

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PERİNATAL DÖNEMDEKİ PSİKIYATRİK HASTALIK
BELİRTİLERİNİN YAPAY ZEKA TABANLI BÜYÜK VERİ
İŞLEME PLATFORMU İLE BELİRLENMESİ

DOKTORA TEZİ

Nur Banu OĞUR

Bilgisayar Mühendisliği Anabilim Dalı

HAZİRAN 2023

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PERİNATAL DÖNEMDEKİ PSİKIYATRİK HASTALIK
BELİRTİLERİNİN YAPAY ZEKA TABANLI BÜYÜK VERİ
İŞLEME PLATFORMU İLE BELİRLENMESİ

DOKTORA TEZİ

Nur Banu OĞUR

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Celal ÇEKEN

HAZİRAN 2023

Nur Banu OĞUR tarafından hazırlanan “Perinatal Dönemdeki Psikiyatrik Hastalık Belirtilerinin Yapay Zeka Tabanlı Büyük Veri İşleme Platformu ile Belirlenmesi” adlı tez çalışması 26.06.2023 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Prof. Dr. Numan Çelebi Sakarya Üniversitesi
Jüri Üyesi :	Prof. Dr. Celal Çeken Sakarya Üniversitesi
Jüri Üyesi :	Dr. Öğr. Üyesi İsmail Öztel Sakarya Üniversitesi
Jüri Üyesi :	Doç. Dr. Ali Çalhan Düzce Üniversitesi
Jüri Üyesi :	Dr. Öğr. Üyesi Faruk Aktaş Kocaeli Üniversitesi

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “Perinatal Dönemdeki Psikiyatrik Hastalık Belirtilerinin Yapay Zeka Tabanlı Büyük Veri İşleme Platformu ile Belirlenmesi” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, etik kurul onay belgesi aldığımı çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(...../...../20.....).

(imza)

Nur Banu OĞUR

Hayatta olsun olmasın, üzerimde emeđi olan tüm sevdiklerime...

TEŞEKKÜR

Yüksek lisans ve doktora eğitimim boyunca değerli bilgi ve deneyimlerinden yararlandığım, araştırmanın planlanmasından yazılmasına kadar tüm aşamalarında yardımlarını esirgemeyen, taviz vermeksizin doğruyu ilke edinen, bilim insanı olarak örnek aldığım değerli danışman hocam Prof. Dr. Celal Çeken'e teşekkürlerimi sunarım.

Tez çalışmamın şekillenmesinde büyük katkıları olan, titizlikle beni yönlendiren kıymetli hocalarım Prof. Dr. Numan Çelebi ve Dr. Öğr. Üyesi İsmail Öztel'e teşekkür ederim. Ayrıca tez çalışmam için gerekli olan tıbbi altyapının oluşması için desteklerini esirgemeyen Prof. Dr. Esra Yazıcı'nın da katkıları büyüktür, kendisine teşekkür ederim.

Akademik hayatımın bana kazandırdığı en özel iki dost; Deniz Balta ve Burcu Çarklı Yavuz. Destekleri ve yardımları çok değerli, teşekkür ederim.

Üzerimdeki haklarını ödeyemeyeceğim, hayattaki en büyük şanslarım olan canım babam Prof. Dr. Ahmet Oğur'a ve canım annem Nesrin Oğur'a, desteğini hiçbir zaman esirgemeyen, varlığını her zaman arkamda hissettiğim canım ağabeyim Dr. Yavuz Selim Oğur'a, neşesinde neşe bulduğum, göz aydınlığım canım kardeşim Nehir Nursu Oğur'a, her daim moral ve motivasyonumu arttıran kuzenim Emircan Cömert'e ve kardeşim Dr. Esin Topkara Oğur'a bu uzun ve zorlu yolculuğumda her zaman yanımda oldukları için minnettarım, sabırları ve fedakârlıkları için çok teşekkür ederim.

Nur Banu OĞUR

İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
TABLO LİSTESİ	xv
ŞEKİL LİSTESİ	xvii
ÖZET	xix
SUMMARY	xxi
1. GİRİŞ ve AMAÇ	1
2. GENEL BİLGİLER	5
2.1. Perinatal Dönemde Görülen Psikiyatrik Tanılar	5
2.1.1. Depresyon	6
2.1.1.1. Depresyonun sıklığı	6
2.1.1.2. Risk etmenleri	7
2.1.1.3. Hastalığın teşhis aşaması	7
2.1.2. Anksiyete (kaygı) bozuklukları	8
2.1.2.1. Anksiyetenin sıklığı	9
2.1.2.2. Risk etmenleri	10
2.1.2.3. Hastalığın teşhis aşaması	11
2.2. Tıp Alanında Büyük Veri Analitiğinin Yeri	11
2.2.1. Perinatal dönemde büyük veri analitiği	12
2.2.1.1. Perinatal dönemde büyük veri mimarisi kullanarak yapılan çalışmalar	14
2.3. Büyük Veri Analitiği	17
2.3.1. Ön işlem	17
2.3.1.1. Kullanılan yazılım ve teknolojiler	18
2.3.2. Özellik seçme algoritmaları	18
2.3.3. Büyük veri teknolojileri	19
2.3.3.1. Apache kafka	20
2.3.3.2. Apache spark	22
2.3.3.3. Docker	25
2.3.3.4. Kubernetes	26
2.3.3.5. Kubernetes apache spark	27
3. GELİŞTİRİLEN YAPAY ZEKA TABANLI BÜYÜK VERİ İŞLEME SİSTEMİ MİMARİSİ VE KULLANILAN YÖNTEMLER	29
3.1. Problemin Tanımı ve Çözüm Önerisi	29
3.2. Yöntem ve Bulgular	29
3.2.1. Veri tanıtımı	31
3.2.1.1. Çalışmada kullanılan ölçekler	32
3.2.2. Veri ön işlem aşaması	33
3.2.2.1. Risk faktörü belirleme	34

3.2.3. Makine öğrenmesi	44
3.2.3.1. Çapraz doğrulama (cross validation).....	44
3.2.3.2. Modelleme.....	44
3.2.3.3. Değerlendirme metrikleri	45
3.2.4. Veri analitiği.....	49
3.2.5. Kümeleme (cluster) teknolojisi	52
3.2.5.1. Ölçeklenebilir mimari	52
3.2.5.2. Kubernetes kümesi	53
3.2.5.3. Kafka producer	62
3.2.5.4. Kafka tüketici-apache spark	64
3.2.5.5. Kafka consumer.....	68
3.2.5.6. Performans karşılaştırması	72
4. SONUÇ VE TARTIŞMA.....	79
KAYNAKLAR.....	83
EKLER.....	91
ÖZGEÇMİŞ.....	107

KISALTMALAR

API	:Application Programming Interface
AUC	:Area Under The Curve
BB	:Bipolar Hastalığı
CLI	:Command Line Interface
DASS	:Depresyon, Anksiyete ve Stres Ölçeği
EDDÖ	: Edinburgh Doğum Sonrası Depresyon Ölçeği
EEG	:Elektroensefalografi
GA	:Genetic Algorithms
GAD-7	:Yaygın Anksiyete Bozukluğu 7 Maddelik Ölçeği
GBT	:Gradient Boosted Tree
HDFS	:Apache Hadoop Distributed File System
IDE	: Integrated Development Environment
KNN	:K-En Yakın Komşular
LDA	:Linear Discriminant Analysis
MDB	:Majör Depresif Hastalığı
NPM	:Node Package Manager
PASS	: Perinatal Anxiety Screening Scale
PASS-TR	:PASS Ölçeğinin Türkçe versiyonu
PCA	:Principal Component Analysis
SDVF	:Sosyo-demografik Veri Formu
SPSS	:Statistical Package for the Social Sciences
SSRI	:Seçici Serotonin Geri Alım İnhibitörleri
STAI	:Spielberg Durumluk-Sürekli Kaygı Envanteri
SVM	:Support Vector Machines

TABLO LİSTESİ

	<u>Sayfa</u>
Tablo 3.1. Sınıflandırma Etiketi	32
Tablo 3.2. Verisetindeki Karşılıkları	43
Tablo 3.3. Performans Karşılaştırmaları	46
Tablo 3.4. Naive Bayes Algoritmasının Değerlendirme Parametreleri.....	50
Tablo 3.5. Sunucu Özellikleri.....	58
Tablo 3.6. Modelin Değerlendirme Metrikleri	66

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Kafka Mimarisi.....	21
Şekil 2.2. Apache Spark Mimarisi (“Spark”, 2023)	23
Şekil 2.3. “Spark Streaming” Genel Akış Süreci (Ogur, Ceken, Ogur, Yuvaci, Yazici ve Yazici, 2023)	23
Şekil 2.4. Konteyner ve Sanal Makine Karşılaştırması (“Konteyner ve Sanal Makine”, 2023)	25
Şekil 2.5. Docker ile Postgresql Kurulumu	25
Şekil 2.6. Docker Konteyner Mimarisi (“Konteyner ve Sanal Makine”, 2023).....	26
Şekil 2.7. Kubernetes Kümesinde Apache Spark Akış Şeması	27
Şekil 3.1. Ana işlem Platformu Diyagramı.....	30
Şekil 3.2. Bilgi Keşfi Süreci (“Keşif Süreci”, 2023)	33
Şekil 3.3. EDDÖ ve PASS Ölçeklerinin Normallik Testleri	35
Şekil 3.4. PASS Verilerinin Korelasyon Performansı	35
Şekil 3.5. EDDÖ Verilerinin Korelasyon Performansı.....	36
Şekil 3.6. Korelasyon Performans Karşılaştırması	36
Şekil 3.7. EDDÖ-SDVF Veri Setinde PCA Kullanımı.....	38
Şekil 3.8. EDDÖ -SDVF Değişken Önemliliği Ağırlık Dağılım Grafiği.....	38
Şekil 3.9. PASS-TR-SDVF Değişken Önemliliği Ağırlık Dağılım Grafiği	39
Şekil 3.10. EDDÖ-SDVF Özellik Seçme Sonuçları.....	39
Şekil 3.11. Önerilen Veri Analitiği Mimarisinin Diyagramı	40
Şekil 3.12. Kullanılan Seçim Metodları.....	41
Şekil 3.13. Sonuç Üzerinde En Etkili Özellikler	42
Şekil 3.14. Kullanılan Makine Öğrenmesi Algoritmaları.....	45
Şekil 3.15. Doğruluk Karşılaştırmaları	46
Şekil 3.16. Hassasiyet Karşılaştırmaları	47
Şekil 3.17. Özgünlük Karşılaştırmaları.....	47
Şekil 3.18. Kesinlik Karşılaştırmaları.....	48
Şekil 3.19. AUC Değeri Karşılaştırmaları	48
Şekil 3.20. Zamanlama Şeması (Ogur ve diğerleri, 2023).....	49
Şekil 3.21. Akış Halindeki Verinin Anlık Sonuçları	50
Şekil 3.22. Her Saniye Gelen Akış Halindeki Verinin İstatistikleri	51
Şekil 3.23. Her Saniye Gelen Akış Halindeki Verinin Çalışma Süresi	52
Şekil 3.24. Kubernetes Mimarisi (“Kubernetes Mimarisi”, 2023)	55
Şekil 3.25. Geliştirilen Büyük Veri İşleme Sistemi İçin Kubernetes Ana İşlem Diyagramı.....	56
Şekil 3.26. Geliştirilen Büyük Veri İşleme Sisteminin Zamanlama Şeması	57
Şekil 3.27. Kurulan Sanal Makineler	58
Şekil 3.28. Sanal Makine IP’leri	58
Şekil 3.29. Master Node	59
Şekil 3.30. Diğer Sanal Makinalar (Node1-Node2).....	59
Şekil 3.31. Kubernetes Token.....	60

Şekil 3.32. Kurulan Node'ların LENS Arayüzünde Görüntüsü.....	60
Şekil 3.33. Kurulan Pod'ların LENS Arayüzünde Görüntüsü	61
Şekil 3.34. Helm ile Zookeeper Kurulumu	61
Şekil 3.35. Helm ile Kafka Kurulumu	62
Şekil 3.36. Helm ile Kafka Kurulumunun Terminal Gösterimi.....	62
Şekil 3.37. Kafka Producer Node.js Kodu	63
Şekil 3.38. Docker File	63
Şekil 3.39. Her Saniye Yollanan Gebe Verisi.....	64
Şekil 3.40. Apache Spark Kodu	65
Şekil 3.41. Model Eğitme Kod Bloğu.....	66
Şekil 3.42. Spark MLLIB Model Eğitimi	66
Şekil 3.43. Ana İşlem Diyagramı.....	68
Şekil 3.44. Kafka Consumer Kodu	69
Şekil 3.45. Docker File	70
Şekil 3.46. Kafka Consumer Ekran çıktısı.....	70
Şekil 3.47. Kubernetes Mimarisi.....	71
Şekil 3.48. Spark WebUI	73
Şekil 3.49. "Executor" Sayısının İşlem Hızına Etkisi.....	74
Şekil 3.50. Kubernetes Dashboard.....	75
Şekil 3.51. "Executor"ların Kaynak Tüketimi (Spark WebUI)	75
Şekil 3.52. CPU Çekirdek İhtiyacı Performans Karşılaştırması	76
Şekil 3.53. Bellek İhtiyacı Performans Karşılaştırması	76
Şekil 3.54. Bellek Kullanımı Performans Karşılaştırması	77

PERİNATAL DÖNEMDEKİ PSİKİYATRİK HASTALIK BELİRTİLERİNİN YAPAY ZEKA TABANLI BÜYÜK VERİ İŞLEME PLATFORMU İLE BELİRLENMESİ

ÖZET

Makine öğrenimi, bilim ve endüstrinin yönlendirici güçlerinden biridir; ancak büyük verinin oluşmasıyla birlikte, bu muazzam veri hacmini işleyebilmek için geleneksel makine öğrenimi algoritmalarında değişiklik yapma ihtiyacı ortaya çıkmaktadır. Bu veri setlerinin boyutu ve karmaşıklığı, geleneksel veri işleme yöntemleri yardımıyla ele alınmasını zor bir hale getirmiştir. İşte burada büyük veri analitiği teknolojileri devreye girmektedir. Büyük veri analitiği teknolojileri sayesinde, yüksek hacimli sağlık veri kümelerinin mevcudiyeti olması durumunda dahi, istenilen hızda ve performansta çeşitli tıbbi durumlar doğru bir şekilde teşhis edilebilmektedir.

Büyük veri analitiği teknolojileri, veri setlerinin içindeki değerli bilgileri ortaya çıkarmak için çeşitli analitik yöntemler kullanmaktadır. İstatistiksel analiz, veri madenciliği, makine öğrenimi ve yapay zeka gibi teknikler, bu büyük veri setlerinden anlamlı sonuçlar elde etmeyi sağlamaktadır. Bunların yanında, büyük veri setlerini paralel ve dağıtık bir şekilde işleyerek hızlı analizler yapılmasını da sağlamaktadır. Büyük veri analitiği teknolojilerinin sağladığı bu yetenekler, birçok alanda büyük bir etki yaratmıştır. Sağlık alanı da bu alanlardan biridir.

Çalışmanın birincil amacı, performansı optimize edilmiş algoritmalar aracılığıyla hamile kadınlarda anksiyete ve depresyonla ilgili en zorlayıcı soruları belirlemektir. Bu sayede daha az soru ile daha kısa sürede sonuca ulaşmak amaçlanmaktadır. Bu çalışmanın bir sonraki hedefi, büyük veri akışında makine öğrenimi modellerini kullanarak hamile kadınlarda depresyon ve anksiyete için anlık bir uzaktan sağlık durumu tahmin sistemi oluşturmaktır. Bu hedef, Apache Spark büyük veri işleme motorunu temel almaktadır. Bu ölçeklenebilir sistemde uygulama, hastanın sağlık durumunu tahmin etmek için gebe kadınlardan veri almaktadır. Bu veriler işlenebilir hale getirilebilmesi adına ön işlem süreçlerinden geçirilmektedir. Performans adına yapılan bir dizi işlemden sonra bu veri kümesi için en iyi sonuçları üreten algoritmanın, Naïve Bayes makine öğrenimi algoritması olduğu tespit edilmiştir.

Çalışmada ayrıca, anlık olarak sonuç üreten tahmin sistemi, Kubernetes kümeleme sistemine entegre edilmiştir. Böylece, oluşturulan ölçeklenebilir mimarinin işlem süresi açısından araştırılması sağlanmıştır. Geliştirilen büyük veri platformunun başarımı dikkate alındığında, hamile kadınlarda anksiyete ve depresyonu tespit etmek için kullanılan geleneksel yöntemler yerine, gerçek zamanlı olarak önemli doğruluk oranında ve hızda çalışan bilgisayar tabanlı sistemlerin kullanılabileceğini söylemek mümkündür.

DETECTION OF PSYCHIATRIC DISEASE SYMPTOMS IN THE PERINATAL PERIOD WITH AN AI-BASED BIG DATA PROCESSING PLATFORM

SUMMARY

One of the forces shaping driving science and industry is machine learning. This technology, which enables computers to learn from data instead of complex algorithms written by humans, has revolutionized many industries.

Machine learning is a tool used to discover patterns and relationships on large data sets. This enables businesses to better understand their data, uncover confidential information and predict future trends. Machine learning also plays a large role in automated decision-making processes. In many areas, tasks that are difficult or time-consuming for humans to do can be delegated to machines. For these reasons, machine learning is an increasingly important technology today. However, the prevalence of big data requires new uses for this large amount of data, which has different speeds than traditional approaches to applying machine learning techniques.

Big data is a valuable concept that plays an important role in machine learning. Today, the rapidly increasing amount of data across various industries enables businesses to make better decisions, make predictions and better understand customer needs. Technologies used in big data analytics are needed to process this large amount of data under the necessary conditions.

Big data analytics technologies play an indispensable role today to process and analyze large data sets and extract valuable information. Big data analytics tools enable businesses to deal with complex data masses, discover patterns and make better decisions. One of its key advantages is the data processing capacity of big data analytics tools. Large datasets are created quickly, often in high volume. Processing and analyzing this data may not be possible with traditional methods. Big data analytics tools enable rapid analysis by processing these large data sets in a parallel and distributed manner. In addition, it offers data visualization and reporting capabilities. Because large data sets are often complex and voluminous, it is important to make sense of and share data. Big data analytics technology provides better understanding of data by visualizing it with graphs, tables and interactive visuals. It also provides users with the ability to create shareable reports that summarize data with its reporting features. In summary, big data analytics technologies are important technologies that help businesses effectively manage, analyze and extract valuable information from large data sets. These technologies facilitate the processing and interpretation of big data with features such as data processing capacity, analysis capabilities, visualization and reporting features, forecasting and forecasting capabilities. There are many areas covered by big data, and one of the most critical areas is medicine, which requires rapid decision-making. Psychiatry, on the other hand, has an important place in the field of medicine as a branch focusing on mental health. Psychiatry, which examines mental, emotional and behavioral disorders,

evaluates people's mental health, makes a diagnosis and applies appropriate treatment methods.

Mental health diseases are health problems that deeply affect the lives of individuals and need to be treated carefully. If psychological disorders such as anxiety and depression occurring in the perinatal period, which includes the pregnancy process, are not detected on time, they will have negative effects on both the mother and the baby. The effects on the perinatal period and pregnant women are of great importance for the mental health of the society. For this reason, rapid diagnosis process is of great importance in psychiatry. Early diagnosis of mental health problems enables early initiation of appropriate treatment and better outcomes for patients.

Although anxiety and depression, which are common in the perinatal period, are diseases known by physicians and whose negative consequences are better understood day by day, they are not sufficiently recognized, and even if recognized, adequate treatment is not applied. Anxiety and depression in the perinatal period need to be diagnosed and treated early, as it causes serious harm to the society. It is possible to solve this need with big data analytics systems.

With big data analytics, it is possible to quickly process pregnant data and transform it into meaningful information. This can enable doctors, healthcare professionals and researchers to provide fast and accurate diagnoses, treatment plans and healthcare services.

In our study, a big data analytics system, which can produce instant results in the diagnosis of anxiety and depression in women in the perinatal period, and which integrates scalable architecture thanks to its cluster structure, has been built. While creating this system, technologies such as Apache Kafka, Apache Spark, Kubernetes were used. Thus, it is aimed to minimize the potential harm caused by depression and anxiety in women in the perinatal period.

The primary aim of this study is to identify the most influential questions on outcome related to anxiety and depression in pregnant women by extracting features through performance-optimized algorithms. In this way, it is aimed to reach the result in a shorter time with fewer questions. The next goal of this study is to find the machine learning algorithm that gives the most optimal result on these cleaned data. Its next goal is to create an instant remote health status prediction module for depression and anxiety in pregnant women, based on the Apache Spark big data processing engine, which focuses on using machine learning models in big data flow. Its final goal is to integrate the modules containing these purposes into the Kubernetes clustering system.

In this study, first of all, the data was made ready for processing by passing through the preprocessing stages. There are sixty questions that pregnant women need to answer. The optimized selection-evolutionary feature selection algorithm, which determined this number as optimum, was reduced to sixteen. These; gestational week, baby's health problem, living place (urban or rural), communication with her spouse, education level, number of people living at home, working status, total income level, emotional support from her husband, presence of people to share her problems with, exercise status, educational status of her husband, whether there is a desired pregnancy, the sex of the baby, smoking status, and the chronic illness of the spouse. Various algorithms (Decision Tree, Naive Bayes, K-Nearest Neighbor (K-NN), Random Forest, Gradient Boosted Tree (GBT), Logistic Regression and Deep Feed Forward Neural Network (DFFNN)) has been run on these data to predict the health status of the pregnant woman. It is concluded that Naïve Bayes machine learning

algorithm produces the best results with 90.8% and 81.71% accuracy and precision, respectively.

All the work done was built on the Kubernetes cluster, and the infrastructure, which includes a scalable architecture suitable for big data that produces results about the disease, has been developed. By performing a performance test on this infrastructure, the effects of the increase in the number of executors on the result were observed. As a result of the performance tests, it was concluded that the spark executor, which is effective in parallelization, contributed positively in terms of speed. In addition, the effect of the increase in the number of executors on resource consumption was also examined, and CPU and memory resource usages were included in the diagrams.

Considering the performance of the developed big data platform, it is possible to conclude that computer-based systems operating in real-time with significant accuracy and speed can be used to detect anxiety and depression in pregnant women, replacing traditional methods.

1. GİRİŞ VE AMAÇ

Büyük veri, hacim, çeşitlilik ve hız açısından geleneksel veri işleme yöntemlerinin sınırlarını zorlayacak kadar büyük ve karmaşık olan veri setlerini ifade eder. Büyük veri, yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış verilerin bir araya gelmesiyle oluşabilir.

Büyük veri analitiği günümüzde giderek daha fazla önem kazanmaktadır. Bu veriler, internet, sosyal medya, mobil cihazlar, sensörler ve diğer teknolojiler aracılığıyla üretilebilir. Büyük veri analitiği, bu verileri toplama, depolama, işleme ve analiz etme yöntemlerini kullanarak anlamlı bilgilere dönüştürmek için kullanılan bir dizi teknik ve araçtır.

Birçok endüstri alanında, büyük veri analitiği farklı şekillerde önemli bir rol oynamaktadır. Örneğin, perakende sektöründe, müşterilerin alışveriş davranışlarına ilişkin verileri analiz etmek, mağaza tasarımı, ürün yerleşimini ve fiyatlandırmayı optimize etmek için kullanılabilir. Finansal hizmetler endüstrisinde, büyük veri analitiği müşterilerin risk profillerini belirlemek, dolandırıcılık faaliyetlerini tespit etmek ve yatırım kararlarını desteklemek için kullanılır.

Büyük veri analitiğinin bir diğer önemli kullanım alanı ise sağlık sektörüdür. Büyük veri analizi, hastalıkların nedenleri ve tedavileri hakkında yeni bilgiler sağlayabilir. Sağlık hizmetleri sağlayıcıları, hastaların tıbbi kayıtlarını toplayarak analiz ederek, tedavi planlarını optimize edebilir ve hastaların sağlık sonuçlarını iyileştirebilir. Bu sebeplerden ötürü büyük veri analitiği günümüzün hızla gelişen teknoloji dünyasında giderek daha önemli hale gelmektedir.

Sağlık alanında büyük öneme sahip disiplinlerden biri psikiyatri alanıdır. Zihinsel ve duygusal sağlık sorunlarına odaklanan psikiyatri, bireylerin ruh hali, düşünce süreçleri, davranışları ve duyguları üzerinde çalışan bir tıp uzmanlık dalıdır.

Psikiyatri, zihinsel ve duygusal sağlık sorunlarıyla mücadele eden bireylerin tanı ve tedavisinde kritik bir rol oynar. Depresyon, anksiyete, şizofreni, bipolar bozukluk gibi zihinsel sağlık sorunları, toplumun geniş bir kesimini ve yaşam kalitesini ciddi şekilde

etkileyebilir. Bu gibi rahatsızlıklar çeşitli nedenlere bağlı olarak ortaya çıkabilir. Her bireyin deneyimlediği psikolojik rahatsızlıklar farklı olabilir, ancak genel olarak aşağıdaki durumlarda ortaya çıkabilir:

Genetik yatkınlık: Genetik faktörler, bazı psikolojik rahatsızlıkların ortaya çıkmasında rol oynayabilir. Ailede psikolojik rahatsızlıkların öyküsü bulunan bireylerin, o rahatsızlıklara yakalanma riski daha yüksek olabilir.

Biyolojik nedenler: Beyin kimyasındaki dengesizlikler, hormonal değişiklikler, beyin hasarları, nörotransmitterlerin düzensiz çalışması gibi biyolojik nedenler, psikolojik rahatsızlıkların ortaya çıkmasında etkili olabilir.

Çevresel faktörler: Travmatik olaylar, çocukluk dönemi travmaları, stresli yaşam, travmatik ilişkiler, zorlu sosyal koşullar, iş kaybı gibi çevresel faktörler, psikolojik rahatsızlıkların ortaya çıkmasına katkıda bulunabilir.

Psikososyal etkenler: Bilişsel, duygusal ve davranışsal faktörler de psikolojik rahatsızlıkların ortaya çıkmasında rol oynayabilir. Olumsuz düşünce kalıpları, duygusal zorluklar, çatışmalar, kişilik yapısı gibi psikososyal etkenler, psikolojik rahatsızlıkların gelişimini etkileyebilir.

Yaşam dönemleri ve geçişler: Yaşamın farklı dönemlerinde, özellikle ergenlik, yetişkinlik, yaşlılık gibi geçiş dönemlerinde, bireylerin karşılaştığı zorluklar ve değişimler, psikolojik rahatsızlıkların ortaya çıkmasında etkili olabilir. Rahatsızlıkları doğrudan etkileyebilecek dönemlerden biri ise gebelik ve doğum sonrası 1 yıllık süreyi kapsayan perinatal dönemdir. Bu dönemde Türkiye’de ve Dünya’da en yaygın görülen ruhsal rahatsızlıklar, depresyon ve anksiyete bozukluklarıdır (Dr Carrie Ladd, Dr Nathalie A Rodriguez McCullough, 2017).

Araştırmalar, perinatal dönemdeki depresyon dünya nüfusunun %10 ila %20'sini etkilediğini göstermektedir (Dr Carrie Ladd, Dr Nathalie A Rodriguez McCullough, 2017). Buna ek olarak, kadınların perinatal dönemde aynı anda hem depresyon hem de anksiyete durumu yaşaması oldukça yaygın bir durumdur ve bazı çalışmalarda bu yaygınlık %40'lara kadar dayanmaktadır (Falah-Hassani, Shiri ve Dennis, 2016). Türkiye’de, gebeliğin farklı dönemlerine ve çalışma yöntemlerine göre değişmekle birlikte, gebelik depresyonu için prevalans %12-36, doğum sonrası depresyonu için %15-35 olarak tespit edilmiştir (Aslan ve ark., 2014 ; Kirkan ve ark., 2015). Anksiyete bozuklukları birden fazla alt kategoriyi kapsadığından tek bir prevalans bildirmek

zordur, bu nedenle önce anksiyete bozukluğu açısından risk gruplarını tespit etmek için yapılan tarama çalışmaları önemlidir. Otuz dört ülkeden 221.974 kadını kapsayan bir metaanaliz çalışmasına göre, perinatal dönemde kaygı semptomlarının, gebelik döneminin başlangıçtaki üç aylık bölümü olan ilk trimesterde %18.2, sonraki 3 aylık bölümü olan ikinci trimesterde %19.1 ve son 3 aylık bölümü olan üçüncü trimesterde %24.6 olduğu görülmektedir. Hamilelik sırasında anksiyete bozukluğu prevalansı %15.2 olarak bildirilmiştir. Doğum sonrası dönemde anksiyete belirtileri % 15.0, anksiyete bozukluğu tanıları ise %9.9 olarak tespit edilmiştir (Dennis, Falah-Hassani ve Shiri, 2017). Ülkemizde bu konuda yapılan sınırlı sayıda çalışmalardan biri olan Ankara araştırmasının sonuçlarına göre, 1. trimesterde gebe kadınların % 13.5'inin anksiyete bozukluğu olduğu, %21.6'sının anksiyete ile birlikte depresif bozukluğunun olduğu saptanmıştır (Yücel, Çayır ve Yücel, 2013).

Perinatal dönemde anksiyete ve depresyon, tespit edilemez ve tedavi edilmez ise; hem anne hem bebek üzerinde olumsuz etkilere neden olabilir. Erken dönemde tespit edilemeyen durumlarda tedavi uyumunda sorunlar yaşanmasına, annelik yaşantısında işlevsel sorunlara, tıbbi hastalıklarda kötüleşmeye, kişilerarası ilişkilerin bozulmasına ve ekonomik kayıplara, sigara ve madde kullanımına, bebekte de gelişimsel sorunlar ve sağlık sorunlarına neden olabilmektedir. Gebelik ve doğum sonrası dönemde depresyon ve anksiyete bozukluklarının prevalansı yüksek olmasına rağmen yeterince tanı alamadığı bu sebeple tedavi süreci için geç kalındığı düşünülmektedir.

Büyük veri analitiği ve ona bağlı teknolojilerin kullanılması ile perinatal dönemde ruhsal bozuklukların erken teşhisi söz konusu olacak ve tedavisi aşamasında önemli ölçüde yol katedilebilecektir. Bu çalışmanın ilk aşamasında, öncelikle risk etmenlerinin en etkili olanları belirlenmiş, sonrasında anlık olarak işleyebilmek adına en yüksek performansa sahip algoritma tercih edilmiş ve anlık olarak işleyebilen hibrit büyük veri analitiği platformu inşa edilmiştir.

Önerilen mimarinin Apache Kafka ve Spark gibi ölçeklenebilir ve yüksek performanslı veri analitiği araçlarını bünyesinde barındırması nedeniyle anlık büyük veri işleme uygulamalarında başarılı bir şekilde kullanılabilmesi sonucuna varılmıştır. Perinatal dönemdeki kadınların sayıca fazla olacağı düşünülerek büyük veri altyapısına hizmet veren teknolojiler kullanılması uygun görülmüştür. Bu bağlamda çok sayıda düğümün bir arada (küme şeklinde) çalıştırılmasına ihtiyaç duyulmaktadır. Çalışmanın sonraki aşamasında, diğer modüllerin de yönetiminde kullanılacak olan

Kubernetes çözümü tercih edilmiştir. Bu teknoloji ile büyük veriler üzerinde çalışabilecek ölçeklenebilir mimari altyapısı elde edilmiş ve işlevselliği değerlendirilmiştir.

Bu altyapının kullanılmasıyla, kaygı ve depresyonun zaman alıcı analizi, makul bir doğruluk oranıyla otomatik bilgisayar tabanlı bir teknikle değiştirilebilir olduğu ve perinatal dönemde anksiyete depresyon tarama ve izlem çalışmalarının daha kolay ve hızlı yapılabilir olduğu sonucuna varılabilmektedir.

2. GENEL BİLGİLER

Aile içinde de kadın genellikle birinci önceliğe sahip olduğu için kadın sağlığının korunması ve geliştirilmesi sağlıklı bir toplumun oluşturulmasında oldukça önem arz etmektedir (Akın ve ark., 2008).

Kadınlar, yaşam döngüleri boyunca farklı biyopsikososyal değişimler yaşarlar. Bu değişimler, fetal dönemden başlayarak çocukluk, erişkinlik ve yaşlılık dönemleri boyunca devam eder. Bu durum, kadınların yaşam kalitesini de etkileyerek ruhsal bozukluklara daha yatkın hale gelmelerine neden olabilir (Piccinelli M., Wilkinson, 2000).

Kadınlarda gebelik dönemi psikiyatrik hastalıkların görülmesi açısından oldukça risk içeren bir zaman dilimidir (De Jesus Silva MM., Peres Rocha Carvalho Leite E., Alves Nogueira D., 2016). Bu dönemde birçok farklı duygu ve düşüncenin bir araya gelmektedir ve anne adayları için oldukça önemlidir. Kadın bedeni bir yandan değişikliklere uyum sağlamaya çalışır. Bir yanda uyum çabaları, diğer yanda aile ve toplumdaki statü kaymasının getirdiği duygusal gerginlik, kadınlarda psikolojik değişimlere yol açmakta ve psikiyatrik hastalıkların ortaya çıkmasına neden olabilmektedir (Kömürcü N., 2012).

Gebelik döneminde, anne adaylarının vücutlarında birçok hormonal değişiklik meydana gelir. Bu hormonal değişimler bazı durumlarda duygu durum değişikliklerine neden olabilir. Anne adayları bu dönemde depresyon, kaygı, stres, endişe ve diğer duygusal bozukluklarla karşılaşabilirler. Bebeklerinin sağlıklı olacağından emin olmak, doğum sırasında ne bekleyecekleri ve yeni ebeveynler olarak nasıl başarılı olacakları gibi konularda endişeler yaşayabilirler. Bu endişeler normaldir, ancak aşırı hale gelirse ve günlük yaşamlarını olumsuz etkilerse, anne adaylarının ruh sağlığı konusunda uzman yardımı almaları önemlidir.

2.1. Perinatal Dönemde Görülen Psikiyatrik Tanılar

Perinatal dönem, hamilelik ve doğum sonrası ilk bir yıl boyunca kadınların yaşadığı bir dönemdir. Bu dönemde kadınlar, fiziksel, hormonal, sosyal ve psikolojik açıdan

birçok deęişim yaşırlar. Bu deęişimler, kadınların duygusal durumlarını da etkileyebilir ve bazı kadınlar psikiyatrik rahatsızlıklar yaşayabilir.

2.1.1. Depresyon

Perinatal depresyon, hamilelik veya doğum sonrası ilk bir yıl boyunca kadınların yaşadığı, şiddetli ve uzun süreli bir depresyon durumudur. Bu durum, anne ve bebek sağlığına olumsuz etkileri olabileceği için ciddi bir durumdur.

Perinatal depresyonun belirtileri arasında umutsuzluk, yorgunluk, kaygı, çaresizlik, değersizlik hissi, uykusuzluk, iştah deęişiklikleri ve odaklanma güçlüğü gibi ruhsal ve fiziksel semptomlar yer alabilir. Bu semptomlar, bebeğe karşı duyarsızlık ve bağlanma sorunları gibi dięer sorunlara da neden olabilir.

Perinatal depresyon, birçok farklı nedeni olabilecek karmaşık bir durumdur. Bu nedenler arasında hormonal deęişimler, stres, yalnızlık, yetersiz sosyal destek, geçmişte yaşanan depresyon gibi psikolojik faktörler, gebelik ve doğum komplikasyonları gibi fiziksel faktörler yer alabilir.

Perinatal depresyon tedavisi, erken teşhis ve doğru bir şekilde uygulanan tedavi ile mümkündür. Tedavi, genellikle psikoterapi, ilaç tedavisi veya her ikisini de içerebilir. Ayrıca, egzersiz, sağlıklı beslenme ve uyku düzeni gibi yaşam tarzı deęişiklikleri de yardımcı olabilir.

Bu nedenle, perinatal dönemde depresyon semptomları yaşayan kadınların, bir sağlık profesyoneline başvurmaları ve tedavi almaları önemlidir. Ayrıca, perinatal depresyonun önlenmesi için gebelik ve doğum sonrası dönemde kadınların sağlıklı yaşam tarzı alışkanlıkları geliştirmeleri, sosyal destek ağlarını genişletmeleri ve stres yönetimi tekniklerini uygulamaları da faydalı olabilir.

2.1.1.1. Depresyonun sıklığı

Yapılan çalışmalar, gebelik dönemi depresyonunun yaygınlığının %7 ila %20 arasında deęişebileceğini göstermektedir(Heather A Bennett, Adrienne Einarson, Anna Taddio, Gideon Koren, 2004).

Doğum sonrası depresyon: Doğum sonrası depresyon, doğum sonrası dönemde annelerde ortaya çıkan bir depresif bozukluktur. Yapılan araştırmalar, doğum sonrası depresyonun yaygınlığının %5 ila %25 arasında deęişebileceğini göstermektedir (Gavin ve ark., 2005).

Doğum sonrası anksiyete ve depresyon: Doğum sonrası dönemde, anksiyete ve depresyon semptomları birlikte görülebilir. Yapılan çalışmalar, doğum sonrası dönemde depresyon ve anksiyete semptomlarının birlikte görülme prevalansının %8 ila %20 arasında değişebileceğini göstermektedir (Falah-Hassani ve diğerleri, 2016).

Bu değerler, farklı popülasyonlar ve çalışma yöntemleri arasında farklılık gösterebilir ve zaman içinde değişebilir. Perinatal depresyonun yaygınlığına dair kesin bir rakam vermek zor olsa da, perinatal dönemde depresyonun yaygın bir sorun olduğunu ve doğum öncesi ve sonrası dönemde psikolojik sağlık açısından dikkate alınması gereken bir konu olduğunu söyleyebiliriz.

2.1.1.2. Risk etmenleri

Perinatal dönemde depresyonu etkileyebilecek çeşitli sayıda risk faktörleri bulunmaktadır. Bunlar aşağıdaki şekilde özetlenebilir.

Önceki psikiyatrik öykü, ailede veya yakın çevrede depresyon öyküsü, stresli yaşam olayları (örneğin, yakın birinin kaybı, evlilik sorunları, maddi sıkıntılar gibi), sosyal destek eksikliği, gebelik veya doğum komplikasyonları (örneğin, erken doğum, doğum sonrası komplikasyonlar), gebelik dönemi fiziksel sağlık sorunları (örneğin, gebelik diyabeti, tiroid hastalığı gibi), yaş: Genç yaşta kadınlar, perinatal dönemde depresyon riskini daha yüksek olabilir, madde kullanımı(alkol, sigara, uyuşturucu), gebelik istememe veya istenmeyen gebelik, hormonal değişiklikler, doğum sonrası hormonal düşüş, zor doğum deneyimi, emzirme problemleri gibi etmenler depresyonun ortaya çıkmasında etkin risk etmenleridir.

2.1.1.3. Hastalığın teşhis aşaması

Depresyon hastalığının teşhis aşamasında, bireylerin depresif semptomlarının şiddetini, sıklığını ve ilerleyişini takip etmek için yapılan süreçler yer almaktadır. Bu süreç, tedavi planının etkinliğini değerlendirmek, semptomların gerilemesini gözlemlemek ve gerektiğinde tedaviyi ayarlamak amacıyla düzenli olarak gerçekleştirilmektedir. Bu bağlamda kullanılan birçok ölçek bulunmaktadır (Milgrom, Westley ve Gemmill, 2004). En önde gelenleri, EDDÖ (Edinburgh Doğum Sonrası Depresyon Ölçeği), BDI (Beck Depression Inventory), PDSS (Postpartum Depression Screening Scale), CES-D (Center for Epidemiological Studies Depression Scale) ölçekleridir.

