



**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
AKILLI SİSTEMLER MÜHENDİSLİĞİ ANA BİLİM DALI**

**DOĞAL DİL İŞLEME VE METİN MADENCİLİĞİ
YÖNTEMLERİYLE TELETIP RANDEVU ROBOTUNUN
TASARLANMASI**

Yüksek Lisans Tezi

Müberra TERZİ

Danışman
Dr. Öğr. Üyesi Naci MURAT

SAMSUN
2021

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
AKILLI SİSTEMLER MÜHENDİSLİĞİ ANA BİLİM DALI



**DOĞAL DİL İŞLEME VE METİN MADENCİLİĞİ
YÖNTEMLERİYLE TELETIP RANDEVU ROBOTUNUN
TASARLANMASI**

Yüksek Lisans Tezi

Müberra TERZİ

Danışman


Dr. Öğr. Üyesi Naci MURAT

Bu tez çalışmasında kullanılan veriler VM Medicalpark Samsun Hastanesinden alınmıştır.

SAMSUN
2021

TEZ KABUL VE ONAYI

Müberra TERZİ tarafından, Dr. Öğr. Üyesi Naci MURAT danışmanlığında hazırlanan “Doğal Dil İşleme ve Metin Madenciliği Yöntemleriyle Teletıp Randevu Robotunun Tasarlanması” başlıklı bu çalışma, jürimiz tarafından 9.6.2021 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı	İmza	Sonuç
Başkan	Prof. Dr. Sermin ELEVLI Ondokuz Mayıs Üniversitesi Endüstri Mühendisliği Anabilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye (Danışman)	Dr. Öğr. Üyesi Naci MURAT Ondokuz Mayıs Üniversitesi Endüstri Mühendisliği Anabilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret
Üye	Dr. Öğr. Üyesi Haydar Koç Çankırı Karatekin Üniversitesi İstatistik Anabilim Dalı		<input checked="" type="checkbox"/> Kabul <input type="checkbox"/> Ret

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

ONAY
... / ... / ...
Prof. Dr. Ali BOLAT
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Yüksek Lisans tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

İmza

03 / 05 / 2021

Müberra TERZİ

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı: Doğal Dil İşleme ve Metin Madenciliği Yöntemleriyle Teletıp Randevu Robotunun Tasarlanması

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 03/05/2021 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 5

Tek kaynak oranı : % 1 çıkmıştır.

İmza

03 / 05 / 2021

Dr. Öğr. Üyesi Naci MURAT

ÖZET

DOĞAL DİL İŞLEME VE METİN MADENCİLİĞİ YÖNTEMLERİYLE TELETIP RANDEVU ROBOTUNUN TASARLANMASI

Müberra TERZİ

Ondokuz Mayıs Üniversitesi

Lisansüstü Eğitim Enstitüsü

Akıllı Sistemler Ana Bilim Dalı

Yüksek Lisans, Mayıs/2021

Danışman: Dr. Öğr. Üyesi Naci MURAT

Teknolojinin gelişmesiyle birlikte çevrimiçi ortamlarda çok sayıda veri oluşmaya başlamıştır. Sayısal halde bulunan veriler yapısal verileri; metin, resim veya ses halinde bulunan veriler ise yapısal olmayan verileri oluşturmaktadır. Yapısal verilerin analizinde veri madenciliği yöntemi kullanılırken, yapısal olmayan verilerin analizinde veri madenciliğinin alt dalı olarak kabul edilen metin madenciliği yöntemi kullanılmaktadır. Özellikle sağlık alanında yapılan çalışmalarda metin madenciliği metodlarının kullanımı son yıllarda artış göstermektedir. Bu çalışmada Ocak-Aralık 2019 tarihleri arasında özel bir hastanenin acil servisine gelen 20.516 hastaya ait sağlık verisi kullanılmış olup, veri setinde hastaların demografik özellikleri, özet epikriz ve yönlendirildikleri poliklinik bilgileri bulunmaktadır. 0-15 yaş arasındaki hastalar yalnızca çocuk hastalıkları polikliniğine yönlendirildiğinden dolayı bu gruba ait veriler analiz dışı tutulmuştur. Zemberek Doğal Dil Kütüphanesi ile kelimeler ek-köklerine ayrılarak kök kelimelerin frekanslarına bakılmıştır. Apriori algoritması ve FP-Growth algoritmaları ile kök kelimeler arasındaki birliktelik ilişkisi incelenmiş ve Apriori algoritmasında 32, FP-Growth algoritmasında 37 tane birliktelik kuralı bulunmuştur. Bu çalışma kapsamında sınıflandırma algoritmalarından Destek Vektör Makineleri (DVM), Karar Ağacı, Rasgele Orman ve K-En Yakın Komşu algoritmaları kullanılmış olup analiz aşamasında verilerin %70'i eğitim-%30'u test verisi olarak ikiye ayrılmıştır. Destek Vektör Makineleri algoritmasında %80,4, budama yapılmış karar ağacı algoritmasında %77,0, budama yapılmamış karar ağacı algoritmasında %74,3, rasgele orman algoritmasında %71,8 ve K-en yakın komşu algoritmasında %73,4 doğruluk oranı bulunmuştur. Sınıflandırma algoritmalarına ait atama matrisleri incelenerek doğru sınıflandırma yapılan şikayetlerdeki polikliniklere atanan kök kelimeler tespit edilmiştir. Kök kelime-poliklinik sınıflandırmaları için chatbot oluşturularak hastaların cinsiyet, yaş ve hastalıklarıyla ilgili şikayetlerini söylediklerinde hangi polikliniğe gideceklerini öğrenebilmeleri sağlanmıştır. Böylece acil servis yoğunluğunun azaltılmasına ve kaynakların doğru bir şekilde yönlendirilmesine olanak sağlayan bir araç oluşturulmuştur.

Anahtar Sözcükler: metin madenciliği, chatbot, doğal dil işleme, sağlık, veri madenciliği

ABSTRACT

DESIGNING A TELETIP APPOINTMENT ROBOT WITH NATURAL LANGUAGE PROCESSING AND TEXT MINING METHODS

Müberra TERZİ

Ondokuz Mayıs University

Institute of Graduate Studies

Department of Intelligent Systems

Master, May/2021

Supervisor: Assist. Prof. Naci MURAT

With the development of technology, a large amount of data has begun to occur in online environments. While the data in digital form creates the structural data, the data in the form of text, picture, or sound creates the unstructured data. While the data mining method is used in the analysis of structured data, the text mining method, which is considered as a sub-branch of data mining, is used in the analysis of unstructured data. The use of text mining methods has increased in recent years, especially in studies in the field of health. In this study, health data of 20,516 patients who came to the emergency department of a private hospital between January and December 2019 were used, and the data set includes demographic characteristics of the patients, summary epicrisis, and outpatient clinic information. The data belonging to the 0-15 years old group were excluded from the analysis because the patients of this group were only referred to the pediatric outpatient clinic. With the Zemberek Natural Language Library, words were separated into affix-roots, and the frequencies of root words were examined. The association between root words was examined with Apriori and FP-Growth algorithms, and 32 association rules were found in the Apriori algorithm and 37 association rules in the FP-Growth algorithm. Within the scope of this study, Support Vector Machines (SVM), Decision Tree, Random Forest, and K-Nearest Neighbor algorithms from classification algorithms were used, and during the analysis phase, the data is divided into two as 70% training data and 30% test data. The accuracy rate was 80.4% in the Support Vector Machines algorithm, 77.0% in the pruned decision tree algorithm, 74.3% in the unpruned decision tree algorithm, 71.8% in the random forest algorithm, and 73.4% in the K-nearest neighbor algorithm. By examining the assignment matrices of the classification algorithms, the root words assigned to the outpatient clinics in the complaints with the correct classification were determined. By creating a chatbot for root word-polyclinic classifications, patients could learn which polyclinic they would go to when they said their age, gender, and complaints about their illness. Thus, a tool was created to reduce the intensity of the emergency services and direct the resources correctly.

Keywords: text mining, chatbot, natural language processing, health, data mining

ÖNSÖZ VE TEŞEKKÜR

Lisans ve yüksek lisans hayatım boyunca her türlü konuda yardımına koşan, her şeyin en iyisini yapmam ve öğrenmem konusunda sürekli yanımda bulunarak yol gösteren danışman hocam Dr. Öğr. Üyesi Naci Murat'a ve Ondokuz Mayıs Üniversitesi Endüstri Mühendisliği bölümünde bulunan tüm hocalarıma teşekkürlerimi sunarım.

Tez kapsamında kullandığım veriler için göstermiş oldukları ilgiden dolayı VM Medicalpark Samsun Hastanesine ve Genel Müdür Dr. Hakan Özcan'a teşekkürlerimi sunarım.

Hayatımın boyunca desteklerini hep yanımda hissettiğim, bugün geldiğim noktada en büyük paya sahip olan aileme ve manevi destekleriyle yanımda olan arkadaşlarıma teşekkür ederim.

Müberra TERZİ

İÇİNDEKİLER

ÖZET	iii
ABSTRACT.....	iv
ÖNSÖZ VE TEŞEKKÜR.....	v
SİMGELER VE KISALTMALAR	viii
ŞEKİLLER DİZİNİ	ix
TABLolar DİZİNİ	x
1. GİRİŞ	1
2. KURAMSAL TEMELLER.....	3
2.1. Metin Madenciliği.....	3
2.1.1. Metin Madenciliği Süreci.....	4
2.1.1.1. Veri Toplama	6
2.1.1.2. Veri Ön İşleme.....	6
2.1.1.3. Vektör Uzay Modeli ve Vektör Oluşturma.....	7
2.1.1.4. Benzerlik Hesaplanması.....	9
2.1.1.5. Özellik Seçimi.....	9
2.1.2. Literatür Taraması.....	10
2.2. Sınıflandırma.....	18
2.2.1. Sınıflandırma Algoritmaları.....	19
2.2.1.1. Karar Ağaçları.....	19
2.2.1.2. Destek Vektör Makineleri (DVM).....	20
2.2.1.3. Ardışık Minimal Optimizasyon Algoritması (SMO)	21
2.2.1.4. K-En Yakın Komşu Algoritması (K-NN).....	21
2.2.1.5. Naive Bayes Algoritması	22
2.2.1.6. Yapay Sinir Ağları	22
2.2.1.7. Rasgele Orman Algoritması.....	23
2.2.2. Sınıflandırma Metodu Performans Ölçütleri.....	23
2.2.2.1. Doğruluk (Accuracy)	24
2.2.2.2. Kesinlik (Precision)	25
2.2.2.3. Duyarlılık (Sensitivity)	25
2.2.2.4. Özgüllük (Specificity).....	25
2.2.2.5. Negatif Prediktif Değer (Negative Predictive Value)	25
2.2.2.6. F-Ölçeği (F-Score).....	25
2.3. Birliktelik Analizi	26
2.3.1. Birliktelik Analizi Algoritmaları.....	26
2.3.1.1. APRIORI	27
2.3.1.2. AIS.....	27
2.3.1.3. SETM.....	27
2.3.1.4. FP-GROWTH.....	28

2.3.1.5. Bölümleme Tekniği	28
2.3.1.6. Örneklememe Tekniği	28
2.3.1.7. CARMA.....	28
2.3.1.8. Dinamik Nesne Sayımı (DNS).....	28
2.3.1.9. Sıra Dışı Aday Belirleme (SDAB).....	29
2.3.2. Birliktelik Analizi Performans Ölçütleri.....	29
2.3.2.1. Destek (Support).....	29
2.3.2.2. Güven (Confidence).....	29
2.3.2.3. Kaldıraç (Lift)	29
2.4. Kümeleme	30
2.5. Doğal Dil İşleme	31
2.6. Teletıp	33
3. MATERYAL VE YÖNTEM.....	35
3.1. KNIME	35
3.2. Dahi.ai.....	37
3.3. Uygulama.....	40
3.3.1. Araştırmanın Amacı ve Problemin Tanımlanması.....	40
3.3.2. Verinin Toplanması	41
3.3.3. Veri Ön İşleme.....	43
3.3.4. Birliktelik Analizi Sonuçları	44
3.3.4.1. Apriori Algoritması Sonuçları	44
3.3.4.2. FP-Growth Algoritması Sonuçları	50
3.3.5. Sınıflandırma Sonuçları	56
3.3.5.1. Destek Vektör Makineleri Algoritması Sonuçları.....	56
3.3.5.2. Karar Ağacı Algoritması Sonuçları.....	60
3.3.5.2.1. Budama Yapılmamış Karar Ağacı Algoritması Sonuçları	60
3.3.5.2.2. Budama Yapılmış Karar Ağacı Algoritması Sonuçları	64
3.3.5.3. Rasgele Orman Algoritması Sonuçları	67
3.3.5.4. K-En Yakın Komşu Algoritması Sonuçları	70
3.3.6. Dahi.ai Chatbot	73
4. TARTIŞMA VE SONUÇ	82
KAYNAKÇA.....	84
ÖZ GEÇMİŞ.....	89
EKLER	90

SİMGELER VE KISALTMALAR

DVM	: Destek Vektör Makineleri
SMO	: Ardışık Minimal Optimizasyon Algoritması
K-NN	: K-En Yakın Komşu Algoritması
NLP	: Doğal Dil İşleme
MDL	: Minimum Açıklama Uzunluğu
BK	: Bilgi Kazancı
DP	: Doğru Pozitif
DN	: Doğru Negatif
YP	: Yanlış Pozitif
YN	: Yanlış Negatif
DNS	: Dinamik Nesne Sayımı
SDAB	: Sıra Dışı Aday Belirleme
YSA	: Yapay Sinir Ağları

ŞEKİLLER DİZİNİ

Şekil 2.1. Metin madenciliği süreci	5
Şekil 2.2. Sınıflandırma algoritmaları	18
Şekil 2.3. Karar ağacı algoritması	19
Şekil 2.4. Yapay sinir ağlarının temel yapısı	23
Şekil 3.1. Dahi.ai chatbot tanımlama	38
Şekil 3.2. Dahi.ai konuşma metni tanımlama	38
Şekil 3.3. Dahi.ai koşul ekleme	39
Şekil 3.4. Kullanılan verinin dağılımı	42
Şekil 3.5. Apriori algoritması ile birliktelik analizi	44
Şekil 3.6. FP-Growth algoritması ile birliktelik analizi	50
Şekil 3.7. KNIME Destek vektör makineleri	57
Şekil 3.8. KNIME Karar ağacı modeli	60
Şekil 3.9. KNIME Rasgele orman algoritması	67
Şekil 3.10. KNIME K-en yakım komşu algoritması modeli	70
Şekil 3.11. Hata tahminlemesi için karar ağacı örneği	75
Şekil 3.12. Chatbot anahtar kelime tanımlama	77
Şekil 3.13. Chatbot başlangıç komutları	77
Şekil 3.14. Chatbot randevu al butonu	78
Şekil 3.15. Chatbot hangi polikliniğe gitmeliyim butonu	78
Şekil 3.16. Chatbot cinsiyet sorusu	79
Şekil 3.17. Chatbot yaş sorusu	80
Şekil 3.18. Chatbot hastalığa ilişkin şikayetler sorusu	80

TABLolar DİZİNİ

Tablo 2.1. Terim-Doküman Matrisi.....	5
Tablo 2.2. Binary vektör için örnek veri.....	7
Tablo 2.3. Frekans vektör için örnek veri.....	8
Tablo 2.4. TFxIDF vektör için örnek veri.....	8
Tablo 2.5. N gram örnek tablo.....	10
Tablo 2.6. Metin madenciliği literatür taraması.....	10
Tablo 2.7. Metin madenciliği örnek çalışmalar.....	15
Tablo 2.8. Sınıflandırma algoritması performans ölçütleri.....	24
Tablo 2.9. Birliktelik analizi algoritmaları.....	26
Tablo 2.10. Kümeleme Algoritmaları (Altunkaynak, 2017).....	30
Tablo 3.1. KNIME düğümleri ve kullanımları.....	35
Tablo 3.2. Dahi.ai değişken tipleri.....	39
Tablo 3.3. Verinin polikliniklere göre dağılımı.....	42
Tablo 3.4. Kök kelime frekansları.....	45
Tablo 3.5. Apriori algoritması analizinde üçlü kelime kullanımı.....	45
Tablo 3.6. Apriori algoritması analizinde ikili kelime kullanımı.....	46
Tablo 3.7. Apriori algoritması analizinde tekli kelime kullanımı.....	48
Tablo 3.8. FP-Growth algoritması analizinde üçlü kelime kullanımı.....	51
Tablo 3.9. FP-Growth algoritması analizinde ikili kelime kullanımı.....	52
Tablo 3.10. FP-Growth algoritması analizinde tekli kelime kullanımı.....	54
Tablo 3.11. Sınıflandırma algoritmalarına göre poliklinik bazlı veri sayıları.....	56
Tablo 3.12. Destek vektör makineleri karışıklık matrisi.....	58
Tablo 3.13. Destek vektör makineleri doğruluk matrisi.....	59
Tablo 3.14. Budama yapılmamış karar ağacı karışıklık matrisi.....	62
Tablo 3.15. Budama yapılmamış karar ağacı algoritması doğruluk matrisi.....	63
Tablo 3.16. Budama yapılmış karar ağacı karışıklık matrisi.....	65
Tablo 3.17. Budama yapılmış karar ağacı doğruluk matrisi.....	66
Tablo 3.18. Rasgele orman algoritması karışıklık matrisi.....	68
Tablo 3.19. Rasgele orman algoritması doğruluk matrisi.....	69
Tablo 3.20. K-en yakın komşu algoritması karışıklık matrisi.....	71
Tablo 3.21. K-en yakın komşu algoritması doğruluk matrisi.....	72
Tablo 3.22. Sınıflandırma algoritmaları için poliklinik tanıları.....	73
Tablo 3.23. Hata tahminlemesi hesaplama.....	76

1. GİRİŞ

Dijital dönüşümün etkisiyle her sektörde milyonlarca veri bulunmakta ve bu veriler işlenerek anlamlı hale getirilmediği sürece kullanılamamaktadır. Verileri anlamlı hale getirmek için veri madenciliği yöntemine başvurulmaktadır. Veri madenciliği; sistemde depolanmış verileri kullanarak matematiksel ve istatistiksel metotlar ile anlamlı bilgiler keşfedebilmektir (Altıntaş, 2010). Metin madenciliği, veri madenciliğinin alt dalı olarak kabul edilmekte olup veri madenciliğinden ayrılan tek farkı yapısal olmayan verilerle ilgileniyor olmasıdır (Gümüştü, 2019). Metin madenciliğinde metin, veri kaynağı olarak kabul edilir. Metin üzerindeki yapısal olmayan verilerden yapılandırılmış veri elde edilmeye çalışılır (Şeker, 2015). Yapısal hale getirilmiş verilerden birliktelik analizi, sınıflandırma veya kümeleme gibi farklı çalışmalar yapılarak karar verme sürecinde kullanılmak üzere bilgiler elde edilir.

Verinin manuel olarak yapılandırılması insan zekâsı için karmaşıktır. Bu nedenle bu işlemler için yapay zekâ uygulamalarından da destek alınmaktadır. Eğitim, sağlık, üretim gibi pek çok alanda yapay zekâ uygulamaları kullanılmaktadır. Özellikle sağlık alanında yapılan çalışmalarla hastaya tanı konulmasından, radyolojik raporların yorumlanmasına kadar pek çok alanda doktorlara yapay zekâ destek olmaktadır. Bu çalışmalara genel olarak teletıp uygulamaları denmektedir. Yapay zekâ ile teletıp çalışmaları sınıflandırma – öngörme, zeki öğretim sistemleri, ölçme değerlendirme, uyumlanabilen sistemler – kişiselleştirme olmak üzere dört ana bölümden oluşmaktadır. Sınıflandırma – öngörme için veri madenciliği ve doğal dil işleme kavramları önemlidir. Uyumlanabilen sistemler ve kişiselleştirmede bir hastalığa tanı koymaya yönelik makine öğrenim modelleri geliştirilmektedir (Nabiyev ve Erümit, 2020).

Bu tez çalışmasında metin madenciliğinin sınıflandırma yöntemi yapay zekâ çalışmalarında yoğun uygulama alanı bulmuş sektörlerden birisi olan sağlık sektöründe uygulanmıştır. Sağlık Bakanlığının hastaların doğru polikliniğe gidebilmelerini sağlamak için yapmak istediği “Sağlıkta Neyim Var” projesine bir temel oluşturabileceği düşünülmektedir (Sağlık Bakanlığı, 2019). Bu tez çalışmasının ana amacı toplanan verilerden oluşturulacak anlamlı bilgileri kullanarak hastaların herhangi bir sağlık personelinden yardım almadan tek seferde doğru polikliniğe gidebilmelerini sağlayacak çevrimiçi bir sistem tasarlayabilmektir. Bu amaç

doğrultusunda Ocak-Aralık 2019 tarihleri arasında özel bir hastanenin acil servisinde yeşil alan muayesine sahip hastalardan bazı veriler toplanmıştır. Yeşil alan muayenesi; hastaların hangi polikliniğe gitmesi gerektiğini bilmemesi, mesai saatlerinden dolayı poliklinikte muayene olamaması gibi nedenlerden dolayı hastalıkları ile ilgili poliklinik branşı yerine acil servise gelmesi durumunu gösteren muayene türüdür. Hastaneden toplanan veri seti hastanın acil servise geliş şikâyeti, doktorun hastayı yönlendirdiği poliklinik ve cinsiyet, yaş gibi demografik özellikleri içeren toplamda on iki parametrelerden oluşmaktadır. Çalışmada Zemberek Doğal Dil Kütüphanesinden, Apriori algoritmasından ve FP-Growth algoritmasından yararlanılarak hasta şikâyetlerine birliktelik analizi yapılmış ve kural çıkarımları bulunmuştur. Kurallar destek, güven ve lift değerleri baz alınarak açıklanmıştır. Bulunan kurallardan elde edilen anahtar kelimeler kullanılarak veri setine destek vektör makineleri, karar ağacı, rasgele orman ve k-en yakın komşu algoritmaları uygulanmıştır. Bu algoritmalar karşılaştırılmış ve sınıflandırmadaki doğruluk oranına göre destek vektör makineleri algoritmasının daha doğru sonuçlar verdiği görülmüştür. Uygulanan algoritmalar için bulunan karışıklık matrisi ve doğruluk matrisi sonuçları tablolar halinde gösterilmiştir. Sınıflandırma algoritmalarının kelime bazlı oluşturduğu binary vektör matrisinden yararlanılarak kurallar bulunmuştur. Bulunan kurallar chatbot yazılımı için poliklinik bazlı anahtar kelimeleri oluşturmakta olup kurallar arasında frekansı 10'un üzerinde olanlar tablolaştırılmıştır. Hata tahminlemesi yoluyla, frekans değerinin minimum 10 olması gerektiğine karar verilmiştir. Tablolaştırılan kurallar çevrimiçi ortamda chatbot haline getirilmiştir. Böylece hastaların acil servise başvuru yapmadan önce hastalıklarına ait şikâyetlerini belirterek hangi polikliniğe başvuru yapacakları yönünde yol gösteren bir uygulama geliştirilmiştir. Bu şekilde acil servisteki acil dışı yoğunluğun önüne geçmek mümkün olabilecek ve sağlık personelleri zamanlarını acil durumdaki hastalara yönlendirebilecektir.

2. KURAMSAL TEMELLER

2.1. Metin Madenciliği

Metin madenciliği, kelime veya kelime öbeklerinin metinsel veri tabanı içerisinde anlamsız halden anlamlı hale getirilme sürecidir. Amaç veri kaynağındaki ilginç örüntüleri bulabilmek ve yararlı bilgiler ortaya koyabilmektir. Literatürde metin madenciliği yerine farklı tanımlamalar yapılmaktadır. Metin veri madenciliği, bilgi keşfi, metin analizi bunlardan bazılarıdır. Farklı tanımlamaları olsa da metin madenciliğinin anlamı birçok araştırmacıya göre benzerlik göstermektedir (Zor, 2021).

1960'lı yıllarda doğal dilin işlenmesine olan istek ve bilgiye erişim için yapılan çalışmalar sonucunda metin madenciliği kavramı insan hayatına girmiştir. Metinlerden kurallı bilgi çıkarımı, metin sınıflandırma – kümeleme, yapısal olmayan verilerin yapısal hale dönüştürülmesi 1990'lı yıllarda yaygınlaşmıştır (Gümüştü, 2019). Teknolojinin gelişmesiyle birlikte 1990'lardan günümüze kadar metin madenciliği sürekli gelişim içerisinde olmuştur. Günümüzde ise başta sağlık sektörü olmak üzere çeşitli sektör ihtiyaçlarına yönelik olarak sıklıkla kullanılmaktadır.

Metin madenciliği yönteminden birçok problemin çözümünde yararlanılmaktadır. Ancak bazı problemlerin çözümünde diğerlerine göre daha çok tercih edilmektedir. Bu problemler aşağıdaki gibidir (Şeker, 2015).

- Enformasyon getirme (Information retrieval): Analizlerdeki ön bilginin toplanma aşamasıdır. Yani çalışma yapılacak olan alanda hangi verilerin toplanması gerektiğine karar verme adımdır. Örneğin; bir web sitesinden veri toplanacaksa web sayfası adı, adresi vb. bilgilerin toplanması ya da hastane bilgi sistemlerinden toplanan verilerde hasta adı, poliklinik ve doktor bilgisi gibi bilgilerin toplanmasıdır.
- Doğal dil işleme (Natural language processing): Metin madenciliğinde özellik çıkarımı yaparken sıklıkla kullanılmaktadır. Metin madenciliğinin aşamalarında bulunmamasına rağmen bilgi çıkarımı yaparken kullanılmaktadır. Cümle parçalama, kelime kökü bulma gibi işlemlerde kullanılır. Bu konu ile ilgili ayrıntılı bilgi bölüm 2.5'de verilmiştir.

