı takip etmiştir. Bununla birlikte önerilen modelde risk grupları arasında en yüksek sınıflandırma başarımı her bir algoritma için yüksek risk sınıfında elde edilmiştir. Bu bağlamda, elde edilen yüksek risk grubuna ait sınıflandırma başarımları kNN, KA ve DVM algoritmaları için %93 iken, RO ve GB algoritmaları içinse %95 olarak

hesaplanmıştır. Bu gruba ait en yüksek başarı oranı ise %97 ile XGB algoritmasıyla elde edilmiştir.

## 4. Sonuç

Gebelik, anne ve bebeğin sağlığı üzerinde önemli etkilere sahip olan karmaşık bir süreçtir. Anne adayının gebelik sürecinde karşılaşılabileceği çeşitli sağlık riskleri, hem anne hem de bebeğin sağlığı açısından ciddi sonuçlar doğurabilmektedir. Bu nedenle, gebelik sürecindeki risklerin erken aşamada tanınması ve etkili bir şekilde yönetilmesi büyük önem taşımaktadır. Ancak, gebelik sırasında ortaya çıkabilecek potansiyel sağlık risklerinin erken aşamada teşhis edilmesi ve etkin bir şekilde yönetilmesi, birçok sağlık sistemi için hala bir zorluktur. Bu bağlamda son yıllarda, gebelik sürecinde uzman hekimlerin karar alma sürecine yardımcı olmak için bilgisayar tabanlı karar destek sistemlerinin geliştirilmesine verilen önem giderek artmaktadır. Gebelik sırasında ortaya çıkabilecek çeşitli risk faktörlerini değerlendirmek ve sınıflandırmak, uzman hekimlerin erken müdahalelerde bulunmasını sağlayarak olası komplikasyonları en aza indirebilir.

Bu çalışmada, gebelik sürecinde anne sağlığı risk seviyelerinin sınıflandırılmasında farklı MÖ algoritmalarının başarımları analiz edilmiştir. Bunun için K-NN, G-NB, KA, RO, LR, XGBoost ve GB algoritmaları kullanılmıştır. Algoritmalarının başarımları hold-out performans değerlendirme kriterine göre elde edilmiştir. Yanı sıra MÖ algoritmalarının başarımları aynı performans değerlendirme kriterine göre veri setindeki dengesizlikler SMOTE örnekleme tekniğiyle giderilmesi durumu için de incelenmiştir. Bununla birlikte çalışmada Chi-kare testi SelectBest özellik seçim yöntemlerine dayanarak gebelik sürecinde anne sağlığı üzerinde etkili (anlamlı) özelliklerin yaş, sistolik kan basıncı ve diyastolik kan basıncı olduğu tespit edilmiştir. Bu özelliklere dayalı önerilen modelin başarımları, tüm özelliklerin kullanıldığı duruma karşılaştırıldığında daha yüksek bulunmuştur. Elde edilen bulgular ışığında hem özellik seçiminin hem de SMOTE tekniğinin gebelik sürecinde anne sağlığı risk seviyesinin (özellikle yüksek risk grubunun) tahmininde sınıflandırıcıların performanslarını önemli ölçüde arttırdığı tespit edilmiştir. Literatürde aynı veri setini kullanılarak daha yüksek başarımların elde edildiği çalışmalar mevcuttur. Bunun temel nedeni eğitim ve test verilerinin bölünmesindeki ayrımların yanı sıra, sınıflandırıcı performans değerlendirme yöntemlerindeki farklılıklarından kaynaklanmaktadır.

Ancak, bu çalışmanın bazı sınırlamaları bulunmaktadır. Örneğin, veri setimizdeki eksiklikler ve kısıtlamalar, makine öğrenmesi algoritmalarının geliştirilebilirliğini etkileyebilmektedir. Bu bağlamda gelecekte daha geniş ve çeşitli veri setleri kullanılarak, modelin performansını arttırmaya yönelik çalışmaların yapılması planlanmaktadır. Bununla birlikte, daha karmaşık MÖ tekniklerinin yanı sıra, modelin klinik uygulamalarda kullanılabilirliğini artırmak için

anlaşılabilirlik açısından daha şeffaf yaklaşımlar araştırılacaktır.

## Katkı Oranı Beyanı

Yazar(lar)ın katkı yüzdesi aşağıda verilmiştir. Tüm yazarlar makaleyi incelemiş ve onaylamıştır.

	İ.Ş.Y.	R.U.A.
K	50	50
T	50	50
Y	30	70
VTI	60	40
VAY	60	40
KT	60	40
YZ	50	50
KI	50	50
GR	100	
PY	40	60

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi.

## Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

## Etik Onay Beyanı

Bu çalışmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

## Kaynaklar

- Abacı İ, Yıldız K. 2023. SMOTE vs. KNNOR: An evaluation of oversampling techniques in machine learning. *Gümüşhane Üniv Fen Bil Derg*, 13(3): 767-779.
- Ahmed M, Kashem MA, Rahman M, Khatun S. 2020. Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT). 5th International Conference on Electrical, Control & Computer Engineering, July 29, Kuantan, Pahang, Malaysia, pp: 357-365.
- Ahmed M, Kashem MA. 2020. IoT based risk level prediction model for maternal health care in the context of Bangladesh. 2nd International Conference on Sustainable Technologies for Industry 4.0, December 19-20, Dhaka, Bangladesh, pp: 1-6.
- Ahmed M. 2023. Maternal health risk. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5DP5D> (erişim tarihi: 2 Şubat 2023).
- Al-Hindi MY, Al Sayari TA, Al Solami R, Baiti AKA, Alnemri JA, Mirza IM, Faden YA. 2020. Association of antenatal risk score with maternal and neonatal mortality and morbidity. *Cureus*, 12(12): e12230.
- Diamantoulaki I, Diamantoulakis PD, Bouzinis PS, Sarigiannidis P, Karagiannidis GK. 2022. Health risk assessment with federated learning. International Balkan Conference on Communications and Networking, Aug. 22-24, Sarajevo · Bosnia and Herzegovina, pp: 57-61.
- Edayath P. 2022. Analysis of factors affecting maternal health using data mining techniques. Doctoral dissertation, The University of Texas at El Paso, US.
- Elen A, Baş S, Közkurt C. 2022. An adaptive gaussian kernel for support vector machine. *Arabian J Sci Eng*, 47(8): 10579-