Doğum öncesi depresyonu değerlendirmek için kullanılan ölçeklerin duyarlılığı ve özgüllüğü, Gaynes ve arkadaşları tarafından yapılan bir meta-analiz çalışmasında incelenmiştir. Tüm ölçümlerin iyi bir özgüllük gösterdiği, EDDÖ ve PDSS'ın BDI'den daha duyarlı olduğu, ve buna ek olarak EDDÖ'nün diğer tüm ölçeklere göre klinik kullanımının daha yararlı olabileceği sonucuna varılmıştır (Gaynes, Gavin, Meltzer-Brody, Lohr, Swinson, Gartlehner, Brody ve Miller, 2005).

Yapılan başka bir çalışmada, depresyonun klinik olarak belirlenmesi zor olabileceği durumlarda, EDDÖ ile yapılan takiplerin önemli bir sonuç sağladığı ve pozitif sonuçlar elde edildiği belirtilmiştir. Ayrıca, EDDÖ'nün kendine zarar verme düşüncesini saptama yeteneği sayesinde, 0 üzerinde bir sonuç elde edilmesinin müdahale gerekliliğine işaret edebileceği vurgulanmıştır (Milgrom ve Gemmill, 2014). Bu sebeple depresyon araştırması yaparken değerlendirmek üzere EDDÖ tercih edilmiştir.

2.1.2. Anksiyete (kaygı) bozuklukları

Anksiyete bozukluğu, bireylerin yaşamlarını etkileyen, sürekli ve aşırı endişe, korku veya kaygı durumlarıyla karakterize edilen bir psikiyatrik rahatsızlıktır. Bu durum, kişinin günlük işlevselliğini, sosyal ilişkilerini ve genel yaşam kalitesini olumsuz yönde etkileyebilir. Anksiyete bozukluğu, farklı semptomlarla kendini gösterebilir ve belirli bir tetikleyici olay veya durumla ilişkili olabileceği gibi, bazen de belirgin bir neden olmadan ortaya çıkabilir.

Anksiyete bozukluğunun temel semptomları arasında sürekli bir kaygı hali, endişe dolu düşünceler, sürekli tedirginlik hissi, huzursuzluk, uyku problemleri, odaklanma güçlüğü, sinirlilik, kas gerginliği, kalp çarpıntısı, solunum problemleri ve sindirim sorunları yer alır. Bu semptomlar bireyin yaşamını olumsuz etkileyerek iş verimliliğini düşürebilir, günlük aktiviteleri zorlaştırabilir ve sosyal ilişkileri olumsuz etkileyebilir.

Gebelik dönemi ise bu semptomların tetiklenebileceği en önemli dönemlerden biridir. Bu dönemde, hem anne adayının hem de bebeğin sağlığını korumak, biyolojik gereksinimlerini karşılamak ve doğum için gerekli anatomik ve fizyolojik uyumu sağlamak amacıyla çeşitli değişiklikler ortaya çıkar. Her kadının bu değişikliklere uyum sağlama şekli kişiden kişiye farklılık gösterebilir. Bu süreçte anne adayının adaptasyon sağlanması, toplumsal ve aile içinde yeni bir statü olan anne rolünü benimsemesi ve artan sorumluluklarına uyum göstermesi beklenir.

Gebelik, kadının hayatında doğal bir süreç olmasına rağmen, her kadının gebelikte yaşanan değişikliklere uyum sağlama şekli kişisel olarak değişebilir. Bu semptomların ortaya çıkabileceği en önemli zamanlardan biri hamilelik dönemidir. Bu süre boyunca, hamile kadının ve doğmamış bebeğin sağlığını garanti altına almak, biyolojik taleplerini karşılamak ve doğum için gerekli olan anatomik ve fizyolojik dengeyi oluşturmak için çeşitli değişiklikler gerçekleşir. Gebe kadının bu değişimlere uyum sağlaması, toplum ve aile içinde değişen statü ve anne rolünü benimsemesi beklenmektedir. Ayrıca değişen ve genişleyen sorumluluğa uyum sağlamaları da beklenmektedir. Yaşanan tüm bu değişiklikler gebelerin ruh sağlığını olumsuz etkileyebilmekte ve bazı psikiyatrik problemlerin çıkmasına sebep olabilmektedir (Özkan Dikmen, 2020).

Çocuklar üzerindeki potansiyel etkileri dikkate alındığında, doğum öncesi ve doğum sonrası dönemde anksiyete bozukluklarının taraması ve tedavisi özellikle önemlidir (Fawcett, E. J., Fairbrother, N., Cox, M. L., White, I. R., & Fawcett, 2019).

Yapılan çalışmalarda, gebelik dönemindeki maternal stresin preeklampsi, düşük doğum ağırlığı ve erken doğum riskini artırdığı, ayrıca anksiyetenin fetal gelişim üzerinde nöroendokrin, immun/inflamatuvar ve vasküler etkileri yoluyla olumsuz bir etkiye sahip olabileceği bulunmuştur. Gebelerde, hipotalamus-hipofiz-adrenal yolakta artmış aktivite ve inhibisyon azalmasıyla birlikte opioid, gaba, serotonin ve dopamin yollarında azalma, aynı zamanda sempatik-adrenal aktivitede artış gözlenmiştir. Yüksek kaygılı annelerin çocuklarıyla etkileşimde daha az bulunduğu, onlara daha az gülümsediği ve daha az oyun oynadığı tespit edilmiştir. Bu annelerin çocukları da daha az gülümsediği ve daha fazla ağladığı, ayrıca annelerine güvensiz bağlanma eğilimi gösterdikleri ve bu şekilde daha güvensiz bir ilişki kurma ihtimallerinin daha yüksek olduğu belirlenmiştir (Y. S. Oğur, 2022), (Ding et al., 2014), (Fawcett, E. J., Fairbrother, N., Cox, M. L., White, I. R., & Fawcett, 2019), (Wadhwa, Glynn, Hobel, Garite, Porto, Chiciz-DeMet, Wigglesworth ve Sandman, 2002).

2.1.2.1. Anksiyetenin sıklığı

Anksiyete bozukluğu, dünya genelinde oldukça yaygın bir durumdur. İstatistiklere göre, her 5 kişiden 1'i hayatının bir döneminde anksiyete bozukluğu yaşamaktadır. Bu durum kadınlarda erkeklere göre daha sık görülmektedir. Genellikle ergenlik döneminden başlayarak yetişkinlik dönemine kadar olan yaş aralığında ortaya çıkmaktadır, ancak çocuklarda da görülebilir. Anksiyete bozukluğu, genetik yatkınlık,

beyin kimyasındaki dengesizlikler, stresli yaşam olayları, travmatik deneyimler, çocuklukta yaşanan olumsuz deneyimler ve kişilik özellikleri gibi çeşitli faktörlerden kaynaklanabilir.

Anksiyete bozukluğunun tedavisi mümkündür ve erken teşhis önemlidir. Tedavide kullanılan yöntemler arasında psikoterapi, ilaç tedavisi ve yaşam tarzı değişiklikleri bulunur. Kognitif davranış terapisi (KDT) ve bilişsel terapi, anksiyete semptomlarını yönetmek ve kişinin düşünce kalıplarını değiştirmek için sıkça kullanılan terapi yöntemleridir. İlaç tedavisinde antidepressanlar ve anksiyolitikler, semptomları hafifletmek ve anksiyete düzeylerini dengelemek için kullanılabilir.

Ayrıca, yaşam tarzı değişiklikleri de anksiyete bozukluğunun yönetiminde etkili olabilir. Düzenli egzersiz yapmak, sağlıklı bir beslenme düzeni takip etmek, uyku düzenine özen göstermek, stresten kaçınmak ve rahatlama teknikleri gibi stratejiler, anksiyete semptomlarını hafifletebilir ve genel yaşam kalitesini artırabilir.

2.1.2.2. Risk etmenleri

Perinatal dönemde ortaya çıkan anksiyetenin birden fazla ilişki sebebi bulunmaktadır. Bunların başında, önceden var olan anksiyete bozukluğu, önceki depresyon veya anksiyete öyküsü, aile öyküsü, stresli yaşam olayları, düşük sosyoekonomik durum, hamilelik komplikasyonları, gebelik sonrası zorluklar gibi sebepler yer almaktadır. Bazı çalışmalarda ise sonuç üzerinde en etkili faktörler değişiklik göstermektedir. Furtado ve ark., öngördürücü risk faktörlerini ilk doğum olması, birden fazla doğum olması, annenin genç yaşı ve düşük sosyal destek olarak belirlemişlerdir (Furtado, Chow, Owais, Frey ve Van Lieshout, 2018). Diğer birçok çalışmada, geçmiş aile öyküsü ya da anksiyete bozukluğunun olması, sonuç üzerinde en etkili özellik olarak saptanmıştır (Faisal-Cury, Menezes, Araya ve Zugaib, 2009), (Martini, Petzoldt, Einsle, Beesdo-Baum, Höfler ve Wittchen, 2015), (Rubertsson, Hellström, Cross ve Sydsjö, 2014). Haßdenteufel ve ark., ise fiziksel hareketlilikte azalmanın, anksiyete ve depresyon için sonuç üzerinde en etkili özellikler olduğu sonucuna varmışlardır (Haßdenteufel, Feißt, Brusniak, Lingenfelder, Matthies, Wallwiener ve Wallwiener, 2020). İnternetin ve dolayısıyla sosyal medya kullanımının artmasının, açık havada geçirilen sürenin kısılmasının, depresyon ve anksiyete üzerine etkisinin araştırılması büyük önem arz etmektedir (Goyal, Gay, Torres ve Lee, 2018), (Marina Vargas-Terrones, Ruben Barakat, Belen Santacruz, Irene Fernandez-Buhigas, 2019), (Y. S. Oğur, 2022).

2.1.2.3. Hastalığın teşhis aşaması

Perinatal dönemde anksiyetenin tespiti için ölçekler kullanılmaktadır. Bunun sebebi anksiyete belirtilerini objektif bir şekilde değerlendirmek ve düzeylerini belirlemektir. Tanısal değerlendirme, semptomların takibi, tedavi planlaması, iyileşme değerlendirmesi, araştırma ve istatistiksel analiz gibi sebepler de ölçek kullanılmasının diğer sebepleri arasında yer almaktadır.

Bu ölçekler arasında önde gelenler, DASS (Depresyon, Anksiyete ve Stres Ölçeği), STAI (Spielberg Durumluk-Sürekli Kaygı Envanteri), GAD-7 (Yaygın Anksiyete Bozukluğu 7 Maddelik Ölçeği), PASS (Perinatal Anksiyete Tarama Ölçeği) olarak özetlenebilir.

Somerville ve ark. PASS'ın tarama kullanımı için geçerliliği kanıtlanmış, kullanımı basit bir ölçek olduğunu ve hamilelik sırasında veya sonrasında anksiyete ile gelen kadınları değerlendirmek için geçerli ve yararlı olduğunu vurgulamışlardır (Somerville, 2014). Bu sebeple gebe kadınların anksiyetelerini değerlendirmek üzere PASS tercih edilmiştir.

2.2. Tıp Alanında Büyük Veri Analitiğinin Yeri

Büyük veri analitiği, tıp dünyasında son yıllarda önemli bir rol oynamaya başlamıştır. Tıbbi alanda verilerin hızla artması, dijital kayıtların yaygınlaşması ve ileri veri analiz yöntemlerinin geliştirilmesi, büyük veri analitiğinin kullanımını artırmıştır. Bu teknolojik ilerlemeler, sağlık hizmetlerinin iyileştirilmesi, hastalıkların daha iyi anlaşılması, teşhis ve tedavi süreçlerinin optimize edilmesi ve hastaların sağlık sonuçlarının geliştirilmesi için büyük potansiyel sunmaktadır (Thirunavukarasu, C, R, Gopikrishnan ve Palanisamy, 2022).

Büyük veri analitiği, tıbbi verilerin toplanması, depolanması, işlenmesi ve analiz edilmesi yoluyla bilgi elde etmeyi amaçlar. Bu veriler, hastaların elektronik sağlık kayıtları, görüntüleme verileri, laboratuvar sonuçları, genetik veriler, sensör verileri, hastane yönetim sistemleri gibi çeşitli kaynaklardan gelir. Büyük veri analitiği, bu verileri anlamlı bilgilere dönüştürerek sağlık hizmeti sağlayıcılarına, araştırmacılara ve karar vericilere değerli öngörüler sunar (Fury, Oh ve Berkson, 2022).

Büyük veri analitiği, tıp dünyasında çeşitli alanlarda kullanılmaktadır (Saurabh Biswas, 2022). İşte bazı örnekler:

Tanı ve tedavi: Büyük veri analitiği, hastaların genetik ve klinik verilerini kullanarak hastalıkların tanı ve tedavisinde daha doğru ve kişiselleştirilmiş yaklaşımların geliştirilmesine yardımcı olur. Örneğin, kanser hastalarının genetik profilleri ve tedavi yanıtları analiz edilerek daha etkili tedavi protokolleri belirlenebilir.

Epidemiyoloji ve hastalık izleme: Büyük veri analitiği, epidemiyolojik verilerin analiziyle hastalık salgınları ve yayılımı hakkında önemli bilgiler sağlar. Bu veriler, hastalık kontrol ve önleme stratejilerinin geliştirilmesinde kullanılabilir.

Sağlık hizmetlerinin iyileştirilmesi: Büyük veri analitiği, hastane ve klinik verilerinin analiziyle sağlık hizmetlerinin etkinliğini ve verimliliğini artırmaya yönelik çalışmalara olanak sağlar. Örneğin, hastane performansı, kaynak yönetimi ve hasta memnuniyeti gibi alanlarda analizler yapılabilir.

İlaç keşfi ve geliştirme: Büyük veri analitiği, ilaç keşfi ve geliştirme süreçlerinde kullanılarak moleküler yapıların analizi, ilaç etkileşimleri ve yan etkilerin değerlendirilmesi gibi konularda bilgi sağlar. Bu da daha etkili ve güvenli ilaçların geliştirilmesine katkıda bulunur.

Hastalık öngörüsü: Büyük veri analitiği, genetik ve klinik verilerin analiziyle bireylerde hastalık risklerinin öngörülmesine yardımcı olur. Örneğin, kalp hastalığı, diyabet veya kanser gibi hastalıkların risk faktörlerini belirleyerek erken müdahale edilebilir ve önleyici tedbirler alınabilir.

Büyük veri analitiği, hastaların sağlık sorunlarını iyileştirme, sağlık hizmetlerini optimize etme, kaynakları etkin bir şekilde kullanma ve hastalıkları daha iyi anlama potansiyeli taşır. Büyük veri analitiği, tıbbi alanda önemli bir araç haline gelmiş olsa da, insan faktörü ve klinik deneyim gibi diğer unsurlarla birlikte değerlendirilerek en iyi sonuçların elde edilmesi sağlanmalıdır (Cirillo ve Valencia, 2019), (Nair, Shetty ve Shetty, 2017), (Saurabh Biswas, 2022).

2.2.1. Perinatal dönemde büyük veri analitiği

İçinde bulunduğumuz çağ, özellikle son yirmi yıl, dijital verilerin araştırma, teknoloji, toplum ve sağlık hizmetleri dahil olmak üzere çeşitli disiplinlerde giderek daha önemli hale geldiği "büyük veri çağı" olarak adlandırılabilir (Nair, Shetty ve Shetty, 2018).

Büyük veri analitiği, değerli bilgilerin hızlı bir şekilde çıkarılmasına yardımcı olurken, aynı zamanda bilgi işlem kaynakları ve analitik çerçevenin önündeki engelleri ortadan

kaldırmaya çalışır. Büyük veri analitiği, büyük veri setlerinin işlenmesini, analizini ve değerlendirilmesini kolaylaştıran yöntemler ve teknolojiler kullanarak bu engelleri aşmayı hedefler.

Sağlık sorunlarını çözmek için Apache Spark (“Apache Spark”, 2023), Apache Hadoop (“Apache Hadoop”, 2023), ApacheStorm (“Apache Storm”, 2023) ve Apache Kafka (“Apache Kafka”, 2023) gibi bir büyük veri analitiği çerçevesi oluşturmada yoğun olarak kullanılan çok sayıda araştırma vardır.

Ruh sağlığı hastalıkları, bireylerin yaşamını derinden etkileyen ve dikkatle tedavi edilmesi gereken sağlık sorunlarıdır. Hamilelik sürecini içeren perinatal dönemde ortaya çıkan psikolojik bozukluklar zamanında tespit edilmezse hem anne hem de bebek üzerinde olumsuz etkiler yaratacaktır. Durum buysa, toplumun ruh sağlığı için büyük önem taşımaktadır.

Makine öğrenimi, hastalık modellerini, hastalığı etkileyebilecek yenilikçi ilaçların keşfini ve geliştirilmesini ve psikiyatride önleme yöntemlerini bilgilendirme potansiyeline sahiptir (Tai, Albuquerque, Carmona, Subramanieapillai, Cha, Sheko, Lee, Mansur ve McIntyre, 2019). Bununla birlikte, beyin ve ruhsal bozuklukların tanınması için kullanılan teknikler bazen eksik veya hatalı temsiller oluşturmuş ve sonuçlar yeterli olmamıştır. Makine öğrenimi algoritmaları, büyük veri analitiği platformunda verileri anlamlandırmak ve çözüm odaklı sonuçlar elde etmek için kullanılabilir. Sağlık alanında, yeni hastalık tedavilerinin keşfi ve geliştirilmesinde de bilgi sağlama potansiyeli bulunmaktadır.

Büyük veri analitiği oluşturularak elde edilecek gelişmelerden psikolojik rahatsızlıkların yararlanması toplum sağlığı açısından önemli bir yere sahiptir. Bireylerin yaşamını derinden etkileyen ve dikkatle tedavi edilmesi gereken sağlık sorunları olan psikolojik bozuklukların, yaşamın bazı dönemlerinde ortaya çıkma olasılığının daha yüksek olduğu görülmektedir. Gebelik sürecini içeren perinatal dönem de bu süreçlerden biridir. Bu çağın zararını azaltmak için, araştırmacılar ve psikiyatrlar makine öğrenimi tekniklerini kullanarak beyindeki, davranışlardaki ve hastalıklardaki karmaşık kalıpları kullanmak için benzeri görülmemiş bir şansa sahipler (Bzdok ve Meyer-Lindenberg, 2018). Bu analiz teknikleri kullanılarak önemli psikiyatrik bozuklukların erken dönemde saptanması, ilişkili faktörlerin belirlenmesi ve önleyici tedbirlerin geliştirilmesi amaçlanmaktadır.

2.2.1.1. Perinatal dönemde büyük veri mimarisi kullanarak yapılan çalışmalar

Sağlık alanında büyük veri analitiği ile entegre olabilen teknolojileri kullanan çok sayıda çalışma bulunmaktadır. Bunlardan en öne çıkanı yapay zeka üzerine yapılan çalışmalardır. Yapay zeka sistemlerinin dış verilerini doğru yorumlamak, bu verilerden ders çıkarmak ve bu öğrenmeleri belirli hedeflere ulaştırmak demektir. Öte yandan, makine öğrenimi, algoritmaların girdi ve çıktı verileri arasındaki bağlantıları keşfettiği bir yapay zeka dalıdır (Bormann, Kanakasabapathy, Thirumalaraju, Gupta, Pooniwala, Kandula, Hariton, Souter, Dimitriadis, Ramirez, Curchoe, Swain, Boehnlein ve Shafiee, 2020). Makine öğrenimi algoritmaları, sağlık hizmetlerinin çeşitli dallarında kullanılmıştır (Arai, Kondo, Fuse, Shibasaki, Masuko, Sugita, Teshima, Uchida, Fukuda, Ohashi, Ozawa, Ichinohe, Kanda ve Atsuta, 2018), (Chen, Wang, Du, Mei, Zhou ve Cai, 2023).

Akut aşı konak hastalığı geliştirme olasılığını hesaplamak için Arai ve arkadaşları 26.695 hastadan oluşan bir Japon popülasyonu kullanmışlardır. Eğitim grubu modelleri, karar ağacı algoritması kullanılarak geliştirilmiştir. Uygulanan işlemler sonunda algoritmalar klinik olarak kabul edilebilir ve yüksek doğrulukta sınıflandırma skorları üretmiştir. Kullanılan algoritmaların doğruluk ve kesinliklerinin yüksek olması özellikle sağlık alanında oldukça önemlidir. Yaseen ve ark. kalp hastalıklarını %97 doğrulukla tanımlayan makine öğrenimi tabanlı bir sistem sunmuştur (Randhawa ve Singh, 2015). Bir diğer çalışma ise sinir ağlarının kan analiz hizmetlerinde kullanılmasıdır. Kullanılan makine öğrenimi sistemi, patolojik örnekler üzerinde test edildiğinde %91'lik bir teşhis verimliliği elde etmiştir (Walter, Pohlkamp, Meggendorfer, Nadarajah, Kern, Haferlach ve Haferlach, 2023). Makine öğrenimi teknikleri tıbbın diğer dallarında olduğu gibi nörolojide de kullanılmaktadır. Nörolojide yararlı bir teşhis unsuru, beynin elektriksel aktivitesi hakkında bilgi sağlayan Elektroensefalografi (EEG) değerleridir. Bu sinyalleri analiz etmek ve bir tahmin sağlamak için birçok makine öğrenimi tekniği uygulanmıştır. Subaşı, EEG veri setindeki hibrit SVM (Support Vector Machines) ve GA (Genetic Algorithms) algoritmalarını kullanarak EEG kayıtlarındaki epileptik nöbetleri tespit etmeyi amaçlamıştır (Subasi, 2007). Sonuç olarak %99,38 doğruluk sağlayan hibrit bir algoritma elde etmiştir. Psikolojik çalışmalarda da yapay zeka teknikleri kullanılmaktadır. Bunlardan biri de Priya ve arkadaşlarının çalışmasıdır. Bu çalışmada makine öğrenmesi algoritmaları kullanılarak kaygı ve depresyonun tespit edilmesi

amaçlanmaktadır. Priya ve ark. bunu başarmak için beş farklı algoritma (Karar ağaçları, SVM, KNN (K-En Yakın Komşular), Naive Bayes ve Random Forest) kullanmışlardır (Priya, Garg ve Tigga, 2020). Sonuç olarak, en iyi sonucu veren modelin Random Forest algoritması ile oluşturulduğu sonucuna varmışlardır. Başka bir çalışma, makine öğrenimi kullanarak denizcilerde kaygı ve depresyonu tespit etmeyi amaçlamaktadır. Bunu yaparken öncelikle özellik seçim algoritmalarından yararlanılmıştır. 470 denizci üzerinde gerçekleştirilen bu çalışmada beş farklı sınıflandırma algoritması (Catboost, Random Forest, Logistic Regresyon, SVM, Naive Bayes) kullanılmıştır. Catboost algoritmasının %82,6 doğruluk oranı ile en doğru sonucu veren algoritma olduğu sonucuna varılmıştır (Sau ve Bhakta, 2019a). Başka bir çalışmada Paulo ve arkadaşları, MDB (majör depresif hastalığı) ve BB (bipolar hastalığı) sınıflandırmak için makine öğrenimi algoritmalarını kullanmıştır (Suen, Goerigk, Razza, Padberg, Passos ve Brunoni, 2021). Doğruluk oranı en yüksek modeli üretmek için 5 katlı çapraz doğrulama tekniği kullanılmış ve makine öğrenmesi algoritması olarak XGBoost algoritması tercih edilmiştir (Luo, Wang ve Wang, 2021). Tedavi edici ilaçların miktarlarını ayarlayarak optimum çözüme ulaşma aşamasında da makine öğrenmesi tekniklerinden faydalanmışlardır. Bunlardan biri Shuzhe Zhou ve diğerlerinin çalışmasıdır. MDB hastaları genellikle SSRI'ler (seçici serotonin geri alım inhibitörleri) ile tedavi edilir; ancak kür oranı yeterli değildir. Bunu optimize etmek için SSRI alan hastalarda 8 haftalık sonuçları tahmin etmek için makine öğrenimi modelleri geliştirmek amaçlanmıştır (Zhou, Ma, Lou, Lv, Tian, Wei, Zhang, Zhu, Chen, Si, Wang, Wang, Zhang, Huang, Liu ve Yu, 2021).

Perinatal dönemde ortaya çıkan psikolojik bozuklukların teşhisinde hastaya çok soru sorulmakta ve bu da hastaların sorularının yanıtlanmasında büyük bir handikap oluşturmaktadır. Hastaların sorulara cevap vermedeki sabırsızlığı tanıyı olumsuz etkilemektedir. Bu sorular arasından hastalığı en çok etkileyen özelliklerin seçilmesi hem yapay zeka algoritmalarının işleme karmaşıklığını (dolayısıyla gecikmesini) azaltacağı gibi, aşırı uyum ve çoklu bağlantı gibi problemlerin etkilerini de azaltacaktır (Remeseiro ve Bolon-Canedo, 2019). Yiye ve arkadaşları doğumdan sonraki bir yılı kapsayan doğum sonrası dönem verilerini incelemiştir (Zhang, Wang, Hermann, Joly ve Pathak, 2021). Bu verilerde doğum sonrası dönemde kadınların depresyonda olup olmadıklarını gözlemleyen bir çerçeve oluşturmayı amaçlamışlardır. Risk tahminini gerçekleştirmek için veri kümelerinden minimum özellik listesi çıkarma, işleme ve

makine öğrenimi algoritması uygulamışlardır. Model en yüksek doğrulukla eğitilirken, ruh sağlığı öyküsü, ultrasonografik komplikasyonlar, ilaç reçeteleri ve hasta demografisi ile ilgili klinik özellikler kullanılarak veri seti azaltılmıştır. Tüm bunlardan sonra, geliştirme ve doğrulama veri setlerinde AUC (area under the curve) değeri tarafından gösterilen model performansları sırasıyla 0,937 ve 0,886 olarak elde edilmiştir (Zhang ve diğerleri, 2021).

Bazı durumlarda, özellik çıkarma gibi yöntemlerle alınan veriler ne kadar azaltılırsa azaltılsın, veri miktarı sistem ve kaynakların çalışmasında performans kayıplarına neden olabilmektedir. Özellikle hacim, hız ve değişkenlik açısından büyük veriye katkıda bulunan birçok kaynaktan eşi benzeri görülmemiş bir veri akışı olan sağlık alanında, verilerin anlık olarak yorumlanmasının performans açısından büyük önem taşıdığı bilinmektedir. Performans ve hızın önemli olduğu bu gibi durumlarda makine öğrenimi (ML) algoritmalarını içeren büyük veri işleme platformları kullanılabilir. Tıp alanında mevcut olan çok büyük miktardaki veri, bu platformlarda büyük veriye yönelik makine öğrenimi yaklaşımları kullanılarak yönetilebilir görünmektedir. Bu tür yöntemler, nesnel, karşılaştırılabilir doğruluk ölçümleri ve hastalık modellerine ilişkin önemli bilgiler sağlayarak, daha önce keşif amaçlı tıbbi araştırma için erişilemeyen olasılıkları ortaya çıkarabilmektedir. Mevcut hastalık modellerini toplayarak ve inceleyerek, büyük veri ve makine öğrenimi psikiyatride birlikte kullanılabilir (Cao, Meyer-Lindenberg ve Schwarz, 2018). Ruhsal hastalıkların anlaşılması ve tedavisi için daha doğru ve uygun hipotezlerin geliştirilmesini destekleyebilmektedir. Manal ve ark. kronik akciğer hastalığını tespit etmek için hibrit modeli geliştirmiştir. Sonuç üzerinde Relief-F ve Ki-Kare teknikleri kullanılarak etkili özellikler seçilmiştir. Daha sonra makine öğrenimi algoritmaları test edilmiş ve performans karşılaştırmaları yapılmıştır. Tüm işler, büyük bir veri işleme platformu olan Apache Spark üzerinde gerçekleşmiştir. Çalışmaya göre en iyi performansa sahip sınıflandırma algoritmalarının SVM, DT ve GBT (Gradient Boosted Tree), en iyi performansa sahip seçim algoritmasının ise Relief-F olduğu görülmüştür (Abdel-Fattah, Othman ve Goher, 2022).

Sağlık alanında büyük veri üzerine yapılan bir başka çalışma da Lekha ve ark. çalışmasıdır. Bu çalışmada, akış halindeki sağlık verileri işlenmiştir ve hasta gerçek zamanlı olarak bir mesaj gönderilerek durumu hakkında bilgilendirilmiştir (Kessler,

van Loo, Wardenaar, Bossarte, Brenner, Cai, Ebert, Hwang, Li, de Jonge, Nierenberg, Petukhova, Rosellini, Sampson, Schoevers, Wilcox ve Zaslavsky, 2016).

2.3. Büyük Veri Analitiği

Büyük veri analitiği, büyük hacimlerde, yüksek çeşitlilikte ve hızlı bir şekilde oluşan verileri anlamak, işlemek ve değerli bilgileri çıkarmak için kullanılan analitik yöntemlerin ve teknolojilerin kullanımını ifade eder.

Büyük veri analitiği, yapılandırılmış veya yapılandırılmamış verilerin toplanması, depolanması, işlenmesi, analiz edilmesi ve yorumlanmasını içerir. Büyük veri analitiği, gelen verileri kullanarak anlamlı bilgileri keşfetmek, eğilimleri ve desenleri tanımlamak, karar verme süreçlerini iyileştirmek ve rekabet avantajı elde etmek için kullanılmaktadır (Ranta, Paroutis, Talaoui ve Kohtam, 2023).

2.3.1. Ön işlem

Makine öğrenmesinde, veri ön işleme süreçleri, makine öğrenimi algoritmalarının daha etkili çalışmasını sağlamak için gerçekleştirilir. Aşağıda, yaygın olarak kullanılan ön işleme adımları yer almaktadır;

- **Veri Temizleme:** Veri setlerinin içerisindeki gürültü, eksik veya yanlış verilerin tespit edilip düzeltilmesini sağlar. Eksik verilerin doldurulması, aykırı değerlerin belirlenmesi ve düzeltilmesi, veri setinin kalitesini artırır ve sonuçların daha güvenilir olmasını sağlar.
- **Veri Standardizasyonu:** Veri setindeki sayısal özelliklerin ölçeklerini birbirine yaklaştırır. Ölçeklendirme işlemi, özelliklerin farklı birimlerde veya aralıklarda ölçülmesi durumunda önemlidir. Bu, makine öğrenimi algoritmalarının daha iyi çalışmasını sağlar ve yanlış sonuçlara yol açan ölçek etkisini ortadan kaldırır.
- **Veri Normalizasyonu:** Veri setindeki sayısal özelliklerin dağılımını normal bir dağılıma dönüştürür. Bu işlem, özelliklerin farklı dağılım özellikleri nedeniyle bazı algoritmaların etkisini azaltmasını engeller. Normalizasyon, veri setindeki değerleri belirli bir aralığa sıkıştırarak algoritmanın daha iyi performans göstermesini sağlar.
- **Özellik Seçimi:** Veri setindeki özellikler arasından en önemli ve anlamlı olanları seçme işlemidir. Özellik seçimi, gereksiz veya tekrarlayan özellikleri

ortadan kaldırarak modelin karmaşıklığını azaltır ve daha az veri ile daha iyi sonuçlar elde etmeyi sağlar. Ayrıca, daha az boyutlu bir veri seti, modelin eğitim ve tahmin sürelerini azaltır.

- **Özellik Dönüşümü:** Var olan özelliklerin dönüştürülerek yeni özelliklerin elde edilmesi işlemidir. Özellik dönüşümü, veri setindeki ilişkileri veya yapıları daha iyi yansıtabilen yeni özelliklerin elde edilmesini sağlar. Örneğin, iki özelliğin kombinasyonu ile yeni bir özellik oluşturulabilir.

Bu ön işleme adımları, makine öğrenimi modellerinin performansını artırmakta, aşırı öğrenmeyi önlemekte, veri setinin güvenilirliğini artırmakta ve sonuçların daha doğru olmasına olanak tanımaktadır. Ayrıca, ön işleme, veri setinin özelliklerini daha uygun hale getirerek, algoritmanın daha hızlı çalışmasını sağlamaktadır (Keykhosravi, Hamednia, Rastegarfar ve Agrell, 2022).

2.3.1.1. Kullanılan yazılım ve teknolojiler

Veri ön işleme sürecinde farklı yazılımlar kullanılabilir. İşlenecek veri türüne ve işlem adımlarına bağlı olarak yazılımlar değişiklik göstermektedir.

- **SPSS(Statistical Package for the Social Sciences):** İlk olarak 1968'de SPSS tarafından başlatılan, ancak 2009'da IBM tarafından satın alınan bir araç olan IBM SPSS, çeşitli araştırmacıların öncelikli olarak istatistiksel ve karmaşık istatistiksel veri analizi için kullandığı bir yazılım paketidir (“SPSS”, 2023).
- **RapidMiner:** Makine öğrenmesi, veri madenciliği, metin madenciliği, tahmin edici analiz ve iş analizi hedefleriyle uyumlu bir yazılım platformu olan RapidMiner, genel olarak iş dünyasında kullanıldığı gibi aynı zamanda araştırma, eğitim, hızlı prototipleme ve uygulama geliştirme gibi çeşitli amaçlar için de kullanılabilir. RapidMiner, veri madenciliği sürecinin her aşamasını desteklediği için veri hazırlama, veri modelleme, doğrulama ve optimizasyon gibi amaçlarla da kullanılmaktadır (“RapidMiner”, 2023).

2.3.2. Özellik seçme algoritmaları

Makine öğrenmesinde özellik seçimi (feature selection), veri setindeki özellikler arasından en önemli ve bilgi taşıyan özellikleri seçmeyi amaçlayan bir süreçtir. Özellik seçimi algoritmaları, bu süreci otomatikleştiren ve optimize eden yöntemlerdir. Özellik seçimi algoritmalarının kullanılmasının en önemli sebepleri arasında, boyut

azaltma, modelin performansını artırma, hesaplama ve bellek maliyetini azaltma, gürültüyü azaltma gibi maddeler yer almaktadır.

Bu nedenlerle, özellik seçimi algoritmaları makine öğrenmesinde yaygın olarak kullanılır. Doğru bir şekilde uygulandığında, veri setinin boyutunu azaltırken modelin performansını artırır ve daha iyi sonuçlar elde etmeyi sağlar (J. Wang, Wang, Nie ve Li, 2023).

2.3.3. Büyük veri teknolojileri

Büyük veri teknolojileri, büyük veri setlerinin yönetimi, depolanması, işlenmesi ve analiz edilmesi için kullanılan özel araçlar ve sistemlerdir. Büyük veri teknolojilerinin temel bileşenleri aşağıda yer almaktadır (Adegoke, 2023).

1. Veri Depolama ve Yönetim Sistemleri: Büyük veri setlerinin etkin bir şekilde depolanması ve yönetilmesi için özel veritabanı sistemleri kullanılır. Bu sistemler, verileri dağıtık bir şekilde depolar ve paralel işleme yetenekleriyle büyük ölçekli veri işleme yapısını destekler. Örneğin, HDFS (Apache Hadoop Distributed File System) ve Apache Cassandra gibi veri depolama sistemleri yaygın olarak kullanılır.
2. Veri İşleme ve Paralel Hesaplama: Büyük veri setlerinin hızlı ve verimli bir şekilde işlenmesi için paralel hesaplama yeteneklerine sahip araçlar kullanılır. Apache Hadoop, büyük veri analitiği için popüler bir açık kaynaklı platformdur. Hadoop MapReduce, büyük veri işleme ve paralel hesaplama modelini uygular. Apache Spark ise dağıtık veri işleme, veri akışı analitiği ve makine öğrenimi için hızlı bir platform olarak kullanılır.
3. Veri Akışı ve Toplama: Büyük veri setlerinin farklı kaynaklardan (sensörler, cihazlar, web trafiği vb.) sürekli olarak toplanması gerekebilir. Bu amaçla, Apache Kafka gibi bir veri akışı sistemi kullanılabilir. Kafka, yüksek hızda veri toplama, depolama ve dağıtma yeteneklerine sahiptir.
4. Veri İşleme ve Analitik Araçları: Büyük veri analitiği için çeşitli araçlar ve yazılımlar bulunmaktadır. Apache Hadoop ekosistemi içinde bulunan Apache Hive ve Apache Pig, büyük veri analitiği için sorgulama ve veri işleme araçlarıdır. Apache Flink ve Apache Storm ise gerçek zamanlı büyük veri analitiği ve akış işleme için kullanılan popüler araçlardır. Ayrıca, Python ve R

gibi programlama dilleri ve ilgili kütüphaneler de büyük veri analitiğinde sıkça kullanılır.

5. Veri Görselleştirme: Büyük veri analizi sonuçlarının etkili bir şekilde sunulması ve görselleştirilmesi önemlidir. Görselleştirme araçları, büyük veri setlerinden gelen verileri grafikler, tablolar ve interaktif görsellerle anlaşılır hale getirir.

Bu büyük veri teknolojileri, veri setlerinin ölçeğini ve karmaşıklığını yönetmek için tasarlanmıştır. Büyük veri analitiği için çeşitli açık kaynaklı ve ticari çözümler bulunmaktadır, ve bu teknolojilerin kombinasyonu, büyük veri projelerinin başarılı bir şekilde yürütülmesini sağlar (Bag, Dhamija, Kumar ve Sabbir, 2023).

2.3.3.1. Apache kafka

Apache Kafka, yüksek veri akış hızlarıyla gerçekleştirilen dağıtık bir veri akışı platformudur. Kafka, ölçeklenebilir, dayanıklı ve olay tabanlı veri akışı için tasarlanmıştır.

Kafka'nın temel amacı, büyük miktardaki veriyi hızlı bir şekilde toplamak, depolamak ve işlemektir. Veri akışını gerçekleştirmek için bir üretici/abone (publisher/subscriber) modeli kullanır. Üreticiler, Kafka'ya veri mesajlarını gönderir ve bu mesajlar bir veya daha fazla abone tarafından tüketilir (Martín, Langendoerfer, Soltani ve Díaz, 2022).

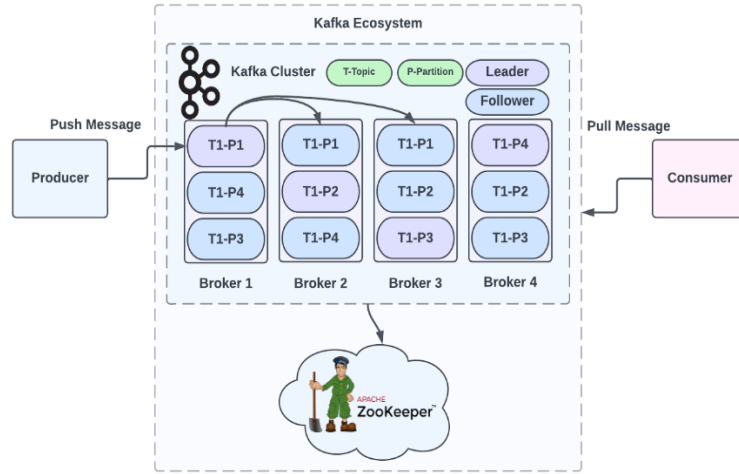
Kafka, aşağıdaki anahtar bileşenlerden oluşur:

1. Broker: Kafka veri akışını sağlayan hizmettir. Birden çok “broker”, bir Kafka kümesi (cluster) oluşturarak yüksek ölçeklenebilirlik ve yüksek kullanılabilirlik sağlar. Her “broker”, mesajları bir veya daha fazla konuda saklar ve iletişim için “ZooKeeper” gibi bir koordinasyon servisiyle etkileşime girer.
2. Topic: Kafka'nın mesajları organize etmek için kullandığı temel birimdir. Bir “topic”, belirli bir konu veya veri türüyle ilişkili mesajların bir araya getirildiği bir kategoriyi temsil eder. Örneğin, "web trafiği", "sensör verileri" veya "işlem günlükleri" gibi konular olabilir.
3. Partition: Her bir “topic”, birden fazla “partition” sekmesine bölünebilir. “Partition” sekmeleri, bir “topic” içindeki verilerin paralel ve dağıtık bir şekilde işlenmesini sağlar. Her “partition”, diskte tutulan bir sıralı günlük log

olarak depolanır ve “offset” yani birincil sıra numarasıyla belirtilen mesajlar içerir.