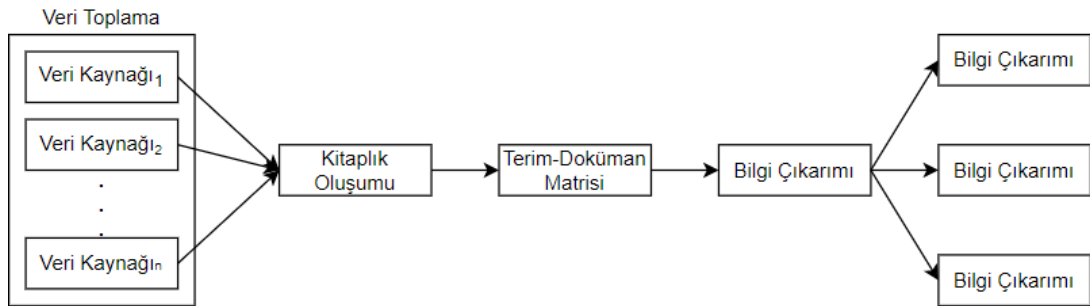
- Adlandırılmış varlık tanıma (Named entity recognition): İstatistiksel bilgi çıkarımı için kullanılmaktadır. Metin madenciliği her zaman düzgün metinlerde yapılmamaktadır. İçerisinde yazım ve noktalama hatalarının olduğu metinlerde de metin madenciliği çalışmaları yapılmaktadır. Adlandırılmış varlık tanıma yönteminde bu tarz hatalar göz önünde bulundurulur. Hedef kelime gruplarının metinde sayılması ve metinden çıkarılması gibi işlemler yapılabilir.
- Örüntüsü tanımlı varlıkların bulunması (Finding assets with a defined pattern): Metin içerisinde bulunan özel bilgilere odaklanmayı esas alır.
- Eş atıf (Co-citation): Bir varlığı belirten kelime gruplarının tanımlanmasıdır.
- İlişki, kural ve olay çıkarımı (Relationship, rule and event inference): Metin içerisindeki bazı bilgilerin çıkarılmasıdır. Bu çıkarımlar belli amaçlar için yapılmalıdır. Kelime grupları arasındaki ilişkisi göstermek için fiil yapılarının çıkarımı örnek olarak verilebilir.
- Duygu analizi (Sentiment analysis): Duygusal ifadelerin metin içerisinden çıkarılmasıdır. Duygusal ifadelerde en çok kutupsallık kullanılır. Yani bir konu hakkındaki pozitif ve negatif fikirlerin iki sınıfa ayrılması ve metindeki ruh halinin çıkarılmasıdır.

Metin madenciliği çalışmalarında kullanılabilen hem ücretsiz hem de ticari amaçlı farklı yazılımlar bulunmaktadır (Gümüşsu, 2019; Melek, 2012). Bu yazılımlardan bazıları SAS Text Mining, SPSS Text Mining, STATISTICA Text Mining, Rapid-Miner, WEKA, R Project, Clementine ve KNIME'dir. Bu programların her biri farklı amaçlar için kullanılabilir. SAS Text Mining, SPSS Text Mining, STATISTICA Text Mining en çok kullanılan yazılımlar arasındadır. Ancak sadece İngilizce dil desteği sağlamaktadırlar ve ücretlidirler. Rapid Miner Java tabanlı bir yazılım olup birçok makine öğrenim algoritmasına sahip bir programdır. R programlama dili açık kaynak kodlu olup birçok grafik özelliğine sahip bir programdır. KNIME ise Java tabanlı bir yazılımdır.

2.1.1. Metin Madenciliği Süreci

Metin madenciliği süreci birçok kaynakta üç aşamalı olarak belirtilmiştir. Şekil 2.1.'de gösterildiği gibi bu aşamalar sırasıyla kitaplık oluşumu, terim-doküman matrisi

oluşturma ve bilgi çıkarımıdır. Her aşamanın çıktısı bir sonraki aşamanın girdisi olarak kabul edilir.



Şekil 2.1. Metin madenciliği süreci

Birinci aşamada veri kaynaklarından gelen veriler girdi olarak kabul edilmektedir. Amaç çalışılan konu hakkındaki tüm bilgilerin toplanmasıdır. Bu nedenle veri kaynağı birden çok olabilir. Toplanan veriler metinsel dokümanlardan ses kayıtlarına, XML dosyalarından hasta epikriz notlarına kadar her türlü yapısal olmayan veriyi içerebilir. Bu aşamada veri toplama ve verilerin ortak forma dönüşümü yapılır. Bunun içinde veri ön işleme yapılmaktadır.

İkinci aşama yapısal olmayan verileri yapısal hale getirmek içindir. Terim-doküman matrisi oluşturulur. Bu matriste belgeler satırları, terimler ise sütunları temsil eder (Tablo 2.1.). Belgedeki terimlere ve bu terimlerin belge içerisindeki frekanslarına göre matrisin iç kısmı doldurulmaktadır. Frekans değerleri ağırlıklandırma yöntemiyle, terim özellikleri ise özellik çıkarım tekniği ile bulunur.

Tablo 2.1. Terim-Doküman Matrisi

Doküman \ Terim	Terim									
	Baş	Ağrı	Bulantı	Kusma	İshal	Karın	Halsizlik	Dönme	...	
Baş Ağrısı Bulantı Kusma	1	1	1	1						
Bulantı Kusma			1	1						
İshal Karın Ağrısı		1			1	1				
Bulantı Kusma Halsizlik			1	1			1			
İshal Kusma				1	1					
Bulantı Kusma İshal Karın Ağrısı		1	1	1	1	1				
Bulantı Kusma Baş Dönmesi Baş Ağrısı	2	1	1	1				1		
Bulantı Kusma Karın Ağrısı Ateş İshal		1	1	1	1	1				
İshal Kusma Karın Ağrısı		1		1	1	1				
Kusma İshal				1	1					
...										

Üçüncü aşamada ise bilgi çıkarımı yapılmaktadır. Bu çıkarımlar metin madenciliğinin sınıflandırma, kümeleme veya birliktelik kuralları metotları ile yapılır (Seçkin, 2011). Metotlara ait detaylı bilgi ilerleyen bölümlerde verilmiştir.

2.1.1.1. Veri Toplama

Yapılacak olan çalışmaya uygun şekilde bir veya daha çok kaynaktan yararlanma sürecinin genel adıdır. Veri toplanırken temelde dört adım takip edilir.

Birinci adımda, çalışmaya uygun başlangıç verisi toplanır. İkinci adımda, toplanılan veri analiz edilir ve çalışma için verinin yeterli olup olmadığına bakılır. Üçüncü adımda, verinin eksik noktaları tespit edilir. Dördüncü adımda ise verinin doğruluğundan ve yeterliliğinden emin olunur (Kaşıkçı ve Gökçen, 2014).

2.1.1.2. Veri Ön İşleme

Bir veri setini analiz edebilmek için öncelikle kelime normalleştirilmesi, büyük – küçük harf farklılıklarının düzenlenmesi, boşlukların temizlenmesi gibi bazı işlemlerin yapılması gerekir (Kahya, 2020). Bu adımda verinin ileriki aşamalarında yapılacak olan metin madenciliği metotlarına hazırlığı söz konusudur. Veri ön işleme dört adımdan oluşur.

- **Dönüştürme:** Web dokümanları HTML veya XML gibi formatlardadır. Veriyi kullanabilmek için düz metin haline getirmek gerekir. Bunun içinde etiket temizlemesi yapılır (Kaşıkçı ve Gökçen, 2014).
- **Tarama ve işaretleme:** Metin içerisindeki kelimeler noktalama işaretlerinden veya farklı simgelerden ayrılır. Her bir kelimenin aynı formatta olması için metinler küçük harfe dönüştürülür (Göker ve Tekedere, 2017; Kaşıkçı ve Gökçen, 2014).
- **Durak kelimelerin çıkarımı:** Metinlerde kelimeleri birbirine bağlarken kullanılan ancak bir şey ifade etmeyen bağlaç ve edatların çıkarılmasıdır.
- **Kök Bulma:** Aynı kökten gelen ancak aldığı eklerle birlikte farklı kelimeler gibi görünen kelimeler vardır. Analiz yaparken bu kelimelerin tek bir kelime gibi görünmesini sağlamak gerekir. Bunun içinde kelime kökleri bulunur (Kaşıkçı ve Gökçen, 2014).

2.1.1.3. Vektör Uzay Modeli ve Vektör Oluşturma

Bilgi çıkarımı, filtreleme gibi işlemlerin yapıldığı adımdır. Öncelikle uzay eksenleri belirlenir. Uzay eksenleri verideki sınıfların kök kelimelerinden oluşur. Her bir kelime bir vektör olarak tanımlanır. Her bir boyut bir kelimeye denk gelmektedir. Bu da her dokümanın bir vektör olarak gösterilmesi anlamına gelir.

Kelimelerin doküman içerisindeki frekansına bakılarak vektör uzayı oluşturulur. Modelleme yapılırken de üç farklı metot kullanılabilir. Bunlar;

- Binary Vektör: Verilerin “1” ve “0” olarak ifade edildiği vektörlerdir. Eğer kelime dokümanın içerisinde bulunuyorsa 1, bulunmuyorsa 0 olarak ifade edilir. Tablo 2.2.’de doküman kısmında bu çalışmadaki hasta epikrizlerinden örnekler verilmiştir. Kelimelerin binary vektöre göre atamaları yapılmıştır. En son dokümanda ağrı kelimesi iki kere geçiyor olmasına rağmen binary kelime matrisinde ağrıya denk gelen kısımda “1” yazmaktadır.

Tablo 2.2. Binary vektör için örnek veri

Doküman	Kelimeler							
	Bulantı	Kus	Karın	Ağrı	Ateş	İshal	Hal	Baş
Bulantı Kusma	1	1	0	0	0	0	0	0
Bulantı Kusma Karın Ağrısı Ateş	1	1	1	1	1	0	0	0
İshal Kusma	0	1	0	0	0	1	0	0
Bulantı Kusma Halsizlik	1	1	0	0	0	0	1	0
Ateş Bulantı	1	0	0	0	1	0	0	0
Karın Ağrısı Bulantı Kusma	1	1	1	1	0	0	0	0
Baş Ağrısı Bulantı Kusma Karın Ağrısı	1	1	1	1	0	0	0	1

- Frekans Vektör: Binary vektörde kelimenin dokümanda olup olmadığı önemli iken frekans vektörde kelimenin doküman içerisinde olmasının yanı sıra kaç kez olduğuna da bakılmaktadır (Göker ve Tekedere, 2017). Tablo 2.3.’te bulunan frekans vektör matrisi oluşturulmuş olup binary vektörden farkı son dokümanda bulunan “ağrı” kelimesinin karşılığının iki olmasıdır.

Tablo 2.3. Frekans vektör için örnek veri

Doküman	Kelimeler							
	Bulantı	Kus	Karın	Ağrı	Ateş	İshal	Hal	Baş
Bulantı Kusma	1	1	0	0	0	0	0	0
Bulantı Kusma Karın Ağrısı Ateş İshal	1	1	1	1	1	1	0	0
İshal Kusma	0	1	0	0	0	1	0	0
Bulantı Kusma Halsizlik	1	1	0	0	0	0	1	0
Ateş Bulantı	1	0	0	0	1	0	0	0
Karın Ağrısı Bulantı Kusma	1	1	1	1	0	0	0	0
Baş Ağrısı Bulantı Kusma Karın Ağrısı	1	1	1	2	0	0	0	1

- TFxIDF Vektör: TF terim frekansıdır. Yani bir terimin doküman içerisinde kaç kez bulunduğudır. IDF ise ters belge sıklığı olarak adlandırılmakta olup bir terimin tüm dokümanlar içerisinde kaç kez bulunduğunu gösterir. TF ve IDF frekans değerleri çarpıldığında ağırlıklandırma değeri bulunmuş olur.

$$TF(t)=TF(d,t) \quad (2.1)$$

$$IDF(t)=\log(N/df(t)) \quad (2.2)$$

$$W(d,t)=TF(d,t)\times IDF(t) \quad (2.3)$$

t: Terim

d: Doküman

TF (d, t): t. terimin j. dokümandaki frekansı

N: Toplam doküman sayısı

df(t): t. terimi içeren doküman sayısı

Örneğin; Binary vektör ve frekans vektör için kullanılan dokümanları kullanarak TFxIDF ağırlıklandırması Tablo 2.4.'te bulunmuştur (Tokunaga ve Iwayama, 1994).

Tablo 2.4. TFxIDF vektör için örnek veri

Doküman	Kelimeler							
	Bulantı	Kus	Karın	Ağrı	Ateş	İshal	Hal	Baş
Bulantı Kusma	0,067	0,067	0	0	0	0	0	0
Bulantı Kusma Karın Ağrısı Ateş İshal	0,067	0,067	0,368	0,368	0,544	0,544	0	0
İshal Kusma	0	0,067	0	0	0	0,544	0	0
Bulantı Kusma Halsizlik	0,067	0,067	0	0	0	0	0,845	0
Ateş Bulantı	0,067	0	0	0	0,544	0	0	0
Karın Ağrısı Bulantı Kusma	0,067	0,067	0,368	0,368	0	0	0	0
Baş Ağrısı Bulantı Kusma Karın Ağrısı	0,067	0,067	0,368	0,736	0	0	0	0,845

2.1.1.4. Benzerlik Hesaplanması

İki doküman arasındaki benzerlik için eğitim vektörü ve sorgu vektörün olduğu uzayda iki vektör arasındaki açının kosinüs değerine bakılır.

$$\cos(d_1, d_2) = \frac{d_1 \cdot d_2}{\|d_1\| \cdot \|d_2\|} \quad (2.4)$$

Formül 2.4 kullanılarak kosinüs değeri hesaplanır. “ $d_1 \cdot d_2$ ” iki vektörün iç çarpım uzunluğudur. “ $\|d_i\|$ ” d_i . dokümanın norm uzunluğudur. Yani dokümanın her boyutunun karesinin toplanarak karekökünün alınması ile bulunan değerdir (Kaşıkçı ve Gökçen, 2014).

2.1.1.5. Özellik Seçimi

Büyük veri setinin küçük boyutlara getirilmesidir. Böylece çalışma için gereken zaman azalır. Gereksiz görülen veriler silinir. Üç tip özellik seçimi yapılabilir.

- Bilgi Kazancı (BK): Sıklıkla terim sayısı azaltırken kullanılan bu teknik bir kelimenin doküman içerisinde olup olmadığına bakar (Yıldız, 2016).

$$BK = \frac{a}{N} \times \log \frac{a \times N}{(a+c)(a+b)} + \frac{b}{N} \times \log \frac{b \times N}{(b+d)(a+b)} + \frac{c}{N} \times \log \frac{c \times N}{(a+c)(c+d)} + \frac{d}{N} \times \log \frac{d \times N}{(b+d)(c+d)} \quad (2.5)$$

N: Doküman Sayısı

a: x kelimesini içeren sınıf içerisindeki doküman sayısı

b: x kelimesini içermeyen sınıf içerisindeki doküman sayısı

c: x kelimesini içeren sınıf dışındaki doküman sayısı

d: x kelimesini içermeyen sınıf dışındaki doküman sayısı

- N-Gram: Dokümanlardaki kelime dizilimlerinin olasılıklarını inceleyen modellerde kullanılmaktadır. Kendinden önce gelen n kelimenin sıralamasına göre kendinden sonra gelecek olan kelimelerin olasılıklarını tahmin etmeye çalışır. N-gram’da bir kelimenin gelme olasılığı kendinden önce gelen (n-1) kelimeye bağlıdır. En çok kullanılan N-gram türleri uni-gram, bi-gram ve tri-gram’dır. Uni-gram kendinden önce gelen herhangi bir kelimeye bağlı değildir. Bi-gram kendinden önce gelen 1 kelimeye, tri-gram ise kendinden önce gelen iki kelimeye bağlıdır (Amanet, 2017). Tablo 2.5’de hasta epikrizinde bulunan şikayetlerin N gram yöntemine göre dağılımı gösterilmektedir.

Tablo 2.5. N gram örnek tablo

Metin	Ateş Üşüme Titreme Bulantı Sabah Başlamış
Uni-gram	'Ateş', 'Üşüme', 'Titreme', 'Bulantı', 'Sabah', 'Başlamış'
Bi-gram	'Ateş Üşüme ', 'Üşüme Titreme', 'Titreme Bulantı', 'Bulantı Sabah', 'Sabah Başlamış'
Tri-gram	'Ateş Üşüme Titreme', 'Üşüme Titreme Bulantı', 'Titreme Bulantı Sabah', 'Bulantı Sabah Başlamış'
N-gram (n=4)	'Ateş Üşüme Titreme Bulantı', 'Üşüme Titreme Bulantı Sabah', 'Titreme Bulantı Sabah Başlamış'

- **Ki-Kare:** Bu yöntem veri setinde bulunan herhangi bir X ve Y kelimesi arasında ilişki olup olmadığını inceler. Özellik seçimi yaparken bulunan ki-kare değerleri büyükten küçüğe sıralanır. Büyük olan değerler üzerinden seçim yapılır. Formül 2.6'daki gibi hesaplanır (Budak, 2018).

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(\frac{N_{ij} - \frac{N_i \times N_j}{N}}{\frac{N_i \times N_j}{N}} \right)^2}{\frac{N_i \times N_j}{N}} \quad (2.6)$$

N_{ij} : Y'nin i. ve X'in j. düzeyinde bulunan veri sayısı

N_i : Y'nin i. düzeyinde beklenen veri sayısı

N_j : X'in j. düzeyinde beklenen veri sayısı

2.1.2. Literatür Taraması

Metin madenciliği çalışmaları farklı platformlarda aranmış olup Tablo 2.6.'da özetlenmiştir. Bulunan 4729 tane çalışmadan 3347 tanesi sağlık sektörüne aittir.

Tablo 2.6. Metin madenciliği literatür taraması

Platform	Toplam Çalışma Sayısı	Sağlık Sektörü Çalışma Sayısı
Ulusal Tez Merkezi	165	13
Dergipark	137	8
Web Of Science	1240	139
PubMed	3187	3187
Toplam	4729	3347

Erişim Tarihi-Saat:16.04.2021-14:50

Tablo 2.6.'da verilen çalışmalardan ağırlıklı olarak sağlık sektöründen seçilmiş altı tez ve yirmi dört makaleden oluşan otuz çalışma detaylı olarak anlatılmıştır.

“Akut Solunum Yolu Hastalığının Tespiti için ICD-9 Kodlu Başlıca Şikayetlerin ve Teşhislerin Doğruluğu” adlı çalışmada acil servis epikrizleri incelenerek hastaların sınıflandırılma performansı ölçülmüştür (Espino ve Wagner, 2001). Hastalık tanıları ICD-9 tanılarına dayalı olanlar ve hasta şikayetlerine dayalı olanlar olmak üzere ikiye ayrılmıştır. Dört hekim tarafından yapılan incelemeler sonucunda hastalığa ait gold standartlar oluşturulmuştur. Hastalık düzeltilebilir ve düzeltilemez olarak iki grup altında toplanmıştır. Yapılan sınıflandırmaların güvenilirliği ölçülmüştür. Dört hekimin toplu olarak yaptıkları çalışmalarda güvenilirlik %95 iken tek hekim değerlendirmelerinde güvenilirliğin %84’e düştüğü görülmüştür. Akut solunum yolu hastalığının ICD-9 tanılarına göre tespitinde duyarlılık 0,44 ve özgüllük 0,97 olarak bulunmuştur.

“Kalp Kusurlu Bebeklerin Tanımlanmasında Hastane Taburcu Verilerinin Geçerliliği” adlı çalışmada kalp hastalığı olan bebeklere ait ICD kodlarının geçerliliğini epikriz raporlarına göre belirlenmektedir (Frohnert vd, 2005). Çalışmada 2001 yılına ait veriler kullanılarak epikriz raporlarında bulunan anahtar kelimelere göre ICD kodları tanımlanmıştır. 2697 bebeğin ICD kodlarına göre 66 tanesinde kalp hastalığı bulunmuştur. Doktor incelemesi sonunda 24 bebekte bulunan kalp hastalığının varlığı doğrulanmıştır. ICD kodlarının bebeklerin hastalığını tanımlamadaki pozitif tahmin değeri 0,364 ve duyarlılığı 0,857’dir. Çalışma sonunda elde edilen verilere göre ICD koduna göre bebeklerde kalp hastalığını tanımlarken bulunan sonuçların çoğu yanlış pozitif aittir. Bu nedenle sadece ICD kodu belirli kalp hastalıklarının teşhisi için yetersizdir.

“Elektronik Tıbbi Kayıtlardan Nöropsikiyatrik Bozuklukların Fenotipik İmzalarının Belirlenmesi” adlı çalışmada hastalara ait tıbbi kayıtlar kullanılarak otizm, bipolar bozukluk ve şizofreni gibi hastalıkların fenotipik imzalarının bulunması amaçlanmıştır (Lyalina vd, 2013). Bunun için metin madenciliği ve istatistiksel analiz yapılmıştır. Her hastalığın birbiri olan ilişkisine bakılmış ve hasta kayıtlarına göre kümeleme yapılmıştır. Çıkan sonuçlara göre şizofreni ve bipolar bozukluğun birbiri ile örtüştüğü ancak otizmin onlardan ayrıldığı görülmüştür.

“Hipertansiyona Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi” adlı çalışmada kişilerin demografik özelliklerinin ve kan sonuçlarının bulunduğu bir veri seti kullanılarak kişilerin hipertansiyon hastası olup olmadığının sonucuna metin sınıflandırma ile varılmak istemiştir (Kökver vd, 2014). Bunun için

metin sınıflandırma algoritmaları olan C4.5, Naive Bayes ve Çok Katmanlı Algılayıcı algoritmalarını kullanmıştır. En yüksek sonuç %91,66 ile C4.5 ve Bayes olup hangisinin daha iyi olduğuna karar verebilmek için düzensizlik matrisi yöntemine başvurmuştur. Gerçekte hasta olup algoritmanın sağlam dediği hasta sayısı daha az olduğu için C4.5 algoritmasının daha iyi olduğu kanısına varmıştır.

“FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleriyle Otomatik Değerlendirilmesi” adlı çalışmada proje ile ilgili internetteki yorumların bulunduğu bir veri seti kullanılarak metin madenciliği yöntemlerinin proje ile ilgili otomatik yorumlama yapabilmesi amaçlanmıştır (Göker ve Tekedere, 2017). Veriler yapısal olmadığı için veri ön işleme yöntemleri ile öncelikle yapısal hale getirilmiş ve “txt” yapısı olarak kaydedilmiştir. En fazla makine öğrenim algoritmaları ile çalışılmış olup bunun için WEKA programı kullanılmıştır. Çalışmada Naive Bayes, K-NN, J48, SMO ve RBF Network algoritmaları kullanılarak başarı ölçütleri karşılaştırılmıştır. En yüksek doğruluk %88,73 ile SMO algoritmasında olup sınıflandırma algoritmalarının başarı yüzdelerinin %80’nin üzerinde olduğu görülmüştür.

“Karmaşık Hastalıkların Teşhisinde Veri Madenciliği Yöntemlerinin Başarım Karşılaştırması” adlı çalışmada çok etnikli prostat kanseri ve melenom veri kümesi olmak üzere iki çeşit veri kümesi kullanılarak sınıflandırma algoritmalarının arasında başarı karşılaştırması yapılmak istemiştir (Yücebaş, 2018). Karar ağacı, Naive Bayes ve Destek vektör makineleri algoritmalarını kullanan Yücebaş çok etnikli prostat kanseri verisinde %75,68, melenom veri kümesinde %78,60 ile Destek vektör makinelerinin daha iyi sonuç verdiğini gözlemlemiştir.

“Koroner Arter Hastalığı Riskinin Veri Madenciliği Yöntemleriyle İncelenmesi” adlı çalışmada kalp hastalarına ait veri kümesi kullanılmıştır (Cihan vd, 2018). Bu çalışmanın veri kümesinde cinsiyet, yaş, kan sonuçları vb. veriler bulunmaktadır. Kardiyovasküler hastalıkların ölüm oranı en yüksek hastalık türü olduğunu ve tanısı için anjiyografi gerekli olup bununda maliyetinin yüksek olup komplikasyona neden olduğundan bahsetmiştir. Hastaların veri madenciliği yöntemleri ile teşhisinin mümkün olup olmayacağını araştırmıştır. Veri setine Karar Ağacı, Naive Bayes ve Yapay Sinir Ağları algoritmaları uygulanmıştır. En yüksek doğruluk %83,69 ile yapay sinir ağlarında bulunmuştur.

“Metin Madenciliği ve Makine Öğrenmesi ile İnternet Sayfalarının Sınıflandırılması” adlı çalışmada internet sayfalarının sınıflandırılması amaçlanmıştır (Şahin, 2019). İnternet sayfası – sınıfı şeklinde oluşturulan veriye ikili ve çoklu sınıflandırma yöntemleri uygulanmıştır. İkili sınıflandırma için en yüksek doğruluk oranı Lojistik Regresyonda, çoklu sınıflandırmada ise Destek Vektör Makineleri algoritmasında bulunmuştur. Aynı veri setine yapay sinir ağları yöntemi de uygulanmıştır. Bulunan sonuçlara göre Lojistik Regresyon yönteminin yapay sinir ağlarına göre daha hızlı çözümlendiği görülmüştür.

“Elektronik Tıbbi Kayıtlarda Doğuştan Kalp Hastalığı Olan Yetişkinlerin Tanımlanmasında ve Sınıflandırılmasında İddia Verilerinin Doğruluğu” adlı çalışmada doğuştan kalp hastalığının sınıflandırılmasında ICD kodlarının doğruluğunun tespit edilmesi amaçlanmıştır (Cohen vd, 2019). 2000-2014 yılları arasında muayene olmuş kardiyojji hastalarına ait verilerden rassal olarak seçilerek klinik tanı-ICD kodu ilişkisine bakılmıştır. %95 güven aralığı, 0,81 duyarlılık ve 0,99 özgülük ile veriler analiz edilmiştir. ICD kodlarının kalp hastalıklarını belirlemede uluslararası olabileceği ve yüksek doğruluk verdiği sonucuna varılmıştır.