- 10588.
- Finlayson K, Crossland N, Bonet M, Downe S. 2020. What matters to women in the postnatal period: A meta-synthesis of qualitative studies. *PloS One*, 15(4): e0231415.
- Friedman JH. 2002. Stochastic gradient boosting. *Comput Stat Data Anal*, 38(4): 367-378.
- Gündoğdu S. 2023. Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique. *Multim Tools Appl*, 82(22): 34163-34181.
- Hacıbeyoğlu M, Çelik M, Çiçek ÖE. 2023. K en yakın komşu algoritması ile binalarda enerji verimliliği tahmini. *Necmettin Erbakan Üniv Fen Müh Bil Derg*, 5(2): 28-37.
- Jahan S, Islam MS, Islam L, Rashme TY, Prova AA, Paul BK, Mosharof MK. 2021. Automated invasive cervical cancer disease detection at early stage through suitable machine learning model. *SN Appl Sci*, 3: 1-17.
- Kırlı OA, Sansarcı M, Özkaraca O, Çetin G. 2023. Manyetik rezonans görüntülerinden beyin tümörü tespitinde sınıflandırma algoritmalarının karşılaştırmalı analizi. *Türk Müh Araş Eğit Derg*, 2(2): 113-122.
- Macrohon JJE, Villavicencio CN, Inbaraj XA, Jeng JH. 2022. A semi-supervised machine learning approach in predicting high-risk pregnancies in the Philippines. *Diagnostics*, 12(11): 2782.
- Mutlu HB, Durmaz F, Yücel N, Cengil E, Yildirim M. 2023. Prediction of maternal health risk with traditional machine learning methods. *Naturengs*, 4(1): 16-23.
- Pawar L, Malhotra J, Sharma A, Arora D, Vaidya D. 2022. A robust machine learning predictive model for maternal health risk. 3rd International Conference on Electronics and Sustainable Communication Systems, August 17-19, Coimbatore, India, pp: 882-888.
- Rai SK, Sowmya K. 2018. A review on use of machine learning techniques in diagnostic health-care. *AI Syst Mach Learn*, 10(4): 102-107.
- Ramdhani Y, Maulidia D, Setiadi A, Alamsyah DP. 2022. Feature weighting optimization: Genetic algorithms and random forest for classification of pregnant potential risk. *International Conference on Information Technology Research and Innovation*, November 10, Virtual, pp: 95-100.
- Selvakuberan K, Indradevi M, Rajaram R. 2008. Combined feature Selection and classification-A novel approach for the categorization of web pages. *J Info Comput Sc*, 3(2): 083-089.
- Şahin F, Tulum G, Karaca Ş. 2023. Anne sağlığı riski için makine öğrenmesi modellerinin performans karşılaştırması. *Dicle Üniv Müh Fak Müh Derg*, 14(4): 547-553.
- Togunwa TO, Babatunde AO. 2023. Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest. *Front AI*, 6: 1213436.
- Tokmak M. 2023. Anne sağlık durumu riskinin makine öğrenmesi yöntemleri ile belirlenmesi1. In: Bozdemir M, uzun Arslan R, editörler. *Müh Alanında Akademik Analiz ve Tartışmalar*, Editörler, Özgür Yayınları, Gaziantep, Türkiye, pp: 75.
- Umoren I, Chigozirim F, Silas A, Ekong B. 2022. Modeling and prediction of pregnancy risk for efficient birth outcomes using decision tree classification and regression model. URL: [https://www.researchgate.net/profile/Imeh-Umoren/publication/359742528\\_Modeling\\_and\\_Prediction\\_of\\_Pregnancy\\_Risk\\_for\\_Efficient\\_Birth\\_Outcomes\\_Using\\_Decision\\_Tree\\_Classification\\_and\\_Regression\\_model/links/624c5b31ef01342066596129/Modeling-and-Prediction-of-Pregnancy-Risk-for-Efficient-Birth-Outcomes-Using-Decision-Tree-Classification-and-Regression-model.pdf](https://www.researchgate.net/profile/Imeh-Umoren/publication/359742528_Modeling_and_Prediction_of_Pregnancy_Risk_for_Efficient_Birth_Outcomes_Using_Decision_Tree_Classification_and_Regression_model/links/624c5b31ef01342066596129/Modeling-and-Prediction-of-Pregnancy-Risk-for-Efficient-Birth-Outcomes-Using-Decision-Tree-Classification-and-Regression-model.pdf) (erişim tarihi: 15 Ocak 2024)
- Yakut Ö, Bolat E. 2020. Arrhythmia diagnosis from ECG signal using tree-based machine learning methods. *Int J Math Eng Nat Sci*, 4(16): 954-964.
- Yıldız İ, Kotan AE, Altınel AB. 2023. Türkçe faturaların sınıflandırılmasında farklı öznelik seçimi yöntemleri ile topluluk öğrenme algoritmalarının etkilerinin incelenmesi. *Avrupa Bil Teknol Derg*, (52): 272-278.