4. Producer: Üretici (producer), veri mesajlarını Kafka'ya gönderen uygulamalardır. Üreticiler, belirli bir “topic” ve “partition” sekmesine mesajları yayımlar ve veri akışını başlatır.
5. Consumer: Tüketiciler (consumer), Kafka'dan mesajları okuyan ve işleyen uygulamalardır. Tüketiciler, belirli bir “topic” ve “partition” sekmesinden mesajları alır ve bu mesajları kullanarak çeşitli işlemler gerçekleştirir (“Kafka”, 2023).

Apache Kafka'nın genel mimarisi Şekil 2.1.'de yer almaktadır.



Şekil 2.1. Kafka Mimarisi

Kafka'nın avantajlarından bazıları şunlardır:

- Yüksek performans: Kafka, yüksek veri akış hızlarını destekleyerek büyük miktardaki verilerin hızlı bir şekilde işlenmesini sağlar.
- Ölçeklenebilirlik: Kafka, birden fazla broker ve partition kullanarak yüksek ölçeklenebilirlik sağlar. Veri akışını artırmak için kolayca yeni “broker”lar eklemek mümkündür.
- Dayanıklılık: Kafka, veri kaybını önlemek için mesajları diskte tutar ve birbir kopyalama (replication) mekanizmasıyla yüksek kullanılabilirlik sağlar.
- Esnek veri entegrasyonu: Kafka, farklı sistemler ve uygulamalar arasında veri entegrasyonunu kolaylaştırır. Veri akışı için standart bir arabirim sağlar.

Bu özellikleri sayesinde Apache Kafka, gerçek zamanlı veri entegrasyonu, büyük veri analitiği, akış işleme ve diğer birçok senaryoda yaygın olarak kullanılan bir veri akışı platformudur (Ahmed, Younis, Hendawi ve Ali, 2020).

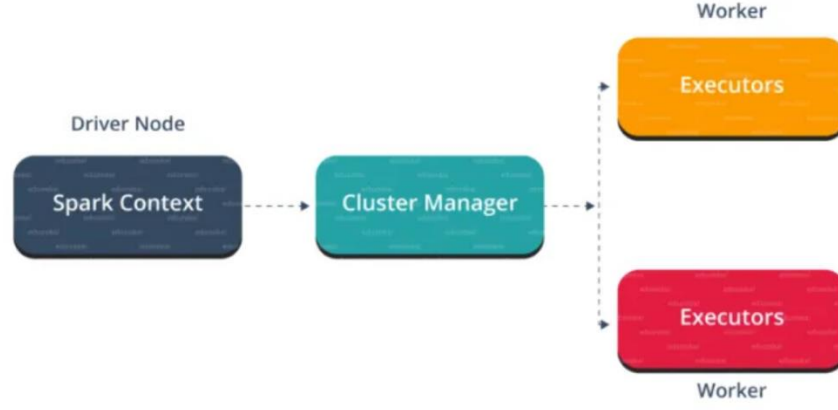
2.3.3.2. Apache spark

Apache Spark, hızlı ve bellek içi bir veri işleme motorudur ve veri çalışanlarının veri kümelerine hızlı yinelemeli erişim gerektiren akış, makine öğrenmesi veya SQL iş yüklerini etkili bir şekilde gerçekleştirebilmeleri için zarif ve etkileyici geliştirme API'lerine (Application Programming Interface) sahiptir (“Apache Spark”, 2023).

Apache Spark, açık kaynak bir küme hesaplama platformudur ve gerçek zamanlı veri işleme için kullanılır. Bellek içi küme hesaplama özelliği, uygulamaların işlem hızını artırmasını sağlar. Spark, veri paralelliği ve hata toleransı ile kümeleri programlamak için bir arayüze sahiptir. Toplu iş uygulamaları, yinelemeli algoritmalar, etkileşimli sorgular ve akış gibi çeşitli iş yüklerini destekleyecek şekilde tasarlanmıştır.

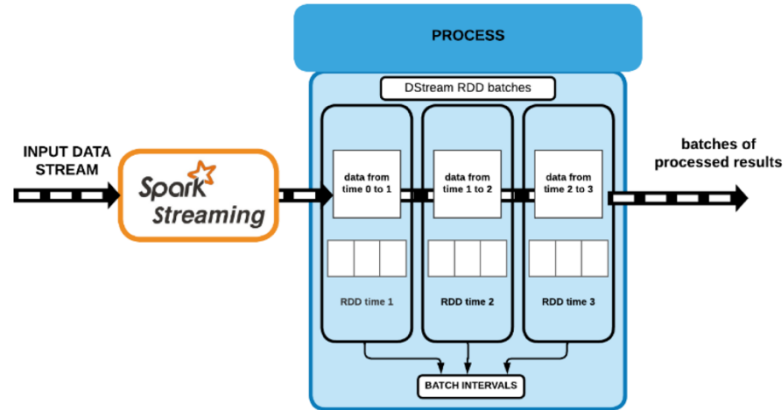
Spark, Hadoop MapReduce'a kıyasla büyük ölçekli veri işleme için kullanıldığında 100 kat daha hızlı çalışan bir platformdur. Bu olağanüstü hız, kontrollü bölümlene yöntemiyle elde edilmektedir. Ayrıca Spark, basit bir programlama katmanıyla güçlü önbellekleme ve disk kalıcılığı özellikleri sunmaktadır. Bellek içi hesaplama sayesinde gerçek zamanlı hesaplama ve düşük gecikme süresi sağlamaktadır. Spark ayrıca Java, Scala, Python ve R gibi dört farklı programlama dili için yüksek seviyeli bir API sunmakta, bu da kullanıcıların tercih ettikleri dili kullanarak Spark'ı kullanmalarını sağlamaktadır (“Spark”, 2023).

Apache Spark, geniş bir ekosisteme sahip olan bir veri işleme ve analitik platformudur. Spark ekosistemi, Spark'ın çeşitli bileşenleri ve ilgili araçlardan oluşmaktadır. Apache Spark'ın genel mimarisi Şekil 2.2.'de yer almaktadır.



Şekil 2.2. Apache Spark Mimarisi (“Spark”, 2023)

- Spark Core: “Spark Core”, Spark projesinin temel motorudur ve diğer tüm bileşenler bu temel üzerine inşa edilir. Cluster ortamında görevlerin dağıtımı, zamanlaması ve hafıza yönetimi gibi işleri yönetir. Spark Core, Spark projelerinin çalışması için gerekli olan temel işlevleri sağlar ve diğer bileşenlerin etkili bir şekilde çalışmasını destekler.
- Spark Streaming: “Spark Streaming”, akış halinde çeşitli kaynaklardan gelen verilerin gerçek zamanlı olarak analiz edilmesini mümkün kılar. “Spark Streaming”, bu verilerin hızlı bir şekilde işlenmesini ve analiz edilmesini sağlayarak, kullanıcılara anlık veri akışlarından değerli bilgiler elde etme imkanı sunar. Bu sayede, gerçek zamanlı veri analizi ve karar alma süreçleri hızlandırılır ve etkili bir şekilde yönetilebilir. “Spark Streaming” genel akış süreci Şekil 2.3.’te resmedilmiştir.

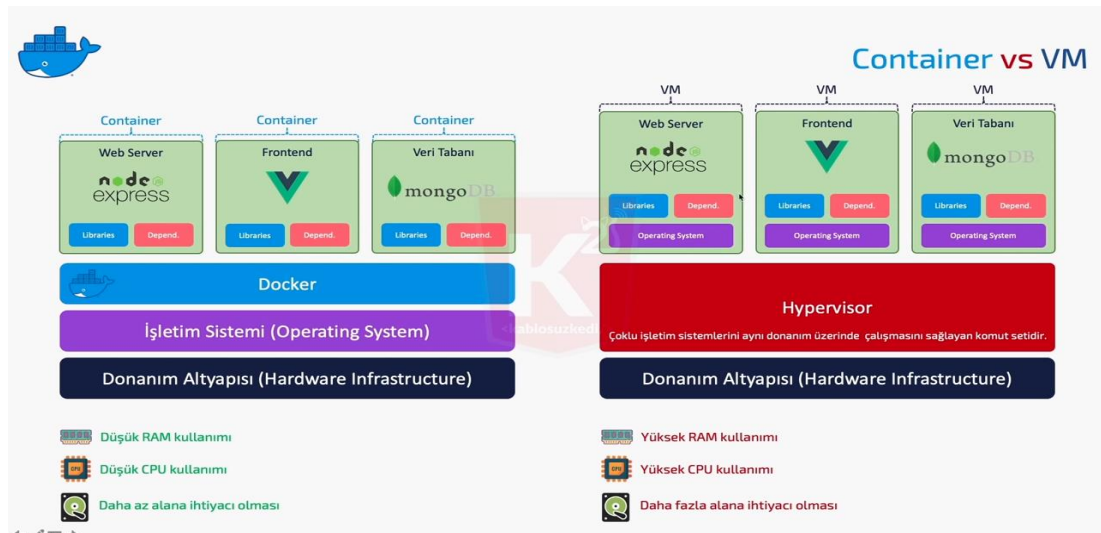


Şekil 2.3. “Spark Streaming” Genel Akış Süreci (Ogur, Ceken, Ogur, Yuvaci, Yazici ve Yazici, 2023)

- Spark SQL: Spark SQL, Spark'a özgü SQL benzeri sorgular oluşturmayı sağlamaktadır. Verilerin JSON, CSV gibi dosya formatlarından alınarak dataframe'lere dönüştürülmesi ve yapısal bir forma getirilmesiyle, Spark SQL üzerinde SQL sorgularının çalıştırılmasına imkan tanımaktadır. Bu sayede, verileri yapısal hale getirerek SQL sorgularıyla analiz etmek ve işlemek mümkün hale gelmektedir. İlişkisel veritabanları için “Spark SQL” performans artırıcı bir çözüm sunmaktadır.
- GraphX: “GraphX”, Spark üzerinde graf tabanlı veri yapılarının oluşturulmasını ve graf algoritmalarının uygulanmasını sağlamaktadır. Bu kütüphane, graf tabanlı veri analizi, ağ trafiği analizi, sosyal ağ analizi gibi çeşitli uygulamalarda kullanılabilir. Spark'ın büyük ölçekli ve dağıtık veri işleme yetenekleriyle birleşen GraphX, karmaşık graf ve ağ analizlerinin verimli bir şekilde gerçekleştirilmesine olanak sağlamaktadır.
- MLLib (Machine Learning): MLLib, Spark MLLib, Apache Spark'ın makine öğrenimi kütüphanesidir. Spark MLLib, büyük ölçekli veri kümeleri üzerinde makine öğrenimi modelleri oluşturmak, eğitmek ve kullanmak için kullanılan bir araç seti sağlar. Spark'ın dağıtık hesaplama yetenekleri sayesinde, MLLib kullanıcılarına paralel ve ölçeklenebilir bir şekilde makine öğrenimi işlemlerini gerçekleştirme imkanı sunar. MLLib, çeşitli makine öğrenimi algoritmalarını ve kullanışlı veri işleme işlevlerini içerir. Sınıflandırma, regresyon, kümeleme, öneri sistemleri, boyut indirgeme ve derecelendirme gibi çeşitli makine öğrenimi görevleri için algoritmalar sunar. Ayrıca, veri ön işleme, özellik çıkarımı, model değerlendirme ve hiperparametre ayarlaması gibi yaygın veri işleme görevlerini de destekler. Spark MLLib, kullanıcılarına kullanımı kolay, ölçeklenebilir ve performanslı bir şekilde makine öğrenimi modelleri oluşturma ve dağıtılmış ortamlarda uygulama imkanı sağlar.
- SparkR: SparkR, Apache Spark ile R programlama dilini entegre eden bir pakettir. SparkR, büyük ölçekli veri analizi ve makine öğrenimi işlemlerini R programlama diliyle gerçekleştirmeyi sağlar. R kullanıcılarına, Spark'ın dağıtık hesaplama yeteneklerini ve performansını kullanarak büyük veri kümeleri üzerinde R kodlarını çalıştırma imkanı sunar. (K. Wang, Fu ve Wang, 2016).

2.3.3.3. Docker

Docker, yazılımların bir konteyner içinde çalıştırılmasını sağlayan açık kaynaklı bir platformdur. Konteynerleme teknolojisi kullanılarak, bir yazılım uygulamasının tüm bağımlılıkları ve çalışma zamanı ortamı bir arada paketlenerek, taşınabilir ve tekrarlanabilir bir şekilde dağıtılmasına imkan sağlar (“Docker”, 2023). Docker, uygulama geliştirme sürecini kolaylaştırır ve hızlandırır. Yazılım geliştiriciler, Docker konteynerlerini kullanarak uygulamalarını kendi çalışma ortamlarından izole edebilir, bağımlılıkları yönetebilir ve uygulamaları farklı bilgisayarlarda sorunsuz bir şekilde çalıştırabilirler (Şekil 2.6). Docker konteynerleme teknolojisi, uygulama geliştirme, dağıtım ve çalıştırma süreçlerini daha kolay, hızlı ve taşınabilir hale getirir. Konteyner ve sanal makine (virtual machine) arasındaki farklara Şekil 2.4.’te yer verilmiştir. Docker kurulumu ve kullanımını gösteren arayüze Şekil 2.5.’te yer verilmiştir.



Şekil 2.4. Konteyner ve Sanal Makine Karşılaştırması (“Konteyner ve Sanal Makine”, 2023)

```
root@nurbanu-N552VW: /home/nurbanu
6a570e943b2a confluentinc/cp-ksqldb-server:6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (137) 17 minutes ago ksqldb-server
88199f1baa4b confluentinc/cp-kafka-rest:6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (1) 17 minutes ago rest-proxy
1846d0b13085 confluentinc/cp-server-connect-datagen:0.4.0-6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (1) 17 minutes ago connect
e8468fe93587 confluentinc/cp-schema-registry:6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (1) 17 minutes ago schema-registry
5b3acae32947 confluentinc/cp-server:6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (143) 18 minutes ago broker
b74fbaf10c17 confluentinc/cp-zookeeper:6.1.0 "/etc/confluent/dock..." 4 hours ago Exited (143) 18 minutes ago zookeeper
0e444c3b5c9d hello-world "hello" 5 hours ago Exited (0) 5 hours ago lucid_nightingale

nurbanu@nurbanu-N552VW:~$ sudo docker run --name postgres -e POSTGRES_PASSWORD=postgres -d postgres:13.2
c1869a5cd3ed2b484693182f945529e43b1b7a07431f1797d5ccd8d05ffc1

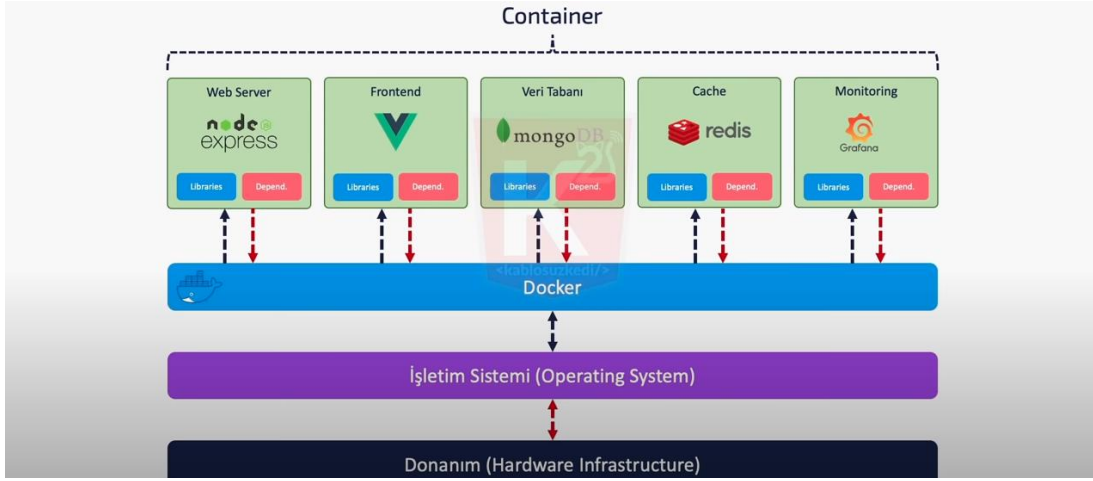
nurbanu@nurbanu-N552VW:~$ sudo docker ps -a
CONTAINER ID        IMAGE               COMMAND                  CREATED              STATUS              PORTS              NAMES
6ee62565aeb3      postgres:13.2     "docker-entrypoint.s..." 0 seconds ago       Up 6 seconds       5432/tcp           postgres
3786d35f48ee      confluentinc/ksqldb-examples:6.1.0 "bash -c 'echo wait..." 4 hours ago        Exited (137) 18 minutes ago ksqldb-ctl
7298982c689e      confluentinc/cp-ksqldb-ctl:6.1.0 "/bin/sh"              4 hours ago        Exited (137) 18 minutes ago control-center
6a570e943b2a      confluentinc/cp-ksqldb-server:6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (137) 17 minutes ago ksqldb-server
88199f1baa4b      confluentinc/cp-kafka-rest:6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (1) 18 minutes ago rest-proxy
1846d0b13085      confluentinc/cp-server-connect-datagen:0.4.0-6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (1) 17 minutes ago connect
e8468fe93587      confluentinc/cp-schema-registry:6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (1) 18 minutes ago schema-registry
5b3acae32947      confluentinc/cp-server:6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (143) 18 minutes ago broker
b74fbaf10c17      confluentinc/cp-zookeeper:6.1.0 "/etc/confluent/dock..." 4 hours ago        Exited (143) 18 minutes ago zookeeper
0e444c3b5c9d      hello-world       "hello"                  5 hours ago        Exited (0) 5 hours ago lucid_nightingale

nurbanu@nurbanu-N552VW:~$ sudo docker ps
CONTAINER ID        IMAGE               COMMAND                  CREATED              STATUS              PORTS              NAMES
c1869a5cd3ed      postgres:13.2     "docker-entrypoint.s..." 12 seconds ago     Up 12 seconds       5432/tcp           postgres

nurbanu@nurbanu-N552VW:~$ sudo docker exec -it postgres bash
root@postgres:/# ./bin/psql -d postgres postgres
psql (13.2 (Debian 13.2-1.pgdg100+1))
Type "help" for help.

postgres=#
```

Şekil 2.5. Docker ile Postgresql Kurulumu



Şekil 2.6. Docker Konteyner Mimarisi (“Konteyner ve Sanal Makine”, 2023)

2.3.3.4. Kubernetes

Mikro servisler, tüm uygulamayı oluşturmak için bir araya gelen bir grup iş olarak düşünülebilir. Bu mikro servisler, iyi işlemleri amaçlanan belirli bir dizi faaliyete sahiptir ve kapsamları sınırlıdır. Gelen istekler, çeşitli özel isteklere bölünür ve daha sonra ayrı süreçler olarak çalışan uygun mikro servislere gönderilir.

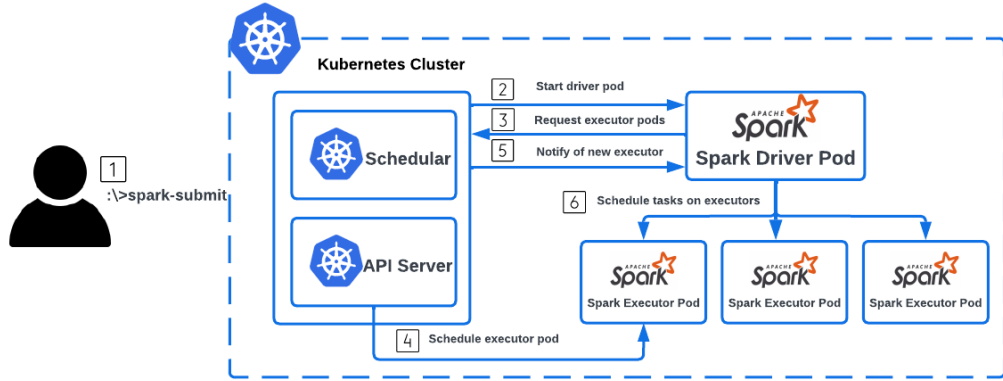
Uygulama, her birini mikro hizmet mimarisinde Docker konteynerlarına yerleştirerek ana bilgisayar ortamından bağımsız hale getirilebilir. Docker, geliştiriciler için uygulamaların konteynerlarda gruplandırılmasını destekler. Docker ile her mikro servis desteklenebilir.

Konteynerlar, sanal makineye göre daha fazla avantajı olan bulut platformları oluşturmak için giderek daha popüler hale gelmektedir. Bir konteyner düzenleme çözümü olan Kubernetes, bilgi işlem kaynaklarının çok sayıda işlem arasında verimli bir şekilde paylaşılmasına izin vererek altyapı kullanımını artırmaktadır (“Kubernetes”, 2023). Bu, bilgi işlem kaynaklarını talebe göre dinamik olarak atayarak yapılmaktadır. Kubernetes, hizmet yükü arttığında ve düştüğünde dinamik kapsayıcı genişletmeyi ve geri dönüşümü otonom olarak uygulayabilmektedir ve donanım kaynaklarının makul bir şekilde tahsis edilmesini sağlamaktadır. Büyüyen eşzamanlı hacmi karşılamak için her mikro servis bir Docker görüntüsü olarak paketlenir ve mevcut kaynakları dinamik olarak genişletmek için Kubernetes kümesi kullanılabilir (Gangula, Jagadeesh Kumar, Anand ve Krishna Prasad, 2021).

2.3.3.5. Kubernetes apache spark

Büyük miktarda veriyi aynı anda işlemek için küme yapısı gereklidir. Apache Spark dört küme yöneticisini destekler: Standalone, Apache Mesos, Hadoop YARN ve Kubernetes (“Apache Spark”, 2023), (Fernández, Gutiérrez-Avilés, Troncoso ve Martínez-Álvarez, 2020).

Kubernetes'te, tüm Spark sürücüsü (driver) ve uygulayıcıları (executor) aslında bölmelerde (pod) çalışır ve sistemin zamanlayıcısı tarafından planlanır. Bir Spark sürücüsü, bir Spark uygulaması bir Kubernetes kümesine gönderildikten sonra ilk olarak bir bölme içinde üretilir ve yürütülür. Sürücü daha sonra bölmelerin içinde de çalıştırılan Spark uygulayıcılarını üretir, onlara bağlanır ve son olarak uygulama kodunu çalıştırır. Uygulayıcı bölmeleri, uygulama tamamlandıktan sonra sonlandırılır ve temizlenir, ancak sürücü bölme, log'ları tutar ve nihayet çöp toplanana veya manuel olarak temizlenene kadar Kubernetes API'sinde "tamamlandı" olarak gösterilmeye devam eder (Manconi, Gnocchi, Milanesi ve Marullo, 2023). Ayrıntılı bilgiye Şekil 2.7'den ulaşılabilmektedir.



Şekil 2.7. Kubernetes Kümesinde Apache Spark Akış Şeması

3. GELİŞTİRİLEN YAPAY ZEKA TABANLI BÜYÜK VERİ İŞLEME SİSTEMİ MİMARİSİ VE KULLANILAN YÖNTEMLER

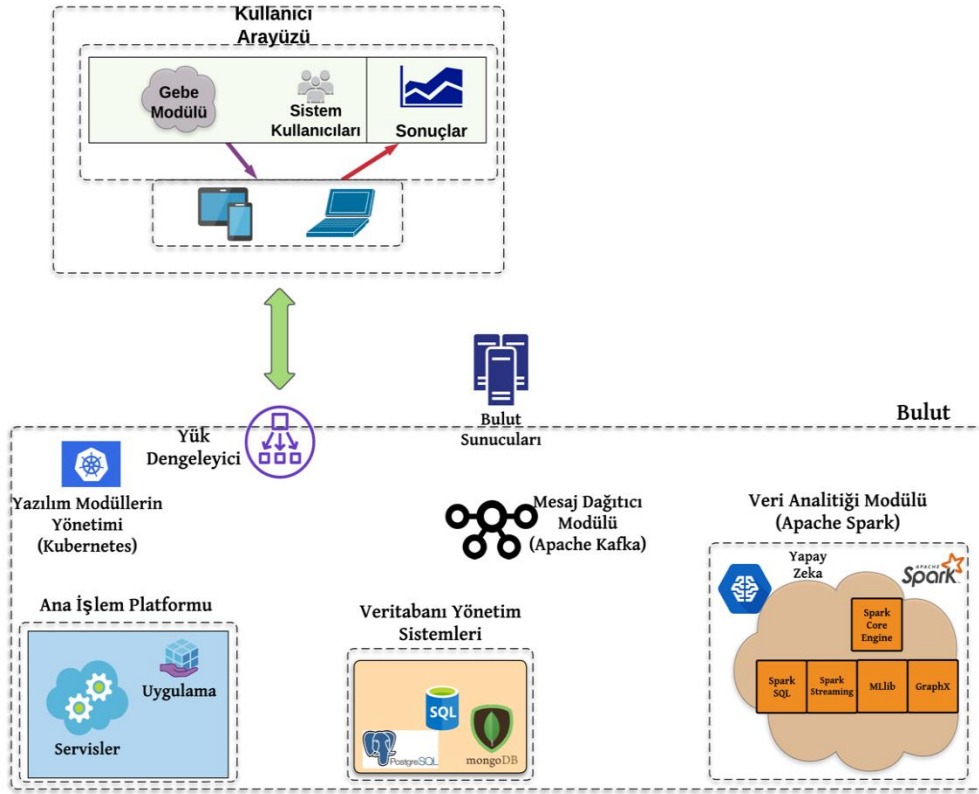
3.1. Problemin Tanımı ve Çözüm Önerisi

Perinatal dönemde anksiyete bozuklukları ile depresyon eş zamanlı veya tek başına seyredebilir (Yücel ve diğerleri, 2013). Perinatal dönemde depresyon ve anksiyete bozukluklarının taranması ve tanınması önemlidir. Bu dönemde tespit edilmez ve tedavi edilmez ise; hem anne hem bebekte olumsuz etkilere, tedavi uyumunda sorunlar yaşanmasına, annelik yaşantısında işlevsel sorunlara, tıbbi hastalıkların kötüleşmesine, kişilerarası ilişkilerin bozulmasına ve ekonomik kayıplara, sigara ve madde kullanımı ile ilgili girişimlere, bebekte gelişimsel sorunlar ve sağlık sorunlarına neden olabilir (Kendig, Keats, Hoffman, Kay, Miller, Simas, Frieder, Hackley, Indman, Raines, Semenuk, Wisner ve Lemieux, y.y.), (Krause, Elrashidi, Halvorsen, McDonald ve Oxentenko, 2017), (Slomian, Honvo, Emonts, Reginster ve Bruyère, 2019). Perinatal dönemde anksiyete ve depresyon hastalığı, bu gibi sebeplerden ötürü toplumda ciddi zararlara neden olabilen bir durumdur. Bu nedenle, hastalığın erken teşhis edilmesi ve tedavi edilmesi büyük önem taşımaktadır. Büyük veri analitiği sistemleri, bu problemleri çözümlenebilmekte kullanılabilmektedir.

Büyük veri analitiği ve ona bağlı teknolojilerin kullanılması ile perinatal dönemde ruhsal bozuklukların erken teşhisi söz konusu olacak ve tedavisi aşamasında önemli ölçüde yol katedilebilecektir.

3.2. Yöntem ve Bulgular

Çalışmamız, Perinatal dönemdeki kadınların anksiyete ve depresyon teşhisinde anlık olarak sonuç üretebilmekte, bunu ölçeklenebilir mimarinin de entegre edilebildiği büyük veri analitiği platformunda gerçekleştirmektedir. Bu altyapıyı oluştururken Apache Kafka, Apache Spark, Kubernetes gibi teknolojilerden faydalanılmıştır. Böylece perinatal dönemdeki kadınlarda depresyon ve anksiyetenin yol açabileceği potansiyel zararın en aza indirilmesi amaçlanmıştır.



Şekil 3.1. Ana İşlem Platformu Diyagramı

Çalışmamızda ilk aşamada, hastaya dair her girilen verinin uygun bir şekilde depolanması sağlanmıştır. Depolanan bu veriler, yapay zekâ algoritmasının kullanılabilmesi için uygun hale getirilmiştir. Öncelikle verinin temizlenmesi, eksik olan verilerin tamamlanması ve türlerin uyumsuzluğunun giderilmesi gibi ön-işlem süreçlerinde geçirilmiş, sonrasında, özellik seçimi algoritmaları kullanılmıştır. Özellik seçimi için ise ileri beslemeli teknik (Forward Selection) performans bakımından en yüksek sonucu ortaya çıkarmıştır. Bu sayede hem geliştirilecek algoritmaların zaman ve bellek karmaşıklığının azaltılması, hem de aşırı öğrenme (overfitting) ve doğrusal bağlantı (multicollinearity) gibi sorunların üstesinden gelinmesi amaçlanmıştır (Şekil 3.1).

Tüm bu işlemlerden sonra, probleme uygun yapay zekâ algoritmaları kullanılarak, kabul edilebilir hassasiyete sahip sonuçları üreten modeller oluşturulmuştur. Geliştirilecek modellerin doğrulanmasında 10-Fold cross validation tekniğinden faydalanılmıştır. Böylelikle anksiyete veya depresyon psikolojik rahatsızlığı içinde olan hamile kadınları belirlemek için, özellik seçme teknikleri ve sınıflandırma makine öğrenimi algoritmalarını içeren hibrit makine öğrenimi teknikleri kullanılmıştır. Bu

işlemler sonunda elde edilen işlenebilir veriseti ve optimum sonucu veren özellik seçme algoritması ve makine öğrenmesi algoritması belirlenmiştir. Hibrit metodolojilerin optimum sonuçlarına göre büyük veri mimarisi geliştirilmiştir.

Bu çalışmadan elde edilen optimum veriseti ile ölçeklenebilir mimariye sahip anlık olarak çıktı üreten büyük veri analitiği inşaa edilmiş ve işlevselliği değerlendirilmiştir. Bu sistemin, daha büyük veriler için bir platform ve altyapı görevi görme yeteneğine sahip olduğunu söylemek mümkündür.

3.2.1. Veri tanıtımı

Bu çalışmada kullanılan veriler, Sakarya Üniversitesi Tıp Fakültesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu kararı (E 71522473 050.01.04-112781-38) ile alınmıştır ve Sakarya Üniversitesi Eğitim ve Araştırma Hastanesi Kadın Doğum Polikliniği'ne başvuran gebelerden toplanmıştır.

Çalışmaya katılmayı kabul eden ve formları dolduran iki yüz elli kadın hedefine ulaşmadan önce üç yüz doksan beş kişiye sorulması gerekmiştir. Gebelerden SDVF (Sosyo-demografik Veri Formu), PASS-TR (Perinatal Anksiyete Semptom Ölçeği) (Yazıcı, Pek, Yuvacı, Köse, Cevrioglu, Yazıcı, Çilli, Erol ve Aydın, 2019) ve EDDÖ (Edinburgh Doğum Sonrası Depresyon Ölçeği) (Cox, Holden ve Sagovsky, 1987) testlerini doldurmaları istenmiştir. SDVF, bu çalışma için uzman doktorlar tarafından oluşturulmuş sorulardan oluşmaktadır. PASS-TR ve EDDÖ sırasıyla depresyon ve anksiyeteyi tespit eden ölçeklerdir.

Bir hasta, etiketleme için kullanılan "etiket" sütununa göre anksiyete veya depresyona sahip olarak sınıflandırılmıştır. Veri seti oluşturulurken sırasıyla EDDÖ ve PASS-TR çıktıları etiket değeri olarak alınmıştır.

Depresyon düzeyi EDDÖ ölçeği kullanılarak değerlendirilmiştir. 13'ün altında puan, anlamlı olmayan depresyon olarak kabul edilmiştir. 13'ten düşük bir puan, önemli olmayan depresyon ("Depresyon Yok" anlamına gelen "0" ile etiketlendi), aksi takdirde depresyon ("Depresyon" anlamına gelen "1" ile etiketlendi) olarak kabul edilmiştir.

Anksiyete derecesini ölçmek için PASS-TR ölçeği kullanılmıştır. 17'den düşük puan, önemli olmayan kaygı olarak kabul edilmiştir ("Anksiyete Yok" anlamına gelen "0" ile etiketlendi); aksi takdirde önemli anksiyete bozukluğundan muzdarip ("Anksiyete

" anlamına gelen "1" ile etiketlenmiştir) olarak kabul edilmiştir (Sau ve Bhakta, 2019b).

Bir kişi, EDDÖ ve PASS-TR puanlarına göre " Anksiyete Yok-Depresyon Yok", " Anksiyete Yok-Depresyon", " Anksiyete-Depresyon Yok" veya " Anksiyete-Depresyon" olarak sınıflandırılabilir. "Anksiyete Yok-Depresyon Yok" dışındaki tüm gruplar, ruh sağlığı bozukluklarının bir psikiyatrist tarafından tedavi edilmesini gerektirir. Bu araştırmanın amacı makine öğreniminin uygulanabilirliğini ve uygunluğunu değerlendirmek olduğundan, tüm çıktıları iki etikette gruplandırmaya karar verilmiştir: "Anksiyete Yok-Depresyon Yok (0)" ve " Anksiyete ve/veya Depresyon yani (1)". Grubun "Anksiyete ve/veya Depresyon" üyeleri, anksiyete, depresyon veya her ikisini yaşıyor olabilecekleri için tanısız bir değerlendirme ve uygun yönetim için bir psikiyatriste sevk edilecektir. Bu durumu daha iyi ifade edebilmek adına Tablo 3.1 eklenmiştir.

Tablo 3.1. Sınıflandırma Etiketi.

Anksiyete(PASS)	Depression (EDDÖ)	Label
0	0	0
1	0	1
0	1	1
1	1	1

Perinatal gebelik veri seti 250 örnek, 60 özellik ve 1 sınıf etiketi içermektedir. Sınıf etiketinin iki değeri vardır: "Anksiyete Yok-Depresyon Yok (0)" ve " Anksiyete ve/veya Depresyon (1)".

3.2.1.1. Çalışmada kullanılan ölçekler

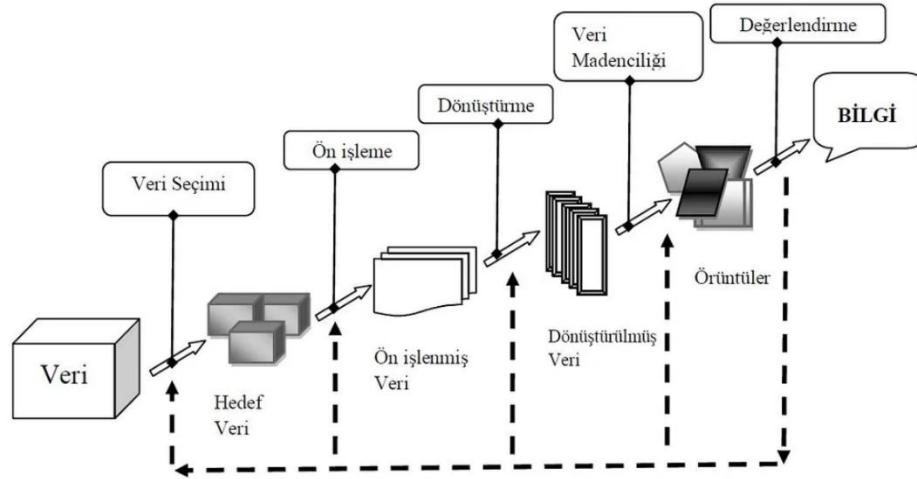
- Sosyo-demografik veri formu (SDVF) gebenin yaşı, medeni durumu, fiziksel aktivite düzeyi, eğitim yılı, çalışma durumu, hanenin toplam gelir düzeyi, psikolojik öyküsü, ilk gebeliğinin yol açtığı psikolojik travmaları, eşinin ve ailesinin psikolojik öyküsü gibi kişiyi temel anlamda tanımada önemli soruları barındıran formdur.

- PASS-TR, mevcut anksiyete belirtilerinin varlığını ve yoğunluğunu değerlendirerek anksiyete düzeylerini ölçen, evrensel olarak test edilmiş ve onaylanmış ölçektir.
- EDDÖ ise depresif semptomların şiddetini ölçmek için kullanılan ölçektir. Ölçek alanında uzman kişiler tarafından bir çok kez test edilmiş ve geçerli, güvenilir olduğu gösterilmiştir.

Tüm veriler anksiyete ve depresyon tanısında kullanılmıştır. Tüm klinik değerlendirmeler deneyimli ve eğitilmiş psikiyatristler tarafından çalışma esas alınarak yapılmıştır.

3.2.2. Veri ön işlem aşaması

Sakarya Üniversitesi Eğitim Araştırma Hastanesine gelen gerçek hastalardan hastalıklarına dair sorular içeren ölçek ve formlar vasıtasıyla tüm gerekli bilgileri toplanmıştır. Bu verilerin bazı ön işlem aşamalarından geçmesi makine öğrenmesi modelinin doğruluğu için oldukça önem arz etmektedir. Bu sebeple toplanan veriseti veri kategorikleştirme, eksik veri tamamlama, normalizasyon gibi bazı ön işlem aşamalarından geçirilmiş ve temiz veri elde edilmiştir (Şekil 3.2).



Şekil 3.2. Bilgi Keşfi Süreci (‘Keşif Süreci’, 2023)

Yapılan çalışmada oluşturulan veriseti için bağımlı ve bağımsız değişkenler aşağıdaki gibi tanımlanmıştır.

- Bağımlı Değişkenler

Edinburgh Doğum sonrası Depresyon Ölçeği puanı ve Perinatal Anksiyete Tarama Ölçeği puanı bağımlı değişkenler olarak kullanılmışlardır.

Sınıf etiketinin iki deęeri vardır: "Anksiyete Yok-Depresyon Yok (0)" ve " Anksiyete ve/veya Depresyon (1)".

- Baęımsız Deęiřkenler

Baęımsız deęiřkenler olarak kullanılan özellikler ise ařaęıdaki gibidir;

yař, medeni durum, fiziksel aktivite düzeyi, eęitim yılı, alıřma durumu, hane halkının toplam gelir düzeyi, psikolojik öykü, ilk gebelięinin neden olduęu psikolojik travmalar, eřinin psikolojik öyküsü gibi temel soruları ieren SDVF ile deęerlendirilmiřtir.

3.2.2.1. Risk faktörü belirleme

Toplanan verilerde hastalıęın teřhis edilebilmesi iin gebelerden 60 tane soru cevaplamařı istenmiřtir. Bu sorular özellikle kiři sayısı arttıka veride büyüme ve sistemi hız anlamında yavařlatmaya olduęa elveriřlidir. Bu sebeple verisetinde sonucu en fazla etkileyecek soruları bulmak olduęa önem arz etmektedir.

Yapılan literatür taraması neticesinde birden fazla yöntemle karřılařılmıřtır. Bunlardan ilki verilerde korelasyon incelemesidir.