“Metin Madenciliği Teknikleri ile Sosyal Medya Gönderilerinin Analiz Edilmesi: Ankilozan Spondilit Hastalığı Örneği” adlı çalışmada Facebook’ta ankilozan hastalarına ait olan bir sayfadan hastaların hastalıkları ile ilgili yazdıkları metinlerin analizinin yapılması amaçlanmıştır (Gümüşsu, 2019). Toplanan metinler farklı kategoriler etrafında toplanmıştır. Metin sınıflandırma algoritmalarından K-En Yakın Komşu Algoritması, Karar Ağaçları, Naive Bayes Algoritması ve Ardışık Minimal Optimizasyon Algoritması kullanılmıştır. En iyi sonuç %53,3 ile Ardışık Minimal Optimizasyon Algoritmasına aittir. Ayrıca kategorilere ait kelimeler için birliktelik analizi yapılmıştır. Birliktelik analizi için Apriori algoritmasından yararlanılmıştır. Anahtar kelime tespiti yapılmış ve bulunan bu anahtar kelimelerin yıllar bazında farklılaştığı görülmüştür. Bu nedenle kelime bulutları oluşturulmuştur.

“Uluslararası Haber Raporlarının Rapor İçeriklerinde Kullanılan İfadelere Göre Makine Öğrenmesi Yöntemiyle Sınıflandırılması ve Denetlenmesi” adlı çalışmada uluslararası yayın yapan bir yerin haber ve bilgi raporlarının sınıflandırılması amaçlanmıştır (Durnagöl, 2020). 600 adet rapor üç farklı sınıfa ayrılmaya çalışılmıştır. Bunun için Rasgele Orman Karar Ağacı, Naive Bayes gibi algoritmalar kullanılmıştır.

En iyi sınıflandırma sonucunun %91 ile Rasgele Orman Karar Ağacı algoritmasında olduğu bulunmuştur.

“Depresif Semptomları Çıkarmak ve Majör Depresif Bozukluğun Tanısını Elektronik Sağlık Kayıtlarından Doğrulamak İçin Metin Madenciliği Kullanma” adlı çalışmada psikiyatri kliniğine ait epikrizlerin ICD kodlarının doğruluğunun analiz edilmesi ve bireylere ait hastalık belirtilerinden metin madenciliği teknikleri ile bilgi çıkarımı yapılması amaçlanmıştır (Wu vd, 2020). 2006-2016 yılları arasındaki epikriz notlarından bazıları seçilmiş, hastalıklara göre gruplama yapılmıştır. Epikriz notlarından gold standart oluşturmak için toplamda 15.271 cümle metin madenciliği yöntemleri ile analiz edilmiştir. Analiz sonuçlarına göre majör depresif bozuklukların tanılarında 0,86 kesinlik ile emin olunmuştur. Minör bozuklukların kesinlik değerlerinden emin olunamamıştır. Minör bozukluklarda bipolar bozukluğun kesinlik değeri 0,64, minör depresyonun kesinlik değeri 0,61 olarak bulunmuştur.

“İsveç'te Konjenital Kalp Hastalığı Olan Hastalarda Miyokard Enfarktüsü Tanısının Doğrulanması” adlı çalışmada doğuştan kalp hastalığı olan yetişkinler için kalp krizi teşhisinin zorluğundan kaynaklı teşhislerde tıbbi kaynak kullanımının doğruluğu hesaplanmak istenmiştir (Fedchenko vd, 2020). 1970-2012 yılları arasında doğmuş olan ve kalp hastalığı-kalp krizi teşhisi olan hasta verileri toplanmıştır. Normal şartlar altında kalp krizi teşhisinin konulabilmesi için EKG, ventriküler hipertrofi gibi klinik müdahaleler gerekmektedir. Ancak bu müdahaleler yapılmadan sadece tıbbi verilerin analizi ile yapılan çalışmada %88 doğruluk oranı bulunmuştur. Test edilen veri setinin %74,8'inde kalp hastalığı tanısı olup bunlardan %70,2'si doğru sınıflandırılmıştır. Doğru tanımlanan kalp hastalığı tanısı olan bireylerin %88'inde kalp krizi teşhisi de bulunmaktadır. Yapılan çalışmalar sonucunda kalp hastalığı olan bireylerin kalp krizi teşhisine sahip olduğu kanısına varılmıştır.

Tablo 2.7.'de literatürde bulunan bazı çalışmalar ve çalışmaların yapılaş amaçları verilmiştir.

Tablo 2.7. Metin madenciliği örnek çalışmaları

Çalışma Adı	Yazar- Yıl	Amaç
Türkçe Radyoloji Raporlarında Metin Madenciliği	Kocatekin (2013)	Elle yapılan rapor sınıflandırma işleminin metin madenciliği sınıflandırma yöntemi ile anatomik bölgelere göre sınıflandırılmasıdır.
Yumurtalık Kanseri Veri Kümesindeki Gen İfadelerinin Veri Madenciliği ile Analizi	Demircoğlu ve Bilge (2015)	Yumurtalık kanseri verisine ait öz niteliklerin seçimi ve sınıflandırma algoritmalarının doğruluk oranlarının karşılaştırılmasıdır.
Vaka Tespitini İyileştirmek İçin Elektronik Tıbbi Kayıt Metninden Bilgi Çıkarma: Sistemik Bir İnceleme	Ford vd (2016)	Klinik vaka tespiti için kullanılan elektronik kayıtlar ve makaleler yardımıyla kural çıkarımı yapmak ve vaka algoritması geliştirebilmektir.
Kalp Yetmezliği Yeniden Kabulleri için Makine Öğrenimi Tekniklerinin Analizi	Mortazavi vd (2016)	Kalp yetmezliği bulunan hastaların hastaneye yatışlarını tahminlemek ve metin madenciliği algoritmalarının tahminlerini karşılaştırmaktır.
Sağlıkla İlgili Sosyal Medya Forum Verilerinde İçeriğinin Bilgilendirici Özellikleri: İlaç Yan Etkilerinin Araştırması Üzerine Bir Çalışma	Kaya (2018a)	İnternet sitesinde bulunan hasta yorumlarının toplanarak ilaç yan etkileri için metin madenciliği çalışması yapmak ve parametreler arasındaki gizli ilişkileri tespit edebilmektir.
Veri Madenciliği Tekniklerini Kullanan Etkili Kalp Hastalığı Tahmin Sistemi	Singh vd (2018)	Metin madenciliği yöntemlerinden yapay sinir ağları yöntemi ile kişilerin kalp hastalığı riskini tespit edebilmektir.
Türk Kulak Burun Boğaz Taburcu Özetlerinden İlişkilendirme Kurallarının Çıkarılması	Yolcular vd (2018)	KBB branşına ait epikrizlerinin metin madenciliği yöntemleriyle yapılandırılması ve ARM kullanılarak ilişkisel kurallar çıkarabilmektir.

Tablo 2.7. (devam)

Çalışma Adı	Yazar- Yıl	Amaç
Sağlık Verilerinde Metin Madenciliği Algoritmalarının Analiz Amaçlı Kullanılması	Akis (2019)	Hastaların geçmiş hikayeleri ve teşhis edilen hastalıklarının birbirleri ile ilişkisine bakılarak tanı koymaya yardımcı sistem oluşturmaktır.
Metin Madenciliği ile Ruhsal Bozukluklarla İlgili Genlerin Bulunması	Wu vd (2019)	Birliktelik analizi ve kümeleme yöntemleri ile kelimeler arası birlikteliği bulmak, ruhsal bozuklukların etkilediği ortak genetik faktörlerini tespit edebilmektir.
Koroner Arter Hastalığını Tahmin Etmek: İki Veri Madenciliği Algoritması Arasında Bir Karşılaştırma	Ayatollahi vd (2019)	Koroner arter hastalığı bulunan hastalara ait veri setinde DVM ve YSA metotlarının karşılaştırılmasıdır.
Veri Madenciliği ile Kalp Hastalığı Teşhisi	Taşçı ve Şamli (2020)	Kalp hastalığı tanısı koyulmuş olan hastaların verilerine metin madenciliği sınıflandırma algoritmaları uygulanarak kalp hastalığı teşhisinin koyulmasını sağlamaktır.
Sağlık Alanında Metin Madenciliği Yöntemi	Toplu ve Cangür (2020)	PubMed veri tabanından toplanan dokümanlardaki kanser ifadelerini insan ve fare için iki farklı kategoride sınıflandırmaktır.
Covid-19 Salgınında Yaşanan Haftalık Duygusal Değişimler: Türkiye Örneği	Silahtaroglu vd (2020)	#EvdeKal hashtagi altında toplanan yazışmaların metin madenciliği ile analizi yapılarak kelimeler arası ilişkiyi incelemektir. Bulunan sonuçlara göre duygu analizi yapılmıştır.

Tablo 2.7. (devam)

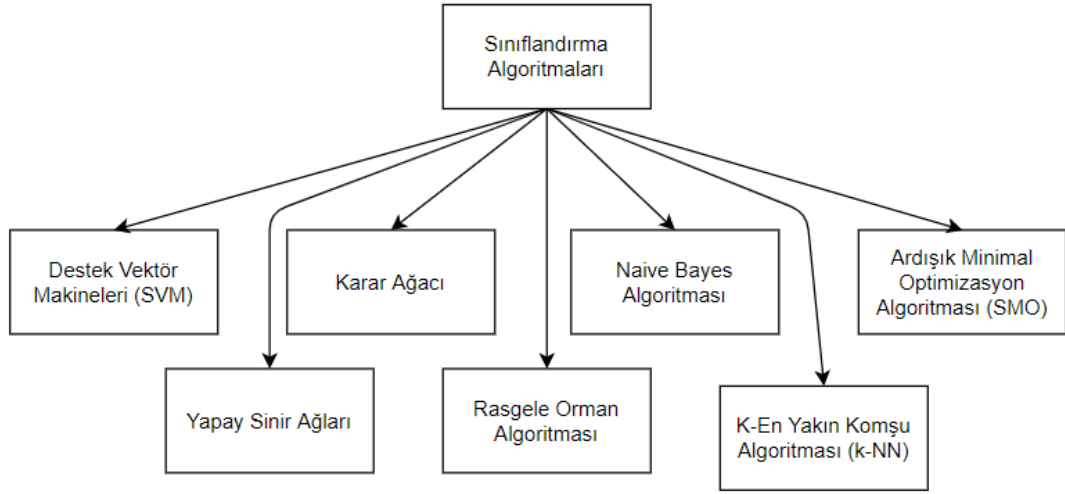
Çalışma Adı	Yazar- Yıl	Amaç
Meme Kanserinin Nüksünü Tahmin Etmek İçin Veri Madenciliği Algoritmalarını Modelleme ve Karşılaştırma	Mosayebi vd (2020)	Minimum beş kez kanser izleme tanısı olan meme kanserine sahip hastaların hastalıklarının tekrar nüksetme durumunu sınıflandırma algoritmaları ile tahminleyebilmektir.
Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı	Akgül vd (2020)	Tiroit hastalığının teşhisinde metin madenciliğini kullanabilmek ve girişimsel işlem riskini azaltabilmektir.
Kovid-19 Salgını ve Mültecilere Yönelik Dijital Nefret Söylemi: Büyük Veri Perspektifinden Metin Madenciliği Tekniği ile Kullanıcı Kaynaklı İçeriklerin Analizi	Kuş (2021)	Youtube'daki mülteci videolarının altına yapılan yorumların toplanması ve bu yorumlarda bulunan nefret söylemlerinin aralarındaki benzerliklerin bulunması, belli kategorilere göre sınıflandırılmasıdır.
Metin Madenciliği ile Tıbbi Tedavi Alanlarının Yakınlıklarının Ölçülmesi	Kurban (2021)	Bakanlığın web sitesinden toplanan verilerle birlikte tıbbi tedavilerde kullanılan terimlerin birlikteliklerini incelemek ve alanlara göre kümeleme yaparak polikliniklere ait randevu sayısını azaltabilmektir

Yapılan çalışmalar incelendiğinde genellikle sağlık alanında belirli hastalıkların verilerinin toplanarak hastalara tıbbi müdahale yapılmaksızın hasta veya sağlıklı tanısı koymaya ve doğruluk performansını ölçmeye yönelik olduğu görülmüştür. Tanı düzeyi geniş olan çalışmaların yapılmıyor olması eksik bulunmuş ve bu alanda çalışma yapılmasına karar verilmiştir.

2.2. Sınıflandırma

İnternetin yaygınlaşması, kitap – makale gibi yayınların artması gibi nedenlerden dolayı bilgiye erişim artmış ve verilerde yığılma olmuştur (Haltaş vd, 2015). Çok miktarda verinin yararlı yönleri olduğu gibi sorun yaratan yönleri de vardır. Bu sorunlardan en önemlisi verilere hızlı erişim sağlayabilmek için verileri uygun kategorilere göre sınıflandırmaktır (Erten, 2015).

Sınıflandırma; sınıfı belli olan verilerden yararlanarak sınıfı belli olmayan verileri belli sınıflara atayabilmektir. Amaç sınıfı bilinen verileri kullanarak modeli eğitebilmek ve bilinmeyen verileri modelin otomatik sınıflandırmasını sağlamaktır. Veri seti öncelikle eğitim ve test kümesi olarak ikiye ayrılır. Eğitim kümesi kullanılarak modelin öğrenmesi, test kümesi kullanılarak atama işleminin yapılması sağlanır. Bu işlemler için Şekil 2.2.'de verilen farklı sınıflandırma algoritmaları kullanılmaktadır (Çalış vd, 2013; Sukanya ve Biruntha, 2012).



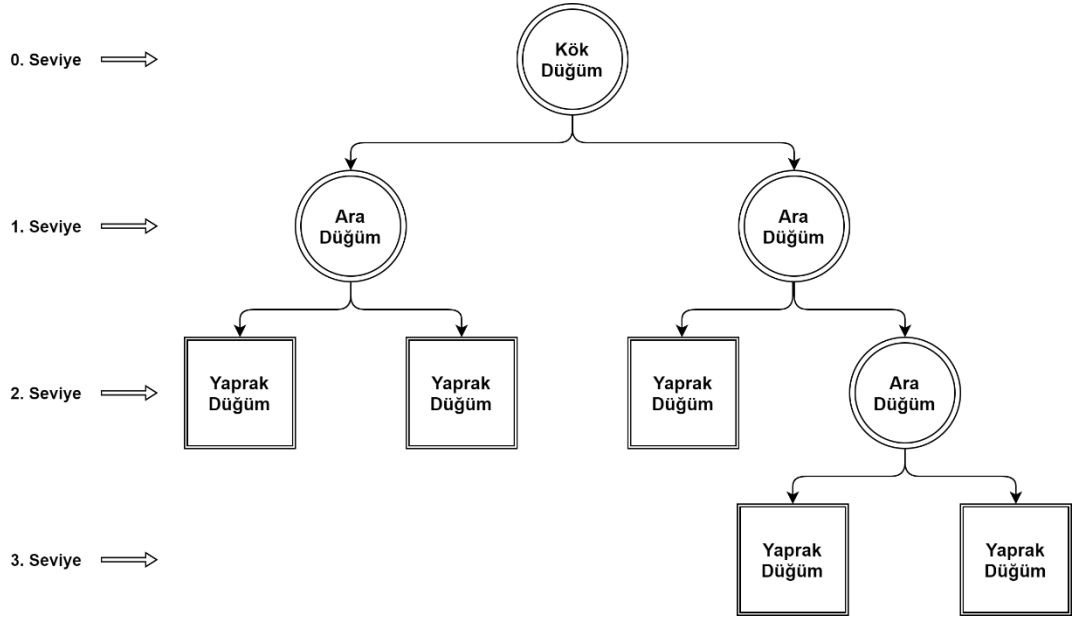
Şekil 2.2. Sınıflandırma algoritmaları

Karar ağacı, destek Vektör Makineleri, Ardışık Minimal Optimizasyon algoritması, K-En Yakın Komşu Algoritması, Naive Bayes Algoritması, Yapay Sinir Ağları ve Rasgele Orman algoritması bu algoritmalar arasındadır. Algoritmalara ilişkin detaylı bilgiler başlık 2.2.1'in altında verilmiştir.

2.2.1. Sınıflandırma Algoritmaları

2.2.1.1. Karar Ağaçları

Karar ağaçları, sınıflandırma metodunda en çok kullanılan algoritmalarından bir tanesidir. Yapısından dolayı anlaşılır kurallar ortaya koyar. Kolay entegre olması ve kolay yorumlanabilmesi nedeniyle sıklıkla kullanılmaktadır. Karar ağaçları genelden özele giden bir yapıya sahip olup aşağıya doğru indikçe eğitilmiş verilerden oluşmaktadır (Demirel, 2019). Yapı olarak ağacı andıran algoritmada düğüm ve dallar vardır. Düğümler kök düğüm, ara düğüm ve yaprak düğüm olarak üçe ayrılır. Karar ağaçları kök düğümle başlar (Tokalakçioğlu, 2020). Aşağı doğru indikçe küçük gruplara bölünen bir yapı söz konusudur (Demirel, 2019). Artık bölünemeyen düğümler yaprak düğüm adını alır ve arada bölünmeyi sağlayan düğümler ara düğümlerdir. Kök düğüm ve yaprak düğümleri birbirine bağlayan yapılar ise dallar olup dal-düğüm yapısı Şekil 2.3.'te gösterilmiştir (Tokalakçioğlu, 2020).



Şekil 2.3. Karar ağacı algoritması

Karar ağaçları oluşturulurken belli sorular sorularak yukarıdaki yapı oluşturulur. Kök düğümden yaprak düğüme inerken sorulan sorulardan alınan cevaplara göre oluşturulan yapı “Eğer” ile oluşturulur. Her yanıt yeni düğümü oluşturur ve devamında soru sorulamayacak yapılar yaprak düğümleri meydana getirir. Her yaprak için farklı kurallar söz konusudur (Demirel, 2019). Literatürde öne çıkan karar ağacı algoritmaları ID3, C4.5, C5.0, CART ve CHAID algoritmalarıdır (Tokalakçioğlu, 2020).

Bir veri setinde bulunan veri sayısı arttıkça karar ağaçlarında gereksiz düğümler oluşabilmektedir. Bu durumu engelleyerek optimum ağacı elde etmek ve karar ağaçlarının veri setini ezberlemesinin önüne geçebilmek için budama yapılır (Gümüşçü vd, 2016). Ön budama ve sonradan budama olmak üzere iki tip budama çeşidi vardır. Sonradan budama daha başarılı sonuçlar verir. Sonradan budama yaparken sık kullanılan iki farklı yöntem vardır. Bu yöntemlerden biri hata tahminini tespit etmek, diğeri ise minimum açıklama uzunluk ilkesi ile budama yapmaktır. Hata tahminini kullanarak budama yaparken bir alt ağacın hata tahmin değeri ağaçtaki yaprakların ağırlıklı hata tahminlerinden oluşur. Formül 2.7'deki gibi hesaplanır. Son düğümün hata değeri bir önceki düğümün hata değerinden büyük olduğunda ağaçta budama yapılır (Saedsayad, 2021).

$$e = \frac{(f + \frac{z^2}{2N} + z \sqrt{\frac{f^2}{N} + \frac{z^2}{4N^2}})}{(1 + \frac{z^2}{N})} \quad (2.7)$$

e: Hata tahmini

f: Eğitim verisi hata değeri

N: Yapraklarda gözlenen örneğin sayısı

z: Normal dağılım değeri

Minimum Açıklama Uzunluk İlkesi ile budamada ise yaprak düğümün doğruluğu test edilir. Bu yöntemde karar ağacının budanacağı nokta son yaprakta bulunan örnek sayısı ile bulunur. Mümkün olduğunca az sayıda örnek bulunduran en anlamlı noktada ağacı budamayı öngörür (Kuzey, 2012).

2.2.1.2. Destek Vektör Makineleri (DVM)

Destek Vektör Makineleri verileri sınıflandırmak veya veriyi tahmin etmek için kullanılır. 1960 yılında Vladimir Vapnik ve Alexey Chervonenkis'in geliştirdiği makine öğrenim algoritmasıdır. Sınıflandırma algoritması olarak kullanılmasına rağmen sağlık sektöründe tanı yöntemi olarak da kullanılmaktadır (Akşehirli vd, 2013).

Destek Vektör Makinelerindeki amaç verilerin en iyi şekilde sınıflandırılmasını sağlayan ve farklı verileri birbirinden ayıran ideal ayırıcı düzlemin oluşturulmasıdır. Her veri bir vektör olarak tanımlanır ve farklı sınıflarda bulunan vektörlerin arasındaki mesafenin maksimum olması hedeflenir. Her sınıfın en ucunda bulunan vektörler

destek vektör olarak seçilir ve bu vektörler üzerinden işlem yapılır. Farklı sınıflar arasında birçok ayırıcı düzlem bulunmaktadır. Bu düzlemlerden biri optimum ayırıcı düzlemdir. Optimum ayırıcı düzlem iki farklı sınıfın destek vektörleri arasındaki mesafeyi maksimum yapan düzlemdir (Çomak, 2004).

2.2.1.3. Ardışık Minimal Optimizasyon Algoritması (SMO)

Diğer adı “Sıralı Minimal Optimizasyon Algoritması” olan Ardışık Minimal Optimizasyon Algoritması destek vektör makineleri algoritmasının eğitim veri setinde kullanılan optimizasyon algoritmasıdır (Uzer, 2014). Bu algoritma formül 2.8 ve 2.9’da verildiği üzere mümkün çözüm kümeleri arasında minimum sonucu bulmayı hedeflemektedir (Elmas, 2012). Büyük veri yapısında bulunan problemleri daha hızlı çözmeye çalışan denetimli sınıflandırma algoritmasıdır (Uzer, 2014).

$$\max_{\alpha} \psi(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j k(x_i, x_j) \alpha_i \alpha_j \quad (2.8)$$

$$\sum_{i=1}^N y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq c \quad \text{ve} \quad i=1, \dots, n \quad (2.9)$$

Denklemden x_i eğitim setinde bulunan örnek veri, y_i 1 ve -1 arasında değer alabilen hedef değer, α_i Lagrange çarpanı ve c maliyettir (Uzer, 2014). Destek vektör makinelerinde kullanarak ikili Lagrange çarpanlarını seçer ve optimizasyon için çarpanların uygun değerlerini bularak modeli günceller (Kılıç vd, 2015).

2.2.1.4. K-En Yakın Komşu Algoritması (K-NN)

Veri setinde bulunan bir verinin hangi sınıfta olacağına en yakınındaki k komşusuna bakarak karar veren algoritmalarıdır. Yoğunluğa dayalı metin sınıflandırma algoritmaları arasında sayılmaktadır. K-NN algoritması kullanılarak bir verinin sınıfı belirlenirken öncelikle veri düzlem içerisine yerleştirilir. Önceden sınıfı belli olan k tane komşu ile yeni veri arasındaki mesafeler ölçülür. Uzaklık hesaplanırken öklit, mahalnobis, manhattan, haming, kosinüs benzerlik gibi farklı hesaplama yöntemlerinden yararlanılabilir. Hesaplama sonucuna göre yeni veri kendisine en yakın k sınıfa dahil edilmektedir.

K-NN algoritmasında en önemli kısımlardan biri k sayısının belirlenmesiyle ilgilidir. Eğer k sayısı olması gerekenden küçük bir sayı alırsa sınıflandırmada kullanılan test verisi gürültüden etkilenecek ve sonuçlar yanıltıcı olacaktır. k sayısı olması gerekenden büyük olursa veri genişliği artacağı için çok uzak komşu sınıflar veriyi etkileyecektir. k sayısı veri setinde bulunan bağımsız değişken sayısından

etkilenmektedir. Bu nedenle k sayısını bağımsız değişken sayısına göre belirlemek gerekmektedir (Bayazıt vd, 2020).

2.2.1.5. Naive Bayes Algoritması

Naive Bayes, hedef değişken ve bağımsız değişken arasındaki ilişkiyi analiz eden tanımlayıcı bir sınıflandırma algoritmasıdır. Verilerin sınıflandırılmasında kullanılan olasılık temelli bir algoritmadır (Abdullahi, 2018; Çalış vd, 2013). Verinin bulunduğu sınıfı tahmin etmek için verinin ait olabileceği sınıfının koşullu olasılığı hesaplanır. Koşullu olasılık normal şartlar altında formül 2.10'a göre hesaplanır (Pawlak, 2003).

$$P(A/B)=\frac{P(B/A)\times P(A)}{P(B)} \quad (2.10)$$

$P(A)$ = Bağımsız A olayının olma olasılığı (ön olasılık)

$P(B)$ = Bağımsız B olayının olma olasılığı (ön olasılık)

$P(B/A)$ = A olayı olduğunda B olayının olma olasılığı (koşullu olasılık)

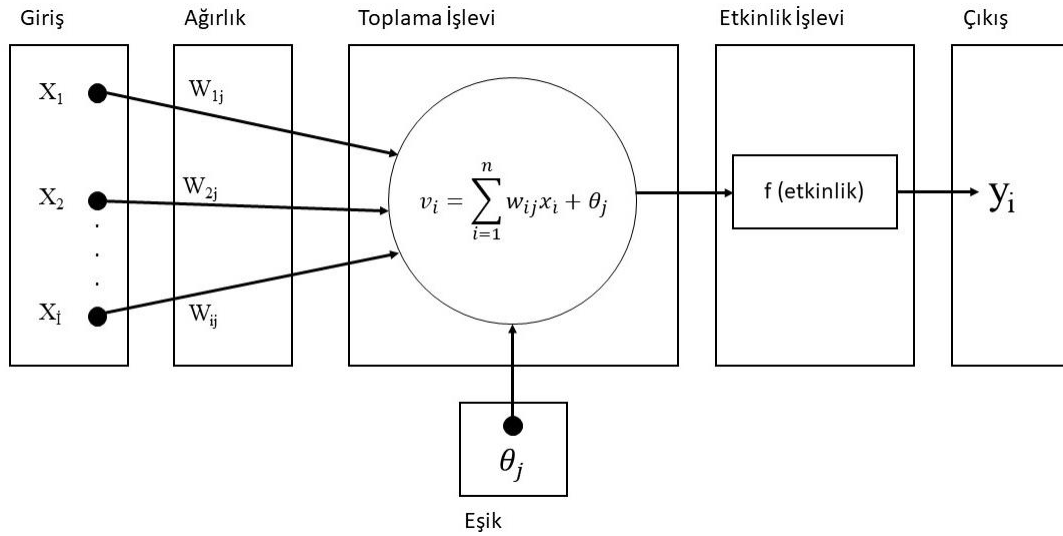
$P(A/B)$ = B olayı olduğunda A olayının olma olasılığı (koşullu olasılık)

2.2.1.6. Yapay Sinir Ağları

İnsanoğlu 1950 yılının sonlarından beri bazı problemlerin çözümlerini doğada aramaya başladı. Bilgi teknolojilerinin zamanla gelişmesiyle birlikte bilim insanları sorunlarına doğada buldukları nesnelere taklit ederek çözüm bulmaya çalıştı. Zamanla insanların biyolojik sinir sistemlerini taklit eden bilişim sistemlerine verilen genel ad yapay sinir ağları oldu (Rençber, 2018). Yapay Sinir Ağları; insan beyninden yola çıkılarak geliştirilen, bağlantılar aracılığıyla birbirine bağlanan, her birinin kendine ait belleğinin olduğu işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarına verilen addır (Elmas, 2018).