Elde edilen verilerin ölekleri üzerinde korelasyon iliřkisi incelenmiřtir. Bunun iin öncelikle öleklerin deęerleri arasında normal daęılımın olup olmadıęı arařtırılmıřtır. Veriler Kolmogorov Smirnov testi ve Shapiro-Wilk sınamasına tabii tutulmuřtur. Bu kapsamda p deęerinin "0" ve ok yakın ıkması dahilinde normal daęılıma sahip olmadıęı kanısına varılabilmektedir. Bu gibi durumlarda korelasyon iliřkisi iin non-parametrik olan testlerin (Spearman, Pearson Korelasyon) uygulanması gerekmektedir. řekil 3.3'te iki öleęin normal daęılım sonuçları bulunmaktadır.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
EPDS1	.441	33	.000	.524	33	.000
EPDS10	.513	33	.000	.393	33	.000
EPDS11	.381	33	.000	.680	33	.000
EPDS2	.510	33	.000	.431	33	.000
EPDS3	.256	33	.000	.807	33	.000
EPDS4	.288	33	.000	.767	33	.000
EPDS5	.243	33	.000	.835	33	.000
EPDS6	.310	33	.000	.774	33	.000
EPDS7	.420	33	.000	.638	33	.000
EPDS8	.258	33	.000	.761	33	.000
EPDS9	.324	33	.000	.706	33	.000
TOPLAMEPDS	.098	33	.200*	.943	33	.082

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PASS1	.325	31	.000	.773	31	.000
PASS10	.376	31	.000	.697	31	.000
PASS11	.299	31	.000	.755	31	.000
PASS12	.284	31	.000	.839	31	.000
PASS13	.257	31	.000	.851	31	.001
PASS14	.324	31	.000	.707	31	.000
PASS15	.309	31	.000	.745	31	.000
PASS16	.429	31	.000	.591	31	.000
PASS17	.284	31	.000	.775	31	.000
PASS18	.361	31	.000	.711	31	.000
PASS19	.466	31	.000	.550	31	.000
PASS2	.265	31	.000	.788	31	.000
PASS20	.482	31	.000	.509	31	.000
PASS21	.325	31	.000	.730	31	.000
PASS22	.466	31	.000	.550	31	.000
PASS23	.241	31	.000	.807	31	.000
PASS24	.506	31	.000	.445	31	.000
PASS25	.386	31	.000	.685	31	.000
PASS26	.431	31	.000	.540	31	.000
PASS27	.434	31	.000	.614	31	.000
PASS28	.412	31	.000	.607	31	.000
PASS29	.492	31	.000	.485	31	.000
PASS3	.302	31	.000	.765	31	.000
PASS30	.301	31	.000	.752	31	.000
PASS31	.401	31	.000	.661	31	.000
PASS4	.288	31	.000	.729	31	.000
PASS5	.350	31	.000	.704	31	.000
PASS6	.401	31	.000	.661	31	.000
PASS7	.357	31	.000	.718	31	.000
PASS8	.326	31	.000	.736	31	.000
PASS9	.401	31	.000	.661	31	.000

Şekil 3.3. EDDÖ ve PASS Ölçeklerinin Normallik Testleri

Uygulanan normal dağılım testlerinde iki ölçeğin de normal dağılıma sahip olmadığı gözlemlenmiştir. Bu sebeple non-parametrik test olan Spearman Korelasyonu, ölçeklere uygulanmaya uygun görülmüştür. Korelasyonun 1'e yaklaşması aralarındaki ilişkinin güçlü olduğunu ve değerlerin artmasının sonucun da artmasına sebep olacağını gösterirken, negatif değer için ise tam tersi söz konusudur. Şekil 3.4 ve Şekil 3.5.'te sırasıyla PASS-TR ve EDDÖ verilerinin korelasyon performans çıktıları yer almaktadır. Şekil 3.6'da ise farklı uygulamalarda gerçekleştirilen korelasyon çıktıları karşılaştırılmaktadır.

		PASS1	PASS10	PASS11	PASS12	PASS13	PASS14	PASS15	PASS16	PASS17	PASS18	PASS19	PASS2	PASS20	
Spearman's rho	PASS1	Correlation Coefficient	1.000	.245	.332	-.024	-.088	-.093	.001	.022	.137	-.040	.600**	.078	
		Sig. (2-tailed)	.	.185	.068	.898	.639	.620	.997	.905	.926	.461	.830	.000	.676
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS10	PASS10	Correlation Coefficient	.245	1.000	.397*	.301	.148	.265	.350	.197	.169	.147	.268	.280	.139
		Sig. (2-tailed)	.185	.	.027	.100	.426	.150	.053	.289	.362	.431	.144	.127	.457
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS11	PASS11	Correlation Coefficient	.332	.397*	1.000	.042	.062	.332	.062	.177	.141	.385*	.266	.428*	.414*
		Sig. (2-tailed)	.068	.027	.	.821	.621	.068	.741	.341	.449	.032	.148	.016	.021
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS12	PASS12	Correlation Coefficient	-.024	.301	.042	1.000	.661**	.381*	.438*	.029	.224	.181	.059	.098	.012
		Sig. (2-tailed)	.898	.100	.821	.	.000	.034	.014	.878	.226	.330	.752	.601	.948
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS13	PASS13	Correlation Coefficient	-.088	.148	.062	.661**	1.000	.536**	.403*	.247	.386*	.232	.128	.145	.034
		Sig. (2-tailed)	.639	.426	.742	.000	.	.002	.025	.180	.032	.209	.492	.437	.857
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS14	PASS14	Correlation Coefficient	-.093	.265	.332	.381*	.536**	1.000	.438*	.468**	.401*	.374*	.040	.225	.253
		Sig. (2-tailed)	.620	.150	.068	.034	.002	.	.014	.008	.026	.038	.830	.223	.170
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS15	PASS15	Correlation Coefficient	.001	.350	.062	.438*	.403*	.438*	1.000	.506**	.540**	.560**	.260	.080	.206
		Sig. (2-tailed)	.997	.053	.741	.014	.025	.014	.	.004	.002	.001	.158	.670	.265
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS16	PASS16	Correlation Coefficient	.022	.197	.177	.029	.247	.468**	.506**	1.000	.495**	.323	.265	.159	.337
		Sig. (2-tailed)	.905	.289	.341	.878	.180	.008	.004	.	.005	.076	.149	.393	.064
		N	31	31	31	31	31	31	31	31	31	31	31	31	31
PASS17	PASS17	Correlation Coefficient	.017	.169	.141	.224	.386*	.401*	.540**	.495**	1.000	.618**	.457**	.385**	.480**
		Sig. (2-tailed)	.926	.362	.449	.226	.032	.026	.002	.005	.	.000	.010	.032	.006
		N	31	31	31	31	31	31	31	31	31	31	31	31	31

Şekil 3.4. PASS Verilerinin Korelasyon Performansı

			EPDS1	EPDS10	EPDS11	EPDS2	EPDS3	EPDS4	EPDS5	EPDS6	EPDS7	EPDS8	EPDS9	TOPLAMEPDS
Spearman's rho	EPDS1	Correlation Coefficient	1.000	.196	.389*	.143	.223	.295	.288	.592**	.448*	.520*	.501**	.637**
		Sig. (2-tailed)		.274	.025	.428	.212	.096	.104	.000	.009	.002	.003	.000
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS10	Correlation Coefficient	-.196	1.000	.505**	.094	-.074	-.076	.001	.155	.371*	.171	.439*	.255
		Sig. (2-tailed)	.274		.003	.603	.684	.674	.997	.389	.033	.342	.011	.153
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS11	Correlation Coefficient	.389*	.505**	1.000	-.005	.364*	.287	.406*	.104	.417*	.218	.538**	.633**
		Sig. (2-tailed)	.025	.003		.977	.037	.106	.019	.563	.016	.223	.001	.000
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS2	Correlation Coefficient	-.143	.094	-.005	1.000	.397*	.429*	.140	.355*	-.044	.435*	.398*	.414*
		Sig. (2-tailed)	.428	.603	.977		.022	.013	.437	.043	.809	.011	.022	.017
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS3	Correlation Coefficient	.223	-.074	.364*	.397*	1.000	.671**	.503*	.436*	.174	.374*	.349*	.710**
		Sig. (2-tailed)	.212	.684	.037	.022		.000	.003	.011	.332	.032	.046	.000
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS4	Correlation Coefficient	.295	-.076	.287	.429*	.671**	1.000	.658**	.370*	.251	.350*	.491**	.770**
		Sig. (2-tailed)	.096	.674	.106	.013	.000		.000	.034	.158	.046	.004	.000
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS5	Correlation Coefficient	.288	.001	.406*	.140	.503*	.658**	1.000	.075	.221	.184	.543**	.707**
		Sig. (2-tailed)	.104	.997	.019	.437	.003	.000		.678	.215	.306	.001	.000
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS6	Correlation Coefficient	.592**	.155	.104	.355*	.436*	.370*	.075	1.000	.219	.458**	.284	.556**
		Sig. (2-tailed)	.000	.389	.563	.043	.011	.034	.678		.221	.007	.109	.001
		N	33	33	33	33	33	33	33	33	33	33	33	33
	EPDS7	Correlation Coefficient	.448**	.371*	.417*	-.044	.174	.251	.221	.219	1.000	.280	.239	.478**
		Sig. (2-tailed)	.009	.033	.016	.809	.332	.158	.215	.221		.115	.180	.005
		N	33	33	33	33	33	33	33	33	33	33	33	33

Şekil 3.5. EDDÖ Verilerinin Korelasyon Performansı

The screenshot shows a software interface for a correlation matrix. The main window displays a table with columns for variables: EPDS, SDWF3-1, SDWF3-2, SDWF3-3, SDWF3-4, SDWF3-5, SDWF3-6, SDWF3-7, SDWF3-8, SDWF3-9, and SDWF3-10. The table contains correlation coefficients, significance levels (Sig. (2-tailed)), and sample sizes (N). The interface also includes a 'Correlation Matrix (Correlation Matrix)' window and an 'ExampleSet (Select Attributes)' window.

Şekil 3.6. Korelasyon Performans Karşılaştırması

Korelasyon verilerin arasındaki ilişkiyi bulmakta kullanılan bir yöntemdir fakat ölçeğin sonucunu en fazla etkileyen özelliği bulmadan kullanılan yöntemler arasında başlıca yer almamaktadır. Her bir kronik hastalığa yönelik sosyal, ekonomik, demografik, coğrafi vb. tüm değişkenler dikkate alınarak, hastalığın ortaya çıkışında etkisi olan değişkenlerin PCA (Principal Component Analysis), faktör analizi (Factor Analysis), lojistik regresyon, özellik çıkarım (Feature Extraction) ve değişken önemliliği belirleme yöntemi ile belirlenmesi mümkündür.

PCA: PCA (Principal Component Analysis), veri setindeki değişkenlerin boyutunu azaltmak için kullanılan bir istatistiksel yöntemdir. Temel amacı, veri setindeki değişkenler arasındaki ilişkileri anlamak ve daha az sayıda yeni değişken (bileşen)

oluşturarak verinin temsilini sağlamaktır. PCA aynı zamanda veri görselleştirme, veri sıkıştırma, veri ön işleme ve desen tanıma gibi birçok alanda da kullanılır. Özellikle makine öğrenimi ve veri madenciliği gibi disiplinlerde veri analizinde yaygın bir yöntemdir.

LDA: LDA (Linear Discriminant Analysis), bir sınıflandırma ve boyut indirgeme yöntemidir. LDA, veri setindeki sınıflar arasındaki farkları en üst düzeye çıkarmak ve sınıfları birbirinden ayıran en iyi diskriminant fonksiyonlarına dayalı bir alt uzay bulmak için kullanılır. LDA, sınıflandırma problemlerinde kullanılarak, veri setindeki değişken sayısını azaltırken sınıflar arasındaki ayrımı en üst düzeye çıkarmayı hedefler. Bu şekilde, veri setindeki karmaşıklığı azaltır ve sınıflandırma performansını artırır.

PCA ve LDA Arasındaki Farklar

- PCA gözetimsiz bir öğrenme algoritmasıdır ve veri setindeki değişkenler arasındaki ilişkileri anlamak için kullanılır. LDA ise gözetimli bir öğrenme algoritmasıdır ve sınıflandırma problemlerinde sınıflar arasındaki ayrımı maksimize etmek için kullanılır.
- PCA, sınıflandırma yapma amacıyla doğrudan kullanılmaz. PCA, veri setindeki örüntüleri ve ilişkileri anlamak için kullanılır. LDA ise, sınıflandırma problemlerinde kullanılmak üzere tasarlanmış bir yöntemdir. Sınıfları birbirinden ayıran diskriminant fonksiyonları oluşturarak sınıflandırma yapar.
- PCA, tüm veri setini dikkate alır ve veri setindeki değişkenler arasındaki kovaryans matrisine dayanır. LDA ise, sınıflar arasındaki farkı en üst düzeye çıkarmak için sınıf ortalamaları ve sınıfların dağılımını dikkate alır.
- PCA boyut azaltma ve veri analizi için kullanılırken, LDA sınıflandırma problemlerinde sınıflar arasındaki ayrımı maksimize etmek için kullanılır.
- PCA, veri setindeki değişkenler arasındaki ilişkileri anlamak için genel bir yöntemken, LDA sınıflandırma odaklı ve sınıflar arasındaki ayrımı önemseyen bir yöntemdir (“PCA-LDA”, 2023).

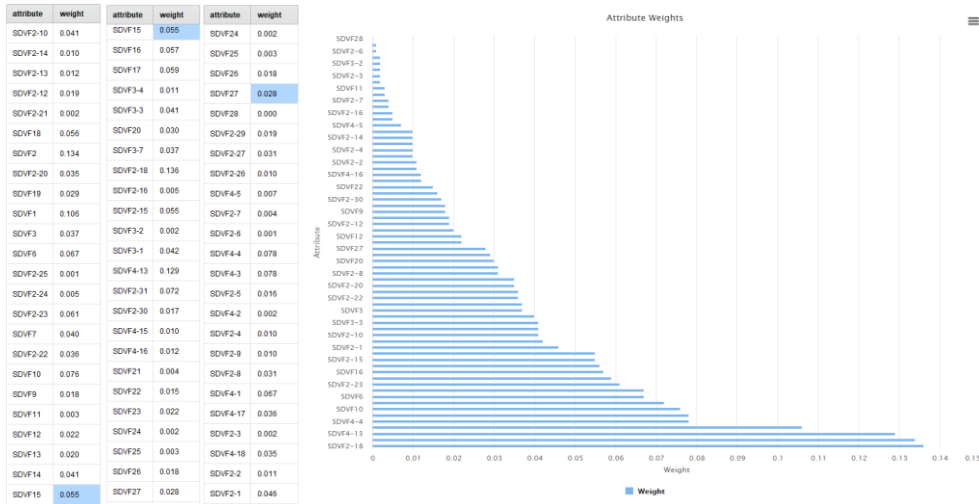
Şekil 3.7’de EDDÖ-SDVF verisetinde PCA kullanımının çıktıları yer almaktadır.

Row No.	pc_1	pc_2	pc_3	pc_4	pc_5	pc_6	pc_7
1	3.303	-0.012	0.739	-0.100	-1.244	0.153	0.753
2	-3.189	-0.361	-4.382	-1.823	-0.512	-0.918	0.736
3	4.410	0.187	1.522	0.159	0.115	-1.065	0.412
4	-1.960	-0.311	-0.053	-0.156	0.717	0.515	-0.783
5	-4.701	5.496	-0.172	-0.418	1.382	0.419	-0.609
6	-1.092	-0.299	-2.702	-0.098	-0.332	-1.910	-0.617
7	1.158	0.041	-0.210	0.150	1.456	0.121	0.813
8	0.075	-0.070	-1.808	4.604	-0.804	0.377	-0.695
9	-15.464	-1.898	1.460	-0.109	-0.702	0.488	0.410
10	0.373	-0.311	1.384	-0.531	-2.155	0.411	0.697
11	7.501	0.558	0.366	-0.108	-0.412	-0.119	-0.117
12	4.423	-10.810	0.204	-0.062	1.297	-0.054	-0.027
13	2.204	0.008	0.565	0.597	0.872	-0.880	0.456
14	-4.113	1.312	0.659	-0.338	-0.385	0.668	0.114
15	-5.030	-0.826	-0.018	-1.337	-1.109	-1.065	-0.995
16	7.501	0.558	0.366	-0.108	-0.412	-0.119	-0.117
17	5.351	0.260	0.627	-0.313	0.124	-0.711	-0.282
18	2.212	0.261	-0.997	-0.622	0.158	1.626	-1.059
19	7.501	0.558	0.366	-0.108	-0.412	-0.119	-0.117
20	-0.923	-0.077	-0.967	-0.292	1.066	1.228	0.137
21	4.341	0.139	0.596	-0.002	-0.687	-0.035	0.154
22	-1.952	-0.277	-0.769	-0.509	0.056	1.318	1.564
23	1.487	7.090	0.227	0.326	0.515	-0.553	0.817
24	-8.427	0.798	1.605	0.072	0.849	-0.930	-1.260

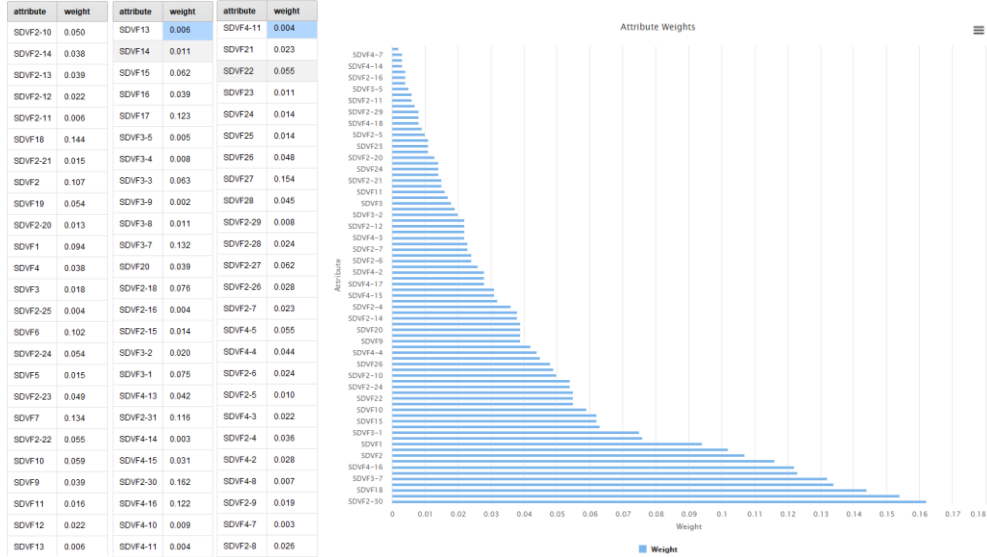
ExampleSet (32 examples, 0 special attributes, 7 regular attributes)

Şekil 3.7. EDDÖ-SDVF Veri Setinde PCA Kullanımı

Değişken Önemliliği: Veriler üzerinde yapılan işlemlerde değişken önemliliği, bir değişkenin analiz veya modelleme sürecinde ne kadar etkili olduğunu ifade eder. Bu, değişkenin sonuçları ne ölçüde açıkladığı, tahmin gücü, ilişki derecesi veya modelin performansı gibi faktörlere dayanır. Değişkenlerin önemliliği, analiz veya modelleme sürecinde belirli bir amaca hizmet eden değişkenleri tanımlamayı ve vurgulamayı amaçlar. Şekil 3.8 ve Şekil 3.9’da sırasıyla EDDÖ-SDVF ve PASS-TR-SDVF veriseti üzerinde değişken önemliliği ağırlık dağılım grafiği yer almaktadır.



Şekil 3.8. EDDÖ -SDVF Değişken Önemliliği Ağırlık Dağılım Grafiği



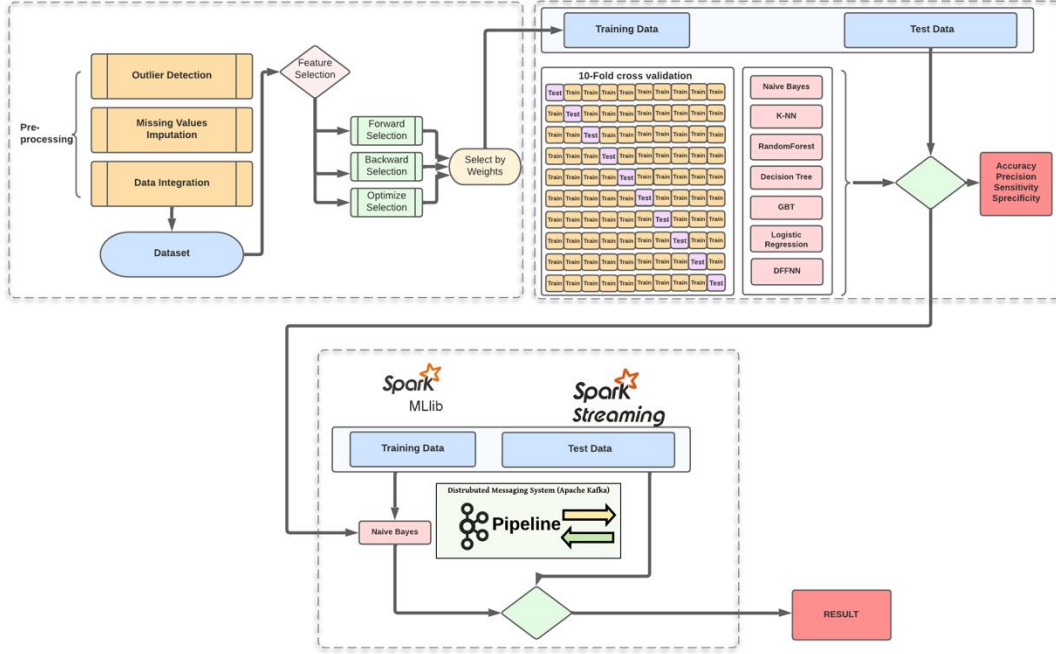
Şekil 3.9. PASS-TR-SDVF Değişken Önemliliği Ağırlık Dağılım Grafiği

Özellik seçimi: Tahmine dayalı bir model geliştirirken girdi değişkenlerinin sayısını azaltma işlemidir. Fazla değişkenlerin eklenmesi, modelin genelleme kapasitesini azaltır ve ayrıca bir sınıflandırıcının genel doğruluğunu azaltabilir. Hem modellemenin hesaplama maliyetini azaltmak hem de bazı durumlarda modelin performansını iyileştirmek için girdi değişkenlerinin sayısını azaltmak gereklidir. Bu sebeple eldeki hasta verilerine Evrimsel strateji ile optimize eden özellik seçme yöntemi uygulanmıştır. Elde edilen sonuçlar Şekil 3.10’de gösterilmektedir.

attribute	weight	Row No.	EPDS	SDVF3-4	SDVF3-5	SDVF3-6	SDVF3-7
		1	0	2	3	3	2
SDVF3-1	0	2	0	1	1	1	2
		3	0	2	3	3	2
SDVF3-2	0	4	0	3	3	3	4
		5	0	2	3	3	2
SDVF3-3	0	6	0	2	3	3	2
		7	0	2	3	3	1
		8	0	1	2	2	2
SDVF3-4	1	9	0	3	3	3	4
		10	0	2	3	3	2
SDVF3-5	1	11	0	3	3	3	4
		12	0	3	3	3	4
SDVF3-6	1	13	1	2	3	3	2
		14	0	2	3	3	2
SDVF3-7	1	15	0	2	3	3	2
		16	0	2	3	3	1
SDVF3-8	0	17	0	2	3	3	2
		18	0	3	3	3	4
SDVF3-9	0	19	0	2	3	3	1
		20	0	2	3	3	2
SDVF3-10	0	21	0	2	3	3	2
		22	0	2	3	3	1
		23	0	2	3	3	2
		24	0	2	3	3	2

ExampleSet (41 examples, 1 special attribute, 4 regular attributes)

Şekil 3.10. EDDÖ-SDVF Özellik Seçme Sonuçları

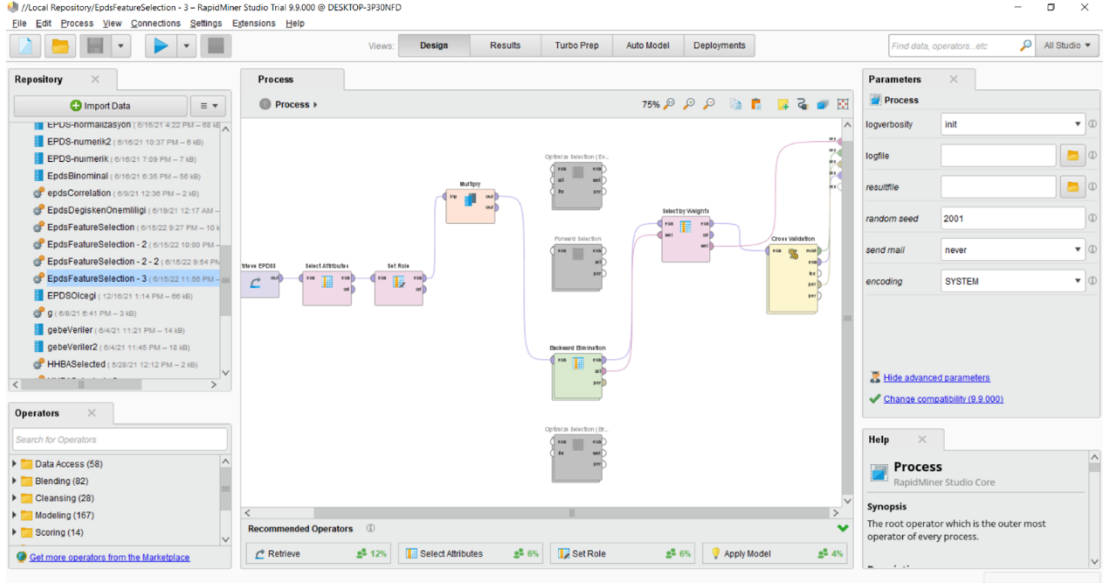


Şekil 3.11. Önerilen Veri Analitiği Mimarısının Diyagramı

Özellik seçimi, yani sınıflandırma veya regresyon problemleri için en alakalı özellikler sorusu, ana veri madenciliği görevlerinden biridir. Özellik seçimi için kullanılan birçok algoritma bulunmaktadır. Hangisinin performans olarak en iyi sonucu vereceğini analiz etmek adına algoritmalar, uygulanan formun çeşitli kombinasyonlarıyla eğitilmiştir. Tüm kombinasyonlar Şekil 3.11’de gösterilmektedir.

Test edilen tüm kombinasyon ayrıntılı bir şekilde anlatılmaktadır;

Ardışık ileri seçim (Sequential Forward Selection), geri yönlü arama seçimi (Backward Elimination) ve optimize edilmiş seçim-evrimsel özellik seçimi (Optimize-Evolutionary Selection).

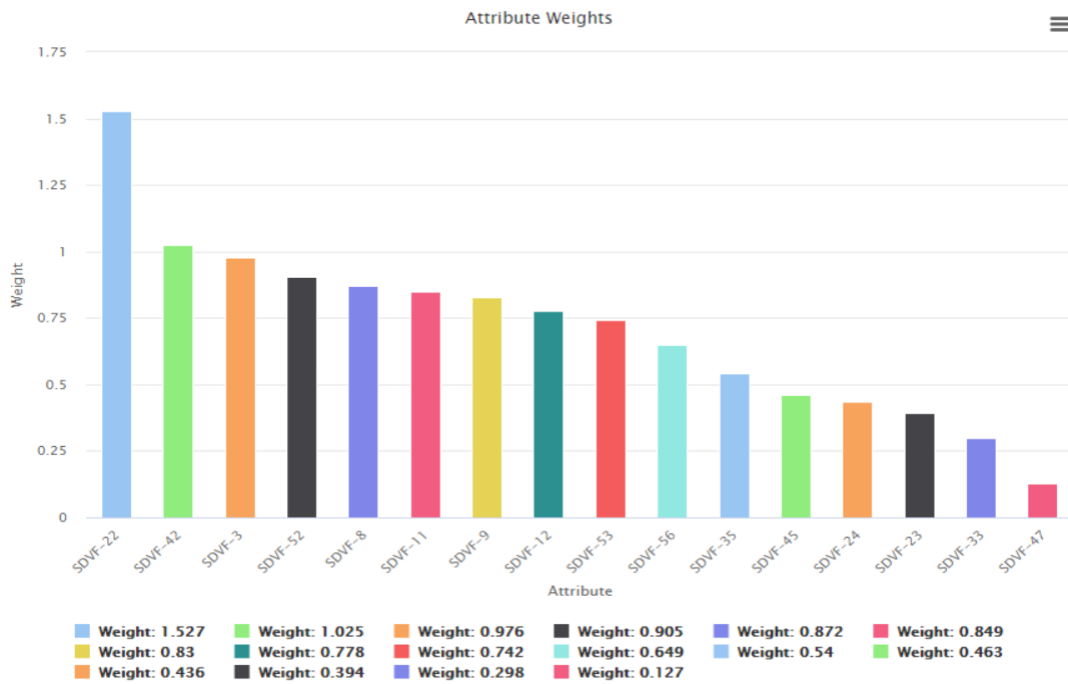


Şekil 3.12. Kullanılan Seçim Metodları

- **Ardışık İleri Seçim:** Bu yöntemde ilk olarak tüm özellikleri hedef değişken ile arasındaki ilişkiye bakarak en fazla hedef değişkeni etkileyen yani açıklayan özelliği seçerek başlanmaktadır. Tüm özellikleri teker teker model sürecine sokmaktadır. En iyi performansı gösteren özellik bu ilk seçilen özellik olmaktadır. Bu ilk özellik yanına hangi diğer değişkeni koyarsak en iyi sonucu elde edilir sorusunun peşine düşerek tekrar tüm değişkenleri deneyerek en iyi 2’li performans gösteren değişkenleri bulmaya çalışılmaktadır. Sonrasında en iyi 3. değişken ne olacaktır diye aynı şekilde denenmeye devam etmektedir. Önceden verilen durdurma kriterine gelene kadar bu döngü devam eder.
- **Geri yönlü arama seçimi:** Bu yöntem ilk olarak tüm özellikleri aynı anda kullanarak bir model performansı hesaplanmaktadır. Sonrasında sırasıyla her bir değişkeni çıkartarak model kurmayı denemekte ve en kötü performansa sebep olan değişkeni elemektedir. Sonrasında tekrar değişkenleri sırasıyla çıkartarak model performansını en kötü etkileyen özelliği tespit etmekte ve o değişkeni elemektedir. Ne zaman model performansı artık değişken çıkardığında artış değil azalış gösterirse durma işlemi gerçekleştirir.
- **Optimize edilmiş seçim-evrimsel özellik seçimi:** Bu operatör, verilen örnek kümesinin en alakalı özelliklerini seçmektedir. Öznitelik özeti için iki deterministik “greddy öznitelik seçim algoritması” olan ileri seçim ve geriye doğru eleme kullanılmaktadır. Deterministik bir algoritma, gayri resmi terimlerle tahmin edilebilir şekilde davranan bir algoritmadır. Belirli bir girdi

verildiğinde, her zaman aynı çıktıyı üretecek ve temeldeki makine her zaman aynı durum dizisinden geçecektir. Greedy algoritma, her adımda yerel olarak en uygun seçimi yaparak global bir optimum bulma umuduyla bir problemi çözmek için kullanılan bir buluşsal yöntemdir. Bazı problemlerde, greedy bir strateji optimal bir çözüm üretmeyebilir, ancak yine de greedy bir buluşsal yöntem, küresel bir optimal çözüme yaklaşan yerel olarak optimal çözümler verebilir (“RapidMiner”, 2023).

Genetik algoritma doğal evrimin seyrini taklit eden optimize edilmiş seçim-evrimsel teknik, model performansı üzerinde en büyük etkiyi gösterdiği tespit edilmiştir. SDVF veri seti üzerinde yapılan özellik seçimi algoritmalarına göre en yüksek önceliğe sahip özellikler Şekil 3.13’te görülmektedir.



Şekil 3.13. Sonuç Üzerinde En Etkili Özellikler

Tablo 3.2. Verisetindeki Karşılıkları

Özellikler	Verisetindeki Adı	Özelliklerin Ağırlıkları
Gebelik haftası	SDVF-22	1.527
Bebeğin sağlık problemi	SDVF-42	1.025
Yaşam yeri (kent veya kırsal)	SDVF-3	0.976
Eşiyle iletişimi	SDVF-52	0.905
Eğitim seviyesi	SDVF-8	0.872
Evinde yaşayan insan sayısı	SDVF-11	0.849
Çalışma durumu	SDVF-9	0.830
Toplam gelir seviyesi	SDVF-12	0.778
Eşinden gördüğü duygusal destek	SDVF-53	0.742
Problemlerini paylaşacağı insan varlığı	SDVF-56	0.649
Egzersiz durumu	SDVF-35	0.540
Eşinin eğitim durumu	SDVF-45	0.463
İstenen gebelik olup olmadığı	SDVF-24	0.436
Bebeğin cinsiyeti	SDVF-23	0.394
Sigara içme durumu	SDVF-33	0.298
Eşinin kronik rahatsızlık durumu	SDVF-47	0.127

Altmış özellikten en önemlileri çoktan aza doğru ağırlıklarına göre sıralanmıştır. Bu özelliklerin SDVF formundaki karşılıkları ve ağırlıkları Tablo 3.2'de gösterilmiştir.

Bu işlemler sonunda işlenebilir veriseti ve optimum sonucu veren özellik seçme algoritması belirlenmiştir.

3.2.3. Makine öğrenmesi

3.2.3.1. Çapraz doğrulama (cross validation)

Çapraz Doğrulama, makine öğrenimi modelinde gerçekleştirilen bir testin hatasını daha iyi tahmin etmek için model seçiminde kullanılan bir algoritmadır. Çalışmamızda, modeli eğitim ve test verileri olarak ayırmak için 10 katlı çapraz doğrulama algoritması kullanılmıştır. 10 katlı çapraz doğrulama algoritması, veri setini on eşit parçaya bölmektedir. Bu parçalardan biri test veri seti olarak, kalan dokuz parça ise eğitim veri seti olarak kullanılmaktadır. Modelin doğruluğu, doğruluk değerlerinin ortalaması alınarak belirlenir (Saccenti ve Camacho, 2015).

3.2.3.2. Modelleme

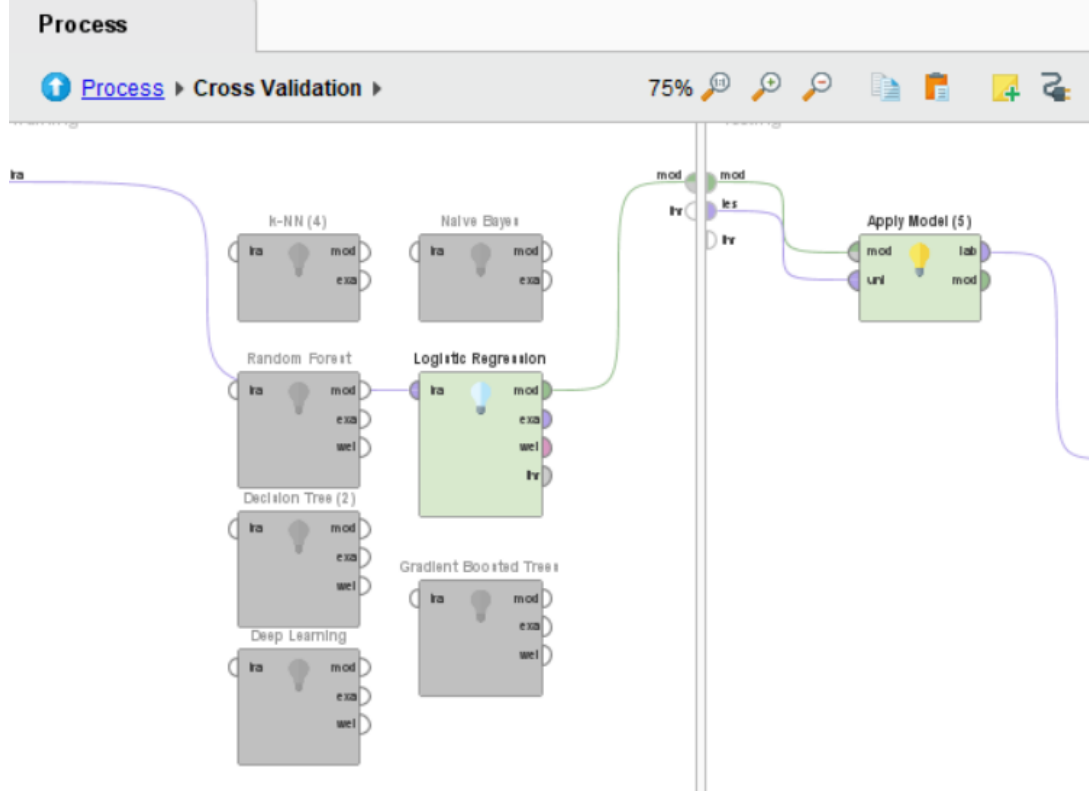
Çalışmada aşağıdaki makine öğrenimi algoritmaları kullanılmıştır: Karar Ağaçları (Decision Tree), Naive Bayes (NB), K-Nearest Neighbor (K-NN), Random Forest, GBT (Gradient Boosted Tree), Logistic Regresyon ve DFFNN (Deep Feed Forward Neural Network). Tablo 3.3.' de görüldüğü gibi Naive Bayes algoritması en yüksek performansı verdiği için çalışmanın devamında Naive Bayes algoritmasının kullanılması uygun görülmüştür.

- NB: Naive Bayes sınıflandırma algoritması, istatistiksel bir sınıflandırma yöntemidir ve Bayes teoremi prensibine dayanır. Algoritmanın çalışma prensibi, veri noktalarını belirli bir sınıfa sınıflandırmak için sınıf etiketlerinin olasılıklarını hesaplar. veri noktasının özelliklerini kullanarak, her bir sınıf için veri noktasının bu sınıfa ait olma olasılığını hesaplamaktır. Ardından, en yüksek olasılığa sahip sınıf, veri noktasının sınıflandırıldığı sınıf olarak belirlenir. Bu algoritma, veri setindeki değişkenler arasındaki bağımlılıkları ve olasılıkları dikkate alarak sınıflandırma yapar.

$C = \{c_1, c_2, \dots, c_m\}$ 'nin bir sınıflar kümesi olduğu ve $X = \{x_1, x_2, \dots, x_n\}$ 'nin bir özellikler kümesi olduğu kabul edilmektedir. Denklem (1), Bayes teoreminden belirli bir özellik verildiğinde her bir sınıf değişkeni için son olasılığı belirlemek için kullanılabilir.

$$P(C_i|X_j) = \frac{P(X_j|C_i)P(C_i)}{P(X_j)} \quad (3.1)$$

$i = 1, 2, \dots, m$ ve $j = 1, 2, \dots, n$, varsayılarak, $P(X_j|C_i) =$ Verilen C_i sınıfı ile X_j özelliğinin olasılığı, $P(X_j) = X_j$ özelliğinin önceki olasılığı ve $P(C_i) = C_i$ sınıfının önceki olasılığı olarak kabul edilmektedir. Sınıflandırıcı daha sonra sınıf değişkenlerinin en yüksek olasılığını belirler. Sınıflandırma sonucu, sonsal olasılık değeri en yüksek olan sınıf olacaktır (Sau ve Bhakta, 2019b).



Şekil 3.14. Kullanılan Makine Öğrenmesi Algoritmaları

3.2.3.3. Değerlendirme metrikleri

Modeller, denklem 2-6'da belirtildiği gibi beş ortak ölçü kullanılarak değerlendirilir: doğruluk, duyarlılık, özgüllük, kesinlik ve ROC'nin eğri altındaki alan (AUC); burada TP(True Pozitif) gerçek pozitif, TN(True Negatif) gerçek negatif, FP(False Positive) ise yanlış pozitif ve yanlış negatif için FN(False Negatif) (Sau ve Bhakta, 2019b).