1943 yılında Warren CcCullach ve Walter Pitts tarafından ilk yapay sinir ağı modeli geliştirdi. Bu model hesaplama yapan elektrik devresiyle ilgilidir. 1948'de Wiener sinirlerin çalışma prensibine, 1949'da Hebb sinirlerin öğrenmesine değindi. Hebb'in çalışması kural olarak kabul edilerek 1950 yılından sonra sinirlerin hesaplama yeteneği artırılmaya çalışıldı. 1950'lerden günümüze kadar Frank Rosentblatt, John Von Neuman, Anderson gibi pek çok bilim adamı bu alanda çalışmaya devam etti (Elmas, 2018). Günümüzde mühendislik, tıp, fen bilimleri gibi farklı alanlarda kullanılan yapay sinir ağları girdi ve çıktı katmanlarından oluşur (Rençber, 2018).

Katmanlar sinir sisteminin en küçük yapı taşları olan nöronların birleşmesinden oluşur. Farklı katmanların birleşmesiyle de sinir ağları oluşur (Gümüştü, 2019). Yapay sinir ağının yapısı en temel hali ile bir Şekil 2.4.'de gösterilmiştir.



Şekil 2.4. Yapay sinir ağlarının temel yapısı

Giriş verileri x_i ile gösterilmektedir. x_i 'ler kendi ağırlıkları (w_i) ile çarpılır ve eşik değeri (θ_j) ile toplanır. Elde edilen sonuç ile etkinlik işlemi yapılır ve sonuçta çıktı değeri (y_i) bulunur (Elmas, 2018).

2.2.1.7. Rasgele Orman Algoritması

Leo Breiman tarafından 2001 yılında geliştirilerek sınıflandırma algoritmalarından biri olan rasgele orman algoritması sınıflandırma yaparken verilerin önceden yüklenmesi ile oluşturulup her sınıf için değişken aday kümelerini bulmaya çalışan bir algoritmadır. Çoklu sınıf problemi bulunan veri setlerinde ağırlıklı olarak kullanılan bu algoritma aynı zamanda regresyon yöntemi olarak da bilinmektedir (Diaz-Uriarte ve Alvarez de Andres, 2006). Rasgele orman algoritması karar ağaçlarında sıklıkla görülen aşırı öğrenme durumunun önüne geçmekte ve tüm değişkenler için en iyi sınıfı bulmaya çalışmaktadır (Malkoçoğlu ve Malkoçoğlu, 2020).

2.2.2. Sınıflandırma Metodu Performans Ölçütleri

Metin sınıflandırma algoritmalarında performansı ölçmek için çeşitli kriterlere bakılır. Bu kriterlerden bazıları matrislerde görünür. Karışıklık matrisi ve doğruluk matrisi olmak üzere iki tip matris vardır. Performans ölçüm kriterleri doğruluk

matrisinde bulunan doğruluk, kesinlik, duyarlılık, özgüllük, negatif prediktif değer, F ölçөгüdür. Bu kriterler karışıklık matrisinde bulunan Doğru Pozitif, Yanlış Pozitif, Doğru Negatif ve Yanlış Negatif değerlerinden elde edilir. Bir modelin başarı oranının yüksek olması için Doğru Pozitif ve Doğru Negatif değerlerinin yüksek olması gerekmektedir (Haltaş vd, 2015; Yang, 1998). Doğru Pozitif (DP) gerçekte doğru sınıfta olan verinin tahminde de doğru sınıfa atanması, Doğru Negatif (DN) gerçekte yanlış sınıfta olan verinin tahminde de yanlış sınıfa atanması, Yanlış Pozitif (YP) gerçekte yanlış sınıfta olan verinin tahminde de doğru sınıfa atanması ve Doğru Negatif (YN) gerçekte doğru sınıfta olan verinin tahminde de yanlış sınıfa atanmasıdır.

Tablo 2.8.'de sınıflandırma algoritması performans ölçütlerinin gerçek ve tahmin durumuna göre girdiği gruplar gösterilmektedir.

Tablo 2.8. Sınıflandırma algoritması performans ölçütleri

		Tahmin	
		Doğru	Yanlış
Gerçek	Doğru	Doğru Pozitif	Yanlış Negatif
	Yanlış	Yanlış Pozitif	Doğru Negatif

2.2.2.1. Doğruluk (Accuracy)

Sınıflandırma metodunda modelin başarı oranını ölçen en yaygın yöntem doğruluktur. Doğru sınıfa atanmış olan verilerin sayısının tüm duruma bölünmesi ile elde edilir (Amanet, 2017). Kullanılan veri setindeki sınıflandırmanın homojen dağılmadığı durumlarda doğruluk yanıltıcı bir parametre haline gelir. Örneğin; sağlıklı bireyler sınıfındaki veri sayısı, hastalıklı bireyler sınıfındaki veri sayısından fazla olduğunda bütün test örnekleri sağlıklı bireyler olarak sınıflandırılacağından sınıflandırma işleminin doğruluk oranı yüksek çıkar. Bu nedenle veri sayılarının birbirlerinden baskın olduğu durumlarda sadece doğruluk parametresini değerlendirmek doğru olmaz. Doğruluk ile birlikte diğer performans ölçütlerinin kullanılması da gereklidir (Gümüşsu, 2019).

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (2.11)$$

2.2.2.2. Kesinlik (Precision)

Kesinlik, sınıfı gerçekte doğru iken tahminde de doğru olarak tahmin edilmiş örnek sayısının (DP) tahminde doğru sınıfına atanan tüm örnek sayısına (DP + YP) bölümüdür. Doğru cevapların yüzdesi olarak da ifade edilir (Karasu, 2019).

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (2.12)$$

2.2.2.3. Duyarlılık (Sensitivity)

Duyarlılık, sınıfı gerçekte doğru iken tahminde de doğru olarak tahmin edilmiş örnek sayısının (DP) gerçekte doğru sınıfına atanan tüm örnek sayısına (DP + YN) bölümüdür. Diğer bir ifade ile olası doğru cevapların yüzdesidir (Karasu, 2019).

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (2.13)$$

2.2.2.4. Özgüllük (Specificity)

Özgüllük, sınıfı gerçekte yanlış iken tahminde de yanlış olarak atanmış örnek sayısının (DN) gerçekte yanlış sınıfına atanan tüm örnek sayısına (DN+YP) bölümüdür. Diğer bir ifade ile yanlış tahminleri ayırabilme yeteneğidir.

$$\text{Özgüllük} = \frac{DN}{DN+YP} \quad (2.14)$$

2.2.2.5. Negatif Prediktif Değer (Negative Predictive Value)

Negatif Prediktif Değer, sınıfı gerçekte yanlış iken tahminde de yanlış olarak atanan örnek sayısının (DN) tahminde yanlış sınıfına atanan tüm örnek sayısına (DN+YN) bölümüdür. Diğer bir ifade ile gerçek yanlışlar içerisindeki doğruluk oranıdır (Çamlıca ve Dişçi, 2008).

$$\text{Negatif Prediktif Değer} = \frac{DN}{DN+YN} \quad (2.15)$$

2.2.2.6. F-Ölçeği (F-Score)

F Ölçeği, kesinlik ve duyarlılığın harmonik ortalamasıdır. 0 ve 1 arasındaki değerleri alır. Aşağıdaki gibi hesaplanır (Karasu, 2019).

$$\text{F ölççeği} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (2.16)$$

2.3. Birliktelik Analizi

Yakın geçmişin yakın gelecekte çok farklı olmadığını kabul ederek geçmişteki verilerin analizi ile bulunan kuralların, gelecekteki verilerin tahminlemede kullanılmasına birliktelik analizi denir (Erpolat, 2012). Yani veri setinde bulunan değişkenlerin birbiri ile olan bağlantısıdır. Bir marketten şeker alan kişinin çay alma olasılığı, bir hastada hem tansiyon hem de şeker hastalığının bulunma olasılığı, tansiyon hastalarının kalp krizi geçirmiş olma olasılığı gibi farklı örnekler verilebilir (Altunkaynak, 2017). Birliktelik analizi yaparken farklı algoritmalar kullanılabilir.

2.3.1. Birliktelik Analizi Algoritmaları

Birliktelik Analizi algoritmaları ardışık algoritmalar ve paralel – dağıtılmış algoritmalar olarak ikiye ayrılmaktadır. Ardışık algoritmalar ürün kümelerinin oluşturulup sayıldığı mantıksal ifadelerden oluşur. Paralel – Dağıtılmış algoritmalar ise ürün kümeleri arasında paralellik oluşmasını sağlar. Tablo 2.9.'da bu algoritmaların altında bulunan algoritmalara değinilmiştir (Erpolat, 2012).

Tablo 2.9. Birliktelik analizi algoritmaları

Ardışık Algoritmalar	Paralel- Dağıtılmış Algoritmalar
APRIORI	CD (Sayım Dağılımı)
AIS	PDM (Paralel VM)
SETM	CCPD (Ortak Aday Bölünmüş Veri Tabanı)
FP-GROWTH	DD (Veri Dağılımı)
Bölümleme Tekniği	HPA (Bağlantı Kurallarının Çırpı Temelli Paralel Madenciliği)
Örnekleme Tekniği	PAR (Paralel Bağlantı Kuralları)
CARMA	DMA (Dağıtılmış Madencilik Algoritması)
DNS (Dinamik Nesne Sayımı)	SH (Çarpık Taşıma)
SDAB (Sıra Dışı Aday Belirleme)	HD (Hibrid Dağılımı)

Bu çalışmada veriler arası paralellikten ziyade verilerin içerisinde bulunan kelimelerin birlikte kullanım sayılarına bakılacağından dolayı birliktelik analizi algoritmalarından ardışık algoritmalar kullanılmıştır. Başlıklar halinde ardışık algoritmalarla ilişkin bilgiler verilmiştir.

2.3.1.1. APRIORI

Birliktelik Kuralları algoritmaları arasında en çok kullanılan algoritmadır. Veri setinin x . seviyesindeki sık geçen kelimelerini tespit etmek için $(x-1)$. seviyeyi kontrol eden düzey türü bir algoritmadır (Çelik, 2019). Agrawal ve Srikant tarafından 1994 yılında yapılan bir çalışmada apriori algoritması ortaya çıkmıştır. Algoritma minimum güven ve minimum destek için belirlenen sınırlar dahilinde verilerden birliktelik kuralları çıkarmaya çalışır.

Algoritmada frekansı yüksek kelimeleri bulmak için öncelikle veri seti taranır. İlk taramadan itibaren minimum destek sınırını sağlayan kelime kümeleri bulunur. İleriki taramalarda hep bir önceki tarama baz alınır ve yeni kelime grupları bulunur. Algoritmada geçen “prior” kelimesi de önceki anlamına gelip buradan gelmektedir. Yani her tarama sonucunda bulunan kelime kümesi bir sonraki tarama için aday kümeyi oluşturur. En sık kullanılan veri kümesi bulununcaya kadar döngü devam eder ve bulunduğu biter.

2.3.1.2. AIS

Veri tabanında sürekli taramalar yaparak geniş veri kümesi bulmayı hedefleyen algoritmalarıdır (Karamaşa, 2018). Verilerin alfabetik sıraya göre olması gerekmektedir. Veri tabanında bulunan her bir veri tek tek sayılarak frekansları belirlenir ve frekanslarına göre aday veri kümeleri oluşturulur. Veriler oluşturulan aday veri kümelerine dahil edilir. Bir sonraki taramada kümelere atanan veriler arasındaki ortak veriler belirlenir. Belirlenen aday veri kümeleri için budama tekniği ile gereksiz veri kümelerinin kaldırılması sağlanır. Elde kalan veri kümelerinde destek değeri bulunur ve eşik değeri ile karşılaştırılır.

Diğer algoritmalara göre olumsuz yönü veri tabanında çok kez tarama yapıldığı için çok sayıda aday veri kümesi oluşur. Bu durum zaman ve veri kümesi sayısında israfa yol açmaktadır.

2.3.1.3. SETM

AIS algoritmasına benzeyen SETM algoritması AIS algoritmasının eksik ve bazı problemlere sahip olması nedeniyle 1995 yılında Hautsmal tarafından geliştirilmiştir (Karamaşa, 2018; Tokyürek, 2019). AIS algoritmasının eksikliği çok sayıda aday veri kümesi oluşturmasıdır. SETM algoritması da AIS gibi veri tabanını çok sayıda tarar ve geniş veri kümeleri bulmaya çalışır. Veri kümeleri veri sayısı ve özellik numarası

olmak üzere iki parametreden oluşur. AIS algoritmasının adımlarına göre tek fark sık kullanılan veri kümelerini oluşturmak için SQL veri tabanının kullanılmasıdır. AIS gibi çok sayıda tarama yaptığı için sıkça kullanılan algoritmalar arasında değildir.

2.3.1.4. FP-GROWTH

Apriori algoritmasından farklı olarak “sık desen ağacı madenciliği algoritması” tabanlıdır (Aydemir, 2018). Tüm verilerin ikili nominal değerlerden oluştuğunu varsayar. Kullanıcı bu ikili değerlerden hangisinin pozitif olacağını belirler. Market-sepet analizindeki gibi düşünülecek olursa ürünün sepette olması gerektiğini belirtir. FP-Growth algoritmasında kullanılan parametreler çoğunlukta apriori ile aynıdır.

2.3.1.5. Bölümleme Tekniği

Veri tabanının iki kez taranması ile gerçekleşen birliktelik analizi algoritmasıdır. İlk taramada veri tabanı mantıksal olarak çakışmayan birkaç büyük veri setine ayrılır. İkinci taramada ise veri setlerine ait destek değerleri bulunur (Savasere vd, 1995).

2.3.1.6. Örnekleme Tekniği

Örnekleme tekniğine ait yaklaşımlarda, genelde bir kişi örnekleme kullanarak aday veri setini tanımlar (Chakaravarthy vd, 2009). Sonra adaylar arasındaki seyrek olan veri setlerini filtrelemek için veri tabanına geçiş yapar. Sık olan veri setlerini tespit edebilmek için seyrek olan veri setlerini iyi analiz edebilmeyi gerektirmektedir.

2.3.1.7. CARMA

Online olarak nesne kümelerinin hesaplanmasının yapıldığı birliktelik analizi algoritmasıdır. Veri kümesinin iki kez taranması ile sonuçlar bulunmaktadır. İlk taramada veri kümesinin destek ve güven değerleri kullanıcılar tarafından online olarak değiştirilebilir. İkinci taramada ise destek değerini sağlayan veriler sayılır (Deliismail, 2019).

2.3.1.8. Dinamik Nesne Sayımı (DNS)

Geçişlerde veri setlerinin sayısını düşük tutarak, veriler üzerindeki geçişlerin sayısını azaltmayı amaçlayan bir birliktelik analizi algoritmasıdır. DNS algoritmasının dezavantajı işlemcinin veri tabanından her geçişinde tüm işlemciler yayılmasıdır (Vyawahare vd, 2015).

2.3.1.9. Sıra Dışı Aday Belirleme (SDAB)

Gereksiz veri kümelerini analiz dışı tutmak için eski döngülere ait analiz sonuçlarını kullanan algoritmadır. Küçük veri kümelerinin daha yüksek doğruluk vereceğine inanır (Taş, 2018).

2.3.2. Birliktelik Analizi Performans Ölçütleri

Destek, güven ve lift değeri olmak üzere üç adet birliktelik analizi performans ölçütü vardır.

2.3.2.1. Destek (Support)

Bir seçeneğin tüm gözlemler arasındaki oranıdır. Yani klasik olasılık da diyebiliriz. Destek değeri 0 ve 1 arasında değer alır. Değerin sıfıra yaklaşması birliktelik kurallarının sayısını artırırken, bire yaklaşması kural sayısını azaltmaktadır. Birliktelik kuralı analizlerinde destek değeri önceden belirlenir. Bir X gözlemi için destek değeri eşitlik 2.17.'deki gibi hesaplanmaktadır.

$$P(X) = \frac{n(X)}{N} \quad (2.17)$$

$n(X)$: X'in görülme sayısı

N : Gözlem sayısı

2.3.2.2. Güven (Confidence)

X ve Y'nin birlikte kullanımınıdır. Yani X kelimesini kullanan bir hastanın Y kelimesini de söyleme olasılığına güven denir. Koşullu olasılık şeklinde hesaplanır. Güven değeri 0 ve 1 arasında bir değer alırken literatürde 0,50 ve üzeri alınması önerilmektedir.

$$P(Y/X) = \frac{P(Y,X)}{P(X)} \quad (2.18)$$

$P(Y, X)$: X ve Y'in birlikte görülme olasılığı

$P(X)$: X'in görülme olasılığı

2.3.2.3. Kaldıraç (Lift)

Birliktelik kurallarının veri setinde ne sıklıkta görüldüğüdür. Yani X kelimesi ile tüm veri seti içerisinde karşılaşma olasılığıdır. Lift ve kaldıraç kelimelerinin dışında ilginçlik olarak da kullanılmaktadır. Destek değerinin güven değerine bölünmesi ile elde edilir. X'in lift değeri $K(X)$ olarak gösterilir.

$$K(X) = \frac{P(X/Y)}{P(X)} \quad (2.19)$$

Kaldıraç değerinin 1'den büyük olması istenir. Değer 1'den ne kadar büyükse o kuralın veri üzerindeki etkisinin o kadar büyük olduğu söylenebilir (Altunkaynak, 2017).

2.4. Kümeleme

Kümeleme yöntemi bir veri setinde bulunan değişkenlerden yararlanılarak veriler arası benzerlik-farklılığın bulunması ve benzer verileri aynı gruba toplama işlemidir (Altunkaynak, 2017). Ana amaç büyük veriler içerisinde doğal kümeler oluşturmaktır. Birbirine benzeyen ancak diğer verilerden farklı olan tüm veriler aynı kümeye koyulmaya çalışılır (Ersöz, 2019). Analizin en başında her bir veri bir küme olarak kabul edilir. Aynı kümeye koyulacak veriler arasındaki uzaklık minimize edilirken, aynı kümedeki verilerle farklı kümelereki veriler arasındaki uzaklık maksimize edilmeye çalışılır (Altunkaynak, 2017).

Kümeleme yöntemi denetimsiz öğrenme modelleri arasında sayılmaktadır. Sınıflandırma metodu ile arasındaki en büyük fark, sınıflandırma yapılırken veri setinde atama yapılacak olan sınıfların belli olmasıdır. Kümeleme analizinde atama yapılacak olan sınıflar belli değildir, benzer veriler kendi kümelerini oluşturur. Oluşturulan farklı kümeler arasındaki uzaklık miktarı arttıkça kümelemenin performansı artar.

Literatürde kümeleme analizi yapmak için birçok metot bulunmaktadır. K-ortalama, hiyerarşik kümeleme ve self-organizing map (SOM) en çok kullanılan metotlardan olup Tablo 2.10.'da bu yöntemler gösterilmektedir (Amasyalı vd, 2012).

Tablo 2.10. Kümeleme Algoritmaları (Altunkaynak, 2017)

Yaklaşım	Yöntem
Bölünmeli	K-ortalama, K-Medoids, CLARANS
Hiyerarşik	DIANA, AGNES, BRICH, ROCK, CAMELEON
Yoğunluğa Dayalı	DBSCAN, OPTICS, DenClue
Grid Tabanlı	STING, CLIQUE
Model Tabanlı	EM, SOFM, COBWEB
Örüntü Tabanlı	pCluster
Kısıt Tabanlı	COD

2.5. Doğal Dil İşleme

İnsanoğlu büyük miktardaki verileri analiz etmede başarısızdır. Bilişim sistemleri ise yapısal verilerin analizinde başarılıdır. İnsanoğlu büyük miktardaki verileri analiz ederken öncelikle yapısal hale getirir, daha sonra verilerin analiz edilmesi için bilişim sistemlerini kullanır. Bunu yaparken de metin ön işleme, doğal dil işleme gibi metin madenciliğinin temel adımlarını kullanırlar (Oğuzlar, 2011).

İnsanlar arasındaki en önemli iletişim aracı dildir. Dilin modellenmesi ile bilgi sistemleri ve insan arasında yeni bir iletişim yolu oldu. Bu iletişim yoluna doğal dil işleme adı verilmektedir (Nabiyev, 2016). Doğal dil işleme (NLP); bilgisayarların insanların dilinden anlıyor hale gelmesi veya bilgisayarların insan dilini nasıl analiz edip anlayacağını öğrenilmesi olarak tanımlanabilir (Kaya, 2018b). Daha genel bir ifadeyle NLP; dilin çözümlenmesi, anlaşılması, yorumlanması ve bilişim sistemlerine tanıtılabilmesini sağlayan mühendislik alanıdır. Temelini dilbilim, matematik, bilgi-bilgisayar teknolojileri, elektronik, yapay zeka, robotik çalışmalar ve psikoloji gibi farklı birçok olan oluşturmaktadır (Chowdhury, 2003).

Doğal dil işleme ile ilgili araştırmalar ilk olarak 1940'lı yıllara dayanmaktadır. Doğal dilin işlenmesine yönelik ilk çalışma makine çevirisi olarak tanımlanmaktadır. İlk bilgisayar çevirisi Weaver ve Booth tarafından 1946 yılında II. Dünya Savaşı sırasında düşmana karşı savunma amaçlı kullanılmıştır. İlerleyen yıllarda ise ABD konu ile ilgili araştırmalar yapmaya başlamıştır. Yapılan ilk araştırmalarda çeviriler, sözcük bazlı ve sözlük tabanlı yapılmaktaydı. Ancak sözcüksel belirsizlik dikkate alınmadığı için başarısız olmuştur. Bu durum 1957 yılında Chomsky tarafından sözdizimsel yapıların yaygınlaşmasını sağlamıştır. Böylece bilgisayarlar artık tam çeviri yapabilmeye başlamıştır. 1965 yılında Chomsky tarafından dil yeterlilik dönüşüm modeli tanıtılarak çeşitli teorik çalışmalar yapılmış ve prototipler geliştirilmeye başlanmıştır. Weizenbaum tarafından geliştirilen ELIZA bilinen ilk doğal dil işleme robotudur. Bu dönemde farklı prototiplerde geliştirilmiştir. Örneğin; PARRY anahtar kelime kullanımına yönelik çalışırken, Woods tarafından geliştirilen LUNAR ara yüz modeli olarak kullanılmıştır. Özellikle son yıllarda artan bu gelişim teknolojik metinlerin ve bilgisayar hafızalarının artışı, internetin yaygınlaşması ile birlikte günümüze kadar devam etmiştir (Liddy, 2001). Günümüzde NLP metodu kullanılarak oluşturulmuş birçok modern uygulama bulunmaktadır. Bunlardan bazıları;

- Doğal Dil Asistanları (Siri, Google Assistant, Google Duplex gibi)
- Google Translate
- Google, Yahoo, Bing gibi çeşitli arama motorlarıdır (Kaya, 2018b).

NLP farklı alanlarda kullanılabilir. Yazım hatalarının düzeltilmesi, metin okuma, bul ve değiştir, metin özetleme, metinden bilgi çıkarımı, metin işleme ve çeviri gibi alanlarda sıklıkla kullanılmakta olan NLP sanıldığı kadar kolay bir süreç olmayıp karmaşık bir süreçtir (Ay, 2019). Bu yöntemin sağladığı en büyük avantaj belgeler analiz edilirken kelimelerin ek-kök şeklinde ayrılması ve anlam kaybının önlenmesidir. Anlamı aynı olan kelimeler dil bilgisi yazımından dolayı farklı şekillerde bulunabilirler. Aynı şekilde aynı yazıma sahip kelimeler farklı anlamlarda kullanılabilir. Kelimelerin ana anlamına ulaşırken kullanılan çözümleme yöntemleri üç ana başlık altında toplanmaktadır (Yelmen, 2016).

- Morfolojik Çözümleme: Biçim bilimsel çözümleme olarak bilinmektedir. Bir cümlede bulunan kelimeleri ek – köklerine ayırmaktır. Türkçede yapım ve çekim eki olmak üzere iki tip ek vardır. Çekim eki sözcüklerin farklı şekillerde kullanılmasını sağlar. Yapım eki ise eklendikleri kelimelerden yeni anlamlar üreten eklerdir.
- Sentatik Çözümleme: Sözdizimsel çözümleme olarak bilinmektedir. Bir cümlenin anlamından ziyade biçimine odaklanır. Cümlenin biçimsel olarak kurallı olup olmadığını analiz etmek için kullanılır. Örneğin; bir cümlede bulunan “büyük koş mavi” kelimeleri peş peşe anlamsızdır. Ancak “mavi gök bak” kelimeleri anlamlıdır.
- Semantik Çözümleme: Anlamsal çözümleme olarak bilinmektedir. Yani cümlenin duygusal olarak ne anlattığını anlamaktır. Anlamsal analiz yaparken önce kelimeler tek tek veri tabanında kontrol edilir. Cümle içerisindeki kelimelerin doğru anlamları bulunmaya çalışılır. Türkçede aynı yazılıp farklı anlamlar veren kelimelerin olduğundan daha önce ifade edilmişti. Bu aşamada kelimelerin anlamlarını doğru tahmin edebilmek önemlidir. Örneğin; göz kelimesi Türkçede iki farklı anlamda kullanılabilir. “Gözüm ağırıyor.” cümlesini analiz ederken kullanılan göz kelimesi beş duyu organdan birini işaret etmektedir. Ancak “Kâğıdı dolabın gözüne koydum.” cümlesinde bulunan göz kelimesi çekmece anlamına gelmektedir. NLP sırasında bu tarz kavramsal farklılıkları ayırmak önem teşkil etmektedir.

Kelimeler köklerine ayrılarak kelimelerin analizleri yapılırken bazı dil bilim sözlüklerinin kullanılması gerekmektedir. NLP yazılımlarında genellikle İngilizce tabanlı dilbilim sözlükleri bulunmaktadır. Ancak İstanbul Teknik Üniversitesi tarafından geliştirilen Kemik Dilbilim Sözlüğü veya OpenOffice eklentisi olan Zemberek Doğal Dil Kütüphanesi gibi Türkçe tabanlı sözlüklerde bulunmaktadır (Gümüştü, 2019; Wikipedia, 2021). Bu çalışmada KNIME programının içerisinde bulunan Zemberek Doğal Dil Kütüphanesi kullanılmıştır.

2.6. Teletıp

1950'li yıllardan itibaren telefon ve video teknolojisinin gelişmesiyle tıpta telekomünikasyon sistemleri daha aktif rol almaya başlamıştır. İnsanların sağlık personellerine ulaşımını kolaylaştıran bu sistemler teletıp ve telesağlık olmak üzere iki temel başlık altında toplanabilir (Dilbaz vd, 2020). Başlangıcı 1960'lara dayanan teletıp teknolojiiden yararlanarak uzak merkezler arasında sağlık hizmeti sunma olarak tanımlanabilir (Işık ve Güler, 2010). Daha geniş bir tanım ile teletıp tıbbi bakım ve hizmetlerin sağlandığı video konferanstan, uzaktan cerrahi müdahalelere kadar her türlü telekomünikasyon işlemleridir. Telesağlık ise sağlık personellerinin kullandığı hastalıkların erken teşhis ve tedavisinde kullanılan hasta-sağlık personeli iletişiminin sağlandığı uzaktan erişim portalıdır. Telesağlık sistemi; teletıp, internet tabanlı bilgi sistemleri, video konferans gibi uygulamaları içeren daha genel bir yapıdır. Kardiyoloji, radyoloji ve patoloji gibi alanlarda sıklıkla kullanılmaktadır. Ancak Dünya Sağlık Örgütüne (WHO) göre Teletıp en çok radyolojik görüntüler için kullanılmaktadır (Dilbaz vd, 2020). Ekonomik Kalkınma ve İş Birliği Örgütü (OECD)'nin 2013 yılında yayınladığı bir çalışmaya göre bilgisayarlı tomografi ve manyetik rezonans görüntüleme sayılarında ciddi bir artış söz konusudur. Bu da her sağlık kuruluşunun bünyesine ilgili ekipmanları alması gerektiği anlamına gelmektedir. Teletıp uygulamaları ile önceden çekilen radyolojik görüntülerin eş zamanlı olarak birçok yerde görülüyor olması özellikle radyoloji alanındaki gereksiz harcamaların azaltılabileceği anlamına gelmektedir (Paksoy, 2017).

Sağlık için harcanan maliyetlerin nüfus artışıyla beraber artması ve hastalıkların tedavilerinde uzman doktorlara olan ihtiyacın çoğalmasıyla sağlık sektöründe yapay zekayı kullanan teknolojik aletler hayatımızın daha da içinde olmaya başlamıştır (Işık ve Güler, 2010). Günümüzde teletıp ile tüm Dünya'da uzak yerler arasında sağlık hizmeti verilmeye başlanmıştır. Ülkemizde ise 2006'da yayınlanan eylem planındaki

“E-dönüşüm Türkiye Projesi” ile uygulanmaya başlanmıştır. Bu proje ülkemizde radyolojik görüntü yorumlayabilen uzman sayısının azlığından, ileri vakalarda ikinci konsültasyon gereksiniminden, memnuniyetin artırılma isteğinden ve doğru tanı-tedavi işlemlerinden kabul görmüştür (Ertek, 2011). Uzaktan hasta takibi, hastalık tanı ve tedavisi, tıbbi eğitim gibi konular teletıp’ın kullanım alanlarındandır (Işık ve Güler, 2010).

Uzaktan hasta takibi sistemlerine tele-evde bakım/sağlık hizmeti, kanser hastası veya verem hastası gibi hastaların ilaç kontrollerinin yapımı, kablosuz sensörler yardımıyla hastalara ait nabız-solunum-tansiyon gibi verilerin takibi gibi örnekler verilebilir (Küçüköner ve Yavuz, 2016).

Hastalık tanı ve tedavisi teletıp’ın en çok kullandığı alanlardandır. Kendi içerisinde tele-radyoloji, tele-dermatoloji, tele-patoloji, tele-konsültasyon, tele-psikiyatri gibi farklı uygulama alanlarına bölünebilir. Tele-radyoloji hastaların radyolojik raporların ve görüntülerin farklı doktorlar tarafından aynı anda görüntülenmesini sağlamaktadır. Böylece farklı doktor görüşlerine aynı anda ulaşım sağlanmış olur. En büyük avantajı belli alanlarda uzman olan kişilere daha hızlı erişim sağlanabiliyor olmasıdır. Tele-dermatoloji teknoloji kullanarak cilt hastalıkları ve bakımı konusunda kentten kırsala her türlü hastanın uzman kişilerden görüş ve tedavi almasını sağlamaktadır. Tele-patoloji mikroskobik görüntülerin bilgisayara aktarılmasını sağlar. Böylece mikroskopta çok küçük görüntülenebilen görüntüler monitör aracılığı ile daha büyük görülür ve yorumlanmasında farklı bölgelerdeki uzmanlar katkıda bulunabilir. Tele-psikiyatri psikolojik sorunları olan hastaların doktor yanına gitmeden sağlık hizmeti almasını sağlamaktadır. Tele-konsültasyon ile benzerlik gösterir. Tek farkı tele-konsültasyon tüm branşlardan doktorların hastaları ile video konferans yardımıyla görüşmesine verilen genel addır. Tanı ve tedavi için kullanılan ucuz ve kullanışlı bir yoldur. Tele-psikiyatri tele-konsültasyonun bir alan için özelleştirilmiş halidir (Korkmaz ve Hoşman, 2018).


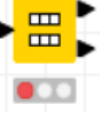
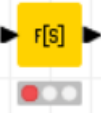

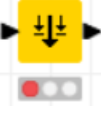


3. MATERYAL VE YÖNTEM

3.1. KNIME

KNIME (Versiyon 4.2.2.) veri biliminin hızlı, kolay ve sezgisel olarak analizinin sağlandığı bir yazılımdır. Büyük veri setlerinde bulunan gizli verileri tespit etmek, veri madenciliği yapmak, geleceğe yönelik veri tahminlemesi yapabilmek gibi işlemleri yapmaktadır. KNIME ile karmaşık ve basit istatistiklerden büyük veri analizlerine kadar her türlü veri işlemleri yapılabilmektedir.

KNIME belli düğümleri kullanarak analiz yapılmasını sağlamaktadır. Tablo 3.1.'de çalışmada kullanılan düğümler ve işlevleri verilmiştir (KNIME, 2021).





Tablo 3.1. KNIME düğümleri ve kullanımları

DÜĞÜM	AD	İŞLEV
	Excel Reader (XLS)	Excel sayfasındaki verileri okur ve sütunları verilerle uyumlu hale getirir.
	Partitioning	Verileri eğitim ve test verisi olarak ikiye böler.
	String Manipulation	Ara ve değiştir işlemidir. Veri setinde bulunan ve farklı halde yazmak istenilen kısımları düzenler.
	Strings To Document	Her satırda bulunan veriyi kendi içinde belge haline getirir ve yeni satıra ekler.
	Column Filter	Sütun filtresidir.
	Punctuation Erasure	Veri setinden noktalama işaretlerini kaldırır.
	N Chars Filter	Belge halindeki terimleri belirlenen N sayısından daha az karakterle filtreler.

Tablo 3.1. (devam)

DÜĞÜM	AD		İŞLEV
	Number Filter		Rakamlardan oluşan verileri filtreler.
	Case Converter		Veri setinin tamamını büyük ya da küçük harf yapar.
	Stop Word Filter		Kelime listelerine göre veri tablosunda bulunan kelimelerde filtreleme yapar.
	Zemberek Tagger	POS	Belge halindeki veri setinin her bir terimine etiket atar.
	Tag Filter		Etiketli verilerde filtreleme yapar.
	Zemberek Stemmer		Zemberek Kök Belirleme Algoritması ile Türkçe kelimelerin köklerine ayrılmasını sağlar.
	Bag Of Words Creator		Dokümanlarda bulunan kelimeleri sütunlara bölerek kelime çantası oluşturur.
	Replacer		Belli bir ifadeyi başka bir ifade ile değiştirmek için kullanılır.
	DF		Her kelime için frekans hesaplar.
	Keygraph Extractor	Keyword	Belgeleri analiz eder. Grafik tabanlı yaklaşım ile anahtar kelimeleri belirler.
	Document Vector		Belgeleri vektörel hale getirir.
	Document Vector Applier	Vector	Her belge için belge vektörü oluşturur.
	Association Rule Learner	Rule	Veriler üzerinde birliktelik analizi yapar.

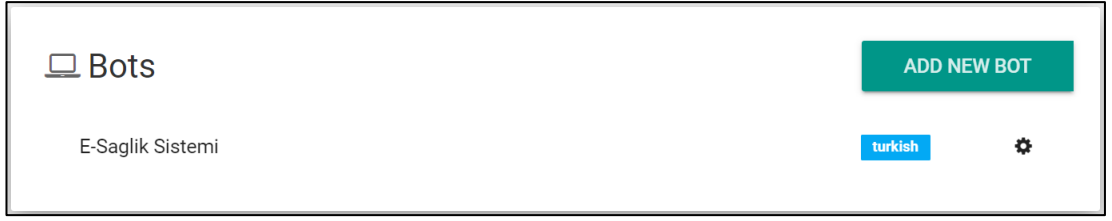
Tablo 3.1. (devam)

DÜĞÜM	AD	İŞLEV
	Category To Class	Belgeleri kategorilerine göre sınıflara ekler. Sınıfa tanımlanamayan değerleri “tanımsız” olarak ayarlar.
	SVM Learner	Verileri “Destek Vektör Makineleri Algoritmasına” göre eğitir.
	SVM Predictor	Destek Vektör Algoritması modelini kullanarak çıktıyı tahmin etmeye çalışır.
	Scorer	Karışıklık matrisi, Doğruluk matrisi gibi sonuçları gösterir.

3.2. Dahi.ai

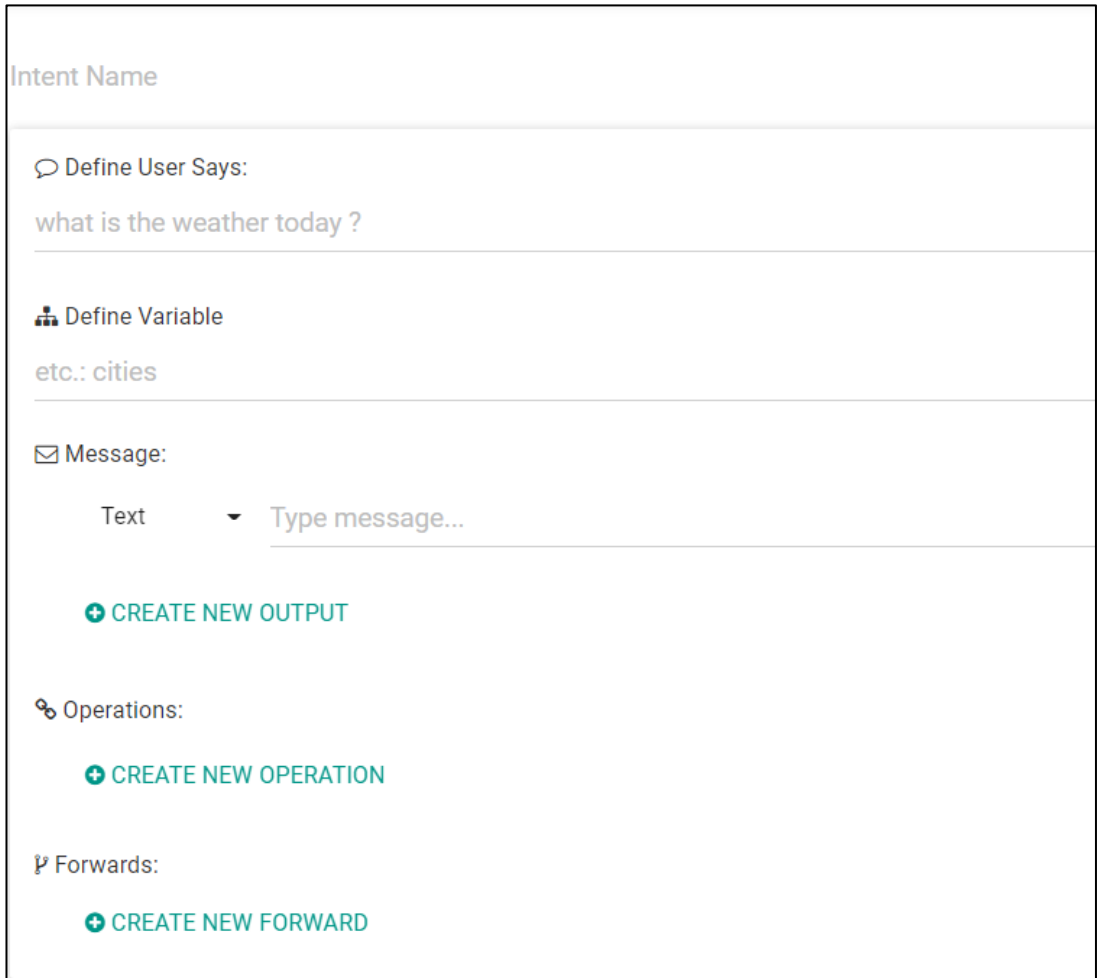
Dahi.ai Yapaytech şirketi tarafından geliştirilen online Chatbot oluşturma platformudur. Türk yazılımcılar tarafından kodlanan bir yazılımdır. Günümüzde yaygınlaşan Chatbot yazılımlarıyla beraber birçok firma kendi online asistanlarını oluşturmaya başlamıştır. ADL, Hotiç, Turkcell, Allianz, Unilever gibi firmalar asistanlarını oluştururken Dahi.ai’yi kullanmıştır (Dahi.ai, 2021). Dahi.ai içerisinde canlı sohbet, chatbot gibi farklı birkaç sistem bulundurmaktadır. Bu çalışmada web sitesinin chatbot özelliği kullanılmıştır. Chatbot kullanıcının sorduğu sorulara otomatik olarak hangi cevapların verileceğinden sorumludur. Canlı sohbetten tek farkı canlı sohbette müşteri temsilcileri de müşterilere cevap verebilirken chatbotta sadece online yazılım cevap vermektedir.

Şekil 3.1.’de görüldüğü gibi öncelikle platformda chatbot tanımlanmalıdır. “E-Sağlık Sistemi” bu çalışma için tanımlanan bottur. Dahi.ai’de İngilizce ve Türkçe olmak üzere iki dil desteği bulunmaktadır. Bu çalışmada Türkçe dil desteğinden yararlanılmıştır. Dahi.ai yazım hatalarını algılamakta ve bu sayede kullanıcının söylemek istediği ifadeyi yanlış yazılmış olsa bile anlayabilmektedir.



Şekil 3.1. Dahi.ai chatbot tanımlama

Konuşma metinleri için tanımlamaların yapıldığı ekran Şekil 3.2.’deki gibidir. “Define user says” kısmında chatbot kullanıcısının hangi sözü söylemesi gerektiği tanımlanabilmektedir. Bu kısımda kelimeler halinde tanımlamalar yapılabilmektedir. Bunun nedeni Dahi.ai platformunda bulunan bir chatbot kelime analizi yaparak kullanıcının söylediği cümleye hangi cevabı vermesi gerektiğini bulabilmesidir.



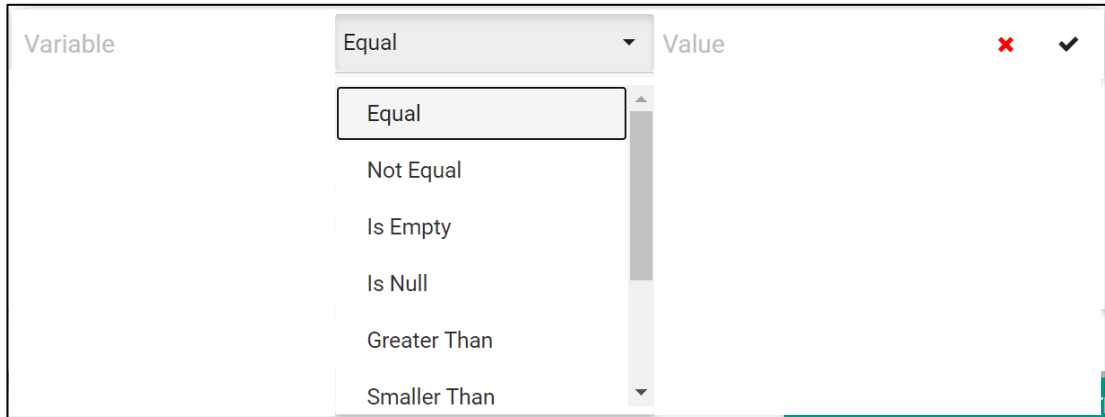
Şekil 3.2. Dahi.ai konuşma metni tanımlama

Chatbotun kullanıcıya vereceği cevaplar farklı değişkenler ile ifade edilebilmektedir. Chatbota tanımlanabilecek değişkenler Tablo 3.2.’de gösterilmiştir.

Tablo 3.2. Dahi.ai deęişken tipleri

Deęişken Tipi	Kullanım Amacı
Text	Metinsel ifadelerin cevap olarak kullanılması için seçilir.
Typing	Verilecek cevabın yazım hızı ayarlanır. En geç 5 en erken 0,1 saniye içerisinde cevap verilebilir.
Array	Kullanıcının belli diziler içerisinde seçim yapması sağlanır.
Horizontal Array	Array işleminden tek farklı dizinin yatay şekilde olmasıdır.
Image	Kullanıcıya yanıt olarak resim gönderileceğinde seçilir.
Video	Kullanıcıya yanıt olarak video gönderileceğinde seçilir.
Custom	Özel bir eşleşme yapılması istenilen durumlarda kullanılır.
Random	Randomize şekilde cevap verilmesini sağlar.
Json	Aynı amaç için çoklu yanıt verilmek istendiğinde kullanılır.

Belirli durumlarda koşula baęlı atamaların yapılması gerekebilir. Örneęin; 18 yaş altında olan bireylerin resmi hükümlülükleri olmadığı için onlara bazı bilgiler verilmemektedir. Bu tarz durumlarda chatbotun kişinin yaşını algılaması sağlanabilir. Bunun için tanımlanan konuşmaya koşul eklenir. Eklenebilecek koşullar Şekil 3.3.'te gösterilmiştir.



Şekil 3.3. Dahi.ai koşul ekleme

Dahi.ai'de oluşturulan bir chatbot Facebook Messenger, Test Console sayfası gibi farklı kısımlara entegre edilerek kullanılabilir. Web sitesinde bulunan analiz sayfası ile de ne zaman kaç kullanıcının chatbotu kullandığı, en çok hangi kelimeleri arattığı gibi farklı analizler görülmektedir (Dahi.ai, 2021).

3.3. Uygulama

3.3.1. Araştırmanın Amacı ve Problemin Tanımlanması

Nüfus artışı, çevre kirliliği, hareketsizlik, dengesiz beslenme gibi nedenlerden dolayı zamanla insanların yaşadıkları sağlık problemleri artış göstermeye başlamıştır. Gün geçtikçe artan sağlık sorunları ile beraber hastane yoğunluğu da artış göstermiştir. İnsanların bilinçli olarak hastaneye gitmemelerinden kaynaklı sağlık sisteminde yoğunluktan kaynaklanan yetersizlikler oluşabileceği gerçeği gündeme gelmiştir.

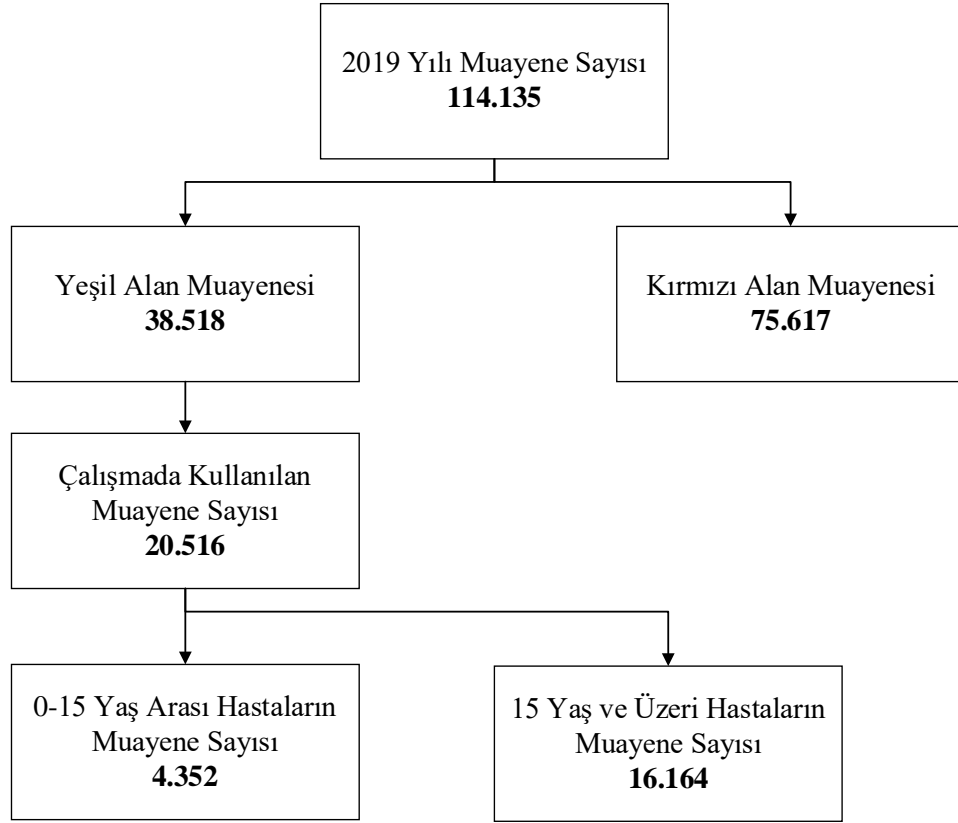
Sağlık bakanlığının yaptığı bir çalışmaya göre Türkiye de bir günde hastaneye giden hastaların %6'sı yanlış poliklinikte muayene olmaktadır. Bu da günlük 120.000 ile 140.000 arasında hastaya denk gelmektedir (Hürriyet Gazetesi, 2019; Sağlık Bakanlığı, 2019). Samsun'da bulunan bir hastanenin acil servisine 2019 yılının Ocak – Aralık ayları arasında başvuran hastalar incelendiğinde hastaların %34'ünün yeşil alan muayenesi kapsamında değerlendirildiği yani hastanın acil servis hastası değil poliklinik hastası olduğu görülmüştür. Hastaneye yıl içerisinde gelen 114.135 hastanın 38.518'i doğru polikliniğe gitmek yerine acil servise başvurmuştur. Bu durum hastaların ilk seferde doğru bir şekilde tedavi hizmetinden yararlanmalarını engellediği gibi acil servis yoğunluğunu da artırmaktadır. Bu gerçek sağlık bakanlığının çalışması ile örtüşmektedir. Bu nedenle hastaların şikayetlerine göre gitmeleri gereken polikliniğin belirlenebileceği bir uygulama yapılmasına karar verilmiştir.

Yapılan çalışma temelde iki aşamada gerçekleştirilmiştir. İlk aşamada metin madenciliği yöntemi uygulanmış olup bu aşamada kendi içinde ikiye ayrılmıştır. Metin madenciliğinin ilk aşamasında hasta şikayetleri üzerinden birliktelik analizi yapılarak kök kelimelerden kural çıkarımları yapılmıştır. Metin madenciliğinin ikinci aşamasında ise ilk aşamada bulunan kelimeler baz alınarak sınıflandırma metodu uygulanmıştır. Böylece hasta şikayetlerine göre yönlendirilen polikliniklerin hangisi olması gerektiği ile ilgili sonuçlar elde edilmiştir. Bu sonuçlar destek vektör makineleri, K-en yakın komşu ve rasgele orman algoritmaları kullanılarak bulunmuştur. Ayrıca karar ağacı algoritması kullanılarak kuralların ne şekilde dallandığına bakılmıştır. Çalışmanın ikinci aşamasında ise bulunan kurallarla birlikte çevrimiçi Chatbot yazılımı yapılmıştır.