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.2)$$

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (3.3)$$

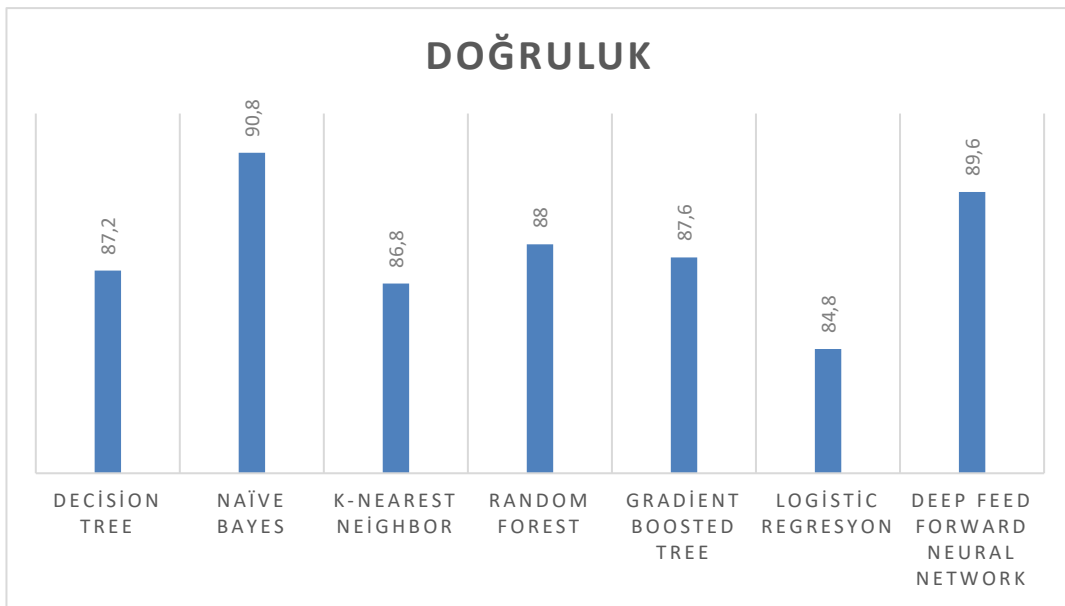
$$\text{Hassasiyet} = \frac{TP}{TP+FN} \quad (3.4)$$

$$\text{Özgüllük} = \frac{TN}{TN+FP} \quad (3.5)$$

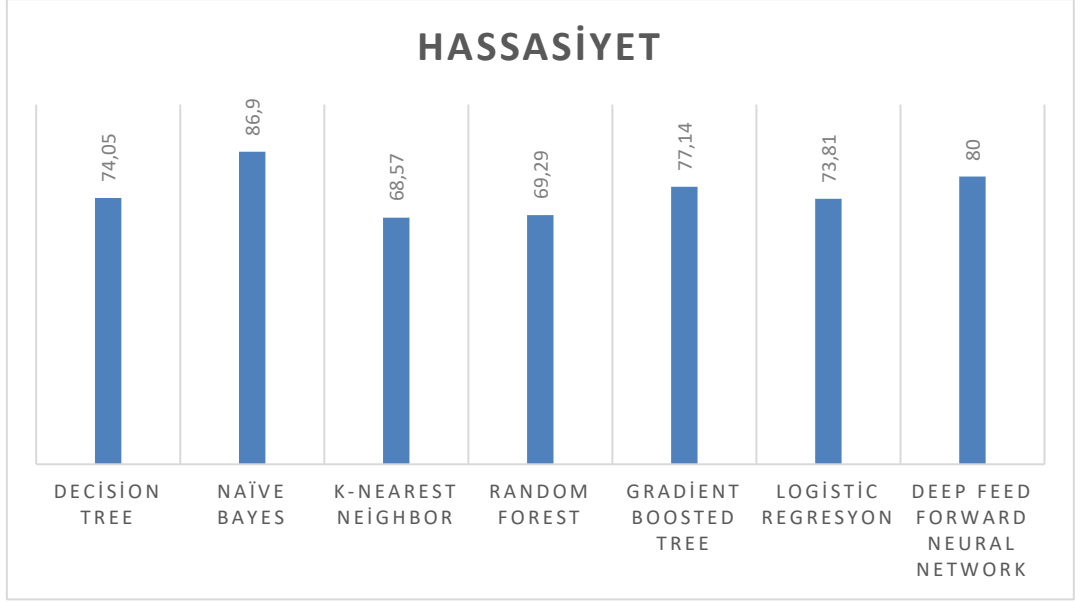
$$\text{AUC} = \text{ROC eğrisinin altındaki alan, yani TP'ye karşı FP} \quad (3.6)$$

Tablo 3.3. Performans Karşılaştırmaları

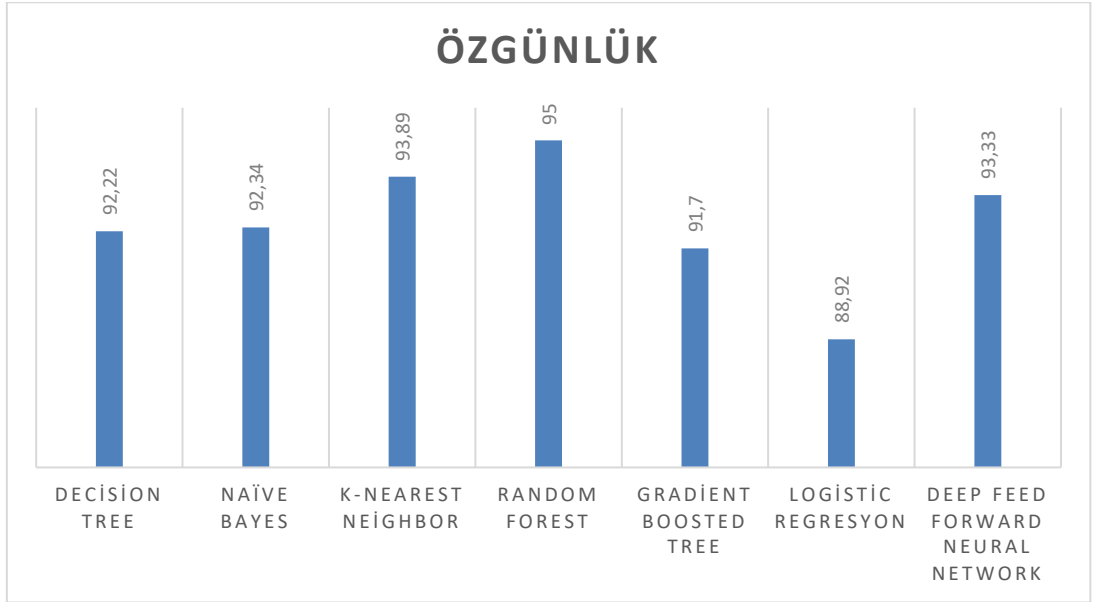
Sınıflandırıcı	Doğruluk	Hassasiyet	Özgünlük	Kesinlik	AUC
Decision Tree	87.20%	74.05%	92.22%	80.05%	0.822
Naive Bayes	90.80%	86.90%	92.34%	81.71%	0.966
K-Nearest Neighbor	86.80%	68.57%	93.89%	81.95%	0.900
Random Forest	88.00%	69.29%	95.00%	86.10%	0.945
Gradient Boosted Tree	87.60%	77.14%	91.70%	77.14%	0.938
Logistic Regresyon	84.80%	73.81%	88.92%	74.58%	0.908
Deep Feed Forward Neural Network	89.60%	80.00%	93.33%	83.26%	0.962



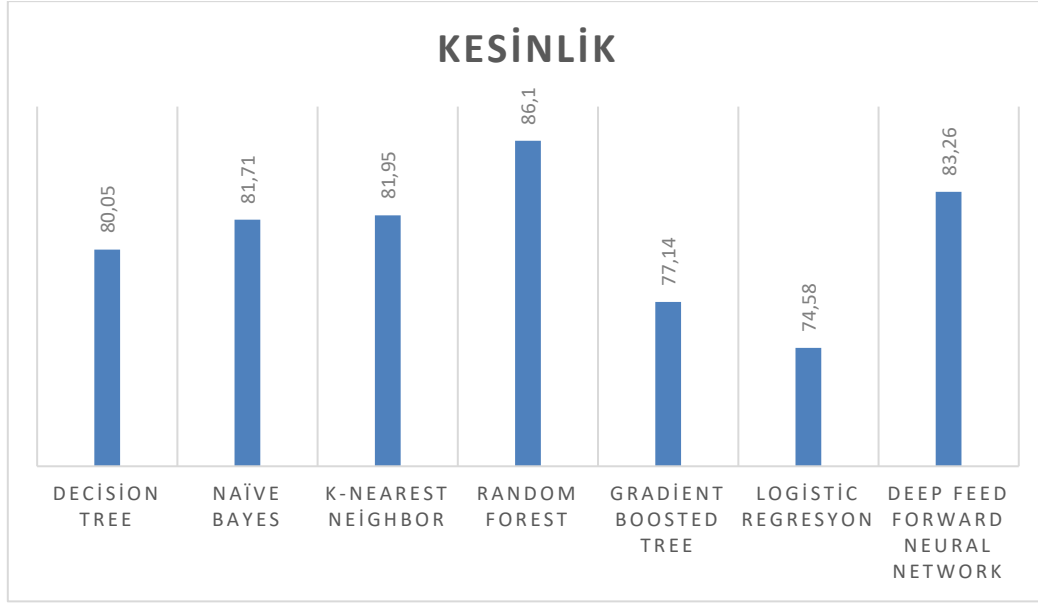
Şekil 3.15. Doğruluk Karşılaştırmaları



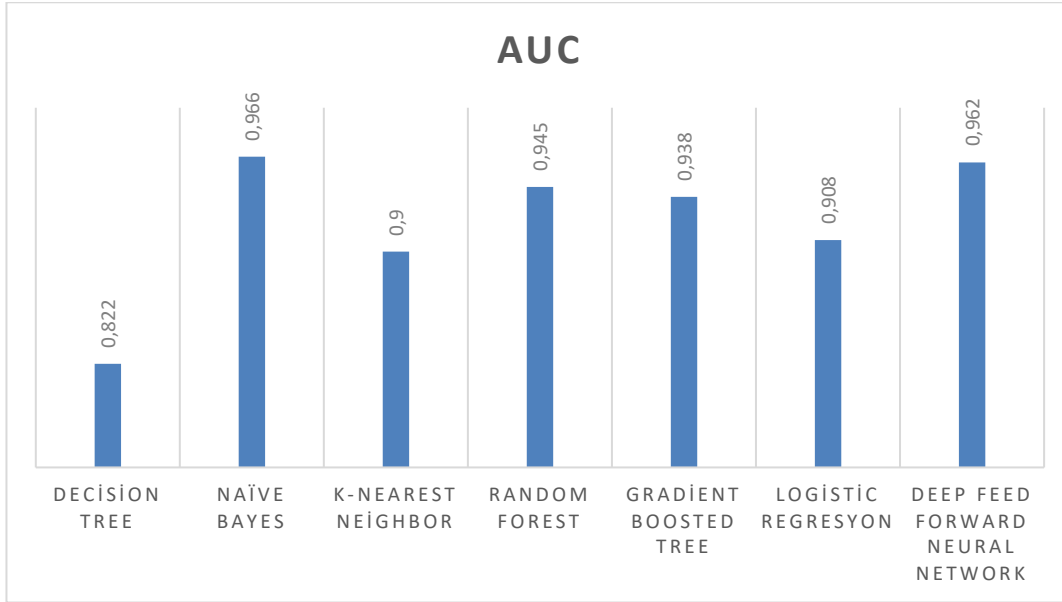
Şekil 3.16. Hassasiyet Karşılaştırmaları



Şekil 3.17. Özgünlük Karşılaştırmaları



Şekil 3.18. Kesinlik Karşılaştırmaları



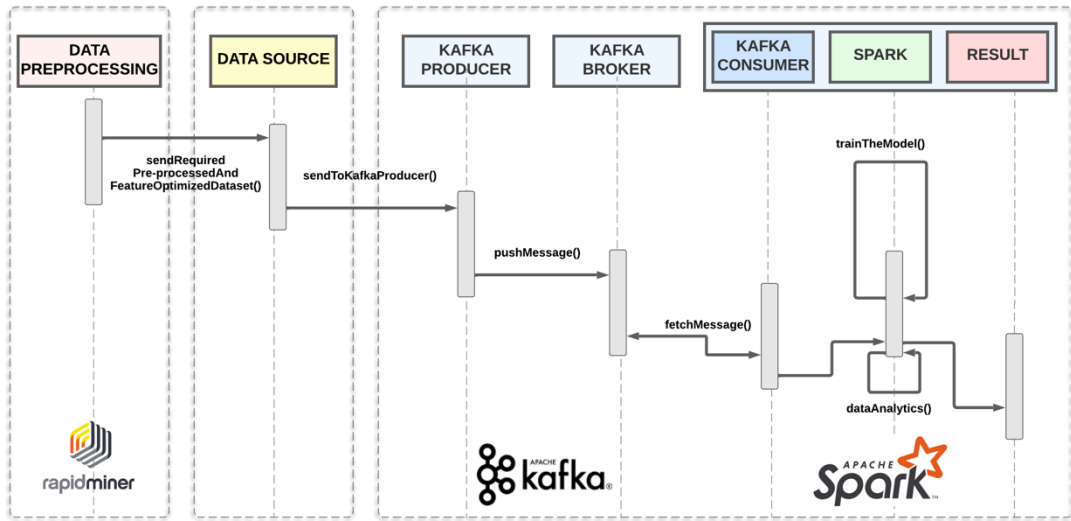
Şekil 3.19. AUC Değeri Karşılaştırmaları

Farklı özellik seçme algoritmalarından geçerek verisetinin özellik sayısı 60 özellikten 16 özelliğe indirgenmiştir. Bu sonucu elde ederken performansının en yüksek olduğu gözlemlenen evrimsel stratejiyi benimseyen optimize etme yönteminden faydalanılmıştır.

Farklı seçim algoritmalarından geçen veriler çapraz doğrulama yardımıyla performansı en yüksek belirlenen on eşit parçaya bölünüp “stratified sampling” uygulama tipi ile modeli eğitmişlerdir. Model eğitiminde ise yedi farklı makine öğrenmesi metodu (Naive Bayes, KNN, Random Forest, Decision Tree, Gradient

Boosted Tree, Deep Learning, Logistic Regresyon) kullanılarak performans karşılaştırılmaları gözlemlenmiştir (Tablo 3.3). Burada performansı en yüksek olan makine öğrenmesi metodu %90,80 doğruluk, %86,90 duyarlılık, %92,34 özgünlük, %81,71 kesinlik ve 0,966 AUC ile “Naive Bayes” algoritması olduğu gözlemlenmektedir (Şekil 3.15-3.19). Performans olarak en yüksek oranı vermesinden dolayı Naive Bayes algoritması, Spark MLLIB API’inde kullanılmak üzere tercih edilmiştir (Şekil 3.11).

3.2.4. Veri analitiği



Şekil 3.20. Zamanlama Şeması (Ogur ve diğerleri, 2023)

Tüm ön işlemlerden geçirilerek elde edilen veriseti %80 eğitim veriseti %20 test veriseti olarak bölünmüştür. Bu %80 olarak bölünen eğitim veriseti performansı en yüksek çıkan algoritma olan Naive Bayes algoritması ile model eğitilmiş ve tekrar kullanılmak üzere kaydedilmiştir.

Diğer kalan test veriseti ise akış halinde Kafka üretici ile satır satır “Kafka Broker”a gönderilmiştir. Kaydedilen model, “Streaming API” sayesinde saniyede bir gelen veri ile etkileşime girmesi sağlanmıştır. Bu sayede eğitilen modelin streaming halde gelen verilere dair verdiği çıktılar ve tüm performans kıyaslamaları “console”da ekrana bastırılmıştır. Akış halinde gelen veriler önceden oluşturulan Naive Bayes makine öğrenmesi modelinden geçirilerek hastalık tahmini yapılması sağlanmıştır. Gelen verinin tahminine göre modelin doğruluk oranı değişmektedir.

Apache Spark, yaygın makine öğrenimi algoritmalarını destekleyen bir MLlib modülüne sahiptir. Tablo 3.3.’te gösterilen diğer sınıflandırmalardan daha yüksek

performans verdiği için bu çalışmada hasta verilerini sınıflandırmak için Naive Bayes algoritması seçilmiştir. Apache Spark'ta kullanılan Naive Bayes algoritmasının değerlendirme ölçütleri Tablo 3.4'te gösterilmektedir. Tüm bu işlem adımlara Şekil 3.20'de yer alan zamanlama şemasında yer verilmiştir.

Tablo 3.4. Naive Bayes Algoritmasının Değerlendirme Parametreleri

Parametre	Değer
Doğruluk	92.45%
Kesinlik	97.29%
Hassasiyet	92.30%
F1	92.60%

Geliştirilen Naive Bayes modelini eğitmek ve test etmek için hasta veri setinin %80'i kullanılmıştır. Bu model, veri kümesinin %20'si tarafından oluşturulan akış verileriyle etkileşime girmiştir.

```

Batch: 44
-----
|SDVF-22|SDVF-42|SDVF-3|SDVF-52|SDVF-8|SDVF-11|SDVF-9|SDVF-12|SDVF-53|SDVF-56|SDVF-35|SDVF-45|SDVF-24|SDVF-23|SDVF-33|SDVF-47|label|prediction|
-----
| 20| 1| 0| 5| 3| 3| 0| 1| 1| 1| 0| 5| 1| 1| 0| 1| 1| 1.0|
-----

Batch: 45
-----
|SDVF-22|SDVF-42|SDVF-3|SDVF-52|SDVF-8|SDVF-11|SDVF-9|SDVF-12|SDVF-53|SDVF-56|SDVF-35|SDVF-45|SDVF-24|SDVF-23|SDVF-33|SDVF-47|label|prediction|
-----
| 11| 1| 0| 10| 3| 2| 1| 1| 10| 1| 2| 5| 0| 2| 0| 1| 0| 1.0|
-----

```

Şekil 3.21. Akış Halindeki Verinin Anlık Sonuçları

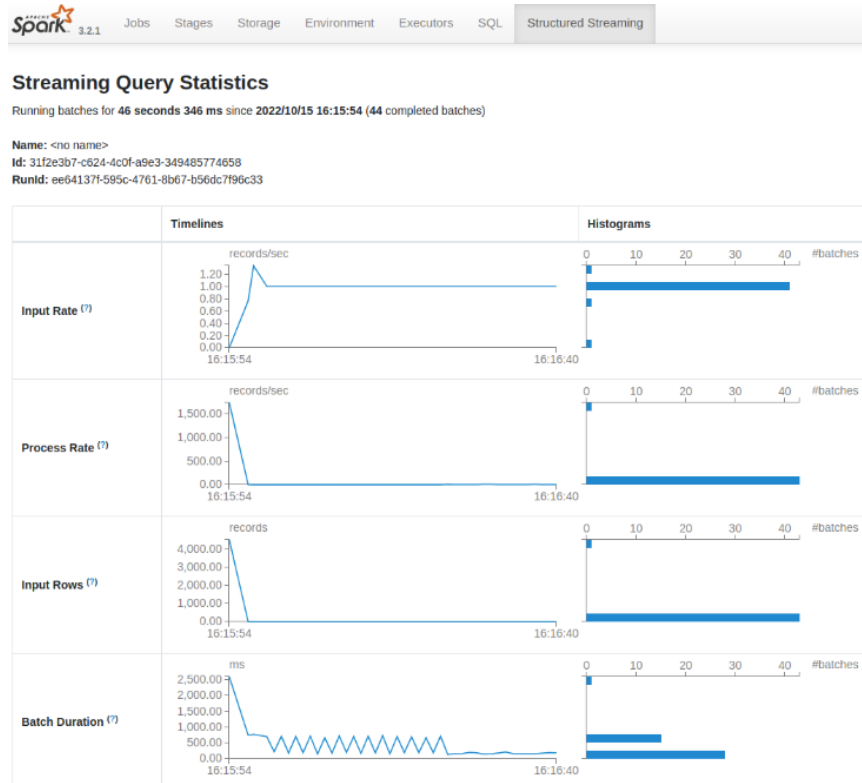
Şekil 3.21, gelen test verilerinin yanı sıra verilerin anlık olarak Naive Bayes algoritması sınıflandırmasını ve veri eğitim modeline göre tahmin edilen sonuçları göstermektedir. Hasta verileri, özellik seçim algoritmalarının sonucu üzerinde en etkili olan, Tablo 3.2'de gösterilen sosyo-demografik verilerdir. Etiketlemede kullanılan "label" sütunu, sınıf derecelendirmesidir. Etiket sütununda iki değer vardır: "Anksiyete Yok-Depresyon Yok (0)" ve "Anksiyete ve/veya Depresyon yani (1).

İki hastanın en önemli özellikleri birer saniye aralıklarla incelenmiş ve bulgular gösterilmiştir (Şekil 3.21). Örneğin, SDVF-22'nin değerini 20 ve SDVF-42'nin değerini 1 olarak gösteren ilk satırdaki batch numarası 44 için, etiket değeri Naive Bayes'in beklenen değerine eşittir. Naive Bayes, kişiyi hastalık için pozitif olarak sınıflandırır.

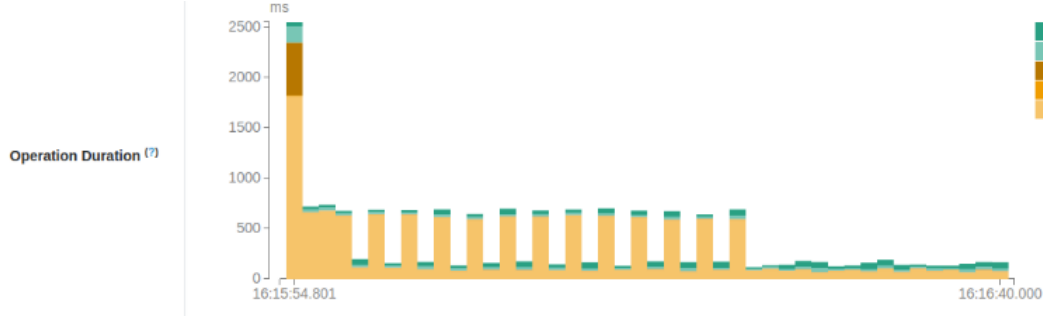
Etiket değeri, SDVF-22'nin değerini 11 ve SDVF-42'nin değerini 1 olarak gösteren ikinci satırdaki batch numarası 45'teki beklenen Naive Bayes değerine eşit değildir. Burada, Naive Bayes modeli yanlış bir şekilde bu kaydı hastalık için pozitif olarak sınıflandırır.

Akış işleme, kısa bir süre içinde (birkaç milisaniyeden dakikalara kadar) sürekli veri akışı işleme gerektirir. Sonuçlar, akış verilerini işleyen bir spark uygulayıcısı kullanılarak elde edildi.

Şekil 3.22 ve şekil 3.23, giriş hızı, işlem hızı, girdi satırları, batch süresi ve işlem süresi dahil olmak üzere spark streaming performans sonuçlarını görüntüleyen spark web kullanıcı arabirimini (WebUI) göstermektedir. Bunlar, gerçek zamanlı olarak veri işleme kapasitesini incelemek için değerli sonuçlardır (N. B. Oğur, Al-Hubaishi ve Çeken, 2022).



Şekil 3.22. Her Saniye Gelen Akış Halindeki Verinin İstatistikleri



Şekil 3.23. Her Saniye Gelen Akış Halindeki Verinin Çalışma Süresi

Girdi hızı (input rate), tüm kaynaklarda gelen veri hızını toplar. İşlem hızı (process rate), verilerin ne kadar hızlı analiz edildiğini gösterir. Giriş satırları (input rows), bir tetikleyici tarafından işlenen toplam kayıt sayısını gösterir. Toplu işlem süresi (Batch duration), her toplu işlemin süresini gösterir. Operasyon süresi (Operation duration), çeşitli işlemlerin milisaniye cinsinden gerçekleştirilmesi için geçen süre anlamına gelir.

Bu çalışma zihinsel sağlık durumlarını tespit etmek için makine öğrenimi tekniklerini kullanan hibrit büyük veri platformlarına odaklanmaktadır. Önerilen mimarinin, Apache Kafka ve Apache Spark gibi yüksek performanslı veri analitiği araçlarını bünyesinde barındırması nedeniyle anlık büyük veri işleme uygulamalarında başarılı bir şekilde kullanılabilmesi sonucuna varılmıştır.

Geliştirilen model aracılığıyla, hekime yeni başvuran bir gebede, girilen veriler sonucunda verdiği bilgiler neticesinde benzer özelliklere sahip gebelerin hastalık tahminleri olasılıksal olarak, anlık hesaplanabilecektir.

3.2.5. Kümeleme (cluster) teknolojisi

3.2.5.1. Ölçeklenebilir mimari

Hastanelerin kalabalık olması ve hastaların yeterli sonucu zamanında alamaması, sağlık hizmetleriyle ilgili sorunlara işaret edebilir. Bu sorunların çözümü için hastaneler ve sağlık hizmeti sağlayıcıları çeşitli adımlar atabilir. Bunlar arasında teknolojik çözümlerin kullanılması yer alabilir. Teknolojik çözümlerin baş sıralarında ölçeklenebilir mimari yer almaktadır (Y. Wang, Krishna ve Elyasi, 2023).

Ölçeklenebilir mimarinin faydaları oldukça fazladır. Bunlardan bazıları aşağıda sıralanmaktadır;

- **Yüksek Kullanılabilirlik: Ölçeklenebilirlik**, yüksek kullanılabilirlik sağlamak için önemlidir. Dağıtık sistemlerde, tek bir nokta arızalandığında veya yüksek taleplerle karşılaşıldığında, sistemin çökmemesi için ölçeklenebilirlik önemli bir faktördür. Ölçeklenebilir bir mimari, yükü dağıtarak ve yedeklilik sağlayarak sistemdeki bir arıza durumunda bile hizmetin sürekli olarak sunulmasını sağlar.
- **Esneklik ve Uyumluluk: Ölçeklenebilir bir mimari**, değişen iş gereksinimlerine kolayca adapte olabilir. Talepler ve iş yükü dalgalanmaları olduğunda, sistem kaynaklarını artırarak veya azaltarak esneklik sağlayabilmektedir. Ayrıca, farklı platformlar ve teknolojiler arasında uyumluluk sağlama yeteneği sayesinde, sistemde yeni bileşenler eklenebilir veya kaldırılabilir
- **Performans İyileştirmesi: Ölçeklenebilir bir mimari**, artan taleplere daha iyi yanıt verebilir ve daha yüksek işlem kapasitesi sunabilir. İş yükü arttıkça, sistem kaynaklarını doğru şekilde ölçeklendirerek performansı artırabilir. Bu sayede, kullanıcılara daha hızlı yanıtlar sağlanır ve uygulama performansı iyileştirilir.
- **Geleceğe Hazırlık: Ölçeklenebilir bir mimari**, gelecekteki büyüme ve gelişmelere hazırlıklı olmanızı sağlar. İşletmenizin ihtiyaçları değiştikçe ve kullanıcı sayısı arttıkça, sistemler kolayca ölçeklendirebilir ve daha fazla talebi karşılayabilir. Bu, uzun vadede sistemleri yeniden tasarlamak veya değiştirmek zorunda kalmadan büyümeyi destekler.
- **Maliyet Tasarrufu: Ölçeklenebilir mimariler**, kaynakları daha etkin ve verimli bir şekilde kullanmayı sağlar. Talepler arttıkça, sistemi ölçeklendirerek mevcut kaynakların kullanımını optimize edilebilir. Bu, gereksiz kaynak harcamalarını önler ve maliyetleri azaltır. Ayrıca, ölçeklenebilirlik sayesinde, sistemi başlangıçta büyük bir ölçekte kurmak yerine ihtiyaç duyuldukça ölçeklendirebilir ve maliyetleri dağıtılabılır.

Tezin devamında Kubernetes üzerinde Apache Spark kullanımı ile ölçeklenebilir mimarinin sisteme entegre edilmesi ve performans testleri yer alacaktır.

3.2.5.2. Kubernetes kümesi

Çalışma kapsamında büyük veri yönetim platformu olarak Apache Spark kullanılmıştır. Apache Spark ise dört temel dört küme yöneticisini destekler: Standalone, Apache Mesos, Hadoop YARN ve Kubernetes. Hangisinin tercih

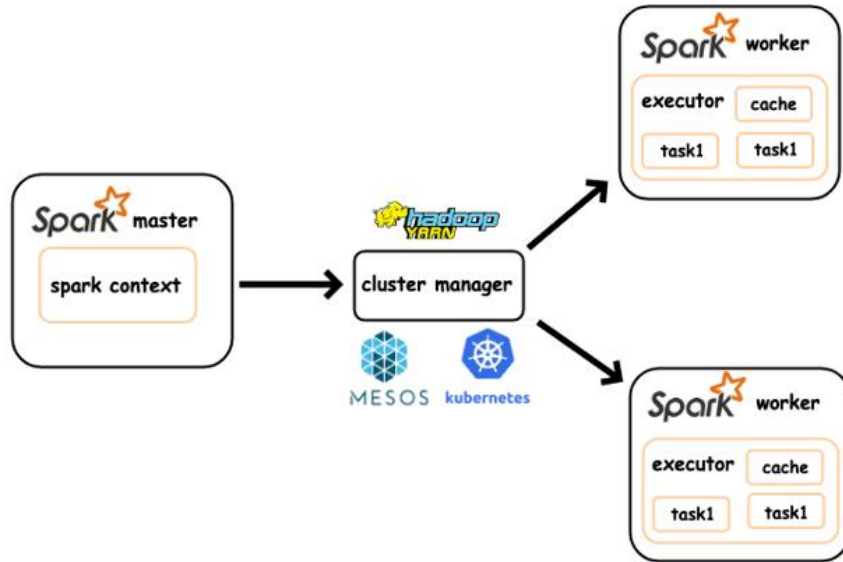
edileceđi sistemin ihtiyalarına ve beklentilerine gre deđiřiklik gstermektedir. Bu alıřmada kullanım kolaylıđından ve iřlevselliđinden dolayı Kubernetes tercih edilmiřtir.

Kubernetes: Kubernetes, aık kaynaklı bir konteyner ynetim ve dađıtım platformudur. Byk lekli uygulamaların ve mikrohizmet mimarilerinin sorunsuz bir řekilde alıřtırılmasını ve ynetilmesini sađlar. Kubernetes, uygulamaları birden ok konteyner halinde alıřtırarak, kaynak ynetimi, otomatik leklendirme, yk dengeleme, hata toleransı ve hızlı dađıtım gibi iřleri kolaylařtırır.

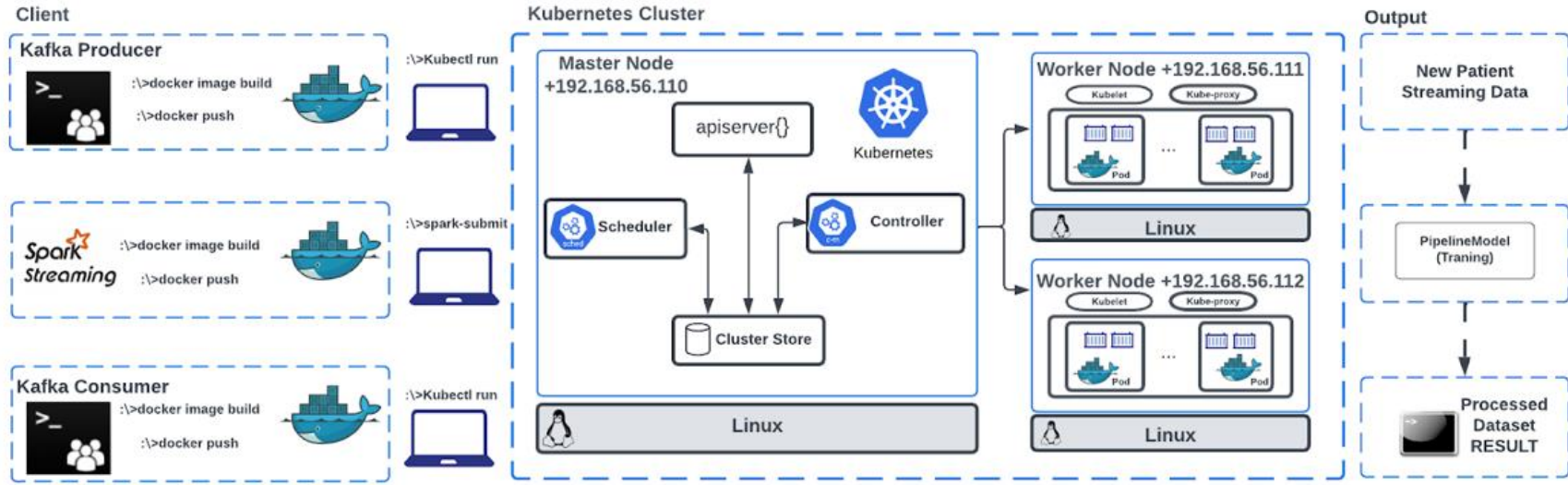
Kubernetes, konteyner teknolojisi olarak en popler olan Docker ile entegre alıřır. Konteynerler, bir uygulamanın tm gerekli kod, bađımlılıklar ve yapılandırma dosyalarını bir arada paketler. Bu sayede uygulama, farklı ortamlarda tutarlı bir řekilde alıřtırılabilir hale gelir.

Kubernetes, konteyner tabanlı uygulamaları birleřtiren ve bunları dzenleyen bir yapıya sahiptir. Bu yapı, "cluster" olarak adlandırılan bir grup bilgisayardan oluřur. Kubernetes, bu kme zerinde konteynerleri otomatik olarak dađıtır, ynetir ve izler. Kaynak tahsisi, yk dengelemesi, otomatik leklendirme ve hata toleransı gibi iřleri ynetir.

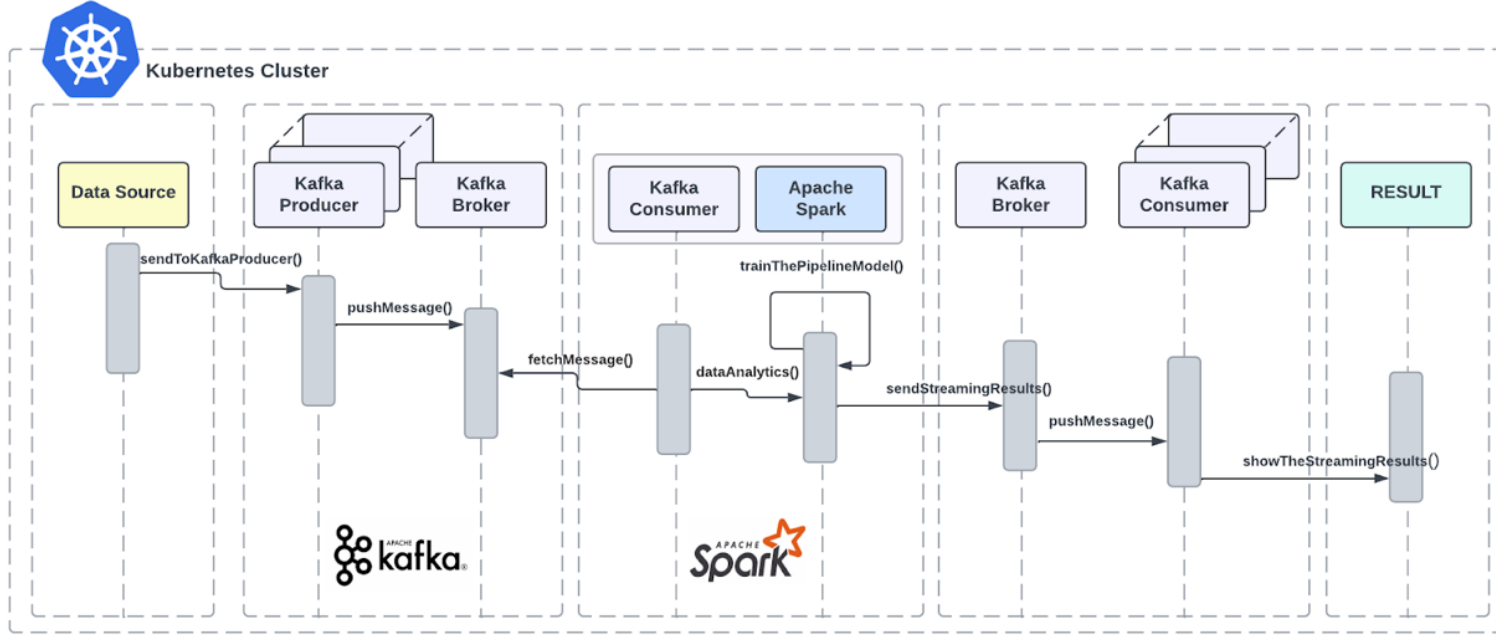
Kubernetes'in en nemli bileřenleri kmeler (cluster), dđmler (node) ve kontrol dzlemdir (control-plane). Kmeler dđmleri ierir. Her dđm en az bir adet iři makine (worker-node) kmesinden oluřur. Dđmler, konuřlandırılan uygulamanın unsurlarını ieren hizmet beklerini (pod) barındırır. Kontrol dzlemi, yksek eriřilebilirlik iin genellikle ok sayıda bilgisayarda bulunan dđm ve hizmet beklerini ynetir (řekil 3.24).



Şekil 3.24. Kubernetes Mimarisi (“Kubernetes Mimarisi”, 2023)



Şekil 3.25. Geliştirilen Büyük Veri İşleme Sistemi İçin Kubernetes Ana İşlem Diyagramı

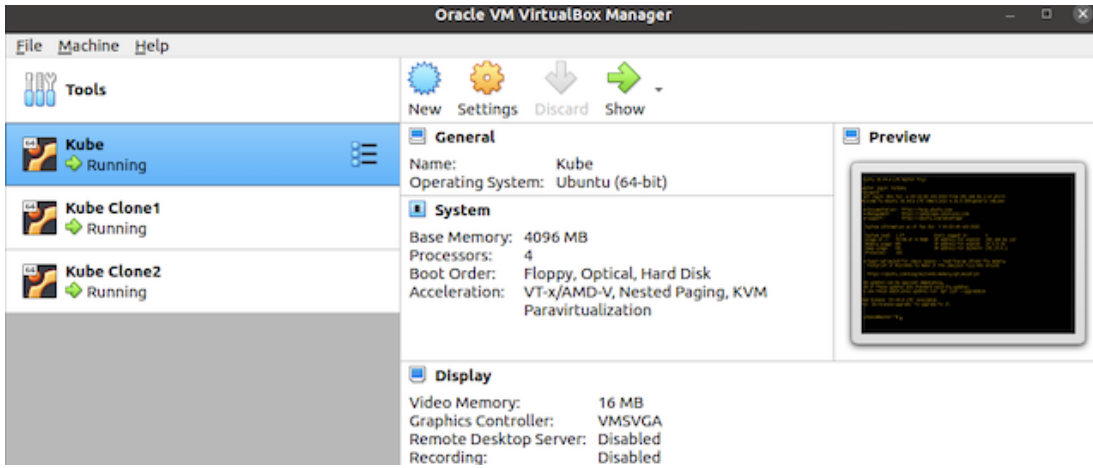


Şekil 3.26. Geliştirilen Büyük Veri İşleme Sisteminin Zamanlama Şeması

Şekil 3.25 Kubernetes kümesinin genel şemasına, Şekil 3.26 ise zamanlama şemasına yer verilmiştir. Şekillerin içinde varolan tüm modüller ayrıntılı bir şekilde aşağıda anlatılacaktır.

Bu çalışmada bir master iki adet node olmak üzere Debian temelli Ubuntu Linux 18.04.6 LTS işletim sistemine sahip üç adet sanal makine kurulmuştur (Şekil 3.27). Bu makinelere atanan IP'ler Şekil 3.28'de gösterilmiştir. Genel özellikleri ise aşağıdaki gibidir;

CPU (master)→4 RAM (master)→4GB Storage (master)→20GB
 CPU (node)→2 RAM (node)→ 4GB Storage (node)→10GB



Şekil 3.27. Kurulan Sanal Makineler

```
master 192.168.56.110
node1 192.168.56.111
node2 192.168.56.112
```

Şekil 3.28. Sanal Makine IP'leri

Sanal makineler Tablo 3.5.'te verilen özelliklere sahip sunucu üzerine kurulmuştur.