3.3.2. Verinin Toplanması

Hastanede Medin ve İş Zekâsı adında iki farklı program kullanılmaktadır. Medin günlük iş akışlarının yapıldığı ve sahadaki personelin kullandığı kurumsal kaynak planlama programıyken İş Zekâsı toplu veri çekme, raporlama vb. işlemler için kullanılan programdır. Çalışmada kullanılacak veri iki aşamada çekilmiştir. Öncelikle İş Zekâsından 2019 yılının Ocak-Aralık ayları arasında hastanenin acil servisine gelen hastalarına ait veriler çekilmiştir. Veride bulunan parametreler; hasta protokol numarası, muayene eden doktor, hasta cinsiyeti, hasta yaşı, hastaya tanımlanan ICD kodu, hastaya tanımlanan ICD kodunun açıklaması, hastaya ait özet epikriz, yönlendirilen polikliniklerdir.

Yukarıdaki parametrelerden özet epikriz ve yönlendirilen poliklinik dışındakiler iş zekâsından çekilmiş olup toplamda çalışma için 114.135 hasta verisi alınmıştır. Bu hastaların sadece 38.518'i yeşil alan muayenesi olarak belirtilmiştir. Kalan 75.617 tane veri kırmızı alan muayenesine sahip olduğu için analiz dışı tutulmuştur. Kırmızı alan muayenesi hastanın hayati tehlikesinin bulunduğunu gösteren ve hastaya müdahale edilmesi gerektiğini öngören muayene türüdür. Bu çalışma için yeşil alan muayenesine sahip 38.518 tane veriden 20.516 tanesi seçilmiştir. Diğer yeşil alan muayenesi verilerinin alınmamasının nedeni verinin alınması gereken zamanın pandemi dönemine denk gelmesi ve hastanede uzun süre bulunamayacak olumasıdır. Seçilen hastalara ait özet epikriz ve hastanın yönlendirildiği poliklinik verileri her protokol için tek tek Medin'den alınmıştır. Özet epikrizde hasta şikayetleri bulunmaktadır. Her poliklinikten 20 tane örnek veri Ek-1'de sunulmuştur. Şekil 3.4'de görüldüğü gibi 20.516 hastanın 4.352'si 0-15 yaş grubu hastaya denk gelmektedir. Bu gruptaki hastaların yönlendirildikleri baskın polikliniğin "Çocuk Sağlığı ve Hastalıkları" polikliniği olduğu görülmüştür. Metin madenciliği yaparken veri ezberlemesinin önüne geçebilmek için bu verilerin analiz dışı tutulmasına karar verilmiştir. 15 yaş üzeri hastalara ait olan 16.164 hastanın 6.934'ü erkek, 9.230'u kadın hastalara ait verilerdir.



Şekil 3.4. Kullanılan verinin dağılımı

Tablo 3.3.'te tüm veri setinde bulunan 13 polikliniğe ait veri sayıları ve yüzdeleri verilmiştir. En fazla yönlendirme yapılan polikliniğin Kulak Burun Boğaz polikliniği, en az yönlendirme yapılan polikliniğin ise Psikiyatri Kliniği olduğu tespit edilmiştir.

Tablo 3.3. Verinin polikliniklere göre dağılımı

Poliklinik	n (Frekans)	% (Yüzde)	Kümülatif Yüzde
Kulak Burun Boğaz	5374	33,2%	33,2%
Dahiliye	2733	16,9%	50,2%
Kardiyoloji	2330	14,4%	64,6%
Göğüs Hastalıkları ve Cerrahisi	1499	9,3%	73,8%
Nöroloji	1249	7,7%	81,6%
Ortopedi ve Travmatoloji	945	5,8%	87,4%
Onkoloji	385	2,4%	89,8%
Üroloji	350	2,2%	92,0%
Kadın Doğum ve Hastalıkları	346	2,1%	94,1%
Fizik Tedavi	265	1,6%	95,7%
Dermatoloji	254	1,6%	97,3%
Enfeksiyon	228	1,4%	98,7%
Psikiyatri	206	1,3%	100,0%
Toplam	16164	100%	

3.3.3. Veri Ön İşleme

Bu aşama bir veri setinde bulunan verilerin düzenlenmesi ile ilgilidir. Veri ön işleme bu çalışmada KNIME programı üzerinden düğümlerle beraber yapılmıştır. Birliktelik analizi ve sınıflandırma algoritmaları için ayrı ayrı veri ön işleme aşaması anlatılmıştır.

Birliktelik analizi yaparken Excel okuyucu (Excel reader) düğümü ile veri seti KNIME programına yüklenmiştir. Belge dönüştürücü (strings to document) düğümü veriyi doküman olarak tanımlamak için kullanılmıştır. Harf dönüştürücü (case converter) düğümü ile veri setinde bulunan dokümanların hepsi küçük harfe dönüştürülmüştür. Sayı filtresi (number filter) düğümü kullanılarak dokümanların içerisindeki sayısal ifadeler filtrelenmiştir. Bunun nedeni hasta epikrizlerinde bulunan “112” ve “3 kez” gibi sayısal ifadelerin birliktelik analizi yaparken bir anlamının olmamasıdır. Noktalama işareti silme (punctuation erasure) düğümü ile dokümanlarda bulunan noktalama işaretleri kaldırılmıştır. Noktalama işaretlerinin kaldırılmasının nedeni yazım hatalarından kaynaklı hataların önüne geçmektir. Aynı zamanda kök kelime bulurken yanında noktalama işareti olan ve olmayan aynı kelimenin farklı kelime gibi algılanmasını önleyebilmektir. Anlatılanlara ait görsel ve çalışmanın devamında bulunan düğümler birliktelik analizi ile ilgili olduğu için bölüm 3.3.4. altında verilmiştir.

Destek Vektör Makineleri ve Karar Ağacı sınıflandırma algoritmalarının veri ön işleme adımları aynıdır. Öncelikle veri seti Excel okuyucu (Excel reader) düğümü ile okunduktan sonra önce verinin tamamının öğrenilmesi için Bölümleme (partitioning) düğümü ile veri tanımlaması yapılmıştır. Dize manüplasyonu (string manipulation) düğümü ile veri setindeki tanımlamalarda farklılık yapılmak istenen alanlar düzeltilmiştir. Örneğin; bu çalışmada toplanan verilerde bazı hastaların şikayetleri arasına konulan “+” sembolünün boşluk olarak düzeltilmesi dize manüplasyonu (string manipulation) düğümü ile sağlanmıştır. Hasta epikrizleri doküman olarak tanımlanmış ve N karakter filtre (N chars filter) ile en az 3 karakterden oluşan kelimelerin sınıflandırılmaya dahil edilmesi sağlanmıştır. Böylelikle edat, soru eki gibi anlamsız ifadelerin filtrelenmesi sağlanmıştır. Aynı işlem Sayı filtresi (number filter) düğümü ile sayısal ifadeler içinde yapılmış ve harf dönüştürücü (case converter) düğümü ile tüm dokümanların küçük harfle yazılması sağlanmıştır. Anlatılanlara ait görsel ve

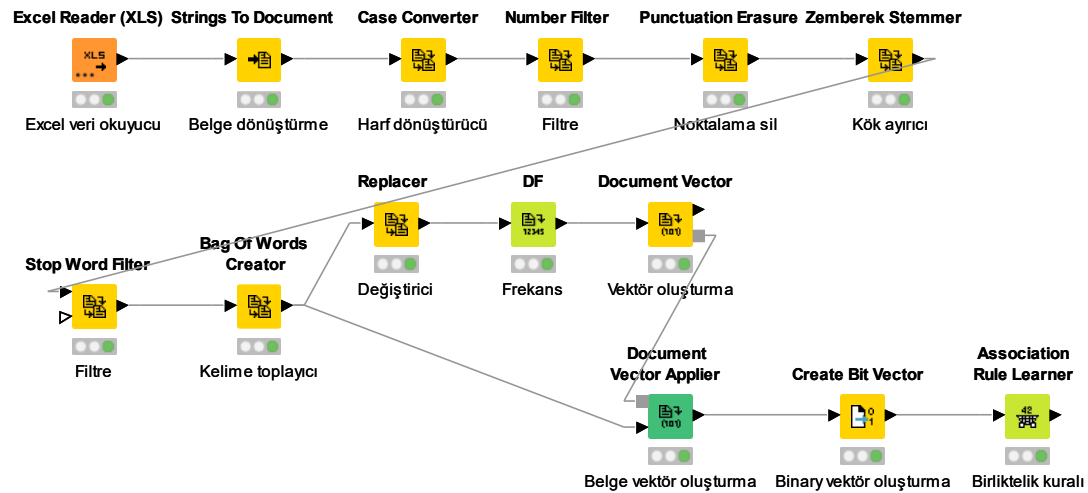
çalışmanın devamında bulunan düğümler sınıflandırma algoritmaları ile ilgili olup bölüm 3.5.5. altında verilmiştir.

3.3.4. Birliktelik Analizi Sonuçları

Hastalar acil servise geldiklerinde hastalıklarına ait şikayetlerini söylemektedirler. Doktor tarafından bu şikayetler özet epikrizlere yazılmaktadır. Çalışmanın bu adımında hasta şikayetlerinden anahtar kelimeler bulunmak istenmiştir. Öncelikle düğümlerden akış oluşturularak veri ön işleme yapılmış ardından anahtar kelimelerin birbirleriyle olan ilişkisi için birliktelik analizi yapılmıştır.

3.3.4.1. Apriori Algoritması Sonuçları

Şekil 3.5.'te birliktelik analizi algoritmalarından Apriori algoritması için KNIME modeli gösterilmiştir.



Şekil 3.5. Apriori algoritması ile birliktelik analizi

Zemberek Kök Ayırıcı (Zemberek Stemmer) düğümüne kadar olan kısımda bulunan düğümler veri ön işleme aşamasında açıklanmıştır. Zemberek Kök Ayırıcı (Zemberek Stemmer) düğümü kullanılarak Zemberek Doğal Dil Kütüphanesinden yararlanılmış ve kelimeler köklerine ayrılmıştır. Bulunan kelime köklerini matris oluşumu için Kelime Çantası Oluşturma (Bag Of Words Creator) düğümü kullanılarak sütunlara bölünmüştür. DF düğümü ile kök kelime frekanslarına bakılmıştır. Örnek olarak “baş ağrısı halsizlik ateş” şikayetine ait kelime frekansları Tablo 3.4.’te verilmiştir. Dokümanda bulunan cümlenin dört terimi vardır ve kök halinde terim kısmında gösterilmiştir. Frekans kısmında ise tüm dokümanlarda bu terimlerin kaç kez bulunduğu gösterilmektedir.

Tablo 3.4. Kök kelime frekansları

Doküman Kök Kelimeleri	Terim	Frekans
"baş ağrı hal ateş"	baş []	2108
	ağrı []	11868
	hal []	328
	ateş []	5407

Bulunan terim frekanslarının ardından doküman vektörleri oluşturulmuştur. Kök kelimelerin birbirleri arasındaki birliktelik analizi için apriori algoritması kullanılmıştır. Apriori algoritması için binary vektöre göre atama yapılmıştır. Yani bir epikrizin notunda bulunan kelimeler tüm kelime matrisinde varsa “1”, yoksa “0” şeklinde kodlanmıştır.

Minimum destek değeri (minimum support) 0,01 ve minimum güven değeri (minimum confidence) 0,80 olarak belirlenmiştir. Apriori algoritmasına göre birliktelik analizi sonucunda toplam 32 kural bulunmuştur. Tablo 3.5.’te üçlü, Tablo 3.6.’da ikili ve Tablo 3.7.’de tekli kullanımlar için birliktelik kuralları verilmiştir. Kurallar anahtar kelimeler arasındaki ilişkinin ne kadar kuvvetli olduğunu gösteren lift değerlerine göre sıralanmıştır. Lift değeri kelimelerin birlikte kullanım sıklıklarını gösterir. Lift değeri arttıkça kuralın ilginçliği artar.

Tablo 3.5.’te üçlü kullanımlar verilmiştir.

Tablo 3.5. Apriori algoritması analizinde üçlü kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[baş, kus, ağrı]	bulantı	1,1	97,9	7,053
2	[kus, ağrı, kar]	bulantı	1,6	88,4	6,369
3	[bulantı, ağrı, kar]	kus	1,6	86,1	6,126
4	[ateş, boğaz, öksürük]	ağrı	1,1	100,0	1,729
5	[kus, bulantı, kar]	ağrı	1,6	100,0	1,729

Kural 1: Lift değeri 7,053 olan “[baş, kus, ağrı] → bulantı” kuralında kelimeler tüm veri setinin %1,1’inde birlikte kullanılmıştır. [baş, kus, ağrı] kelimelerini kullanan hastalar %97,9 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 2: Lift değeri 6,369 olan “[kus, ağrı, kar] → bulantı” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıştır. [kus, ağrı, kar] kelimelerini kullanan hastalar %88,4 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 3: Lift değeri 6,126 olan “[bulantı, ağrı, kar] → kus” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıştır. [bulantı, ağrı, kar] kelimelerini kullanan hastalar %86,1 olasılıkla kus kelimesini kullanmaktadır.

Kural 4: Lift değeri 1,729 olan “[ateş, boğaz, öksürük] → ağrı” kuralında kelimeler tüm veri setinin %1,1’inde birlikte kullanılmıştır. [ateş, boğaz, öksürük] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 5: Lift değeri 1,729 olan “[kus, bulantı, kar] → ağrı” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıştır. [kus, bulantı, kar] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Tablo 3.6.’da ikili kullanımlar verilmiştir.

Tablo 3.6. Apriori algoritması analizinde ikili kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[nefes, öksürük]	darlıg	2,3	99,6	19,971
2	[darlıg, öksürük]	nefes	2,3	100,0	19,210
3	[darlıg, ağrı]	nefes	1,6	99,7	19,150
4	[nefes, ağrı]	darlıg	1,6	95,0	19,043
5	[ağrı, dön]	baş	1,4	98,3	9,562
6	[baş, kus]	bulantı	1,6	95,5	6,882
7	[ishal, bulantı]	kus	2,2	92,2	6,560
8	[kus, ağrı]	bulantı	4,5	89,0	6,415
9	[ateş, bulantı]	kus	2,5	84,8	6,038
10	[ateş, ağrı]	boğaz	12,9	85,3	3,232
11	[kus, kar]	ağrı	1,8	100,0	1,729
12	[bulantı, kar]	ağrı	1,9	100,0	1,729
13	[boğaz, bulantı]	ağrı	1,5	99,7	1,723
14	[ateş, boğaz]	ağrı	12,9	99,6	1,722
15	[boğaz, öksürük]	ağrı	2,3	99,2	1,714
16	[ateş, baş]	ağrı	1,3	95,1	1,645

Çıkan kurallar arasında “darlıg” kelimesi görülmektedir. Bu kelime şikayetlerde bulunan “nefes darlıgı” ifadesinden gelmekte olup köklerine ayrıldığında

“darlıđ” Őeklinde kalmıŐtır. Kütüphanede “darlık” kelimesi ile aynı ifadeye denk gelmektedir.

Kural 1: Lift deđeri 19,971 olan “[nefes, öksürük] → darlıđ” kuralında kelimeler tüm veri setinin %2,3’ünde birlikte kullanılmıŐtır. [nefes, öksürük] kelimelerini kullanan hastalar %99,6 olasılıkla darlıđ kelimesini kullanmaktadır.

Kural 2: Lift deđeri 19,210 olan “[darlıđ, öksürük] → nefes” kuralında kelimeler tüm veri setinin %2,3’ünde birlikte kullanılmıŐtır. [darlıđ, öksürük] kelimelerini kullanan hastalar %100 olasılıkla darlıđ kelimesini kullanmaktadır.

Kural 3: Lift deđeri 19,150 olan “[darlıđ, ađrı] → nefes” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıŐtır. [darlıđ, ađrı] kelimelerini kullanan hastalar %99,7 olasılıkla nefes kelimesini kullanmaktadır.

Kural 4: Lift deđeri 19,043 olan “[nefes, ađrı] → darlıđ” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıŐtır. [nefes, ađrı] kelimelerini kullanan hastalar %95 olasılıkla darlıđ kelimesini kullanmaktadır.

Kural 5: Lift deđeri 9,562 olan “[ađrı, dön] → baŐ” kuralında kelimeler tüm veri setinin %1,4’ünde birlikte kullanılmıŐtır. [ađrı, dön] kelimelerini kullanan hastalar %98,3 olasılıkla baŐ kelimesini kullanmaktadır.

Kural 6: Lift deđeri 6,882 olan “[baŐ, kus] → bulantı” kuralında kelimeler tüm veri setinin %1,6’sında birlikte kullanılmıŐtır. [baŐ, kus] kelimelerini kullanan hastalar %95,5 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 7: Lift deđeri 6,560 olan “[ishal, bulantı] → kus” kuralında kelimeler tüm veri setinin %2,2’sinde birlikte kullanılmıŐtır. [ishal, bulantı] kelimelerini kullanan hastalar %92,2 olasılıkla kus kelimesini kullanmaktadır.

Kural 8: Lift deđeri 6,415 olan “[kus, ađrı] → bulantı” kuralında kelimeler tüm veri setinin %4,5’inde birlikte kullanılmıŐtır. [kus, ađrı] kelimelerini kullanan hastalar %89 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 9: Lift deđeri 6,038 olan “[ateŐ, bulantı] → kus” kuralında kelimeler tüm veri setinin %2,5’inde birlikte kullanılmıŐtır. [ateŐ, bulantı] kelimelerini kullanan hastalar %84,8 olasılıkla kus kelimesini kullanmaktadır.

Kural 10: Lift değeri 3,232 olan “[ateş, ağrı] → boğaz” kuralında kelimeler tüm veri setinin %12,9’unda birlikte kullanılmıştır. [ateş, ağrı] kelimelerini kullanan hastalar %85,3 olasılıkla boğaz kelimesini kullanmaktadır.

Kural 11: Lift değeri 1,729 olan “[kus, kar] → ağrı” kuralında kelimeler tüm veri setinin %1,8’inde birlikte kullanılmıştır. [kus, kar] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 12: Lift değeri 1,729 olan “[bulantı, kar] → ağrı” kuralında kelimeler tüm veri setinin %1,9’unda birlikte kullanılmıştır. [bulantı, kar] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 13: Lift değeri 1,723 olan “[boğaz, bulantı] → ağrı” kuralında kelimeler tüm veri setinin %1,5’inde birlikte kullanılmıştır. [boğaz, bulantı] kelimelerini kullanan hastalar %99,7 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 14: Lift değeri 1,722 olan “[ateş, boğaz] → ağrı” kuralında kelimeler tüm veri setinin %12,9’unda birlikte kullanılmıştır. [ateş, boğaz] kelimelerini kullanan hastalar %99,6 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 15: Lift değeri 1,714 olan “[boğaz, öksürük] → ağrı” kuralında kelimeler tüm veri setinin %2,3’ünde birlikte kullanılmıştır. [boğaz, öksürük] kelimelerini kullanan hastalar %99,2 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 16: Lift değeri 1,645 olan “[ateş, baş] → ağrı” kuralında kelimeler tüm veri setinin %1,3’ünde birlikte kullanılmıştır. [ateş, baş] kelimelerini kullanan hastalar %95,1 olasılıkla ağrı kelimesini kullanmaktadır.

Tablo 3.7.’de tekli kullanımlar verilmiştir.

Tablo 3.7. Apriori algoritması analizinde tekli kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[yüksek]	tansiyon	1,9	94,9	32,664
2	[darlığı]	nefes	5,0	99,6	19,135
3	[nefes]	darlığı	5,0	95,4	19,135
4	[dön]	baş	2,8	97,4	9,483
5	[boğaz]	ağrı	26,2	99,2	1,714
6	[kar]	ağrı	5,7	99,2	1,714
7	[bel]	ağrı	2,0	98,5	1,703

Tablo 3.7. (devam)

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
8	[göğüs]	ağrı	7,4	96,1	1,661
9	[kulak]	ağrı	1,4	89,9	1,555
10	[mide]	ağrı	2,0	86,1	1,488
11	[baş]	ağrı	8,5	82,6	1,429

Kural 1: Lift değeri 32,664 olan “[yüksek] → tansiyon” kuralında kelimeler tüm veri setinin %1,9’unda birlikte kullanılmıştır. [yüksek] kelimesini kullanan hastalar %94,9 olasılıkla tansiyon kelimesini kullanmaktadır.

Kural 2: Lift değeri 19,135 olan “[darlıg] → nefes” kuralında kelimeler tüm veri setinin %5’inde birlikte kullanılmıştır. [darlıg] kelimesini kullanan hastalar %99,6 olasılıkla nefes kelimesini kullanmaktadır.

Kural 3: Lift değeri 19,135 olan “[nefes] → darlıg” kuralında kelimeler tüm veri setinin %5’inde birlikte kullanılmıştır. [nefes] kelimesini kullanan hastalar %95,4 olasılıkla darlıg kelimesini kullanmaktadır.

Kural 4: Lift değeri 9,483 olan “[dön] → baş” kuralında kelimeler tüm veri setinin %2,8’inde birlikte kullanılmıştır. [dön] kelimesini kullanan hastalar %97,4 olasılıkla baş kelimesini kullanmaktadır.

Kural 5: Lift değeri 1,714 olan “[boğaz] → ağrı” kuralında kelimeler tüm veri setinin %26,2’sinde birlikte kullanılmıştır. [boğaz] kelimesini kullanan hastalar %99,2 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 6: Lift değeri 1,714 olan “[kar] → ağrı” kuralında kelimeler tüm veri setinin %5,7’sinde birlikte kullanılmıştır. [kar] kelimesini kullanan hastalar %99,2 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 7: Lift değeri 1,703 olan “[bel] → ağrı” kuralında kelimeler tüm veri setinin %2’sinde birlikte kullanılmıştır. [bel] kelimesini kullanan hastalar %98,5 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 8: Lift değeri 1,661 olan “[göğüs] → ağrı” kuralında kelimeler tüm veri setinin %7,4’ünde birlikte kullanılmıştır. [göğüs] kelimesini kullanan hastalar %96,1 olasılıkla ağrı kelimesini kullanmaktadır.

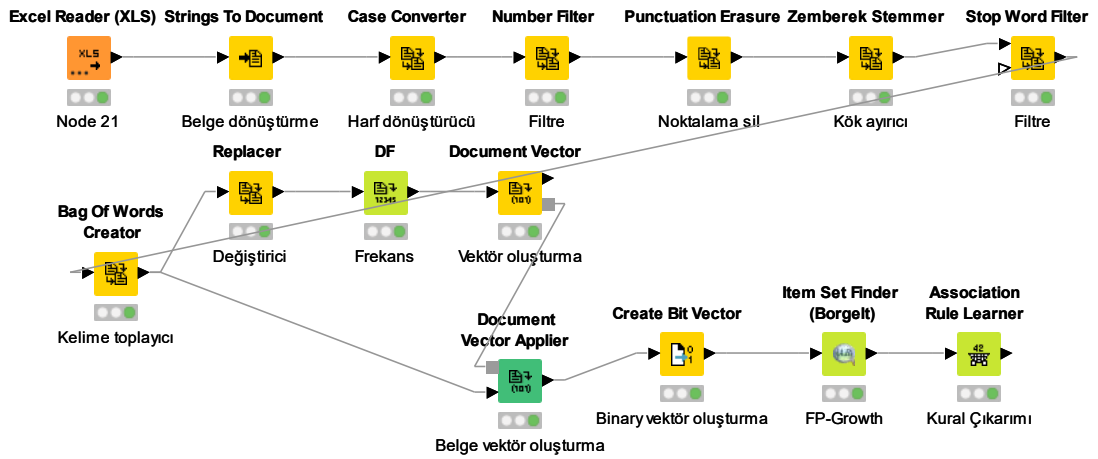
Kural 9: Lift değeri 1,555 olan “[kulak] → ağrı” kuralında kelimeler tüm veri setinin %1,4’ünde birlikte kullanılmıştır. [kulak] kelimesini kullanan hastalar %89,9 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 10: Lift değeri 1,488 olan “[mide] → ağrı” kuralında kelimeler tüm veri setinin %2’sinde birlikte kullanılmıştır. [mide] kelimesini kullanan hastalar %86,1 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 11: Lift değeri 1,429 olan “[baş] → ağrı” kuralında kelimeler tüm veri setinin %8,5’inde birlikte kullanılmıştır. [baş] kelimesini kullanan hastalar %82,6 olasılıkla ağrı kelimesini kullanmaktadır.

3.3.4.2. FP-Growth Algoritması Sonuçları

Şekil 3.6.’da birliktelik analizi algoritmalarından FP-Growth algoritması için KNIME modeli gösterilmiştir.



Şekil 3.6. FP-Growth algoritması ile birliktelik analizi

Modelin FP-Growth düğümüne kadarki kısmı Apriori algoritmasında anlatılmıştır. FP-Growth düğümü ile birliktelik analizi algoritması seçilmiş olup kural çıkarımı düğümü ile kurallar bulunmuştur. Minimum destek değeri (minimum support) 0,01 ve minimum dize boyutu (minimum set size) 2 olarak belirlenmiştir. FP-Growth yöntemi ile birliktelik analizi sonucunda toplam 37 kural bulunmuştur. Tablo 3.8.’de üçlü, Tablo 3.9.’da ikili ve Tablo 3.10.’da tekli kullanımlar için birliktelik kuralları verilmiştir. Kurallar anahtar kelimeler arasındaki ilişkinin ne kadar kuvvetli olduğunu gösteren lift değerlerine göre sıralanmıştır.

Tablo 3.8.’de üçlü kullanımlar verilmiştir.

Tablo 3.8. FP-Growth algoritması analizinde üçlü kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[baş, kus, bulantı]	ağrı	4,2	100	1,412
2	[baş, bulantı, ağrı]	kus	4,2	100	6,000
3	[baş, kus, ağrı]	bulantı	4,2	100	4,000
4	[kus, bulantı, kar]	ağrı	4,2	100	1,412
5	[bulantı, ağrı, kar]	kus	4,2	100	6,000
6	[kus, ağrı, kar]	bulantı	4,2	100	4,000
7	[ateş, boğaz, öksürük]	ağrı	4,2	100	1,412
8	[boğaz, ağrı, öksürük]	ateş	4,2	100	4,800
9	[ateş, boğaz, ağrı]	öksürük	4,2	100	12,000
10	[ateş, ağrı, öksürük]	boğaz	4,2	100	12,000

Kural 1: Lift değeri 1,412 olan “[baş, kus, bulantı] → ağrı” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [baş, kus, bulantı] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 2: Lift değeri 6,000 olan “[baş, bulantı, ağrı] → kus” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [baş, bulantı, ağrı] kelimelerini kullanan hastalar %100 olasılıkla kus kelimesini kullanmaktadır.

Kural 3: Lift değeri 4,000 olan “[baş, kus, ağrı] → bulantı” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [baş, kus, ağrı] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 4: Lift değeri 1,412 olan “[kus, bulantı, kar] → ağrı” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [kus, bulantı, kar] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 5: Lift değeri 6,000 olan “[bulantı, ağrı, kar] → kus” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [bulantı, ağrı, kar] kelimelerini kullanan hastalar %100 olasılıkla kus kelimesini kullanmaktadır.

Kural 6: Lift değeri 4,000 olan “[kus, ağrı, kar] → bulantı” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [kus, ağrı, kar] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 7: Lift değeri 1,412 olan “[ateş, boğaz, öksürük] → ağrı” kuralında kelimeler tüm veri setinin %4,2’inde birlikte kullanılmıştır. [ateş, boğaz, öksürük] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 8: Lift değeri 4,800 olan “[boğaz, ağrı, öksürük] → ateş” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [boğaz, ağrı, öksürük] kelimelerini kullanan hastalar %100 olasılıkla ateş kelimesini kullanmaktadır.

Kural 9: Lift değeri 12,000 olan “[ateş, boğaz, ağrı] → öksürük” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [ateş, boğaz, ağrı] kelimelerini kullanan hastalar %100 olasılıkla öksürük kelimesini kullanmaktadır.

Kural 10: Lift değeri 12,000 olan “[ateş, ağrı, öksürük] → boğaz” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [ateş, ağrı, öksürük] kelimelerini kullanan hastalar %100 olasılıkla boğaz kelimesini kullanmaktadır.

Tablo 3.9.’da ikili kullanımlar verilmiştir.

Tablo 3.9. FP-Growth algoritması analizinde ikili kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[baş, dön]	ağrı	4,2	100	1,412
2	[ağrı, dön]	baş	4,2	100	8,000
3	[ateş, baş]	ağrı	4,2	100	1,412
4	[boğaz, bulantı]	ağrı	4,2	100	1,412
5	[nefes, ağrı]	darlıg	4,2	100	12,000
6	[darlıg, ağrı]	nefes	4,2	100	12,000
7	[ishal, bulantı]	kus	4,2	100	6,000
8	[ishal, kus]	bulantı	4,2	100	4,000
9	[ateş, kus]	bulantı	4,2	100	4,000
10	[nefes, öksürük]	darlıg	4,2	100	12,000
11	[darlıg, öksürük]	nefes	4,2	100	12,000
12	[kus, ağrı]	bulantı	8,3	100	4,000

Kural 1: Lift değeri 1,412 olan “[baş, dön] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [baş, dön] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 2: Lift değeri 8,000 olan “[ağrı, dön] → baş” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [ağrı, dön] kelimelerini kullanan hastalar %100 olasılıkla baş kelimesini kullanmaktadır.

Kural 3: Lift değeri 1,412 olan “[ateş, baş] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [ateş, baş] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 4: Lift deęeri 1,412 olan “[boęaz, bulantı] → aęrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [boęaz, bulantı] kelimelerini kullanan hastalar %100 olasılıkla aęrı kelimesini kullanmaktadır.

Kural 5: Lift deęeri 12,000 olan “[nefes, aęrı] → darlıę” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [nefes, aęrı] kelimelerini kullanan hastalar %100 olasılıkla darlıę kelimesini kullanmaktadır.

Kural 6: Lift deęeri 12,000 olan “[darlıę, aęrı] → nefes” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [darlıę, aęrı] kelimelerini kullanan hastalar %100 olasılıkla nefes kelimesini kullanmaktadır.

Kural 7: Lift deęeri 6,000 olan “[ishal, bulantı] → kus” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [ishal, bulantı] kelimelerini kullanan hastalar %100 olasılıkla kus kelimesini kullanmaktadır.

Kural 8: Lift deęeri 4,000 olan “[ishal, kus] → bulantı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [ishal, kus] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 9: Lift deęeri 4,000 olan “[ateř, kus] → bulantı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [ateř, kus] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

Kural 10: Lift deęeri 12,000 olan “[nefes, öksürük] → darlıę” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [nefes, öksürük] kelimelerini kullanan hastalar %100 olasılıkla darlıę kelimesini kullanmaktadır.

Kural 11: Lift deęeri 12,000 olan “[darlıę, öksürük] → nefes” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıřtır. [darlıę, öksürük] kelimelerini kullanan hastalar %100 olasılıkla nefes kelimesini kullanmaktadır.

Kural 12: Lift deęeri 4,000 olan “[kus, aęrı] → bulantı” kuralında kelimeler tüm veri setinin %8,3’ünde birlikte kullanılmıřtır. [kus, aęrı] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

Tablo 3.10.’da tekli kullanımlar verilmiřtir.

Tablo 3.10. FP-Growth algoritması analizinde tekli kelime kullanımı

Kural	Grup	Sonuç	Destek (%)	Güven (%)	Lift
1	[kulak]	ağrı	4,2	100	1,412
2	[bel]	ağrı	4,2	100	1,412
3	[sağ]	ağrı	4,2	100	1,412
4	[mide]	ağrı	4,2	100	1,412
5	[sol]	ağrı	4,2	100	1,412
6	[göğüs]	ağrı	4,2	100	1,412
7	[yüksek]	tansiyon	4,2	100	12,000
8	[burun]	akıntı	4,2	100	24,000
9	[akıntı]	burun	4,2	100	24,000
10	[idrar]	yan	4,2	100	12,000
11	[boğaz]	ağrı	8,3	100	1,412
12	[nefes]	darlığı	8,3	100	12,000
13	[darlığı]	nefes	8,3	100	12,000
14	[baş]	ağrı	12,5	100	1,412
15	[kus]	bulantı	16,7	100	4,000

Kural 1: Lift değeri 1,412 olan “[kulak] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [kulak] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 2: Lift değeri 1,412 olan “[bel] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [bel] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 3: Lift değeri 1,412 olan “[sağ] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [sağ] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 4: Lift değeri 1,412 olan “[mide] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [mide] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 5: Lift değeri 1,412 olan “[sol] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [sol] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 6: Lift değeri 1,412 olan “[göğüs] → ağrı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [göğüs] kelimelerini kullanan hastalar %100 olasılıkla ağrı kelimesini kullanmaktadır.

Kural 7: Lift deęeri 12,000 olan “[yüksek] → tansiyon” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [yüksek] kelimelerini kullanan hastalar %100 olasılıkla tansiyon kelimesini kullanmaktadır.

Kural 8: Lift deęeri 24,000 olan “[burun] → akıntı” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [burun] kelimelerini kullanan hastalar %100 olasılıkla akıntı kelimesini kullanmaktadır.

Kural 9: Lift deęeri 24,000 olan “[akıntı] → burun” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [akıntı] kelimelerini kullanan hastalar %100 olasılıkla burun kelimesini kullanmaktadır.

Kural 10: Lift deęeri 12,000 olan “[nefes] → darlıę” kuralında kelimeler tüm veri setinin %4,2’sinde birlikte kullanılmıştır. [nefes] kelimelerini kullanan hastalar %100 olasılıkla darlıę kelimesini kullanmaktadır.

Kural 11: Lift deęeri 1,412 olan “[boęaz] → aęrı” kuralında kelimeler tüm veri setinin %8,3’ünde birlikte kullanılmıştır. [boęaz] kelimelerini kullanan hastalar %100 olasılıkla aęrı kelimesini kullanmaktadır.

Kural 12: Lift deęeri 12,000 olan “[nefes] → darlıę” kuralında kelimeler tüm veri setinin %8,3’ünde birlikte kullanılmıştır. [nefes] kelimelerini kullanan hastalar %100 olasılıkla darlıę kelimesini kullanmaktadır.

Kural 13: Lift deęeri 12,000 olan “[darlıę] → nefes” kuralında kelimeler tüm veri setinin %8,3’ünde birlikte kullanılmıştır. [darlıę] kelimelerini kullanan hastalar %100 olasılıkla nefes kelimesini kullanmaktadır.

Kural 14: Lift deęeri 1,412 olan “[baş] → aęrı” kuralında kelimeler tüm veri setinin %12,5’inde birlikte kullanılmıştır. [baş] kelimelerini kullanan hastalar %100 olasılıkla nefes kelimesini kullanmaktadır.

Kural 15: Lift deęeri 4,000 olan “[kus] → bulantı” kuralında kelimeler tüm veri setinin %16,7’sinde birlikte kullanılmıştır. [kus] kelimelerini kullanan hastalar %100 olasılıkla bulantı kelimesini kullanmaktadır.

3.3.5. Sınıflandırma Sonuçları

Sınıflandırma algoritmalarında birliktelik analizinin aksine veri seti %70 eğitim verisi ve %30 test verisi olarak ikiye ayrılmıştır. Toplam veri sayısı 16.164, eğitim veri seti veri sayısı 11.408 ve test veri seti veri sayısı 4.756 olarak belirlenmiştir. Tablo 3.11.'de sınıflandırma algoritmalarının 13 poliklinik bazında test veri seti sayıları verilmiştir.

Tablo 3.11. Sınıflandırma algoritmalarına göre poliklinik bazlı veri sayıları

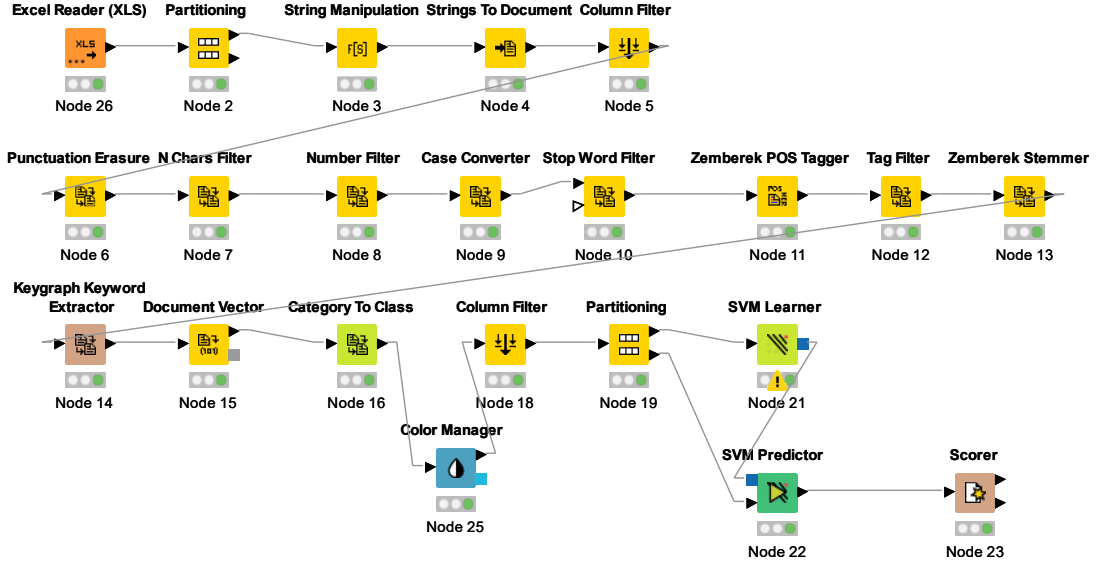
Poliklinik	TF	DVM TF	KA TF	RO TF	K-NN TF
Kulak Burun Boğaz	5374	1606	1606	1606	1606
Dahiliye	2733	813	813	813	813
Kardiyoloji	2330	684	684	684	684
Göğüs Hastalıkları ve Cerrahisi	1499	447	447	447	447
Nöroloji	1249	370	370	370	370
Ortopedi ve Travmatoloji	945	248	248	248	248
Onkoloji	385	115	115	115	115
Üroloji	350	105	105	105	105
Kadın Doğum ve Hastalıkları	346	100	100	100	100
Fizik Tedavi	265	80	80	80	80
Dermatoloji	254	62	62	62	62
Enfeksiyon	228	68	68	68	68
Psikiyatri	206	58	58	58	58
Toplam	16164	4756	4756	4756	4756

TF: Veri setinin toplam frekansı, DVM TF: Destek vektör makineleri algoritması test verisi frekansı, KA TF: Karar ağacı algoritması test verisi frekansı, RO TF: Rasgele orman algoritması test verisi frekansı, K-NN TF: K-En yakın komşu algoritması test verisi frekansı

3.3.5.1. Destek Vektör Makineleri Algoritması Sonuçları

Birliktelik analizi sonucunda bulunan kurallardaki anahtar kelimelerden de faydalanılarak hasta şikâyeti ve yönlendirilen poliklinik arasında sınıflandırma yapılmıştır. Şekil 3.7.'de destek vektör makineleri sınıflandırma modeli gösterilmiştir.

Zemberek POS Etiketleyici (Zemberek POS Tagger) düğümüne kadarki kısım veri ön işleme aşamasında anlatılmıştır. Doğru kelimelerin doğru sınıflara atanabilmesi için doğal dil işleme kütüphanesinden Zemberek Doğal Dil Kütüphanesi kullanılmıştır. Böylece hasta şikayetlerinde kök kelimeler baz alınmış yazım hatası vb. durumların önüne geçilmesi sağlanmıştır. Poliklinikler renklere göre sınıflandırılmış ve görüntüde daha kolay belli olmaları sağlanmıştır. Buraya kadar verinin öğrenilmesi sağlanmıştır. Bölümlenme (Partitioning) ile veri seti %70 eğitim, %30 test olarak ayrılmış ve Destek Vektör Algoritmaları kullanılarak sınıflandırma yapılmıştır.



Şekil 3.7. KNIME Destek vektör makineleri

Toplamda 16.164 veri kullanılmıştır. Çıkan sonuçlar test verisine göre (4.756) elde edilmiştir. Tablo 3.12.'de polikliniklere ait karışıklık matrisi verilmiştir. Bu matriste poliklinikler arasında hangi polikliniğe kaç tane verinin atandığı gösterilmektedir. Örneğin; kadın doğum ve hastalıkları polikliniğine ait olan toplam 100 hasta şikayetinden 64 tanesi kadın doğum ve hastalıkları polikliniğine, 3 tanesi üroloji polikliniğine, 3 tanesi nöroloji polikliniğine, 1 tanesi fizik tedavi polikliniğine, 22 tanesi dahiliye polikliniğine, 3 tanesi kulak burun boğaz polikliniğine ve 4 tanesi ortopedi ve travmatoloji polikliniğine atanmıştır. 4.756 veriden 3.822 tanesi doğru sınıflandırılmıştır. Tablo 3.13.'de doğruluk matrisi verilmiştir. Destek Vektör Makineleri algoritmasında %80,4 doğruluk oranı bulunmuştur. Doğruluk matrisinde poliklinik bazlı inceleme yapıldığında en yüksek doğrulukların %90,2 ile Kulak Burun Boğaz polikliniğine ve %84 ile Dahiliye polikliniğine ait olduğu görülmüştür. En düşük doğruluklar ise %53,7 ile Üroloji polikliniğine ve %52,2 ile Psikiyatri polikliniğine aittir. Bu polikliniklere ait doğruluk oranları veri sayısının artırılması durumunda artırılabilir.

Tablo 3.12. Destek vektör makineleri karışıklık matrisi

	Kadın Doğum ve Hastalıkları	Üroloji	Nöroloji	Fizik Tedavi	Kardiyoloji	Dahiliye	Kulak Burun Boğaz	Psikiyatri	Dermatoloji	Onkoloji	Enfeksiyon Hastalıkları	Ortopedi ve Travmatoloji	Göğüs Hastalıkları ve Cerrahisi	Toplam	
Kadın Doğum ve Hastalıkları	64	3	3	1	0	22	3	0	0	0	0	4	0	100	
Üroloji	5	47	0	3	4	37	3	0	0	0	2	2	2	105	
Nöroloji	1	2	281	11	6	25	15	3	1	4	0	17	4	370	
Fizik Tedavi	0	2	1	65	0	2	1	0	0	0	0	9	0	80	
Kardiyoloji	2	1	56	5	508	50	17	1	1	2	0	15	26	684	
Dahiliye	4	9	8	0	25	726	22	3	1	2	3	7	3	813	
Kulak Burun Boğaz	1	3	36	0	5	24	1415	0	0	1	41	7	73	1606	
Psikiyatri	0	0	8	0	9	2	1	24	0	1	0	9	4	58	
Dermatoloji	0	0	2	0	3	1	0	1	39	1	0	13	2	62	
Onkoloji	1	1	1	0	1	20	10	0	1	58	0	8	14	115	
Enfeksiyon Hastalıkları	0	0	0	0	0	1	3	0	1	0	62	0	1	68	
Ortopedi ve Travmatoloji	1	0	4	4	7	2	2	1	3	0	0	224	0	248	
Göğüs Hastalıkları ve Cerrahisi	1	2	3	0	75	3	40	1	0	0	11	2	309	447	
DP: Doğru Pozitif														DP Toplamı	3822
														Doğruluk Oranı	80,4%

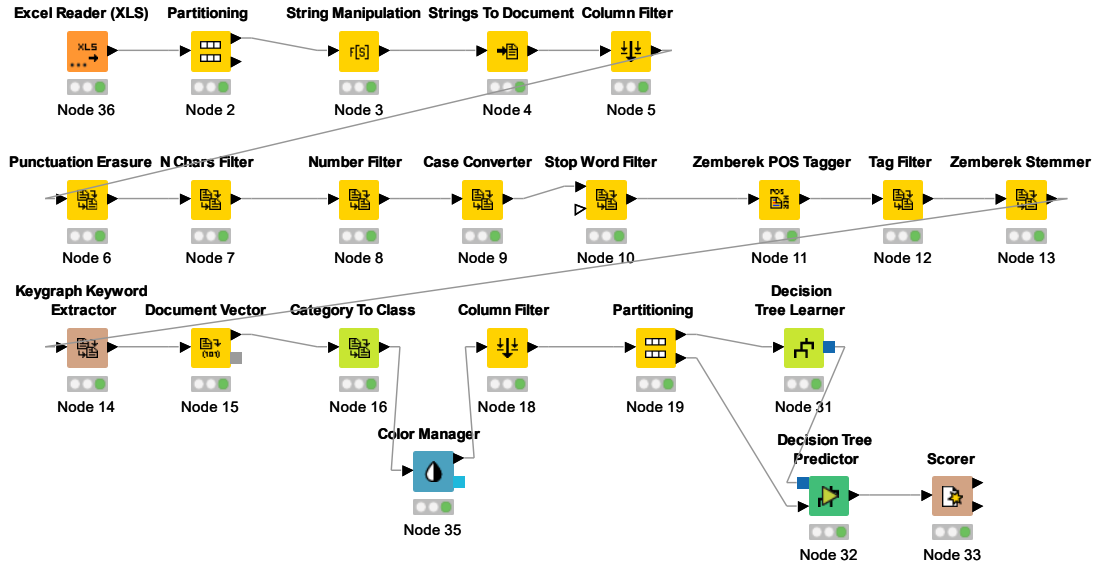
Tablo 3.13. Destek vektör makineleri doğruluk matrisi

Poliklinikler	DP	YP	DN	YN	Kesinlik (%)	Duyarlılık (%)	Özgüllük (%)	F Measure	Doğruluk (%)
Kadın Doğum ve Hastalıkları	64	16	4640	36	80,0	64,0	99,7	0,711	98,9
Üroloji	47	23	4628	58	67,1	44,8	99,5	0,537	98,3
Nöroloji	281	122	4264	89	69,7	75,9	97,2	0,727	95,6
Fizik Tedavi	65	24	4652	15	73,0	81,3	99,5	0,769	99,2
Kardiyoloji	508	135	3937	176	79,0	74,3	96,7	0,766	93,5
Dahiliye	726	189	3754	87	79,3	89,3	95,2	0,840	94,2
Kulak Burun Boğaz	1415	117	3033	191	92,4	88,1	96,3	0,902	93,5
Psikiyatri	24	10	4688	34	70,6	41,4	99,8	0,522	99,1
Dermatoloji	39	8	4686	23	83,0	62,9	99,8	0,716	99,3
Onkoloji	58	11	4630	57	84,1	50,4	99,8	0,630	98,6
Enfeksiyon Hastalıkları	62	57	4631	6	52,1	91,2	98,8	0,663	98,7
Ortopedi ve Travmatoloji	224	93	4415	24	70,7	90,3	97,9	0,793	97,5
Göğüs Hastalıkları ve Cerrahisi	309	129	4180	138	70,5	69,1	97,0	0,698	94,4
Toplam	3822	934	56138	934	80,4	00,4	98,4	0,804	80,4

DP: Doğru Pozitif, YP: Yanlış Pozitif, DN: Doğru Negatif, YN: Yanlış Negatif

3.3.5.2. Karar Ağacı Algoritması Sonuçları

Şekil 3.8.'de Karar ağacı sınıflandırma modeli gösterilmiştir. Destek vektör makineleri ile aynı adımlarda gerçekleştirilmiştir. Tek farkı uygulanan algoritma kısmıdır. Karar ağacı algoritması budanmış ve budanmamış hal arasındaki farklarında görülebilmesi için iki halde verilmiştir. Model her ikisi içinde aynı olup “Karar Ağacı (Decision Tree Learner)” düğümünde değişiklik yapılmıştır. Toplamda 16.164 veri kullanılmıştır. Çıkan sonuçlar test verisine (4.756) göre elde edilmiştir.



Şekil 3.8. KNIME Karar ağacı modeli

3.3.5.2.1. Budama Yapılmamış Karar Ağacı Algoritması Sonuçları

Budama yapılmamış model için karışıklık matrisi Tablo 3.14.'de verilmiştir. Bu matriste poliklinikler arasında hangi polikliniğe kaç tane verinin atandığı gösterilmektedir. Örneğin; kadın doğum ve hastalıkları polikliniğine ait olan toplam 100 hasta şikayetinden 55 tanesi kadın doğum ve hastalıkları polikliniğine, 1 tanesi üroloji polikliniğine, 5 tanesi nöroloji polikliniğine, 1 tanesi fizik tedavi polikliniğine, 2 tanesi kardiyoloji polikliniğine, 25 tanesi dahiliye polikliniğine, 1 tanesi kulak burun boğaz polikliniğine, 9 tanesi ortopedi ve travmatoloji polikliniğine ve 1 tanesi göğüs hastalıkları ve cerrahisi polikliniğine atanmıştır. 4.756 veriden 3.533 tanesi doğru sınıflandırılmıştır. Tablo 3.15.'de doğruluk matrisi verilmiştir. Bu algorithmada %74,3 doğruluk oranı bulunmuştur. Doğruluk matrisinde poliklinik bazlı inceleme yapıldığında en yüksek doğrulukların %89 ile Kulak Burun Boğaz polikliniğine ve %79,3 ile Dahiliye polikliniğine ait olduğu görülmüştür. En düşük doğruluklar ise

%15,2 ile Fizik tedavi polikliniğine ve %49 ile Psikiyatri polikliniğine aittir. Bu polikliniklere ait doğruluk oranları veri sayısının artırılması ile artırılabilir. Budama yapılmamış karar ağacının dallanma modeli Ek-2’te verilmiştir.