Tablo 3.5. Sunucu Özellikleri

HPE ProLiant ML150 Gen.9 Server	
CPU	Intel (R) Xeon CPU E5-2620 v4, 2.1 GHz 16 çekirdek
Bellek	32 GB DDR4
Hard disk	1 TB
Operating system	Linux Ubuntu Server 16.04

```

Kube [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Ubuntu 18.04.6 LTS master tty1
master login: nurbanu
Password:
Last login: Mon Nov 29 11:50:46 +03 2021 from 192.168.56.1 on pts/1
Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 4.15.0-159-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Sat Dec 18 18:51:21 +03 2021

System load:  3.16           Users logged in:      0
Usage of /:   76.2% of 9.78GB IP address for enp0s3: 192.168.56.110
Memory usage: 19%           IP address for enp0s8: 10.0.3.15
Swap usage:   0%            IP address for docker0: 172.17.0.1
Processes:    143

 * Super-optimized for small spaces - read how we shrank the memory
   footprint of MicroK8s to make it the smallest full K8s around.

https://ubuntu.com/blog/microk8s-memory-optimisation

63 updates can be applied immediately.
45 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

nurbanu@master:~$ [ 146.091940] watchdog: BUG: soft lockup - CPU#0 stuck for 23s! [kubelet:1370]
[ 146.091986] watchdog: BUG: soft lockup - CPU#3 stuck for 29s! [kubelet:1476]

```

Şekil 3.29. Master Node

```

Kube Clone1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Ubuntu 18.04.6 LTS node1 tty1
node1 login: nurbanu
Password:
Last login: Mon Nov 29 10:32:54 +03 2021 from 192.168.56.1 on pts/0
Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 4.15.0-159-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Sat Dec 18 18:51:53 +03 2021

System load:  1.65           Users logged in:      0
Usage of /:   64.1% of 9.78GB IP address for enp0s3: 192.168.56.111
Memory usage: 7%            IP address for enp0s8: 10.0.3.15
Swap usage:   0%            IP address for docker0: 172.17.0.1
Processes:    123

Kube Clone2 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 4.15.0-159-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

System information as of Sat Dec 18 18:51:56 +03 2021

System load:  3.17           Users logged in:      0
Usage of /:   78.3% of 9.78GB IP address for enp0s3: 192.168.56.112
Memory usage: 10%           IP address for enp0s8: 10.0.3.15
Swap usage:   0%            IP address for docker0: 172.17.0.1
Processes:    134

 * Super-optimized for small spaces - read how we shrank the memory
   footprint of MicroK8s to make it the smallest full K8s around.

https://ubuntu.com/blog/microk8s-memory-optimisation

44 updates can be applied immediately.
30 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

nurbanu@node2:~$ [ 81.400871] e1000 0000:00:08:0 enp0s8: Detected Tx Unit Hang
[ 81.400871] Tx Queue          <0>
[ 81.400871] TDH                  <93>
[ 81.400871] TDT                  <93>
[ 81.400871] next_to_use           <93>
[ 81.400871] next_to_clean        <91>
[ 81.400871] buffer_info[next_to_clean]
[ 81.400871] time_stamp            <ffff22db>
[ 81.400871] next_to_watch        <92>
[ 81.400871] jiffies               <ffff2492>
[ 81.400871] next_to_watch.status <0>
[ 83.668442] e1000 0000:00:08:0 enp0s8: Reset adapter

```

Şekil 3.30. Diğer Sanal Makinalar (Node1-Node2)

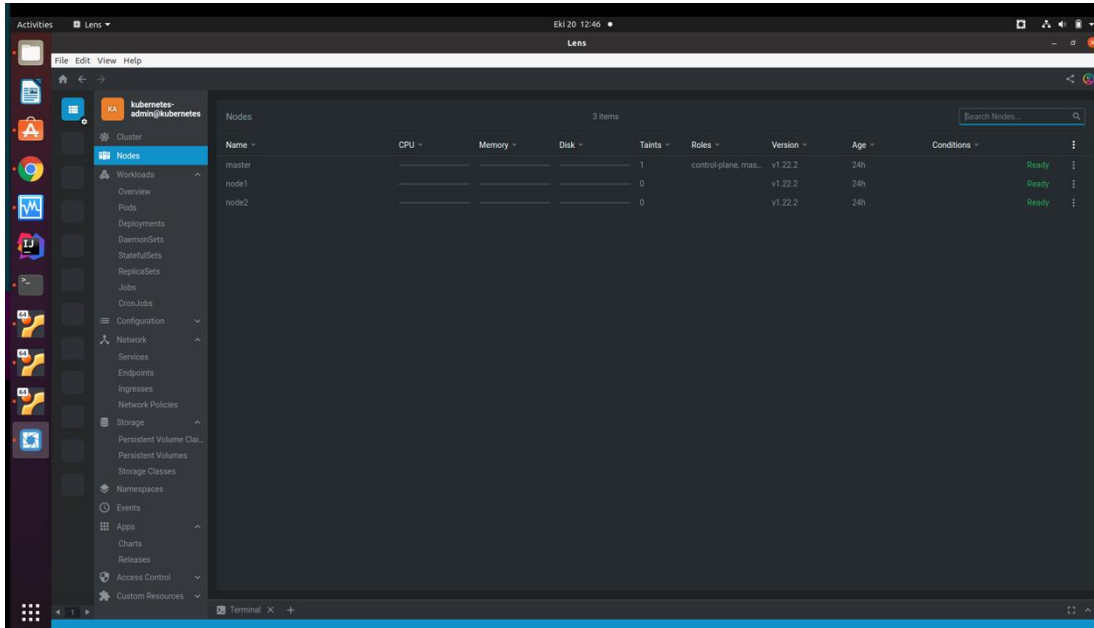
Master bilgisayara (Şekil 3.29) Docker, Kubernetes ve diğer gerekli tüm kurulumlar yapılmıştır. Sanal makineler (Şekil 3.30) klonlanarak diğer makineler de ayağa kaldırılmıştır. Diğer düğümlerin Şekil 3.31.’de yer verilen token aracılığıyla master ile bağlantı kurması sağlanmış ve böylelikle Kubernetes kümesi oluşturulmuştur.

```
* join link - need to be applied to all nodes(workers)
sudo kubeadm join 192.168.56.110:6443 --token bzbwl4.ll5o9x3jjhqwofa --discovery-token-ca-cert-hash sha256:ecb0223a05be3502c2d102f3e56104b10fcd105430eb723d3b3e816618323d73
```

Şekil 3.31. Kubernetes Token

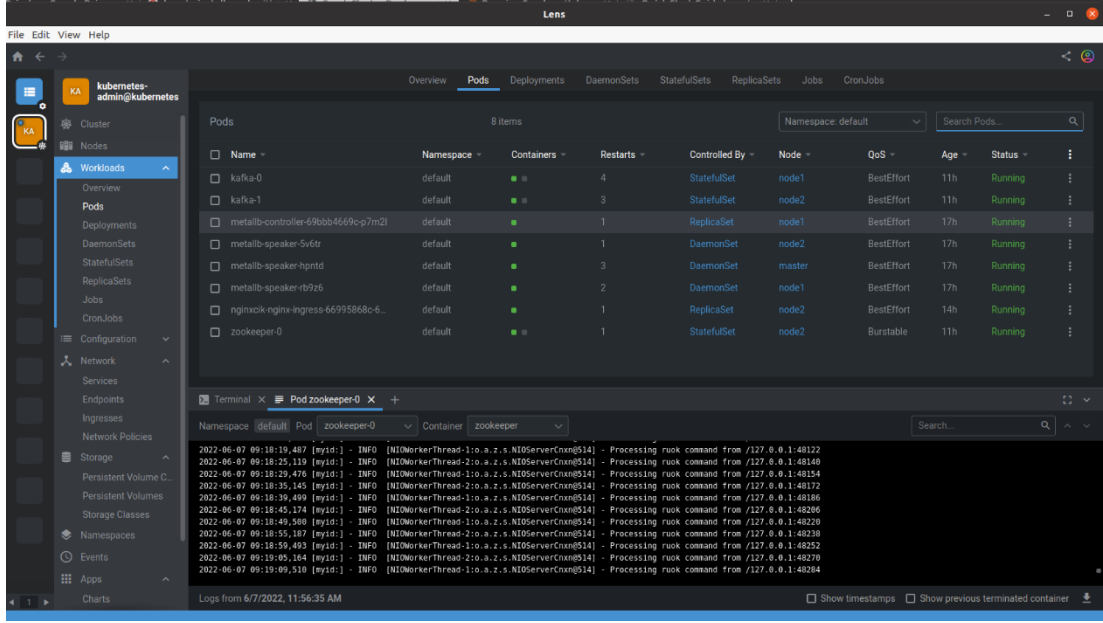
Kümenin görselleştirilmesi için kullanılan IDE’lerin başında gelen LENS platformu ile cluster ve içinde varolan podlar incelenebilir.

- **LENS:** Kubernetes Lens, etkili bir açık kaynaklı IDE olarak hizmet veren bir platformdur. Lens, Kubernetes’i yönetme ve izleme sürecini kolaylaştırarak kümelerinizi gerçek zamanlı olarak kontrol etmenize olanak sağlar. Şekil 3.32’de bir master ve iki node sanal makinenin LENS üzerinde tanımlanmış hali yer almaktadır (“Lens”, 2023).



Şekil 3.32. Kurulan Node’ların LENS Arayüzünde Görüntüsü

Gerekli görülen teknolojilerin gerek manuel gerekse Helm vasıtasıyla kurulumu gerçekleştirilmiştir. Kurulum tamamlandıktan sonra kurulan “pod”ların aktiflik durumu, hangi düğüm üzerinde ayağa kaldırıldığı gibi bilgilere LENS üzerinden erişim sağlanabilmektedir (Şekil 3.33).



Şekil 3.33. Kurulan Pod'ların LENS Arayüzünde Görüntüsü

- **HELM:** Helm, Kubernetes için bir paket yöneticisidir. Kubernetes ortamında uygulama dağıtımını kolaylaştırmak ve yönetmek için kullanılır. Helm, önceden yapılandırılmış uygulama bileşenlerini paketleyerek bunları "chart" adı verilen bir formatta depolar. Bu "chart"lar, bir uygulamanın gereksinim duyduğu tüm kaynakları, yapılandırmaları ve bağımlılıkları içerir ("Helm", 2023).

Helm, yazılım dağıtımını otomatikleştirir ve tekrar kullanılabilir "chart"ları kullanarak uygulamaları hızla oluşturmanıza ve yönetmenize olanak sağlar. "Chart"lar, uygulama sürümleri arasında güncellemeler yapmayı kolaylaştırır ve altyapı değişiklikleri için yeniden yapılandırmaları basitleştirir.

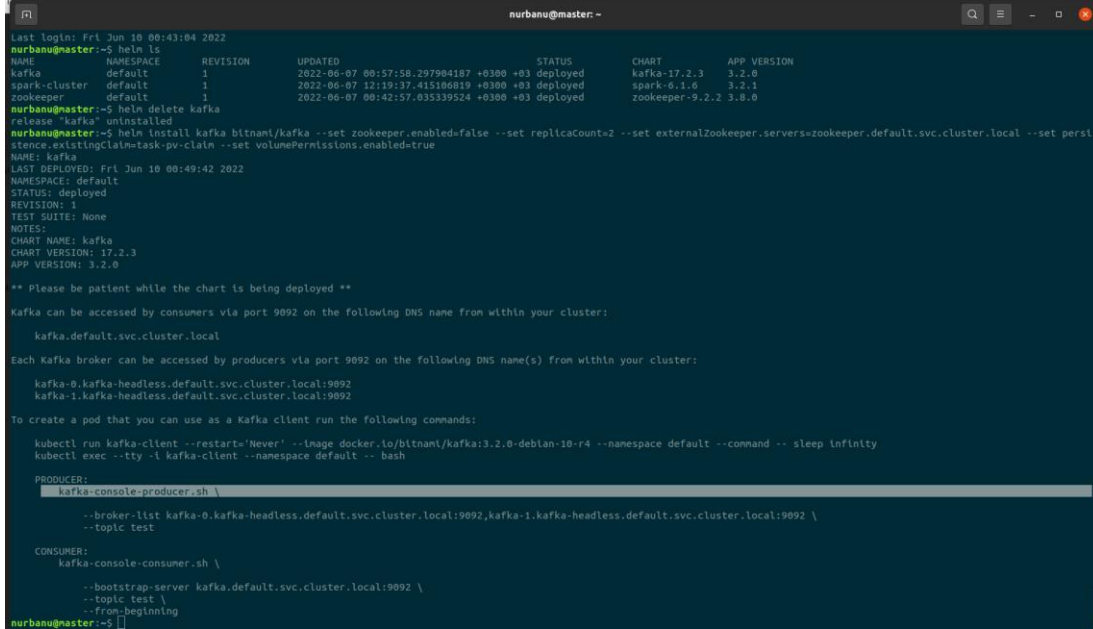
Helm, komut satırı aracılığıyla veya Helm CLI (Command Line Interface) kullanılarak kullanıcılar tarafından yönetilebilir. Ayrıca, Helm topluluğu tarafından geliştirilen birçok hazır "chart" mevcuttur. Bu "chart"lar, popüler uygulamaların hızlı bir şekilde dağıtılmasını sağlar. Şekil 3.34 ve Şekil 3.35'te kafka'nın helm komutlarına Şekil 3.36'da kurulumuna yer verilmiştir.

```
helm install zookeeper bitnami/zookeeper --set replicaCount=1 --set auth.enabled=false --set allowAnonymousLogin=true --set persistence.existingClaim=task-pv-claim --set volumePermissions.enabled=true
```

Şekil 3.34. Helm ile Zookeeper Kurulumu

```
helm install kafka bitnami/kafka --set zookeeper.enabled=false --set replicaCount=2 --set externalZookeeper.servers=zookeeper.default.svc.cluster.local --set persistence.existingClaim=task-pv-claim --set volumePermissions.enabled=true
```

Şekil 3.35. Helm ile Kafka Kurulumu



```
Last login: Fri Jun 10 00:43:04 2022
nurbanu@master:~$ helm ls
NAME          NAMESPACE    REVISION    UPDATED                                 STATUS          CHART          APP VERSION
kafka         default       1           2022-06-07 00:57:58.297904187 +0300 +03 deployed      kafka-17.2.3   3.2.0
spark-cluster default       1           2022-06-07 12:19:37.415106819 +0300 +03 deployed      spark-6.1.6    3.2.1
zookeeper     default       1           2022-06-07 00:42:57.035339524 +0300 +03 deployed      zookeeper-9.2.2 3.8.0
nurbanu@master:~$ helm delete kafka
release "kafka" uninstalled
nurbanu@master:~$ helm install kafka bitnami/kafka --set zookeeper.enabled=false --set replicaCount=2 --set externalZookeeper.servers=zookeeper.default.svc.cluster.local --set persistence.existingClaim=task-pv-claim --set volumePermissions.enabled=true
NAME: kafka
LAST DEPLOYED: Fri Jun 10 00:49:42 2022
NAMESPACE: default
STATUS: deployed
REVISION: 1
TEST SUITE: None
NOTES:
CHART NAME: kafka
CHART VERSION: 17.2.3
APP VERSION: 3.2.0

** Please be patient while the chart is being deployed **

Kafka can be accessed by consumers via port 9092 on the following DNS name from within your cluster:
    kafka.default.svc.cluster.local

Each Kafka broker can be accessed by producers via port 9092 on the following DNS name(s) from within your cluster:
    kafka-0.kafka-headless.default.svc.cluster.local:9092
    kafka-1.kafka-headless.default.svc.cluster.local:9092

To create a pod that you can use as a Kafka client run the following commands:
kubectrl run kafka-client --restart='Never' --image docker.io/bitnami/kafka:3.2.0-debian-10-r4 --namespace default --command -- sleep infinity
kubectrl exec --tty -i kafka-client --namespace default -- bash

PRODUCER:
kafka-console-producer.sh \
    --broker-list kafka-0.kafka-headless.default.svc.cluster.local:9092,kafka-1.kafka-headless.default.svc.cluster.local:9092 \
    --topic test

CONSUMER:
kafka-console-consumer.sh \
    --bootstrap-server kafka.default.svc.cluster.local:9092 \
    --topic test \
    --from-beginning
nurbanu@master:~$
```

Şekil 3.36. Helm ile Kafka Kurulumunun Terminal Gösterimi

3.2.5.3. Kafka producer

Önceki aşamalarda değinildiği üzere, gebe verileri gerekli görülen önışlem aşamalarından geçirilmiş, işlemek için hazır hale getirilmiştir. Bu işlemlerden sonra elde edilen veriseti, node.js ile kodlanan Kafka Üretici modülü vasıtasıyla “perinatal” adı verilen “topic”e yollanması gerçekleştirilmiştir. Bu modül her saniyede bir gebe verisi yollanacak şekilde kodlanmıştır. Böylelikle akış halinde veri aktarımı sağlanmıştır. Yazılan kodun bir parçasına Şekil 3.37.’de yollanan verilere ise Şekil 3.39.’da yer verilmiştir.

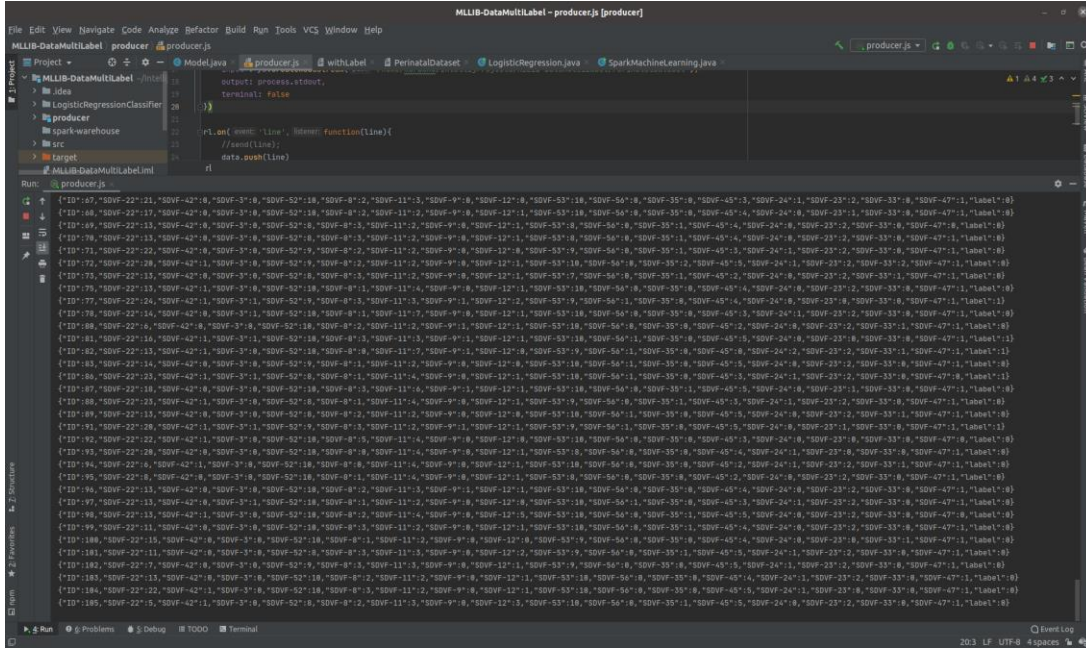
- **NODE.JS:** Node.js, ölçeklenebilir, hızlı ve verimli bir şekilde çalışabilen sunucu tarafı uygulamaların geliştirilmesinde popüler bir seçenektir. NPM (Node Package Manager) adı verilen paket yöneticisi sayesinde Node.js topluluğunda geliştirilen birçok ücretsiz ve açık kaynaklı modül ve kütüphane kullanılabilir. Node.js, olay tabanlı mimariye sahiptir, bu da çoklu işlemler ve paralel işlemler üzerinde etkili bir şekilde çalışabilmesini sağlar. Bu modüller, geliştirme sürecini hızlandırır ve işlevsellik eklemek için kullanılabilir. Bu sebepler dolayısıyla node.js bu çalıştırmada tercih edilmiştir (“Node.js”, 2023).


```
8     brokers: ['kafka:9092', 'localhost:9092']
9   })
10  var sleep=require('system-sleep');
11  const fs = require('fs');
12  const readline = require('readline');
13
14
15  const producer = kafka.producer()
16
17  var data =[];
18
19  var rl = readline.createInterface( options: {
20    input : fs.createReadStream( path: './PerinataDataset.json'),
21    output: process.stdout,
22    terminal: false
23  })
24
25  rl.on( event: 'line', listener: function(line){
26    //send(line);
27    data.push(line)
28    //console.log("Okundu : ", line) //or parse line
29
30  });
intervalFunc()
```

Şekil 3.37. Kafka Producer Node.js Kodu

```
1 # syntax=docker/dockerfile:1
2
3 FROM node:18.12.1
4 ENV NODE_ENV=production
5
6 WORKDIR /app
7
8 COPY ["package.json", "."]
9 # COPY ["package.json", "package-lock.json*", "./"]
10 RUN npm install --production
11
12 COPY . .
13
14 CMD [ "node", "kafkaproducer.js" ]
15
```

Şekil 3.38. Docker File



Şekil 3.39. Her Saniye Yollanan Gebe Verisi

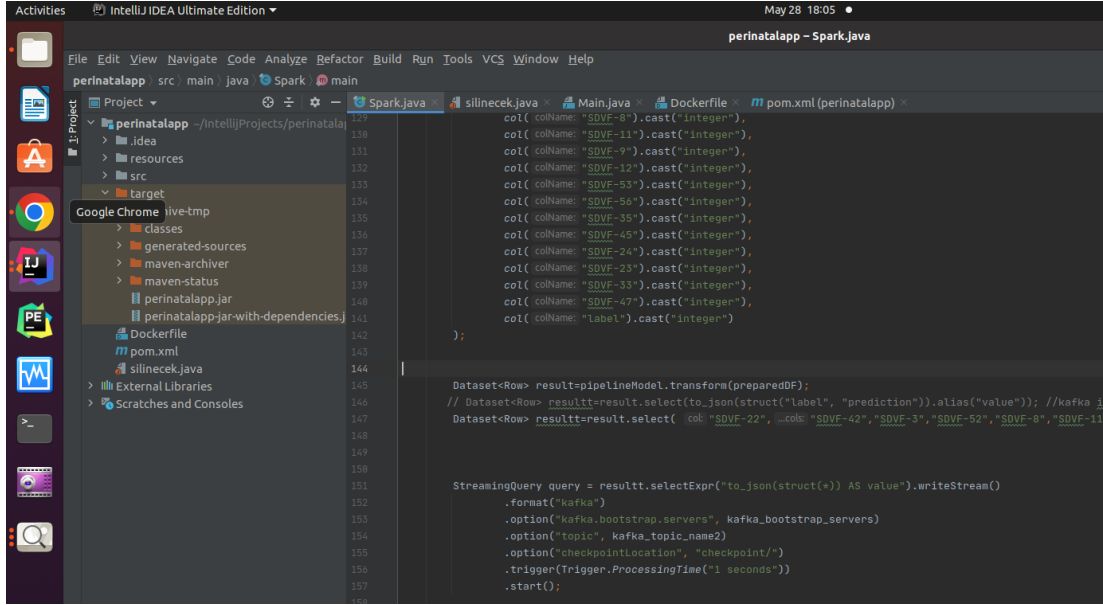
Node.js ile kullanılarak oluşturulan Kafka Üretici Şekil 3.38.’de gösterilen dockerfile vasıtasıyla imaj haline getirilip Docker Hub’a atılmış iki “worker-node” bir “master-node”den oluşan Kubernetes kümesinde ayağa kaldırılmıştır.

- >nurbanu@master:~\$ kubectl run -- image=nurbanu/perinatalproducer2 perinatalproducer

3.2.5.4. Kafka tüketici-apache spark

Çalışmanın bu bölümünde anlık olarak gelen akış halindeki verilerin Apache Spark yardımıyla anlamlandırılması ve elde edilen hastalık sonuçlarının “result” adlı “topic”e aktarması yer almaktadır. Böylelikle gelen gebe verileri anlık olarak sonuç üretebilecektir. Kullanılan teknolojiler sebebiyle büyük verilerin de sistem için elverişli olması adına altyapı oluşturulmuştur. Geliştirilen çalışma, ölçeklenebilir mimari ile performans karşılaştırması yaparak sistemin işlevselliğini test etme imkanı sunmaktadır.

Modül oluşturulduken Java yazılım dilinden faydalanılmıştır. Yazılım projesinin yapısını ve bağımlılıklarını yönetmek için ise Maven otomasyon aracı kullanılmıştır.



Şekil 3.40. Apache Spark Kodu

Apache Spark “Structure Streaming” yapısı ile Apache Kafka’dan gelen veriler alınıp “SparkSQL” şeması yardımıyla görselleştirilmiş ve “map-reduce” işlemleri ile işlenmiştir (Şekil 3.40). Elde edilen verisetinin bir bölümü modeli eğitmek için alınmıştır. Bu veriseti %80 eğitim, %20 test veriseti olarak ayrılmıştır. Eğitim veriseti performans olarak en yüksek değeri veren Naïve Bayes algoritmasının modelini eğitmek için, test veriseti ise modelin doğruluğunu test etmek için kullanılmıştır. Tüm bu süreçler pipeline oluşturularak gerçekleştirilmiştir.

Spark pipeline, Apache Spark’ın veri işleme çatısı içerisindeki bir kavramdır. Bir “Spark pipeline”, veri işleme süreçlerini yönetmek, organize etmek ve otomatikleştirmek için kullanılan bir dizi veri işleme adımını temsil eder. “Spark pipeline”, Spark’ın yüksek performanslı ve ölçeklenebilir veri işleme yeteneklerini kullanarak karmaşık veri işleme iş akışlarının kolayca yönetilmesini sağlar. Ayrıca, Spark’ın sağladığı otomatik optimizasyonlar sayesinde veri işleme işlemleri paralel olarak ve verimli bir şekilde gerçekleştirilir. “Spark pipeline” bu sebeplerden ötürü tercih edilmiştir ve kullanımına Şekil 3.41’de yer verilmiştir.

```

VectorAssembler vectorAssembler=new VectorAssembler()
    .setInputCols(new String[]{"SDVF-22", "SDVF-42", "SDVF-3",
"SDVF-52", "SDVF-8", "SDVF-11", "SDVF-9", "SDVF-12", "SDVF-53", "SDVF-56"});
    .setOutputCol("features");
Dataset<Row> output = vectorAssembler.transform(preparedDF);
output.show();
NaiveBayes nb = new NaiveBayes();
Pipeline pipeline=new Pipeline()
    .setStages(new PipelineStage[]{vectorAssembler,nb});
Dataset<Row>[] splitDF=preparedDF.randomSplit(new double[] {0.8,0.2});
Dataset<Row> trainingData = splitDF[0];
Dataset<Row> testData = splitDF[1];
PipelineModel pipelineModel=pipeline.fit(trainingData);
Dataset<Row> predictionDF=pipelineModel.transform(testData);
predictionDF.show(false);

```

Şekil 3.41. Model Eğitme Kod Bloğu

Spark MLLIB API'sı vasıtasıyla Naive Bayes algoritması modeli eğitilir. Test verileriyle bu modelin doğruluğu değerlendirilir. Oluşturulan modelin verilerine şekil 3.42.'de yer verilmektedir.

ID	SDVF-22	SDVF-42	SDVF-3	SDVF-52	SDVF-8	SDVF-11	SDVF-9	SDVF-12	SDVF-53	SDVF-56	SDVF-35	SDVF-45	SDVF-24	SDVF-25	SDVF-33	SDVF-47	Label	features	rawPrediction
128	10	11	10	12	10	10	10	11	10	12	10	10	11	10	11	10	10	[[0.0,1.0,1.0,0.0,3.0,2.0,0.0,0.0,0.0,0.0]]	[-8.435489967620766, -8.435489967620766]
131	17	10	10	110	11	12	10	11	110	10	10	10	10	12	10	11	10	[[7.0,0.0,0.0,10.0,1.0,2.0,0.0,1.0,10.0,0.0]]	[-3.979398102900795, -3.979398102900795]
136	15	10	10	14	11	15	10	11	14	10	10	10	15	10	12	10	11	[[5.0,0.0,0.0,4.0,1.0,5.0,0.0,1.0,4.0,0.0]]	[[1.8461167245988279, -1.8461167245988279]]
144	121	10	10	10	13	12	10	12	10	10	10	10	15	10	11	10	10	[[21.0,0.0,0.0,0.0,3.0,2.0,0.0,2.0,0.0,0.0]]	[[3.631808116997154, -3.631808116997154]]
154	14	10	10	110	12	12	10	12	110	10	10	10	15	10	12	11	11	[[4.0,0.0,0.0,10.0,2.0,2.0,0.0,2.0,10.0,0.0]]	[[4.805897079362582, -4.805897079362582]]
159	127	10	11	10	13	12	10	12	14	12	10	10	10	11	12	10	11	[[19.0,0.0,1.0,0.0,3.0,2.0,0.0,2.0,0.0,1.0]]	[[0.23048998585646313, -0.23048998585646313]]
165	120	10	10	110	11	16	10	10	15	10	10	10	13	11	12	11	10	[[10.0,3.4,5.8], [20.0,10.0,1.0,4.0,5.8]]	[[2.4828925871788024, -2.4828925871788024]]
168	117	10	10	110	12	12	10	11	110	10	10	10	14	10	11	10	11	[[17.0,0.0,0.0,10.0,2.0,2.0,0.0,1.0,10.0,0.0]]	[[3.9889953844810278, -3.9889953844810278]]
171	122	10	10	19	12	12	10	10	19	10	11	13	11	12	10	11	10	[[10.0,3.4,5.8], [22.0,9.0,2.0,2.0,9.8]]	[[3.689625688228956, -3.689625688228956]]
175	115	10	10	10	13	12	10	11	17	10	11	12	10	12	11	11	10	[[13.0,0.0,0.0,0.0,3.0,2.0,0.0,1.0,7.0,0.0]]	[[3.2474988891356244, -3.2474988891356244]]
179	115	11	10	110	12	14	10	11	110	10	10	10	14	10	12	10	11	[[13.0,1.0,0.0,10.0,1.0,4.0,0.0,1.0,10.0,0.0]]	[[1.213238239884730, -1.213238239884730]]
182	116	11	11	110	12	13	11	11	110	12	10	10	10	10	10	10	11	[[10.0,1.0,1.0,10.0,3.0,2.0,0.0,2.0,10.0,1.0]]	[[2.2588943887767256, -2.2588943887767256]]
187	118	10	10	110	15	16	11	11	110	10	11	15	10	11	10	11	10	[[10.0,0.0,0.0,10.0,3.0,4.0,1.0,1.0,10.0,0.0]]	[[1.8913397140524839, -1.8913397140524839]]
194	16	11	10	110	10	14	10	11	110	10	10	12	11	12	11	11	10	[[4.0,1.0,0.0,10.0,0.0,4.0,0.0,1.0,10.0,0.0]]	[[1.2246643532915227, -1.2246643532915227]]
1117	113	10	10	110	12	13	10	12	110	10	13	14	11	12	10	11	10	[[13.0,0.0,0.0,10.0,2.0,3.0,0.0,2.0,10.0,0.0]]	[[3.9156278628349227, -3.9156278628349227]]
1119	113	10	10	110	11	16	10	11	110	10	13	13	11	12	10	11	10	[[13.0,0.0,0.0,10.0,1.0,4.0,0.0,1.0,10.0,0.0]]	[[3.5954999585158158, -3.5954999585158158]]
1200	102	11	10	19	13	12	10	12	110	10	11	15	10	12	11	11	10	[[22.0,1.0,0.0,7.0,3.0,2.0,0.0,2.0,10.0,0.0]]	[[1.6012935140830261, -1.6012935140830261]]
1221	14	10	11	110	11	12	10	12	110	11	11	13	10	12	10	11	11	[[4.0,0.0,1.0,10.0,3.0,2.0,0.0,2.0,10.0,1.0]]	[[1.7489196470116311, -1.7489196470116311]]
12210	11	10	110	11	12	10	11	11	11	10	11	12	10	11	10	11	10	[[0.0,1.0,0.0,10.0,1.0,2.0,0.0,1.0,10.0,1.0]]	[[0.475679976312301, -0.475679976312301]]
1225	120	10	10	110	10	13	10	11	110	10	10	12	11	10	10	11	10	[[10.0,3.5,7.8], [20.0,10.0,3.0,1.0,10.0]]	[[3.919438919492263, -3.919438919492263]]

Şekil 3.42. Spark MLLIB Model Eğitimi

Tablo 3.6. Modelin Değerlendirme Metrikleri

Parametre	Değer
Doğruluk	92.45%
Kesinlik	97.29%
Hassasiyet	92.30%
F1	92.60%

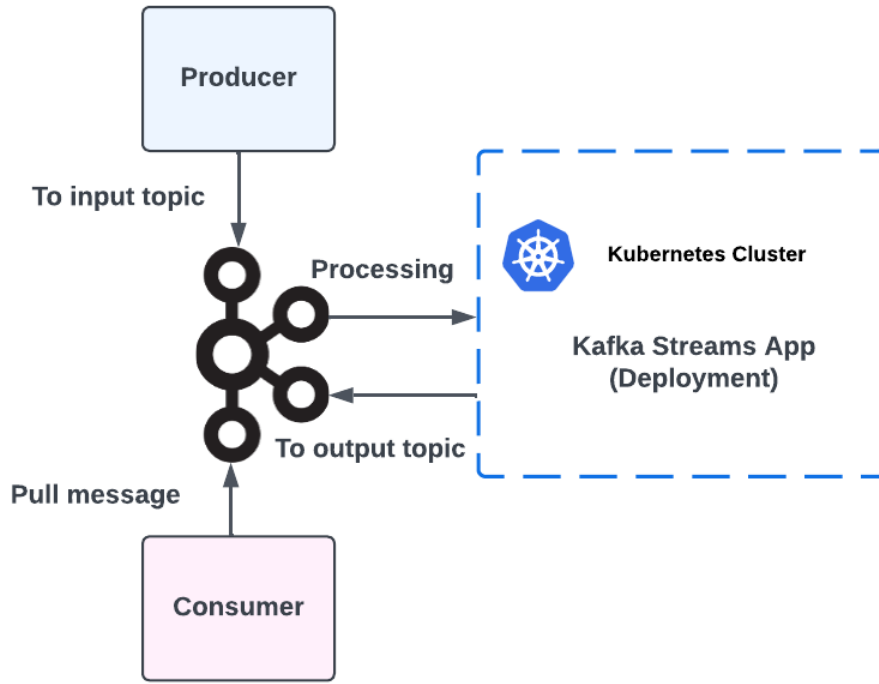
Naïve Bayes modelinin parametrelerine Tablo 3.6’da yer verilmiştir. Akış halinde gelen veriler bu model ile etkileşime girmektedir. Bu sayede anlık olarak gebelerin verilerine dair anksiyete ve depresyon teşhisi yapılmaktadır. Teşhisten sonra ise hastalığa dair elde edilen sonuçlar “result” adlı “topic”e gönderilmektedir.

Java ile kodlanan “Spark Streaming” kodu Spark klasörünün içinde bulunan “docker-image-tool.sh” dosyası yardımıyla imajlanıp Docker Hub’a atılmıştır. İki “worker-node” bir “master-node”den oluşan Kubernetes kümesinde ayağa kaldırılmıştır.

Kubernetes kümesinde ayağa kaldırılması için ise “-spark-submit” komutu kullanılmıştır. Bu komut ile aşağıdaki işlem adımlarında spark kodu çalıştırılmıştır;

- Spark Submit, bir istemciden master düğümdeki Kubernetes API sunucusuna gönderilir.
- Kubernetes, yeni bir “Spark Driver Pod” planlar.
- “Spark Driver Pod”, “Spark executor pod”ları istemek için Kubernetes ile iletişime geçer.
- Yeni “executor pod”lar Kubernetes tarafından planlanır.
- “Executor pod”lar çalışmaya başladığında Kubernetes, “Spark Driver Pod”lara yeni “Spark executor pod”ların hazır olduğunu bildirir.
- “Spark Driver Pod”, “executor pod”ları üzerinde görevleri planlar.

Tüm bu işlem adımlarını uygulayarak “-spark-submit” komutuyla Kubernetes üzerinde Apache Spark uygulaması ayağa kaldırılmıştır. Şekil 3.43.’te Kubernetes kümesinin ana işlem diyagramı resmedilmiştir.



Şekil 3.43. Ana İşlem Diyagramı

- ```

helmspark-master-0:/opt/bitnami/spark$./bin/spark-submit -
-master k8s://https://192.168.56.110:6443 --deploy-mode
cluster --name perinatalspark --class Spark --conf
spark.executor.instances=2 --conf
spark.kubernetes.authenticate.driver.serviceAccountName=hel
mspark --conf
spark.kubernetes.container.image=nurbanu/spark:sparkcode --
conf
spark.kubernetes.driver.volumes.persistentVolumeClaim.check
pointpvc.mount.path=/checkpoint --conf
spark.kubernetes.driver.volumes.persistentVolumeClaim.check
pointpvc.mount.readOnly=false --conf
spark.kubernetes.driver.volumes.persistentVolumeClaim.check
pointpvc.options.claimName=spark-pv-claim2
local:///opt/spark/examples/jars/sparkcode.jar

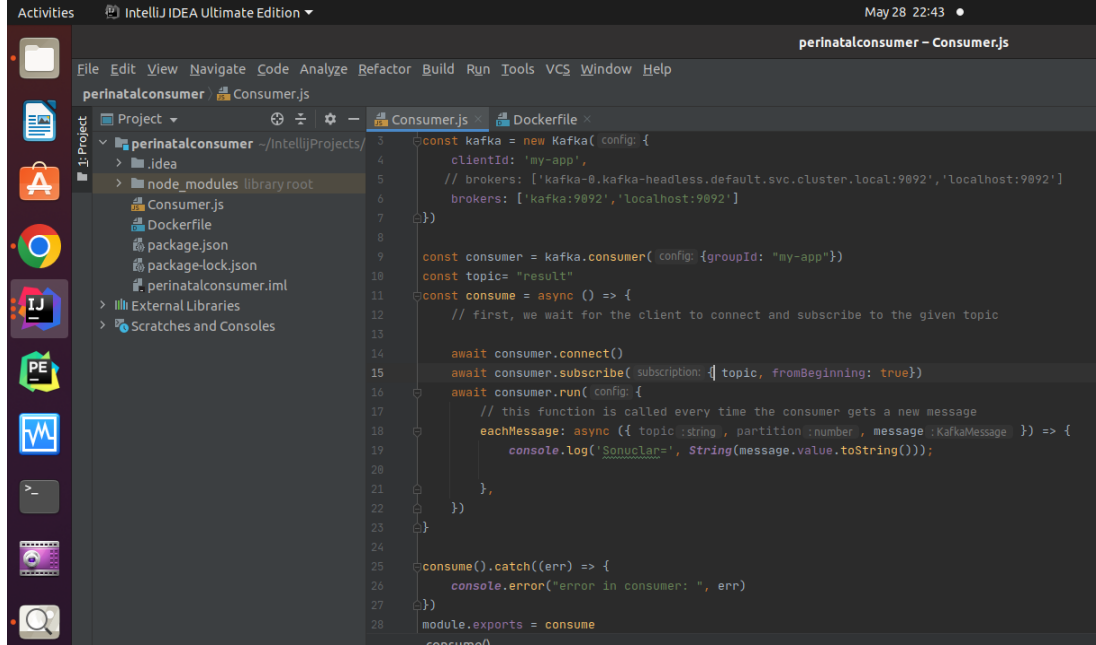
```

### 3.2.5.5. Kafka consumer

Spark modulünde gebe verileri işlenmiş ve hasta olup olmama durumlarını göre etiketlenmişlerdir. Etiketlenen veriler “result” adlı topic’e gönderilmişlerdir. Bu

modülde ise etiketlenen hasta verileri “result” adlı topic’i dinleyip anlık olarak gelen verileri ekrana basma işlemini gerçekleştirmektedir.

Bu modül de node.js ile kodlanmıştır. Şekil 3.44.’te Kafka Consumer kodunun ekran görüntüsü yer almaktadır.

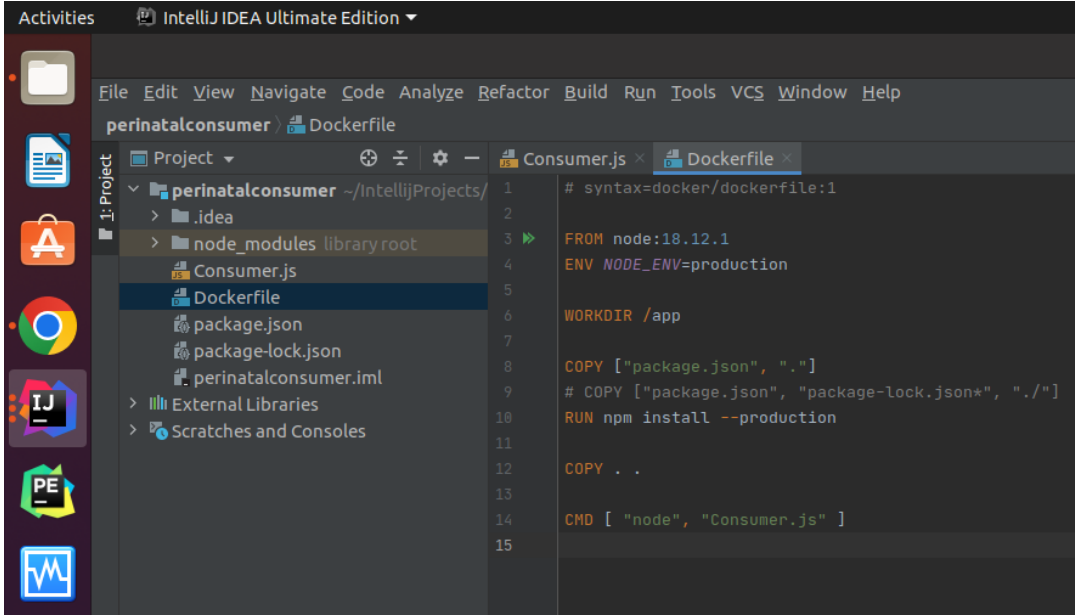


```
3 const kafka = new Kafka(config: {
4 clientId: 'my-app',
5 // brokers: ['kafka-0.kafka-headless.default.svc.cluster.local:9092', 'localhost:9092']
6 brokers: ['kafka:9092', 'localhost:9092']
7 })
8
9 const consumer = kafka.consumer(config: {groupId: "my-app"})
10 const topic= "result"
11 const consume = async () => {
12 // first, we wait for the client to connect and subscribe to the given topic
13
14 await consumer.connect()
15 await consumer.subscribe(subscription: { topic, fromBeginning: true})
16 await consumer.run(config: {
17 // this function is called every time the consumer gets a new message
18 eachMessage: async ({ topic:string, partition:number, message :KafkaMessage }) => {
19 console.log('Sonuclar=', String(message.value.toString()));
20 },
21 },
22)
23 }
24
25 consume().catch((err) => {
26 console.error("error in consumer: ", err)
27 })
28 module.exports = consume
29 consume()
```

Şekil 3.44. Kafka Consumer Kodu

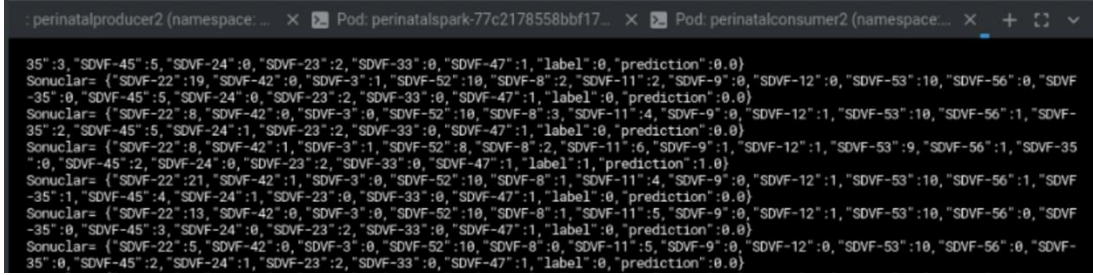
Spark consumer kodu Şekil 3.45.’da görünen “dockerfile” dosyası yardımıyla imaj haline getirilmiş ve Docker Hub’a atılmıştır. İki “worker-node” bir “master-node”dan oluşan Kubernetes Cluster’da ayağa kaldırılmıştır.

- nurbanu@master:~\$ kubectl run --image=nurbanu/perinatalconsumer2 perinatalconsumer



Şekil 3.45. Docker File

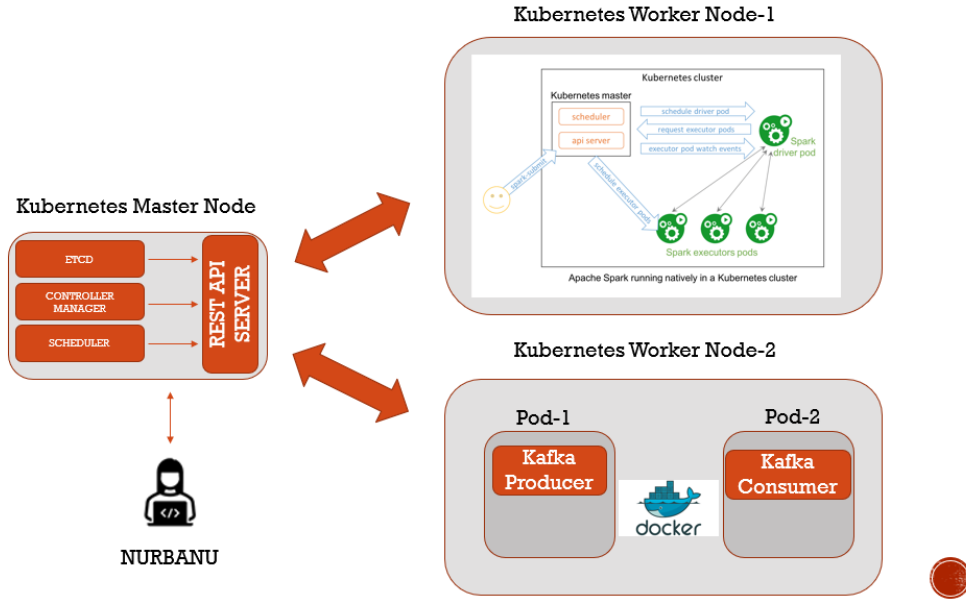
Bu sayede perinatal dönemdeki gebelerin anksiyete ve depresyon hastalıkları teşhisi sonuçları, anlık olarak işlenmiş ve ekrana bastırılmıştır (Şekil 3.46).



Şekil 3.46. Kafka Consumer Ekran çıktısı

Bu sistemin Kubernetes kümesi ortamına taşınması yani ölçeklenebilir olarak çalıştırılması sadece Sakarya’da varolan hastenede değil, tüm illere hatta tüm Türkiye’ye hitap etmesi, geliştirilen sistem ile mümkündür.





Şekil 3.47. Kubernetes Mimarisi

Tüm bu işlemlerin neticesinde Kubernetes kümesinde üç tane “pod” ayağa kaldırılmıştır (Şekil 3.47). Kubernetes “pod”ların hangi “node”larda ayağa kalkacağına kararını vererek kullanıcı dostu bir politika izlemektedir. Talepler ve iş yükü dalgalanmaları olduğunda, sistem kaynaklarını artırarak veya azaltarak esneklik sağlayabilmektedir.

- `nurbanu@master:~$ kubectl run --image=nurbanu/perinatalproducer2 perinatalproducer`
- `helmspark-master-0:/opt/bitnami/spark$ ./bin/spark-submit -master k8s://https://192.168.56.110:6443 --deploy-mode cluster --name perinatalspark --class Spark --conf spark.executor.instances=2 --conf spark.kubernetes.authenticate.driver.serviceAccountName=helmspark --conf spark.kubernetes.container.image=nurbanu/spark:sparkcode --conf spark.kubernetes.driver.volumes.persistentVolumeClaim.checkpointpvc.mount.path=/checkpoint --conf spark.kubernetes.driver.volumes.persistentVolumeClaim.checkpointpvc.mount.readOnly=false --conf spark.kubernetes.driver.volumes.persistentVolumeClaim.checkpointpvc.options.claimName=spark-pv-claim2 local:///opt/spark/examples/jars/sparkcode.jar`

- `nurbanu@master:~$ kubectl run --image=nurbanu/perinatalconsumer2 perinatalconsumer`

Kubernetes kümesinde bir “node”da ayağa kalkan spark kodunu submit ederken parametreleri belirleyerek ölçeklenebilirliği etkileyen “executor” sayısını ayarlamak mümkündür.

“Spark executor” sayısı, Spark uygulamasının performansı ve ölçeklenebilirliği açısından önemlidir.

“Executor”lar, Spark uygulamasının iş yükünü paralel olarak çalıştıran işlem birimleridir. Daha fazla “executor” kullanmak, iş yükünü daha fazla paralel işlemci üzerinde dağıtabilir ve işlerin eşzamanlı olarak çalışmasını sağlayabilir. Bu, işlemlerin daha hızlı tamamlanmasını sağlar.

Her “executor”, Spark uygulaması için ayrılan kaynaklara sahiptir, yani CPU, bellek vb. “Executor” sayısı, toplam kullanılabilir kaynak miktarını belirler. İş yüküne uygun bir “executor” sayısı seçmek, kaynakların etkin bir şekilde kullanılmasını sağlar ve uygulamanın performansını artırır.

“Executor” sayısını doğru bir şekilde ayarlamak, Spark uygulamasının ölçeklenebilirliğini etkiler. Daha fazla “executor” kullanarak, uygulamayı daha büyük veri setleri veya daha karmaşık işlemlerle başa çıkmak üzere ölçeklendirilebilir. Bu, büyük veri analitiği veya makine öğrenimi gibi yoğun iş yükleriyle çalışırken önemlidir.

Ancak, “executor” sayısını seçerken dikkatli olunması gerekmektedir. Her bir “executor”, bellek ve CPU gibi kaynakları kullanır. “Executor” sayısını gereksiz yere artırmak, kaynak tükenmesine ve performans sorunlarına yol açabilir.

Bu çalışmamızda optimum “executor” sayısını ayarlayabilmek adına performans karşılaştırmaları yapılmış ve sistemin işlevselliği test edilmiştir.

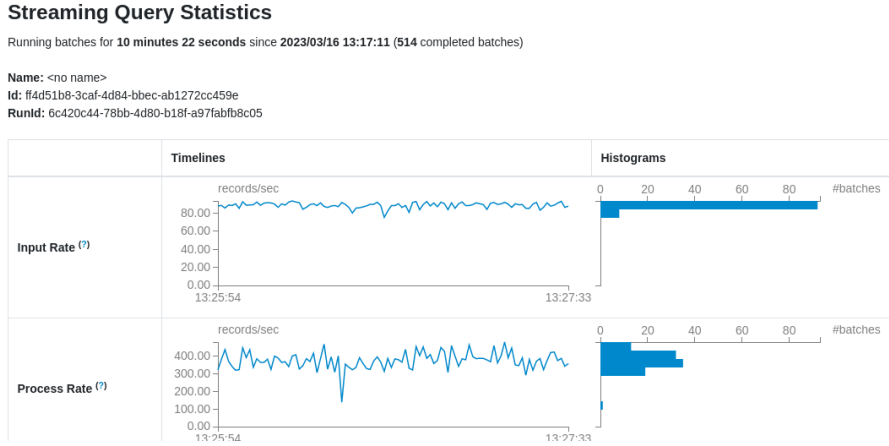
### **3.2.5.6. Performans karşılaştırması**

Streaming halde gelen veri üzerinde çeşitli performans karşılaştırmaları yapılmıştır. “Kafka Producer”dan gelen verilerin sıklığı, yollanan veri boyutu ve “executor” miktarı değiştirilmiştir. Bu esnek mimari, birçok girdiyi test etmek için oldukça uygundur ve geniş bir test kapsamını destekleyebilir.

Her bir “executor” sanal makinelerde ayrılan kaynakları tüketmektedir. Bu sebeple elde edilen çıktılar neticesinde, kaynak arttırıldığı takdirde sistem üzerinde daha yüksek performanslar ortaya çıkaracağını söylemek mümkündür.

İşlem hızının (process rate) aynı girdi hızında (input rate) sabit kalması sistemin hız ihtiyaçlarını karşıladığı anlamına gelmektedir. Bu, işin işleme kapasitesinin girdi verilerini işlemek için yeterli olduğu anlamına gelir. Yüksek gecikme süresinin ana nedeni, her toplu’da (batch) çok fazla veri olmasıdır. Böyle bir durumda da paralellliği arttırarak gecikme azaltılabilir. Bunu test edebilmek için yapılan testin sonuçları Şekil3.49’da verilmiştir.

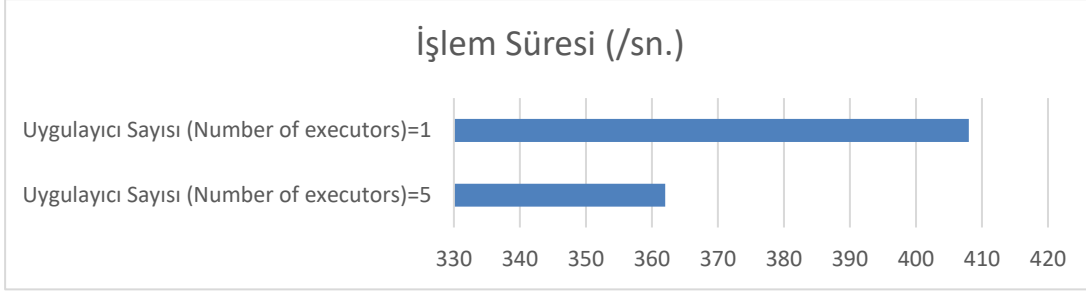
Şekil 3.48.’de localhost:4040’dan hizmet veren Spark Web arayüzü gösterilmiştir. Girdi hızı ve işlem hızı oranlarına bakıldığında, sistemin gelen veriyi gecikme olmadan işleyebildiğini söylemek mümkündür.



**Şekil 3.48.** Spark WebUI

“Executor” sayısının performansa etkisi test etmek için akış halindeki verinin “executor” sayısı değiştirilmiş ve performans çıktılarına bakılmıştır. Bu çıktılar neticesinde işlem süresinin “executor” sayısı arttıkça azaldığı gözlemlenmiştir. Bu ölçeklenebilir mimari ile, özellikle büyük veri işlemlerinde zaman açısından olumlu sonuçlar sağlayarak, sürecin hızlandırılmasına olanak tanıdığı sonucuna varılacağını söylemek mümkündür.

Gerçekleştirilen test düzeneği, on dakika boyunca saniyede bir satır veri yollayarak çalışacak şekilde düzenlenmiştir. “Executor” sayıları bir ve beş olan iki ayrı performansın sonucu değerlendirilmiştir. Bu değerlendirme sonucunda “executor” sayısı fazla olanın olmayana göre daha hızlı işlem yaptığı gözlemlenmiştir. Şekil 3.49’da grafiksel olarak gösterimi mevcuttur.



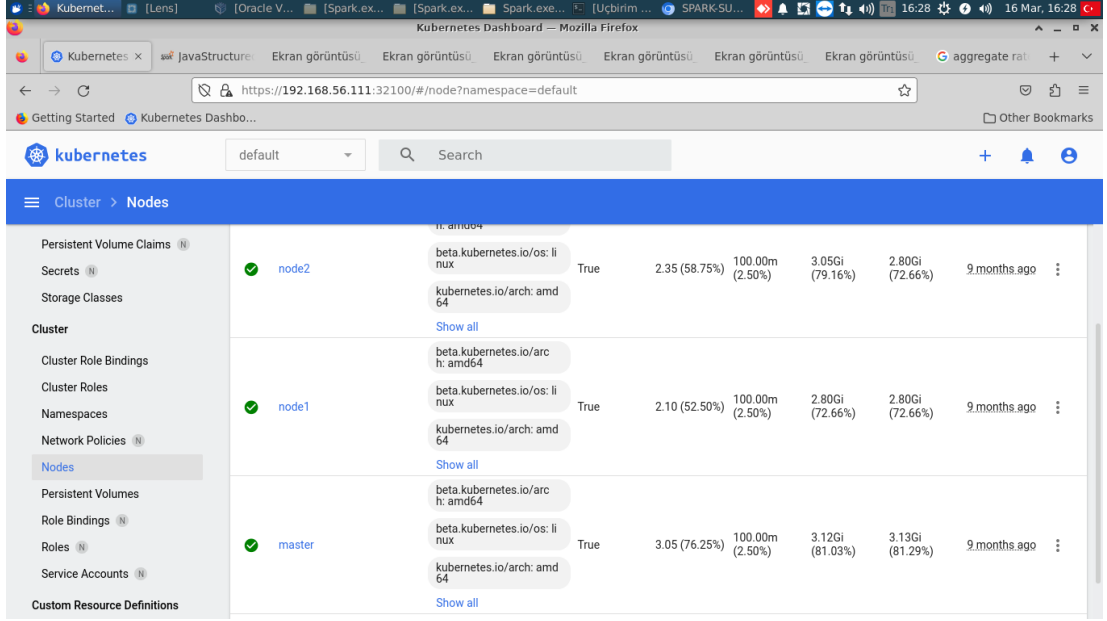
**Şekil 3.49.** “Executor” Sayısının İşlem Hızına Etkisi

Kaynak tüketiminin gösterimi için “Kubernetes Dashboard” kullanılmıştır. “Kubernetes Dashboard”, Kubernetes kümelerini yönetmek için kullanılan bir web tabanlı grafik arayüzdür. “Kubernetes Dashboard”, kullanıcıların kümelerini izlemelerine, uygulamaları yönetmelerine ve kaynakları görüntülemelerine olanak tanır. Ayrıca, cluster üzerindeki “pod”ları, servisleri, replika setlerini, “deployment”ları ve diğer Kubernetes kaynaklarını görüntüleyebilir ve yönetebilir.

“Kubernetes Dashboard”, kullanıcı dostu bir arayüz sağlayarak, karmaşık YAML dosyalarıyla uğraşmadan Kubernetes kümelerini kontrol etmeyi kolaylaştırır. Kullanıcılar, bir web tarayıcısı üzerinden “Dashboard”a erişerek kümelerini izleyebilir, uygulamaları başlatabilir, durdurabilir veya yeniden başlatabilir, kaynak kullanımını izleyebilir ve hata ayıklama yapabilirler.

“Kubernetes Dashboard”, kullanıcıların kaynakları görüntüleme, oluşturma, düzenleme ve silme gibi temel operasyonları gerçekleştirmelerini sağlar. Ayrıca, loglara erişme, “pod”lar arasında bağlantı kurma ve uygulama günlüklerini görüntüleme gibi ek işlevler de sunabilir.

“Kubernetes Dashboard” arayüzünde, kümelerin durumu ve kaynak kullanımı gösterilmektedir. Ayrıca kümenin genel bilgileri, çalışan “pod”ları, servisleri, “deployment”ları, “replicaset”leri ve diğer kaynakları görüntülenebilir. Şekil 3.50’ de Kubernetes Dashboard yer almaktadır.



Şekil 3.50. Kubernetes Dashboard

Her bir “executor”, bellek ve CPU gibi kaynakları kullanır. Bu nedenle, “executor” sayısını belirlerken, küme üzerindeki kaynakların ve iş yükünün gereksinimlerini dikkate alıp duruma göre “executor” sayısı belirlemek önemlidir. Yeterli kaynakların sağlanması, “executor”ların etkin bir şekilde çalışmasını ve performanslarını artırmasını sağlar. Aksi takdirde, kaynak tükenmesi, performans sorunları ve hatalarla karşılaşma riski olabilir.

Yapılan test düzeneğinde “executor” sayısı bir ve beş olarak değiştirilmiş kaynak tüketimi üzerindeki etkisi incelenmiştir (Şekil 3.51). Bu bağlamda “executor” sayısındaki artışın kaynak tüketimi üzerinde oldukça etkili olduğu gözlemlenmiştir (Şekil 3.52, Şekil 3.53, Şekil 3.54).

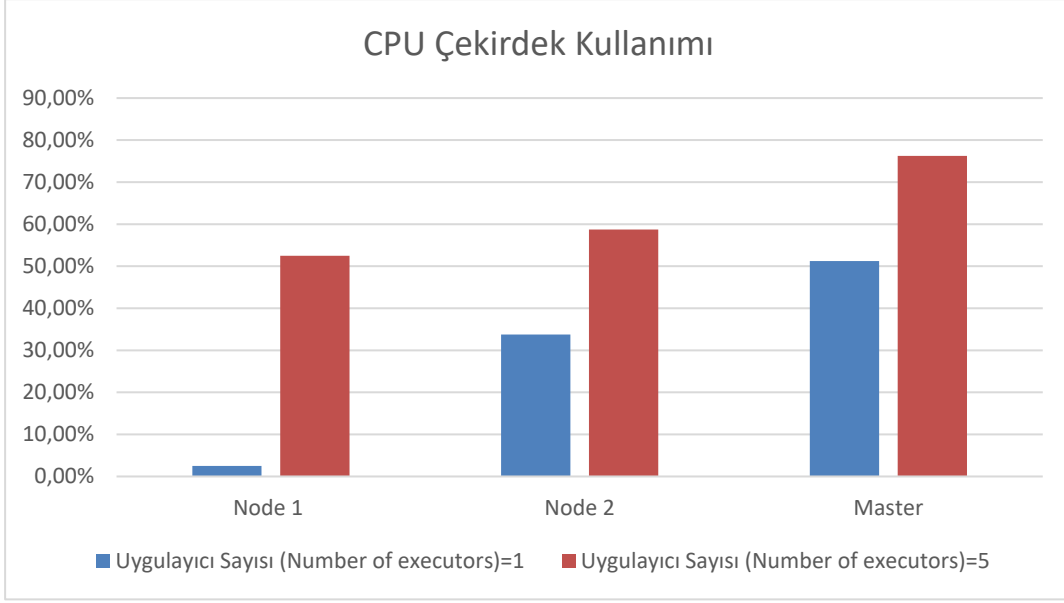
Executors

Show 20 entries

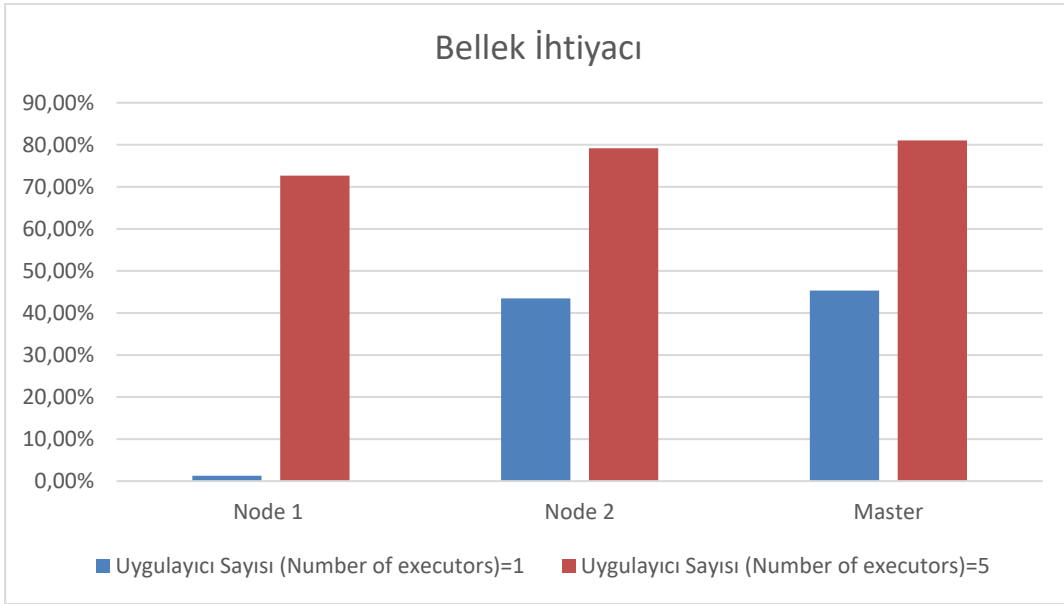
Search:

| Executor ID | Address                                                     | Status | RDD Blocks | Storage Memory       | Disk Used | Cores | Active Tasks | Failed Tasks | Complete Tasks | Total Tasks | Task Time (GC Time) | Input    | Shuffle Read | Shuffle Write | Thread Dump                 |
|-------------|-------------------------------------------------------------|--------|------------|----------------------|-----------|-------|--------------|--------------|----------------|-------------|---------------------|----------|--------------|---------------|-----------------------------|
| driver      | perinatalspark-b8f23986ea9047ad-driver-svc.default.svc:7079 | Active | 0          | 73.1 KiB / 413.9 MiB | 0.0 B     | 0     | 0            | 0            | 0              | 0           | 0.0 ms (0.0 ms)     | 0.0 B    | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |
| 1           | 10.244.1.111:44187                                          | Active | 0          | 0.0 B / 413.9 MiB    | 0.0 B     | 1     | 0            | 0            | 107            | 107         | 20 s (0.3 s)        | 16.5 KiB | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |
| 2           | 10.244.0.154:36331                                          | Active | 0          | 0.0 B / 413.9 MiB    | 0.0 B     | 1     | 0            | 0            | 107            | 107         | 1.6 min (2 s)       | 0.0 B    | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |
| 3           | 10.244.2.26:44769                                           | Active | 0          | 24.4 KiB / 413.9 MiB | 0.0 B     | 1     | 0            | 0            | 106            | 106         | 18 s (0.4 s)        | 1.3 KiB  | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |
| 4           | 10.244.1.112:38805                                          | Active | 0          | 24.4 KiB / 413.9 MiB | 0.0 B     | 1     | 0            | 0            | 100            | 100         | 17 s (0.4 s)        | 0.0 B    | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |
| 5           | 10.244.0.155:42973                                          | Active | 0          | 0.0 B / 413.9 MiB    | 0.0 B     | 1     | 0            | 0            | 116            | 116         | 21 s (0.4 s)        | 0.0 B    | 0.0 B        | 0.0 B         | <a href="#">Thread Dump</a> |

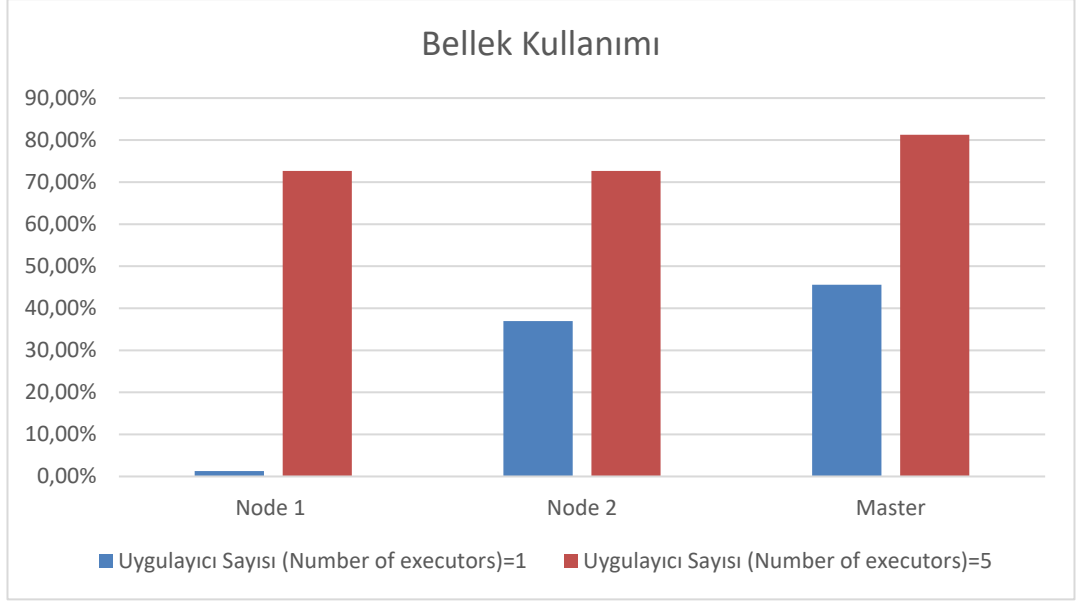
Şekil 3.51. “Executor”ların Kaynak Tüketimi (Spark WebUI)



**Şekil 3.52.** CPU Çekirdek İhtiyacı Performans Karşılaştırması



**Şekil 3.53.** Bellek İhtiyacı Performans Karşılaştırması



**Şekil 3.54.** Bellek Kullanımı Performans Karşılaştırması

“Executor” sayısındaki artışın, kaynak tüketimini artırdığı yapılan çalışmada elde edilen sonuçlar neticesinde görülmüştür. “Executor” sayısındaki artış, iş yükünün daha iyi paralelleştirilmesini sağlayabilmekte ve işlemlerin daha hızlı tamamlanmasına yardımcı olabilmektedir. Ancak, “executor” sayısını belirlerken kaynakların ve iş yükünün gereksinimlerini dikkate almak önemlidir. Yetersiz kaynaklar, performans sorunlarına ve hatalara yol açabilirken, aşırı kaynak kullanımı ise kaynak tükenmesine ve küme dengesizliklerine neden olabilmektedir. “Executor” sayısının doğru bir şekilde ayarlanması, dengeli bir kaynak dağılımını ve optimum performansı sağlamaktadır.





#### 4. SONUÇ VE TARTIŞMA

Bilimi ve endüstriyi yönlendiren güçlerden biri makine öğrenimidir. İnsanlar tarafından yazılan karmaşık algoritmalar yerine, bilgisayarların verilerden öğrenmesini sağlayan bu teknoloji, birçok sektörde devrim yaratmıştır.

Makine öğrenimi, büyük veri setleri üzerinde desenler ve ilişkiler keşfetmek için kullanılan bir araçtır. Bu, işletmelerin verilerini daha iyi anlamalarını, gizli bilgileri ortaya çıkarmalarını ve gelecekteki eğilimleri tahmin etmelerini sağlar. Makine öğrenimi aynı zamanda otomatik karar verme süreçlerinde de büyük bir rol oynar. Birçok alanda, insanların yapması zor veya zaman alıcı olan görevler, makinelere devredilebilir. Bu sebeplerden dolayı, makine öğrenimi günümüzde giderek daha önemli hale gelen bir teknolojidir. Ancak büyük verinin yaygınlaşması, makine öğrenimi tekniklerini uygulamadaki geleneksel yaklaşımlardan farklı hızlara sahip bu büyük miktardaki veriye yeni kullanım alanları gerektirir.

Büyük veri, makine öğrenmesinde önemli bir rol oynayan değerli bir kavramdır. Günümüzde, çeşitli sektörlerde hızla artan veri miktarı, işletmelerin daha iyi kararlar almasını, öngörüler yapmasını ve müşteri ihtiyaçlarını daha iyi anlamasını sağlar. Bu büyük miktarda verinin gerekli koşullarda işlenebilmesi için büyük veri analitiğinde kullanılan teknolojilere ihtiyaç duyulmaktadır.

Büyük veri analitiği teknolojileri, günümüzde büyük veri setlerini işlemek, analiz etmek ve değerli bilgiler çıkarmak için vazgeçilmez bir rol oynamaktadır. Büyük veri analitiği araçları, işletmelerin karmaşık veri kütleleriyle başa çıkmasını, desenleri keşfetmesini ve daha iyi kararlar almasını sağlar. En önemli avantajlarından biri, büyük veri analitiği araçlarının veri işleme kapasitesidir. Büyük veri setleri genellikle yüksek hacimde, hızlı bir şekilde oluşur. Bu verileri işlemek ve analiz etmek, geleneksel yöntemlerle mümkün olmayabilir. Büyük veri analitiği araçları, bu büyük veri setlerini paralel ve dağıtık bir şekilde işleyerek hızlı analizler yapılmasını sağlar. Bunun yanı sıra veri görselleştirme ve raporlama yetenekleri sunar. Büyük veri setleri genellikle karmaşık ve hacimli olduğundan, verilerin anlamlı hale getirilmesi ve paylaşılması önemlidir. Büyük veri analitiği teknolojisi, verileri grafikler, tablolar ve

interaktif görsellerle görselleştirerek daha iyi anlaşılmasını sağlar. Ayrıca, raporlama özellikleriyle kullanıcılara verileri özetleyen ve paylaşılabilir raporlar oluşturma imkanı sunar. Özetle, büyük veri analitiği teknolojileri, işletmelerin büyük veri setlerini etkili bir şekilde yönetmesine, analiz etmesine ve değerli bilgiler çıkarmasına yardımcı olan önemli teknolojilerdir. Bu teknolojiler, veri işleme kapasitesi, analiz yetenekleri, görselleştirme ve raporlama özellikleri, tahmin ve öngörü yetenekleri gibi özellikleriyle büyük verinin işlenmesini ve anlamlandırılmasını kolaylaştırır. Büyük veri kapsamına giren birçok alan bulunmaktadır ve en önemli kritik alanlardan biri hızlı karar verme gerektiren tıp alanıdır. Psikiyatri ise, tıp alanında ruh sağlığına odaklanan bir dal olarak önemli bir yer tutmaktadır. Zihinsel, duygusal ve davranışsal bozuklukları inceleyen psikiyatri, insanların ruh sağlığını değerlendirir, tanı koyar ve uygun tedavi yöntemlerini uygular.

Ruh sağlığı hastalıkları, bireylerin yaşamını derinden etkileyen ve dikkatle tedavi edilmesi gereken sağlık sorunlarıdır. Hamilelik sürecini içeren perinatal dönemde ortaya çıkan anksiyete ve depresyon gibi psikolojik bozukluklar zamanında tespit edilmezse hem anne hem de bebek üzerinde olumsuz etkiler yaratacaktır. Perinatal dönem ve gebe üzerindeki etkileri, toplumun ruh sağlığı açısından büyük önem taşımaktadır. Bu sebeple, psikiyatride hızlı teşhis süreci büyük bir öneme sahiptir. Zihinsel sağlık sorunlarının erken teşhisi, uygun tedaviye erken başlanmasını ve hastaların daha iyi sonuçlar elde etmesini sağlamaktadır.

Perinatal dönemde sık rastlanan anksiyete ve depresyon, hekimler tarafından bilinen, olumsuz sonuçları da her geçen gün daha iyi anlaşılabilir hastalıklar olmasına rağmen yeterince tanınmamakta, tanınsa bile yeterli tedavi uygulanmamaktadır. Toplum üzerinde ciddi zararlara sebep olması nedeniyle de perinatal dönemde anksiyete ve depresyon hastalığı erkenden teşhis edilmeye ve tedavi edilmeye ihtiyaç duymaktadır. Bu ihtiyacı büyük veri analitik sistemleri ile çözümlenebilir hale getirmek mümkündür.

Büyük veri analitiği ile, gebe verilerinin hızlı bir şekilde işlenmesini ve anlamlı bilgilere dönüştürülmesini sağlamak mümkündür. Bu da doktorlara, sağlık uzmanlarına ve araştırmacılara hızlı ve doğru teşhisler, tedavi planları ve sağlık hizmetleri sunma imkanı verebilmektedir.

Çalışmamızda, perinatal dönemdeki kadınların anksiyete ve depresyon teşhisinde anlık olarak sonuç üretebilen, küme yapısı sayesinde ölçeklenebilir mimarinin de

entegre olduđu, büyük veri analitik sistemi gerçeklenmiştir. Bu sistemi oluştururken Apache Kafka, Apache Spark, Kubernetes gibi teknolojilerden faydalanılmıştır. Böylece perinatal dönemdeki kadınlarda depresyon ve anksiyetenin yol açabileceđi potansiyel zararın en aza indirilmesi amaçlanmıştır.

Bu çalışmanın birincil amacı, performansı optimize edilmiş algoritmalar aracılığıyla özellikleri çıkararak hamile kadınlarda anksiyete ve depresyonla ilgili sonuç üzerinde etkili soruları belirlemektir. Bu sayede daha az soru ile daha kısa sürede sonuca ulaşmak amaçlanmaktadır. Bu çalışmanın bir sonraki hedefi, temizlenen bu veriler üzerinde en optimum sonucu veren makine öğrenmesi algoritmasını bulmaktır. Bir sonraki hedefi ise, büyük veri akışında makine öğrenimi modellerini kullanmaya odaklanan Apache Spark büyük veri işleme motorunu temel alan hamile kadınlarda depresyon ve anksiyete için anlık bir uzaktan sağlık durumu tahmin modülü oluşturmaktır. Son hedefi ise, bu amaçları içeren modülleri Kubernetes kümeleme sistemine entegre ederek ayağa kaldırmaktır.

Bu çalışmada öncelikle önışlem aşamalarından geçirilerek veri işlenmeye hazır hale getirilmiştir. Gebelerin cevaplaması gereken altmış adet soru bulunmaktadır. Bu sayı optimum olarak belirlenen optimize edilmiş seçim-evrimsel özellik seçimi algoritması ise onaltıya indirilmiştir. Bunlar; gebelik haftası, bebeđin sağlık problem, yaşam yeri (kent veya kırsal), eşile iletişimi, eğitim seviyesi, evinde yaşayan insan sayısı, çalışma durumu, toplam gelir seviyesi, eşinden gördüğü duygusal destek, problemlerini paylaşacağı insan varlığı, egzersiz durumu, eşinin eğitim durumu, istenen gebelik olup olmadığı, bebeđin cinsiyeti, sigara içme durumu, eşinin kronik rahatsızlık durumu özellikleridir. Bu veriler üzerinde gebenin sağlık durumunu tahmin etmek için çeşitli algoritmalar (Decision Tree, Naive Bayes, K-Nearest Neighbor, Random Forest, Gradient Boosted Tree, Logistic Regresyon ve Deep Feed Forward Neural Network) çalıştırılmıştır. Bu veri kümesi için sırasıyla %90,8 ve %81,71 doğruluk ve kesinlik ile en iyi sonuçları üreten Naive Bayes makine öğrenimi algoritması olduđu sonucuna varılmıştır.

Yapılan tüm çalışmalar Kubernetes kümesinde ayağa kaldırılmış, anlık olarak hastalık hakkında sonuç üreten büyük veri için elverişli ölçeklenebilir mimariyi kapsayan altyapıyı geliştirilmiştir. Bu altyapı üzerinde performans testi yaparak “executor” sayısındaki artışta sonuç üzerinde etkileri gözlemlenmiştir. Performans testleri neticesinde paralelleşmede etkili spark “executor”ın hız anlamında olumlu yönde katkı

sağladığı sonucuna varılmıştır. Bunun yanında “executor” sayısındaki artışın kaynak tüketimindeki etkisi de incelenmiş, CPU ve bellek kaynak kullanımlarına diyagramlarda yer verilmiştir.

Bu bağlamda, altyapının sağlık alanı kullanımının oldukça makul sonuçlara yol açtığı sonucuna varılabilmektedir. Bu ölçeklenebilir büyük veri platformunun yardımıyla, hamile kadınlarda anksiyete ve depresyonu tespit etmek için zaman alan prosedürün yerini, gerçek zamanlı olarak önemli doğruluk ve hızda çalışan bilgisayar tabanlı sistemlere bırakabileceği sonucuna varılabileceğini söylemek mümkündür.

## KAYNAKLAR

- Abdel-Fattah, M. A., Othman, N. A. ve Goher, N. (2022). Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark. *Hindawi Computational Intelligence and Neuroscience*, 2022. doi:10.1155/2022/9898831
- Adegoke, D. (2023). A systematic review of big data and digital technologies security leadership outcomes effectiveness during natural disasters. *Sustainable Futures*, 5(March), 100113. doi:10.1016/j.sfr.2023.100113
- Ahmed, H., Younis, E. M. G., Hendawi, A. ve Ali, A. A. (2020). Heart disease identification from patients ' social posts , machine learning solution on Spark. *Future Generation Computer Systems*, 111, 714–722. doi:10.1016/j.future.2019.09.056
- Akın A., Üner S., Arslan Ş., Yıldırım N., Aslan D., Erdost T., Erbaydar N., Ç. ve E., Coşkun A., K. Ö. (2008). Kadının Statüsü ve Sağlığı ile İlgili Gerçekler. *Kadının Statüsü Genel Müdürlüğü*.
- Apache Hadoop. (2023). <https://hadoop.apache.org/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Apache Kafka. (2023). <https://kafka.apache.org/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Apache Spark. (2023). <https://spark.apache.org/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Apache Storm. (2023). <http://storm.apache.org/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Arai, Y., Kondo, T., Fuse, K., Shibasaki, Y., Masuko, M., Sugita, J., ... Atsuta, Y. (2018). Prediction of Acute Graft-Versus-Host Disease Following Allogeneic Hematopoietic Stem Cell Transplantation Using a Machine Learning Algorithm. *Blood*, 132(Supplement 1), 68–68. doi:10.1182/BLOOD-2018-99-114794
- Bag, S., Dhamija, P., Kumar, R. ve Sabbir, M. (2023). Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain : An empirical study. *Journal of Business Research*, 154(September 2022), 113315. doi:10.1016/j.jbusres.2022.113315
- Bormann, C. L., Kanakasabapathy, M. K., Thirumalaraju, P., Gupta, R., Pooniwala, R., Kandula, H., ... Shafiee, H. (2020). Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *eLife*, 9, 1–14. doi:10.7554/ELIFE.55301

- Bzdok, D. ve Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. doi:10.1016/j.bpsc.2017.11.007
- Cao, H., Meyer-Lindenberg, A. ve Schwarz, E. (2018). Comparative Evaluation of Machine Learning Strategies for Analyzing Big Data in Psychiatry. *International journal of molecular sciences*, 19(11). doi:10.3390/ijms19113387
- Chen, H., Wang, N., Du, X., Mei, K., Zhou, Y. ve Cai, G. (2023). Classification Prediction of Breast Cancer Based on Machine Learning. *Computational intelligence and neuroscience*, 2023, 6530719. doi:10.1155/2023/6530719
- Cirillo, D. ve Valencia, A. (2019). Big data analytics for personalized medicine. *Current Opinion in Biotechnology*, 58, 161–167. doi:10.1016/j.cobio.2019.03.004
- Cox, J. L., Holden, J. M. ve Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry: the journal of mental science*, 150(JUNE), 782–786. doi:10.1192/BJP.150.6.782
- De Jesus Silva MM., Peres Rocha Carvalho Leite E., Alves Nogueira D., C. M. (2016). Depression in pregnancy. Prevalence and associated factors. *Investigación y educación en enfermería*, 34, 342–350.
- Dennis, C. L., Falah-Hassani, K. ve Shiri, R. (2017). Prevalence of antenatal and postnatal anxiety: Systematic review and meta-analysis. *British Journal of Psychiatry*, 210(5), 315–323. doi:10.1192/BJP.BP.116.187179
- Ding, X. X., Wu, Y. Le, Xu, S. J., Zhu, R. P., Jia, X. M., Zhang, S. F., ... Tao, F. B. (2014). Maternal anxiety during pregnancy and adverse birth outcomes: A systematic review and meta-analysis of prospective cohort studies. *Journal of Affective Disorders*, 159(81), 103–110. doi:10.1016/j.jad.2014.02.027
- Docker. (2023). <https://www.docker.com/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Dr Carrie Ladd, Dr Nathalie A Rodriguez McCullough, and D. C. C. (2017). Perinatal mental illness. *InnovAiT*.
- Faisal-Cury, A., Menezes, P., Araya, R. ve Zugaib, M. (2009). Common mental disorders during pregnancy: Prevalence and associated factors among low-income women in São Paulo, Brazil : Dand Anxiety during Pregnancy. *Archives of Women's Mental Health*, 12(5), 335–343. doi:10.1007/s00737-009-0081-6
- Falah-Hassani, K., Shiri, R. ve Dennis, C. L. (2016). Prevalence and risk factors for comorbid postpartum depressive symptomatology and anxiety. *Journal of Affective Disorders*, 198, 142–147. doi:10.1016/j.jad.2016.03.010
- Fawcett, E. J., Fairbrother, N., Cox, M. L., White, I. R., & Fawcett, J. M. (2019). The Prevalence of Anxiety Disorders During Pregnancy and the Postpartum Period: A Multivariate Bayesian Meta-Analysis. *The Journal of Clinical Psychiatry*.
- Fernández, A. M., Gutiérrez-Avilés, D., Troncoso, A. ve Martínez-Álvarez, F. (2020). Automated Deployment of a Spark Cluster with Machine Learning Algorithm Integration. *Big Data Research*, 19–20, 100135. doi:10.1016/j.bdr.2020.100135

- Furtado, M., Chow, C. H. T., Owais, S., Frey, B. N. ve Van Lieshout, R. J. (2018). Risk factors of new onset anxiety and anxiety exacerbation in the perinatal period: A systematic review and meta-analysis. *Journal of Affective Disorders*, 238(June), 626–635. doi:10.1016/j.jad.2018.05.073
- Fury, M. S., Oh, L. S. ve Berkson, E. M. (2022). New Opportunities in Assessing Return to Performance in the Elite Athlete: Unifying Sports Medicine, Data Analytics, and Sports Science. *Arthroscopy, Sports Medicine, and Rehabilitation*, 4(5), e1897–e1902. doi:10.1016/j.asmr.2022.08.001
- Gangula, R., Jagadeesh Kumar, G., Anand, S. ve Krishna Prasad, A. V. (2021). Integration of dynamic Docker containers and kubernetes with advanced cloud and Internet of Things. *Materials Today: Proceedings*, (xxxx). doi:10.1016/j.matpr.2021.07.276
- Gavin, Norma; Gaynes, Bradley N. MD; Lohr, Kathleen N. Meltzer-Brody, Samantha MD, Gartlehner, Gerald, Swinson, T. (2005). Perinatal Depression: A Systematic Review of Prevalence and Incidence. *Obstetrics and Gynecology*, (106), 1071–83.
- Gaynes, B. N., Gavin, N., Meltzer-Brody, S., Lohr, K. N., Swinson, T., Gartlehner, G., ... Miller, W. C. (2005). Perinatal depression: prevalence, screening accuracy, and screening outcomes. *Evidence report/technology assessment (Summary)*, (119), 1–8. doi:10.1037/e439372005-001
- Goyal, D., Gay, C., Torres, R. ve Lee, K. (2018). Shortening day length: a potential risk factor for perinatal depression. *Journal of Behavioral Medicine*, 41(5), 690–702. doi:10.1007/s10865-018-9971-2
- Haßdenteufel, K., Feißt, M., Brusniak, K., Lingenfelder, K., Matthies, L. M., Wallwiener, M. ve Wallwiener, S. (2020). Reduction in physical activity significantly increases depression and anxiety in the perinatal period: a longitudinal study based on a self-report digital assessment tool. *Archives of Gynecology and Obstetrics*, 302(1), 53–64. doi:10.1007/s00404-020-05570-x
- Heather A Bennett, Adrienne Einarson, Anna Taddio, Gideon Koren, T. R. E. (2004). Prevalence of depression during pregnancy: systematic review. *Obstet Gynecol.*, (103), 698–709.
- Helm. (2023). <https://helm.sh/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Kafka. (2023). <https://www.mshowto.org/apache-kafka-nedir-genel-bakis-bolum-1.html> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Kendig, S., Keats, J. P., Hoffman, M. C., Kay, L. B., Miller, E. S., Simas, T. A. M., ... Lemieux, L. A. (y.y.). Journal of Midwifery & Women ' s Health Consensus Bundle on Maternal Mental Health: Perinatal Depression and Anxiety. doi:10.1111/jmwh.12603
- Keşif Süreci. (2023). <https://medium.com/veri-madenciliği/veri-madenciliği-süreci-11243e7966fc> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., ... Zaslavsky, A. M. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular psychiatry*, 21(10), 1366–1371. doi:10.1038/mp.2015.198

- Keykhosravi, K., Hamednia, A., Rastegarfar, H. ve Agrell, E. (2022). Data preprocessing for machine-learning-based adaptive data center transmission. *ICT Express*, 8(1), 37–43. doi:10.1016/j.ict.2022.02.002
- Kömürcü N., D. M. Y. (2012). Doğum öncesi dönem. Kadın Sağlığı Ve Hastalıkları Hemşireliği El Kitabı. *İstanbul: Koç Üniversitesi Yayınları*, 131–166.
- Konteyner ve Sanal Makine. (2023). [https://www.youtube.com/watch?v=4XVfmGE1F\\_w&t=8215s](https://www.youtube.com/watch?v=4XVfmGE1F_w&t=8215s) adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Krause, M. L., Elrashidi, M. Y., Halvorsen, A. J., McDonald, F. S. ve Oxentenko, A. S. (2017). Impact of Pregnancy and Gender on Internal Medicine Resident Evaluations: A Retrospective Cohort Study. *Journal of general internal medicine*, 32(6), 648–653. doi:10.1007/s11606-017-4010-5
- Kubernetes. (2023). <https://kubernetes.io/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Kubernetes Mimarisi. (2023). <https://medium.com/rahasak/spark-cluster-deployment-with-kubernetes-1848d061cfc9> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Lens. (2023). <https://k8slens.dev/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Luo, Y., Wang, Z. ve Wang, C. (2021). Improvement of APACHE II score system for disease severity based on XGBoost algorithm. *BMC Medical Informatics and Decision Making*, 21(1), 1–12. doi:10.1186/s12911-021-01591-x
- Manconi, A., Gnocchi, M., Milanesi, L. ve Marullo, O. (2023). Heliyon Framing Apache Spark in life sciences. *Heliyon*, 9(2), e13368. doi:10.1016/j.heliyon.2023.e13368
- Marina Vargas-Terrones, Ruben Barakat, Belen Santacruz, Irene Fernandez-Buhigas, M. F. M. (2019). Physical exercise programme during pregnancy decreases perinatal depression risk: a randomised controlled trial. *British journal of sports medicine*.
- Martín, C., Langendoerfer, P., Soltani, P. ve Díaz, M. (2022). Kafka-ML : Connecting the data stream with ML / AI frameworks. *Future Generation Computer Systems*, 126, 15–33. doi:10.1016/j.future.2021.07.037
- Martini, J., Petzoldt, J., Einsle, F., Beesdo-Baum, K., Höfler, M. ve Wittchen, H. U. (2015). Risk factors and course patterns of anxiety and depressive disorders during pregnancy and after delivery: A prospective-longitudinal study. *Journal of Affective Disorders*, 175, 385–395. doi:10.1016/j.jad.2015.01.012
- Milgrom, J. ve Gemmill, A. W. (2014). Screening for perinatal depression. *Best Practice and Research: Clinical Obstetrics and Gynaecology*, 28(1), 13–23. doi:10.1016/j.bpobgyn.2013.08.014
- Milgrom, J., Westley, D. T. ve Gemmill, A. W. (2004). The mediating role of maternal responsiveness in some longer term effects of postnatal depression on infant development. *Infant Behavior and Development*, 27(4), 443–454. doi:10.1016/j.infbeh.2004.03.003
- Nair, L. R., Shetty, S. D. ve Shetty, S. D. (2017). Applying spark based machine learning model on streaming big data for health status prediction. *Computers & Electrical Engineering*, 0, 1–7. doi:10.1016/j.compeleceng.2017.03.009



- Nair, L. R., Shetty, S. D. ve Shetty, S. D. (2018). Applying spark based machine learning model on streaming big data for health status prediction. *Computers & Electrical Engineering*, 65, 393–399. doi:10.1016/J.COMPELECENG.2017.03.009
- Node.js. (2023). <https://nodejs.org/en> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Oğur, N. B., Al-Hubaishi, M. ve Çeken, C. (2022). IoT data analytics architecture for smart healthcare using RFID and WSN. *ETRI Journal*, 44(1), 135–146. doi:10.4218/ETRIJ.2020-0036
- Ogur, N. B., Ceken, C., Ogur, Y. S., Yuvaci, H. U., Yazici, A. B. ve Yazici, E. (2023). Development of an artificial intelligence-supported hybrid data management platform for monitoring depression and anxiety symptoms in the perinatal period: Pilot-scale study. *IEEE Access*. doi:10.1109/ACCESS.2023.3262467
- Oğur, Y. S. (2022). *Gebelik Döneminde Anksiyete ve Depresyon İçin Mobil İzleme Programı Geliştirilmesi ve İlk Üç Aylık Sonuçlarının Değerlendirilmesi*.
- Özkan Dikmen. (2020). *Sakarya İlinde Gebelerde Anksiyete ve Depresyon Sıklığı ve İlişkili Etmenler*.
- PCA-LDA. (2023). <https://techbigdatacloud.medium.com/veri-biliminde-pca-ve-lda-kavramları-6039e3aa34de> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Piccinelli M., Wilkinson, G. (2000). Gender differences in depression: Critical review. *The British Journal of Psychiatry*, 486–492.
- Priya, A., Garg, S. ve Tigga, N. P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science*, 167(2019), 1258–1267. doi:10.1016/j.procs.2020.03.442
- Puren Akçalı Aslan, Nazan Aydın, Esra Yazıcı, Ayse Nur Aksoy, Tulay Sati Kirkan, G. A. D. (2014). Prevalence of depressive disorders and related factors in women in the first trimester of their pregnancies in Erzurum, Turkey. *Int J Soc Psychiatry*.
- Randhawa, S. K. ve Singh, M. (2015). Classification of Heart Sound Signals Using Multi-modal Features. *Procedia Computer Science*, 58, 165–171. doi:10.1016/j.procs.2015.08.045
- Ranta, M., Paroutis, S., Talaoui, Y. ve Kohtam, M. (2023). Recovering the divide : A review of the big data analytics — strategy relationship, 56(January), 1–40. doi:10.1016/j.lrp.2022.102290
- RapidMiner. (2023). <https://rapidminer.com/> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Remeseiro, B. ve Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112, 103375. doi:10.1016/J.COMPBIOMED.2019.103375
- Rubertsson, C., Hellström, J., Cross, M. ve Sydsjö, G. (2014). Anxiety in early pregnancy: Prevalence and contributing factors. *Archives of Women's Mental Health*, 17(3), 221–228. doi:10.1007/s00737-013-0409-0

- Saccenti, E. ve Camacho, J. (2015). Chemometrics and Intelligent Laboratory Systems Determining the number of components in principal components analysis : A comparison of statistical , crossvalidation and approximated methods. *Chemometrics and Intelligent Laboratory Systems*, 149, 99–116. doi:10.1016/j.chemolab.2015.10.006
- Sau, A. ve Bhakta, I. (2019a). Screening of anxiety and depression among the seafarers using machine learning technology. *Informatics in Medicine Unlocked*, 16(December 2018), 100149. doi:10.1016/j.imu.2018.12.004
- Sau, A. ve Bhakta, I. (2019b). Screening of anxiety and depression among the seafarers using machine learning technology. *Informatics in Medicine Unlocked*, 16, 100149. doi:10.1016/J.IMU.2018.12.004
- Saurabh Biswas, Y. H. (2022). Big data analytics in precision medicine. *Big Data Analytics for Healthcare* içinde (ss. 63–72).
- Slomian, J., Honvo, G., Emonts, P., Reginster, J.-Y. ve Bruyère, O. (2019). Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. *Women's health (London, England)*, 15, 1745506519844044. doi:10.1177/1745506519844044
- Somerville, S., Dedman, K., Hagan, R., Oxnam, E., Wettinger, M., Byrne, S., ... Page, A. C. (2014). The Perinatal Anxiety Screening Scale: development and preliminary validation. *Archives of Women's Mental Health*, 17(5), 443–454. doi:10.1007/s00737-014-0425-8
- Spark. (2023). <https://medium.com/5bayt/apache-spark-nedir-ne-iş-yapar-5797c28eb95> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- SPSS. (2023). <https://www.ibm.com/products/spss-statistics> adresinden 6 Mayıs 2023 tarihinde alınmıştır.
- Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4), 1084–1093. doi:10.1016/J.ESWA.2006.02.005
- Suen, P. J. C., Goerigk, S., Razza, L. B., Padberg, F., Passos, I. C. ve Brunoni, A. R. (2021). Classification of unipolar and bipolar depression using machine learning techniques. *Psychiatry Research*, 295. doi:10.1016/J.PSYCHRES.2020.113624
- Tai, A. M. Y., Albuquerque, A., Carmona, N. E., Subramaniepillai, M., Cha, D. S., Sheko, M., ... McIntyre, R. S. (2019). Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine*, 99(August), 101704. doi:10.1016/j.artmed.2019.101704
- Thirunavukarasu, R., C, G. P. D., R, G., Gopikrishnan, M. ve Palanisamy, V. (2022). Towards computational solutions for precision medicine based big data healthcare system using deep learning models: A review. *Computers in Biology and Medicine*, 149(August), 106020. doi:10.1016/j.combiomed.2022.106020
- Tulay Sati Kirkan, Nazan Aydin, Esra Yazici, Puren Akcali Aslan, Hamit Acemoglu, A. G. D. (2015). The depression in women in pregnancy and postpartum period: A follow-up study. *Int J Soc Psychiatry*.

- Wadhwa, P. D., Glynn, L., Hobel, C. J., Garite, T. J., Porto, M., Chicz-DeMet, A., ... Sandman, C. A. (2002). Behavioral perinatology: Biobehavioral processes in human fetal development. *Regulatory Peptides*, 108(2–3), 149–157. doi:10.1016/S0167-0115(02)00102-7
- Walter, W., Pohlkamp, C., Meggendorfer, M., Nadarajah, N., Kern, W., Haferlach, C. ve Haferlach, T. (2023). Artificial intelligence in hematological diagnostics: Game changer or gadget? *Blood Reviews*, 58(October 2022), 101019. doi:10.1016/j.blre.2022.101019
- Wang, J., Wang, H., Nie, F. ve Li, X. (2023). Neurocomputing Feature selection with multi-class logistic regression. *Neurocomputing*, 543, 126268. doi:10.1016/j.neucom.2023.126268
- Wang, K., Fu, J. ve Wang, K. (2016). SPARK – A Big Data Processing Platform for Machine Learning. *2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration*, 48–51. doi:10.1109/ICIICII.2016.27
- Wang, Y., Krishna, S. ve Elyasi, I. (2023). Big data analysis using a parallel ensemble clustering architecture and an unsupervised feature selection approach. *Journal of King Saud University - Computer and Information Sciences*, 35(1), 270–282. doi:10.1016/j.jksuci.2022.11.016
- Yazıcı, E., Pek, T. M., Yuvacı, H. U., Köse, E., Cevrioglu, S., Yazıcı, A. B., ... Aydın, N. (2019). Perinatal Anxiety Screening Scale validiy and reliability study in Turkish ( PASS-TR validity and reliability ) Perinatal Anxiety Screening Scale validiy and reliability study in Turkish, 0573. doi:10.1080/24750573.2018.1506247
- Yücel, P., Çayır, Y. ve Yücel, M. (2013). Birinci Trimester Gebelerde Depresyon ve Anksiyete Bozukluğu (Depression and Anxiety Among First Trimester Pregnancies), 16(7–17), 83–87.
- Zhang, Y., Wang, S., Hermann, A., Joly, R. ve Pathak, J. (2021). Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *Journal of Affective Disorders*, 279(June 2020), 1–8. doi:10.1016/j.jad.2020.09.113
- Zhou, S., Ma, Q., Lou, Y., Lv, X., Tian, H., Wei, J., ... Yu, X. (2021). Machine learning to predict clinical remission in depressed patients after acute phase selective serotonin reuptake inhibitor treatment. *Journal of Affective Disorders*, 287(January), 372–379. doi:10.1016/j.jad.2021.03.079



## **EKLER**

### **EK 1: Sosyodemografik Veri Formu**

#### **ANNEYE AİT GENEL BİLGİLER**

**1.Adınız Soyadınız nedir? .....**

**2.Doğum tarihiniz nedir?.....**

**Yaş.....**

**Telefon no: .....**

**3.Nerede yaşıyorsunuz? (İlçe).....**

Kent merkezi  Kırsal

**4.Medeni haliniz nedir?**

Evli  Bekar  Dul  Boşanmış

**5.Boyunuz: .....**

**6.Kilonuz: .....**

**7.Kaç kilo ile gebe kaldınız?.....**

**8.Eğitim durumunuz nedir?**

İlkokul mezunu  Ortaokul mezunu

Lise mezunu  Üniversite mezunu

Yüksek Lisans/Doktora mezunu

**9.Çalışma durumunuz nedir?**

Ev hanımı  Düzensiz olarak çalışıyorum  Düzenli bir işte çalışıyorum

**10.Kiminle birlikte yaşıyorsunuz?**

Yalnız Eşim  Eşim ve çocuklarım

Eşim, çocuklarım ve eşimin ailesi  Eşim, çocuklarım ve kendi ailem

Diğer .....

11.Evinizde sürekli yaşayan toplam kişi sayısı kaçtır? .....

12.Ailenizin toplam gelir düzeyi ne kadar?

3000 tl ve altı  3001-6000

6001-9000  9001-12000

12001-15.000  15.000 üstü

**13. Düşükler, kürtaj, ölü doğum ve şimdiki gebelik dahil TOPLAM KAÇ GEBELİK geçirdiniz?**

.....

**14.Yaşayan kaç çocuğunuz var?.....**

**15.Daha önce ölü doğum yaşadınız mı?**

Evet  Hayır

**Evetse; kaç tane?.....**

**16.Daha önce düşük yaşadınız mı?**

Evet  Hayır

**Evetse; kaç tane?.....**

**17.Daha önce kürtaj yaşadınız mı?**

Evet  Hayır

**Evetse; kaç tane?.....**

**18.Gebelik ÖNCESİNDE kronik bir hastalığınız var mıydı?**

Evet  Hayır

**Evetse belirtiniz;**

Diyabet  Hipertansiyon  Kalp yetmezliği

Böbrek yetmezliği  Tiroit rahatsızlığı

Migren  Kas, iskelet hastalığı

Romatizmal hastalık  Diğer .....

**19. Daha ÖNCE, GEBELİK VE LOHUSALIK DÖNEMİ DIŞINDA herhangi bir psikiyatrik tanı**

**aldınız mı?**

**(Önceki gebelik ve lohusa dönemi sonraki sorularda sorulacaktır.)**

Evet  Hayır

**Evetse belirtiniz;**

Depresyon

Anksiyete bozukluğu (kaygı bozukluğu)

Bipolar bozukluk (iki uçlu duygudurum bozukluğu/manik depresif)

Şizofreni (psikoz)

Panik bozukluk (panik atak)

Obsesif Kompulsif Bozukluk (takıntı hastalığı)

Diğer .....

**20. Daha ÖNCE, GEBELİK VE LOHUSALIK DÖNEMİ DIŞINDA psikiyatrik tanı aldıysanız bunun için tedavi aldınız mı?**

Psikiyatrik ilaç tedavisi Evet  Hayır

Psikoterapi Evet  Hayır

EKT Evet  Hayır

TMS Evet  Hayır

**21. Daha önce hiç sizi fiziksel, cinsel, duygusal açıdan etkileyen, sonradan da etkisinde kaldığımız travmatik bir olay yaşadınız mı?**

Evet  Hayır

**Evetse belirtiniz;**

Ölümcül hastalık tanısı aldım

Hayati tehlike atlattığım bir kaza geçirdim

Ailemden biri (ya da çok yakın arkadaşım) kaza/cinayet/intihar sebebiyle vefat etti

Cinsel istismar/tecavüze uğradım

Çocukluğumda şiddete maruz kaldım

Bıçakla/silahla saldırıya uğradım

Biri saldırıya uğrarken tanık oldum

### **SON GEBELİK DÖNEMİ**

**22.Kaçıncı gebelik haftasındasınız? .....**

**23.Bebeğinizin cinsiyeti nedir?**

Kız  Erkek  Belli değil

**24.Şu anki gebeliğinizi nasıl tanımlarsınız?**

Planlı gebelik

Plansız(sürpriz) istenen gebelik

Hem plansız hem istenmeyen gebelik

**25.Gebelik için herhangi bir tedavi aldınız mı?**

(Tüp bebek, aşılama vb.)

Evet  Hayır  .....

**26.Gebelik takiplerine düzenli gidiyor musunuz?**

Evet  Hayır

**Tahmini kaç kez takibe gittiniz?.....**

**27.ŞU ANKI gebeliğiniz sırasında ortaya çıkan bir hastalığınız oldu mu?**

Evet  Hayır

**Evetse belirtiniz;**



Diyabet (Gebelikte yüksek şeker)

Hipertansiyon (Gebelikte yüksek tansiyon)

Tiroid hastalığı (Guatr)

Migren

Düşük tehditi

Diğer .....

**28. ŞİMDİKİ gebeliğiniz sürecinde psikiyatrik bir tanı aldınız mı?**

Evet  Hayır

**Evetse belirtiniz;**

Depresyon

Anksiyete bozukluğu (kaygı bozukluğu)

Bipolar bozukluk (iki uçlu duygudurum bozukluğu/manik depresif)

Şizofreni (psikoz)

Panik bozukluk (panik atak)

Obsesif Kompulsif Bozukluk (takıntı hastalığı)

Diğer .....

**29- Şimdiki gebeliğiniz süresince psikiyatrik bir tanı aldıysanız bunun için tedavi aldınız mı?**

Psikiyatrik ilaç tedavisi Evet  Hayır

Psikoterapi Evet  Hayır

EKT Evet  Hayır

TMS Evet  Hayır

**30-Şu an psikiyatrik bir tedavi veya terapi görüyor musunuz?**

Evet  Hayır

**31.Gebelik nedeniyle bıraktığınız bir psikiyatrik tedavi oldu mu?**

Evet  Hayır

**32.Gebelik öncesinde sigara kullanıyor muydunuz?**

Hayır hiç kullanmadım

Hayır, gebelikten .....ay önce bırakmıştım

Evet, bazen

Evet, hemen her gün

**33.Gebelik sürecinde sigara kullanıyor musunuz?**

Hayır kullanmıyorum

Evet, bazen

Evet, hemen her gün

**34.Doğum süreciniz ile ilgili tedavi ekibinden bilgi aldınız mı?**

Evet  Hayır

**35.Gebeliğiniz süresinde egzersiz (yürüyüş, dans, yoga vb) yapıyor musunuz?**

Egzersiz yapmıyorum

Haftada 1-2 gün

Haftada 3-4 gün

Evet hemen her gün

**36.Gebeliğiniz süresinde 1 günde yaklaşık ne kadar süre açık havada vakit geçirirsiniz?**

Günde 0-1 saat  2-3 saat  4-5 saat  6-7 saat  8-9 saat  9 saat den fazla

**ÖNCEKİ GEBELİK DÖNEMLERİ (ilk gebeliği olanlar 44. Soruya geçebilirler)**

**37.Daha ÖNCEKİ gebeliğiniz sırasında ortaya çıkan tıbbi bir hastalığınız oldu mu?**

**(Birden fazla seçenek işaretleyebilirsiniz.)**

Evet  Hayır

**Evetse belirtiniz;**

Diyabet (Gebelikte yüksek şeker)

Hipertansiyon (Gebelikte yüksek tansiyon)

Gebelik zehirlenmesi (Preeklampsi)

Tiroid hastalığı (Guatr)  Migren

Plasenta Previa (Yanlış yerleşimli plasenta)

Plasenta Dekolmanı (Plasentanın erken ayrılması)

Diğer .....

**38. Daha ÖNCEKİ GEBELİK döneminizde herhangi bir psikiyatrik tanı aldınız mı?**

Evet  Hayır

Tanı aldıysanız belirtiniz;

Depresyon

Anksiyete bozukluğu (kaygı bozukluğu)

Bipolar bozukluk (iki uçlu duygudurum bozukluğu/manik depresif)

Şizofreni (psikoz)

Panik bozukluk (panik atak)

Obsesif Kompulsif Bozukluk (takıntı hastalığı)

Diğer  .....

**39. Daha ÖNCEKİ GEBELİK döneminde psikiyatrik bir tanı aldıysanız bunun için tedavi aldınız mı?**

Psikiyatrik ilaç tedavisi Evet  Hayır

Psikoterapi Evet  Hayır

EKT Evet  Hayır

TMS Evet  Hayır

**40. Daha ÖNCEKİ LOHUSALIK döneminizde herhangi bir psikiyatrik tanı aldınız mı?**

Evet  Hayır

Tanı aldıysanız belirtiniz;

Depresyon

Anksiyete bozukluğu (kaygı bozukluğu)

Bipolar bozukluk (iki uçlu duygudurum bozukluğu/manik depresif)

Şizofreni (psikoz)

Panik bozukluk (panik atak)

Obsesif Kompulsif Bozukluk (takıntı hastalığı)

Diğer  .....

**41. Daha ÖNCEKİ LOHUSALIK döneminizde psikiyatrik tanı aldıysanız bunun için tedavi aldınız mı?**

Psikiyatrik ilaç tedavisi Evet  Hayır

EKT Evet  Hayır

TMS Evet  Hayır

Psikoterapi Evet  Hayır

**42. ÖNCEKİ gebeliğinizde bebeğinizin sağlık sorunu var mıydı, varsa neydi?**

**(Birden fazla seçenek işaretleyebilirsiniz)**

Evet  Hayır

**Evetse belirtiniz;**

Solunum sıkıntısı

Kardiyak malformasyon (kalp hastalığı)

Persistan pulmoner hipertansiyon

IUGR (gelişme geriliği)

Erken doğum

Düşük doğum ağırlığı

Diğer.....

**43.ÖNCEKİ gebeliğinizde bebeğinizin hastaneye yatması gereken sağlık sorunu oldu mu?**

Evet  Hayır

**Yattıysa kaç gün hastanede kaldı?.....**

**EŞ VE AİLE İLE İLGİLİ BİLGİLER**

**44-Eşinizin doğum tarihi nedir (yıl)? .....**

**45-Eşinizin eğitim durumu nedir?**

Okuryazar değil  Okuryazar

İlkokul mezunu  Ortaokul mezunu

Lise mezunu  Üniversite mezunu

Yüksek Lisans/Doktora mezunu

**46-Eşinizin çalışma durumu nedir?**

İşsiz /Çalışmıyor

Düzensiz olarak çalışıyor

Düzenli bir işte çalışıyor

**47- Eşinizin fiziksel ya da kronik bir hastalığı var mı?**

Evet  Hayır

**Varsa belirtiniz;**

Diyabet  Hipertansiyon  Kalp yetmezliği  Böbrek yetmezliği  Tiroit rahatsızlığı

Migren  Kas, iskelet hastalığı

Romatizmal hastalık  Diğer .....

**48- Eşinizin ek tıbbi hastalığı varsa günlük yaşamını olumsuz etkiliyor mu?**

Evet  Hayır

**49-Eşinizin ruhsal hastalığı var mı? (Depresyon, Panik Bozukluk, Psikoz vb.)**

Evet  Hayır

**50- Eşinizin ruhsal hastalığı varsa günlük yaşamını olumsuz etkiliyor mu?**

Evet  Hayır

**51- Son 1 yılı değerlendirdiğinizde eşiniz alkol ya da madde kullanır mı?**

Hayır kullanmıyor

Evet, bazen

Evet, hemen her gün

**52-Eşinizle iletişiminiz ne kadar iyi diye sorduğumuzda 1'den 10'a kadar puan verseniz kaç puan verirsiniz? (1 en kötü, 10 en iyi puan)**

.....

**53-Doğum sonrası eşinizin size olan duygusal desteğine 1'den 10'a kadar puan verseniz kaç puan verirsiniz? (1 en kötü, 10 en iyi puan) .....**

**54-Gebelik sürecinde şiddete maruz kaldınız mı?**

a. Fiziksel ve cinsel şiddet :( İtme, yumruklama, tokat vurma, tutup fırlatma, boğma, sivri bir alet/bıçak vb. saplama, ateş etme, ilişkiye zorlama vb.)

Evet  Hayır

b. Duygusal şiddet :( Aşırı biçimde eleştirmek, tehdit etmek, aşağılamak, hakaret etmek, utandırmak, hiç iletişime geçmemek, başkalarıyla iletişimini engellemek, duygusal olarak ihmal etmek, yalan söylemek, küçümsemek, görmezden gelmek, kendi çıkarları için kullanmak)

Evet  Hayır

c. Ekonomik şiddet: (Bütün parasal kontrolün eş tarafından kendi istekleri doğrultusunda kontrol edilmesi, paraya ulaşımın engellenmesi)

Evet  Hayır

**55-Ailede (Kan bağı ile bağlı olan) psikiyatrik hastalık tanısı almış olan var mı?**

Evet  Hayır

**Evet ise belirtiniz;**

Depresyon

Anksiyete bozukluğu (kaygı bozukluğu)

Bipolar bozukluk (iki uçlu duygudurum bozukluğu/manik depresif)

Şizofreni (psikoz)

Panik bozukluk (panik atak)

Obsesif Kompulsif Bozukluk (takıntı hastalığı)

Diğer  .....

**56. Sorunlarınız olduğunda çevrenizde paylaşabileceğiniz insanlar var mı?**

Evet  varsa kaç kişi.....

Hayır

## **EK 2: Edinburgh Doğum Sonrası Depresyon Ölçeği (Eddö)**

Yakın zamanlarda bebeğiniz oldu. Sizin son hafta içindeki duygularınızı öğrenmek istiyoruz. Böylelikle size daha iyi yardımcı olabileceğimize inanıyoruz. Lütfen, yalnızca bugün değil son 7 gün içinde, kendinizi nasıl hissettiğinizi en iyi tanımlayan ifadeyi işaretleyiniz.

**Son 7 gündür;**

### **1) Gülebiliyor ve olayların komik tarafını görebiliyorum.**

- Her zaman olduğu kadar
- Artık pek okadar değil
- Artık kesinlikle okadar değil
- Artık hiç değil

**Son 7 gündür;**

### **2) Geleceğe hevesle bakıyorum.**

- Her zaman olduğu kadar
- Her zamankinden biraz daha az
- Her zamankinden kesinlikle daha az
- Hemen hemen hiç

**Son 7 gündür;**

### **3) Birşeyler kötü gittiğinde gereksiz yere kendimi suçluyorum.**

- Evet, çoğu zaman
- Evet, bazen
- Çok sık değil
- Hayır, hiç bir zaman

**Son 7 gündür;**

### **4) Nedensiz yere kendimi sıkıntılı ya da endişeli hissediyorum.**

- Hayır, hiç bir zaman



- Çok seyrek
- Evet, bazen
- Evet, çođu zaman

**Son 7 gündür;**

**5) İyi bir nedeni olmadığı halde, korkuyor ya da panikliyorum.**

- Evet, çođu zaman
- Evet, bazen
- Hayır, çok sık deđil
- Hayır, hiç bir zaman

**Son 7 gündür;**

**6) Her şey giderek sırtıma yükleniyor.**

- Evet, çođu zaman hiç başa çıkamıyorum
- Evet, bazen eskisi gibi başa çıkamıyorum
- Hayır, çođu zaman oldukça iyi başa çıkamıyorum
- Hayır, her zamanki gibi başa çıkabiliyorum

**Son 7 gündür;**

**7) Öylesine mutsuzum ki uyumakta zorlanıyorum.**

- Evet, çođu zaman
- Evet, bazen
- Çok sık deđil
- Hayır, hiç bir zaman

**Son 7 gündür**

**8) Kendimi üzüntülü ya da çökkün hissediyorum.**

- Evet, çođu zaman
- Evet, oldukça sık

Çok sık deęil

Hayır, hi bir zaman

**Son 7 gndr**

**9) ylesine mutsuzum ki aęlıyorum.**

Evet, oęu zaman

Evet, olduka sık

Çok seyrek

Hayır, asla

**Son 7 gndr**

**10) Kendime zarar verme dřncesinin aklıma geldięi oldu.**

Evet, olduka sık

Bazen

Hemen hemen hi

Asla

**EK 3: Perinatal Anksiyete Tarama Ölçeği (Pass-Tr)**

Aşağıdakileri geçtiğimiz ay ne sıklıkla yaşadınız? Lütfen her soru için yaşadıklarınıza en yakın olan yanıtı

işaretleyiniz.

1. Bebek / gebelik ile ilgili endişe 0 1 2 3
2. Bebeğe zarar geleceği ile ilgili korku 0 1 2 3
3. Bir şeylerin kötü gideceğine dair korku hissi 0 1 2 3
4. Pek çok şey hakkında endişelenme 0 1 2 3
5. Gelecek hakkında endişe 0 1 2 3
6. Birşeylerin üstüne fazla yüklendiğini hissetme 0 1 2 3
7. İğne, kan, doğum, ağrı vb. şeyler konusunda çok şiddetli korkular 0 1 2 3
8. Birden bastıran aşırı korku veya huzursuzluk 0 1 2 3
9. Durdurulması veya kontrol edilmesi zor olan, tekrarlayan düşünceler 0 1 2 3
10. Uyumak için fırsatım olsa bile uyumakta zorlanma 0 1 2 3
11. İşleri belirli bir düzen veya sıra ile yapmak zorunda hissetme 0 1 2 3
12. Herşeyin mükemmel olmasını isteme 0 1 2 3
13. Herşeyi kontrol etme ihtiyacı 0 1 2 3
14. Birşeyleri defalarca kontrol etmeyi veya yapmayı durdurmakta zorluk 0 1 2 3
15. Diken üstünde hissetme veya kolayca irkilme 0 1 2 3
16. Tekrarlayan düşüncelerin yol açtığı rahatsızlık /sıkıntı 0 1 2 3
17. Birşeyler için tetikte olma ya da dikkatli olma ihtiyacı 0 1 2 3
18. Tekrarlayan anılar, rüyalar ya da kabuslardan dolayı üzülme, sıkıntı çekme 0 1 2 3
19. Başkalarının önünde kendimi rezil edeceğim endişesi 0 1 2 3
20. Diğer insanların beni olumsuz yargılayacağı korkusu 0 1 2 3
21. Kalabalık içinde fazla rahatsız hissetme 0 1 2 3
22. Huzursuz olacağım korkusu ile sosyal aktivitelerden kaçınma 0 1 2 3

23. Huzursuz eden şeylerden kaçınma 0 1 2 3
24. Kendinizi sanki bir filmde izliyormuş gibi kopuk hissetme 0 1 2 3
25. Zamanın nasıl geçtiğini farkedememe ve ne olduğunu hatırlayamama 0 1 2 3
26. Yakın zamanda olan değişikliklere uyum sağlamakta zorluk 0 1 2 3
27. Birşeyler yapabilmenize engel olan kaygı 0 1 2 3
28. Konsantre olmayı güçleştiren yarışan düşünceler 0 1 2 3
29. Kontrolünü kaybetme korkusu 0 1 2 3
30. Paniklemiş hissetme 0 1 2 3
31. Tedirgin (ajite) hissetme, kıpırdanma 0 1 2 3

## ÖZGEÇMİŞ

Ad-Soyad :Nur Banu OĞUR

### ÖĞRENİM DURUMU:

- **Lisans** : 2015, Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği
- **Yükseklisans** : 2018, Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilgisayar Mühendisliği

### MESLEKİ DENEYİM:

- 2017 yılında Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'ne Araştırma Görevlisi olarak atandı. Halen Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'nde Araştırma Görevlisi olarak görev yapmaktadır.

### TEZDEN TÜRETİLEN ESERLER:

- N. B. Oğur, C. Çeken, Y. S. Oğur, H. U. Yuvaci, A. B. Yazici and E. Yazici, "Development of an Artificial Intelligence-Supported Hybrid Data Management Platform for Monitoring Depression and Anxiety Symptoms in the Perinatal Period: Pilot-Scale Study," in *IEEE Access*, vol. 11, pp. 31456-31466, 2023, doi: 10.1109/ACCESS.2023.3262467.