Tablo 3.14. Budama yapılmamış karar ağacı karışıklık matrisi

	Kadın Doğum Ve Hastalıkları	Üroloji	Nöroloji	Fizik Tedavi	Kardiyoloji	Dahiliye	Kulak Burun Boğaz	Psikiyatri	Dermatoloji	Onkoloji	Enfeksiyon Hastalıkları	Ortopedi Ve Travmatoloji	Göğüs Hastalıkları ve Cerrahisi	Toplam
Kadın Doğum Ve Hastalıkları	55	1	5	1	2	25	1	0	0	0	0	9	1	100
Üroloji	7	53	14	1	0	19	5	0	0	1	2	2	1	105
Nöroloji	1	1	233	3	9	42	15	0	0	1	1	61	3	370
Fizik Tedavi	0	0	59	8	0	1	2	0	0	0	0	10	0	80
Kardiyoloji	1	5	66	4	487	45	18	1	1	3	1	29	23	684
Dahiliye	2	3	60	4	23	658	23	1	1	1	8	26	3	813
Kulak Burun Boğaz	2	2	45	0	4	27	1394	0	1	1	39	28	63	1606
Psikiyatri	0	0	5	0	7	0	0	20	0	1	0	22	3	58
Dermatoloji	0	0	2	1	4	1	2	0	33	0	0	17	2	62
Onkoloji	1	1	4	1	1	23	8	0	0	51	4	7	14	115
Enfeksiyon	0	0	0	0	0	4	2	0	0	1	59	1	1	68
Ortopedi Ve Travmatoloji	0	0	45	2	0	0	3	0	0	0	0	198	0	248
Göğüs Hastalıkları ve Cerrahisi	0	0	3	0	81	1	53	1	0	4	7	13	284	447
													DP Toplamı	3533
													Doğruluk Oranı	74,3%

DP: Doğru Pozitif

Tablo 3.15. Budama yapılmamış karar ağacı algoritması doğruluk matrisi

Poliklinikler	DP	YP	DN	YN	Kesinlik (%)	Duyarlılık (%)	Özgüllük (%)	F Measure	Doğruluk (%)
Kadın Doğum ve Hastalıkları	55	14	4642	45	79,7	55,0	99,7	0,651	98,8
Üroloji	53	13	4638	52	80,3	50,5	99,7	0,620	98,6
Nöroloji	233	308	4078	137	43,1	63,0	93,0	0,512	90,6
Fizik Tedavi	8	17	4659	72	32,0	10,0	99,6	0,152	98,1
Kardiyoloji	487	131	3941	197	78,8	71,2	96,8	0,748	93,1
Dahiliye	658	188	3755	155	77,8	80,9	95,2	0,793	92,8
Kulak Burun Boğaz	1394	132	3018	212	91,3	86,8	95,8	0,890	92,8
Psikiyatri	20	3	4695	38	87,0	34,5	99,9	0,494	99,1
Dermatoloji	33	3	4691	29	91,7	53,2	99,9	0,673	99,3
Onkoloji	51	13	4628	64	79,7	44,3	99,7	0,570	98,4
Enfeksiyon	59	62	4626	9	48,8	86,8	98,7	0,624	98,5
Ortopedi ve Travmatoloji	198	225	4283	50	46,8	79,8	95,0	0,590	94,2
Göğüs Hastalıkları ve Cerrahisi	284	114	4195	163	71,4	63,5	97,4	0,672	94,2
Toplam	3533	1223	55849	1223	74,3	74,3	97,9	0,743	74,3

DP: Doğru Pozitif, YP: Yanlış Pozitif, DN: Doğru Negatif, YN: Yanlış Negatif

3.3.5.2.2. Budama Yapılmış Karar Ağacı Algoritması Sonuçları

Veri ezberlemesinin önüne geçerek daha doğru sonuçlarda karar ağacı elde etmek için budama işlemi yapılmaktadır. KNIME programında MDL yöntemi ile budama yapılmaktadır. Budama yapılmış model için karışıklık matrisi Tablo 3.16.'da verilmiştir. Bu matriste poliklinikler arasında hangi polikliniğe kaç tane verinin atandığı gösterilmektedir. Örneğin; kadın doğum ve hastalıkları polikliniğine ait olan toplam 102 hasta şikayetinden 48 tanesi kadın doğum ve hastalıkları polikliniğine, 1 tanesi üroloji polikliniğine, 2 tanesi nöroloji polikliniğine, 3 tanesi fizik tedavi polikliniğine, 2 tanesi kardiyoloji polikliniğine, 34 tanesi dahiliye polikliniğine, 1 tanesi kulak burun boğaz polikliniğine, 1 tanesi enfeksiyon polikliniğine, 9 tanesi ortopedi ve travmatoloji polikliniğine ve 1 tanesi göğüs hastalıkları ve cerrahisi polikliniğine atanmıştır. 4.756 veriden 3.660 tanesi doğru sınıflandırılmıştır. Tablo3.17.'de doğruluk matrisi verilmiştir. Bu algorithmada %77,0 doğruluk oranı bulunmuştur. Doğruluk matrisinde poliklinik bazlı inceleme yapıldığında en yüksek doğrulukların %89,9 ile kulak burun boğaz polikliniğine ve %81 ile dahiliye polikliniğine ait olduğu görülmüştür. En düşük doğruluklar ise %38,3 ile onkoloji polikliniğine ve %43,9 ile psikiyatri polikliniğine aittir. Bu polikliniklere ait doğruluk oranları veri sayısının artırılması ile artırılabilir. Budama yapılmış karar ağacının dallanma modeli Ek-3'te verilmiştir.

Tablo 3.16. Budama yapılmış karar ağacı karışıklık matrisi

	Kadın Doğum Ve Hastalıkları	Üroloji	Nöroloji	Fizik Tedavi	Kardiyoloji	Dahiliye	Kulak Burun Boğaz	Psikiyatri	Dermatoloji	Onkoloji	Enfeksiyon	Ortopedi Ve Travmatoloji	Göğüs Hastalıkları ve Cerrahisi	Toplam
Kadın Doğum Ve Hastalıkları	48	1	2	3	2	34	1	0	0	0	1	9	1	102
Üroloji	3	48	0	2	2	39	3	0	0	0	2	2	1	102
Nöroloji	0	1	264	15	12	23	13	0	0	0	0	36	2	366
Fizik Tedavi	0	1	0	70	0	1	0	0	0	0	0	15	0	87
Kardiyoloji	1	1	70	13	494	47	21	0	0	2	1	32	18	700
Dahiliye	1	1	9	4	21	693	39	0	0	2	2	19	4	795
Kulak Burun Boğaz	0	2	39	0	6	20	1413	0	0	0	35	4	73	1592
Psikiyatri	0	0	5	0	17	1	1	18	0	0	0	18	4	64
Dermatoloji	0	0	3	1	1	4	2	0	39	0	1	21	2	74
Onkoloji	1	1	2	2	4	46	8	0	0	32	7	8	8	119
Enfeksiyon	0	0	0	0	0	4	5	0	0	0	60	2	0	71
Ortopedi Ve Travmatoloji	0	1	14	7	3	3	6	0	0	0	1	208	0	243
Göğüs Hastalıkları ve Cerrahisi	0	0	2	1	90	1	41	0	0	12	13	8	273	441
DP: Doğru Pozitif													DP Toplamı	3660
													Doğruluk oranı	77,0%

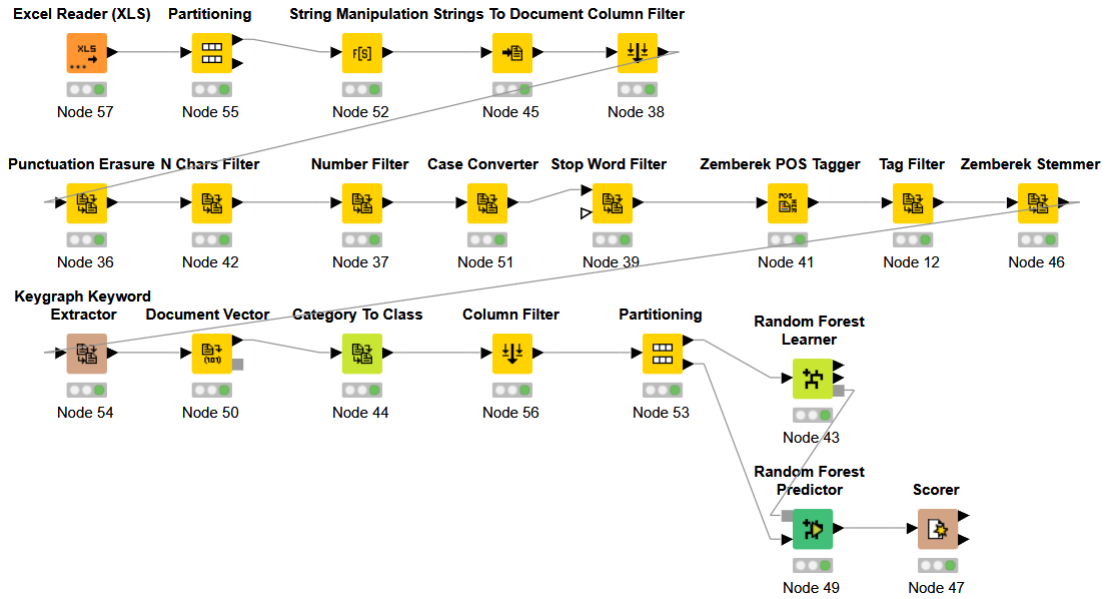
Tablo 3.17. Budama yapılmış karar ağacı doğruluk matrisi

Poliklinikler	DP	YP	DN	YN	Kesinlik (%)	Duyarlılık (%)	Özgüllük (%)	F-measure	Doğruluk (%)
Kadın Doğum ve Hastalıkları	48	6	4648	54	88,9	47,1	99,9	0,615	98,7
Üroloji	48	9	4645	54	84,2	47,1	99,8	0,604	98,7
Nöroloji	264	146	4244	102	64,4	72,1	96,7	0,680	94,8
Fizik Tedavi	70	48	4621	17	59,3	80,5	99,0	0,683	98,6
Kardiyoloji	494	158	3898	206	75,8	70,6	96,1	0,731	92,3
Dahiliye	693	223	3738	102	75,7	87,2	94,4	0,810	93,2
Kulak Burun Boğaz	1413	140	3024	179	91,0	88,8	95,6	0,899	93,3
Psikiyatri	18	0	4692	46	100,0	28,1	100,0	0,439	99,0
Dermatoloji	39	0	4682	35	100,0	52,7	100,0	0,690	99,3
Onkoloji	32	16	4621	87	66,7	26,9	99,7	0,383	97,8
Enfeksiyon	60	63	4622	11	48,8	84,5	98,7	0,619	98,4
Ortopedi ve Travmatoloji	208	174	4339	35	54,5	85,6	96,1	0,666	95,6
Göğüs Hastalıkları ve Cerrahisi	273	113	4202	168	70,7	61,9	97,4	0,660	94,1
Toplam	3660	1096	55976	1096	77,0	77,0	98,1	0,770	77,0

DP: Doğru Pozitif, YP: Yanlış Pozitif, DN: Doğru Negatif, YN: Yanlış Negatif

3.3.5.3. Rasgele Orman Algoritması Sonuçları

Birliktelik analizi sonucunda bulunan kurallardaki anahtar kelimelerden de faydalanılarak hasta şikâyeti ve yönlendirilen poliklinik arasında sınıflandırma yapılmıştır. Şekil 3.9.'da rasgele orman algoritması sınıflandırma modeli gösterilmiştir. Rasgele orman algoritması modeli destek vektör makineleri ve karar ağacı modelleri ile aynı adımlarda gerçekleştirilmiştir.



Şekil 3.9. KNIME Rasgele orman algoritması

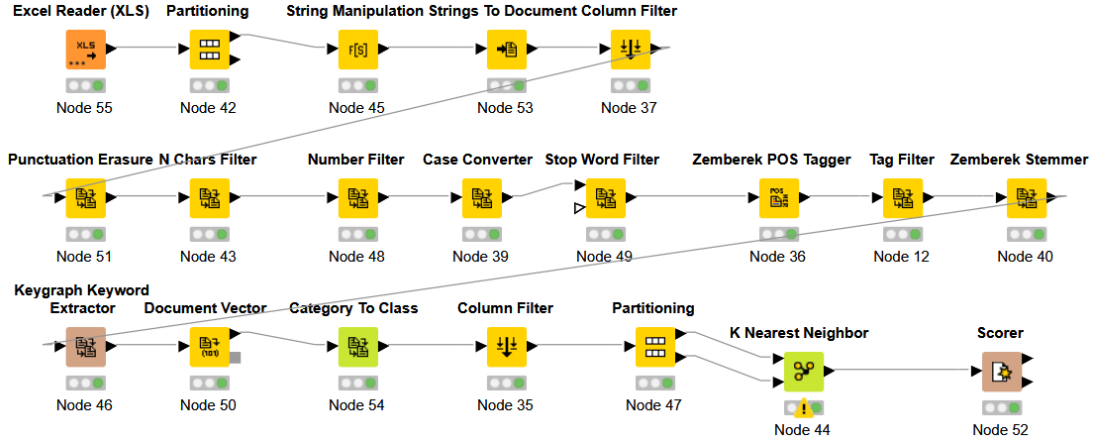
Rasgele orman algoritması için karışıklık matrisi Tablo 3.18'de verilmiştir. Bu matriste poliklinikler arasında hangi polikliniğe kaç tane verinin atandığı gösterilmektedir. Örneğin; üroloji polikliniğine ait olan 105 hasta şikayetinden 38 tanesi üroloji, 41 tanesi nöroloji, 1 tanesi fizik tedavi, 14 tanesi dahiliye, 5 tanesi kulak burun boğaz, 3 tanesi kadın doğum ve hastalıkları, 2 tanesi ise ortopedi ve travmatoloji polikliniğine atanmıştır. 4.756 veriden 3.415 tanesi doğru sınıflandırılmıştır. Tablo3.19'da doğruluk matrisi verilmiş olup rasgele orman algoritmasında %71,8 doğruluk oranı bulunmuştur. Doğruluk matrisinde inceleme yapıldığında en yüksek doğruluklar %87 ile kulak burun boğaz polikliniğine ve %75,5 ile dahiliye polikliniğine ait olduğu görülmüştür. Fizik tedavi ve enfeksiyon poliklinikleri ise sonuç çıkmayan polikliniklerdir.

Tablo 3.19 Rasgele orman algoritması doğruluk matrisi

Poliklinik	DP	YP	DN	YN	Kesinlik (%)	Duyarlılık (%)	Özgüllük (%)	F-measure	Doğruluk (%)
Kadın Doğum ve Hastalıkları	56	17	4639	44	0,767	0,560	0,996	0,647	0,987
Üroloji	55	19	4632	50	0,743	0,524	0,996	0,615	0,985
Nöroloji	276	532	3854	94	0,342	0,746	0,879	0,469	0,868
Kardiyoloji	483	122	3950	201	0,798	0,706	0,970	0,749	0,932
Dahiliye	592	163	3780	221	0,784	0,728	0,959	0,755	0,919
Kulak Burun Boğaz	1425	244	2906	181	0,854	0,887	0,923	0,870	0,911
Psikiyatri	22	11	4687	36	0,667	0,379	0,998	0,484	0,990
Dermatoloji	38	5	4689	24	0,884	0,613	0,999	0,724	0,994
Onkoloji	37	7	4634	78	0,841	0,322	0,998	0,465	0,982
Ortopedi ve Travmatoloji	117	49	4459	131	0,705	0,472	0,989	0,565	0,962
Göğüs Hastalıkları ve Cerrahisi	314	172	4137	133	0,646	0,702	0,960	0,673	0,936
Fizik Tedavi	0	0	4676	80	-	0,000	1,000	-	0,983
Enfeksiyon	0	0	4688	68	-	0,000	1,000	-	0,986
Toplam	3415	1341	55731	1341	0,718	0,718	0,977	0,718	0,718

3.3.5.4. K-En Yakın Komşu Algoritması Sonuçları

Şekil 3.10.'da K-en yakın komşu algoritması sınıflandırma modeli gösterilmiştir. K-en yakın komşu algoritması modeli destek vektör makineleri, karar ağacı ve rasgele orman algoritmaları modelleri ile aynı adımlarda gerçekleştirilmiştir.



Şekil 3.10. KNIME K-en yakın komşu algoritması modeli

K-En yakın komşu algoritması için karışıklık matrisi Tablo 3.20'de verilmiştir. Bu matriste poliklinikler arasında hangi polikliniğe kaç tane verinin atandığı gösterilmektedir. Örneğin; kadın doğum ve hastalıkları polikliniğine ait olan 100 hasta şikayetinden 47 tanesi kadın doğum ve hastalıkları, 5 tanesi nöroloji, 1 tanesi fizik tedavi, 3 tanesi kardiyoloji, 32 tanesi dahiliye, 3 tanesi kulak burun boğaz, 8 tanesi ortopedi ve travmatoloji, 1 tanesi ise göğüs hastalıkları ve cerrahisi polikliniğine atanmıştır. 4.756 veriden 3.492 tanesi doğru sınıflandırılmıştır. Tablo 3.21'de doğruluk matrisi verilmiş olup k-en yakın komşu algoritmasında %73,4 doğruluk oranı bulunmuştur. Doğruluk matrisinde inceleme yapıldığında en yüksek doğruluklar %89,3 ile kulak burun boğaz polikliniğine ve %78,1 ile dahiliye polikliniğine ait olduğu görülmüştür. En düşük doğruluklar ise %15,4 ile Fizik tedavi polikliniğine ve %35,0 ile Onkoloji polikliniğine aittir. Bu polikliniklere ait doğruluk oranları veri sayısının artırılması ile artırılabilir.

Tablo 3.20. K-en yakın komşu algoritması karışıklık matrisi

	Kadın Doğum Ve Hastalıkları	Üroloji	Nöroloji	Fizik Tedavi	Kardiyoloji	Dahiliye	Kulak Burun Boğaz	Psikiyatri	Dermatoloji	Onkoloji	Enfeksiyon	Ortopedi Ve Travmatoloji	Göğüs Hastalıkları ve Cerrahisi	Toplam
Kadın Doğum Ve Hastalıkları	47	0	5	1	3	32	3	0	0	0	0	8	1	100
Üroloji	5	48	14	1	2	22	6	0	0	0	2	2	3	105
Nöroloji	1	2	237	1	8	43	13	0	0	1	1	61	2	370
Fizik Tedavi	0	0	59	8	0	1	2	0	0	0	0	10	0	80
Kardiyoloji	0	0	68	4	485	55	17	0	0	1	1	30	23	684
Dahiliye	5	1	61	4	19	664	21	0	1	0	8	26	3	813
Kulak Burun Boğaz	0	0	46	0	6	28	1397	0	1	0	38	28	62	1606
Psikiyatri	0	0	6	0	9	0	0	18	0	0	0	23	2	58
Dermatoloji	0	0	1	2	4	1	1	0	28	0	2	20	3	62
Onkoloji	1	1	7	1	4	36	9	0	0	25	6	8	17	115
Enfeksiyon	0	0	1	0	0	3	2	0	0	0	60	1	1	68
Ortopedi Ve Travmatoloji	1	0	49	2	0	2	4	0	0	0	0	190	0	248
Göğüs Hastalıkları ve Cerrahisi	0	0	5	0	86	1	48	0	0	1	7	14	285	447
													DP Toplamı	3492
													Doğruluk Oranı	73,4%

DP: Doğru Pozitif

Tablo 3.21. K-en yakın komşu algoritması doğruluk matrisi

Poliklinikler	DP	YP	DN	YN	Kesinlik (%)	Duyarlılık (%)	Özgüllük (%)	F-measure	Doğruluk (%)
Kadın Doğum ve Hastalıkları	47	13	4643	53	0,783	0,470	0,997	0,587	0,986
Üroloji	48	4	4647	57	0,923	0,457	0,999	0,611	0,987
Nöroloji	237	322	4064	133	0,424	0,641	0,927	0,510	0,904
Fizik Tedavi	8	16	4660	72	0,333	0,100	0,997	0,154	0,981
Kardiyoloji	485	141	3931	199	0,775	0,709	0,965	0,740	0,929
Dahiliye	664	224	3719	149	0,748	0,817	0,943	0,781	0,922
Kulak Burun Boğaz	1397	126	3024	209	0,917	0,870	0,960	0,893	0,930
Psikiyatri	18	0	4698	40	1,000	0,310	1,000	0,474	0,992
Dermatoloji	28	2	4692	34	0,933	0,452	1,000	0,609	0,992
Onkoloji	25	3	4638	90	0,893	0,217	0,999	0,350	0,980
Enfeksiyon	60	65	4623	8	0,480	0,882	0,986	0,622	0,985
Ortopedi ve Travmatoloji	190	231	4277	58	0,451	0,766	0,949	0,568	0,939
Göğüs Hastalıkları ve Cerrahisi	285	117	4192	162	0,709	0,638	0,973	0,671	0,941
Toplam	3492	1264	55808	1264	0,734	0,734	0,978	0,734	0,734

3.3.6. Dahi.ai Chatbot

Budanmış-Budanmamış Karar Ağaçları ve Destek Vektör Makineleri algoritmalarına göre chatbotta kullanılmak için terim bazlı poliklinikler elde edilmiştir. Tablo 3.22.'de karşılaştırmalı olarak hangi algorithmada hangi anahtar kelimelerin ilgili polikliniğe atandığı görülmektedir. Budama yapılmamış karar ağacında doğru atama yapılan 3.533 tane verinin 2.749 tanesi, budama yapılmış karar ağacında doğru atama yapılan 3.660 tane verinin 2.615 tanesi, Destek Vektör Makinelerinde doğru atama yapılan 3.822 tane verinin 1.465 tanesi, rasgele orman algoritmasında doğru atama yapılan 3.415 tane verinin 2.666 tanesi, K-en yakın komşu algoritmasından 3.492 tane verinin 2.725 tanesi tabloya alınmıştır. Frekans sayısı 10'un altında olan kelimeler veri yetersizliğinden dolayı Chatbot kuralları içerisine alınmamıştır. Verinin yetersiz olarak düşünülmesinin nedeni hata tahminlemesidir.

Tablo 3.22. Sınıflandırma algoritmaları için poliklinik tanıları

Poliklinik	Kelime	Frekans				
		KA	BKA	DVM	RO	K-NN
Kulak Burun Boğaz	grip	58	45	24	58	58
	burun	10	-	-	10	10
	ağrı boğaz	528	508	276	528	528
	ağrı ateş	26	-	-	26	26
	ağrı kulak	-	15	11	-	-
	ateş	-	-	-	37	-
	ağrı ateş boğaz	488	443	222	488	488
	ağrı bulantı boğaz	23	20	11	23	23
	ağrı boğaz güçlüğü	19	12	12	11	19
	ağrı ateş baş	-	11	12	-	-
	ağrı baş boğaz	-	10	11	-	-
	ağrı boğaz öksürük	24	37	24	24	24
	ağrı ateş boğaz öksürük	32	34	19	32	32
	ağrı ateş bulantı boğaz	23	17	10	23	23
	ağrı ateş boğaz güçlüğü	11	-	-	11	11
ağrı ateş baş boğaz	-	34	14	-	-	
Dahiliye	ishal	97	89	61	97	97
	bulantı	88	79	67	88	88
	ağrı kar	100	91	63	100	100
	bulantı ishal	75	75	35	75	75
	ateş bulantı	37	30	19	-	37
	ağrı bulantı	22	-	-	22	22
	ağrı mide	21	20	12	21	21
	ağrı yan	-	42	-	-	-
	ağrı bulantı baş	-	15	12	-	-

Tablo 3.22. (devam)

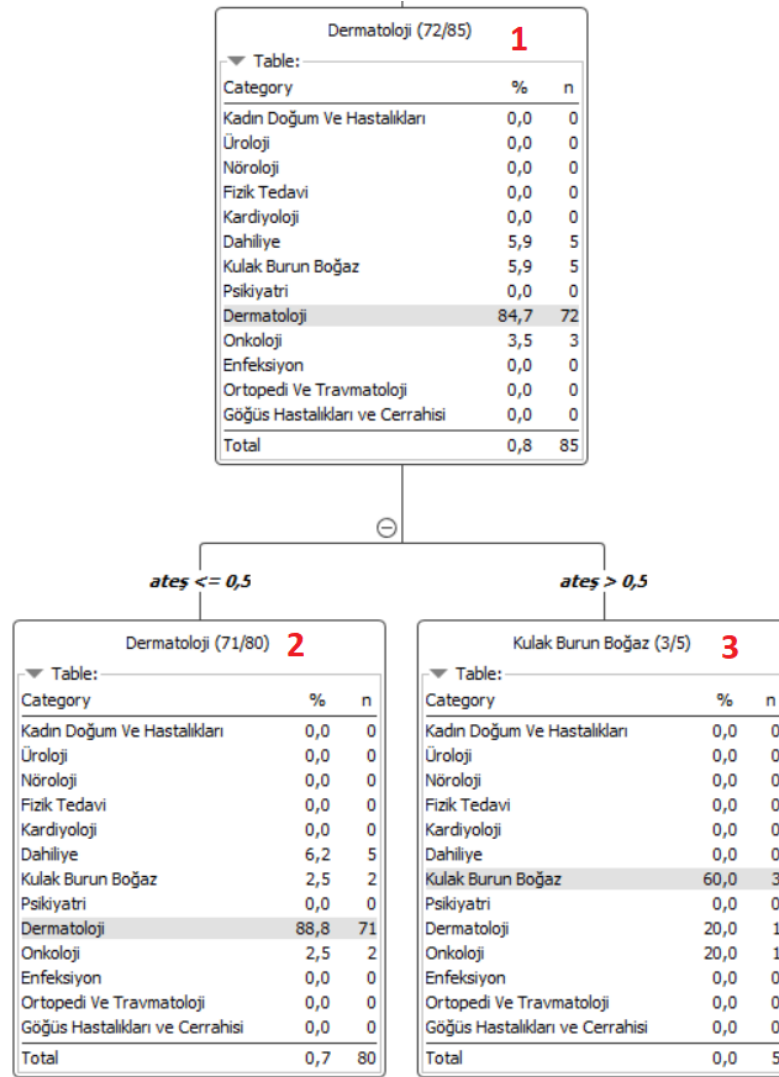
Poliklinik	Kelime	KA	BKA	DVM	RO	K-NN
Dahiliye	ağrı bulantı mide	18	19	14	18	18
	ateş bulantı ishal	12	13	-	12	12
	ağrı bulantı kar	66	51	43	66	66
	ağrı bulantı kar ishal	-	10	-	-	-
Göğüs Hastalıkları ve Cerrahisi	öksürük	46	36	26	46	46
	ateş öksürük	67	82	41	67	67
	nefes darlığı	46	31	18	46	46
	ateş öksürük balgam	11	10	-	11	11
	nefes öksürük darlığı	48	42	23	48	48
	ateş nefes öksürük darlığı	12	16	-	12	12
Kardiyoloji	tanı	24	17	-	-	-
	çarpıntı	26	24	19	26	26
	tansiyon	76	86	48	76	76
	ağrı göğüs	205	140	63	205	205
	ağrı göğüs tansiyon	52	54	24	52	52
	ağrı kar göğüs	10	-	-	10	10
	ağrı göğüs baş	-	21	11	-	-
	ağrı göğüs nefes darlığı	10	-	-	10	10
Kadın Doğum ve Hastalıkları	adet sancı	19	21	19	19	19
	ağrı kasık	11	-	11	11	11
Nöroloji	ağrı	197	-	-	197	197
	baş	-	-	11	-	-
	baş ağrı	-	180	123	-	-
Enfeksiyon	ateş	59	60	28	-	59
Üroloji	idrar	27	21	12	27	27
Fizik Tedavi	ağrı bel	-	54	16	-	-
Dermatoloji	kaşıntı	13	-	-	13	13
Ortopedi ve Travmatoloji	ağrı ayak	12	-	-	12	12
Toplam		2.749	2.615	1.465	2.666	2.725

KA: Budanmamış karar ağacı, BKA: Budanmış karar ağacı, DVM: Destek vektör makineleri, RO: Rasgele orman algoritması, K-NN: K-en yakın komşu algoritması

Algoritmelerde ortaya çıkan kurallar birbiri ile uyumlu olup sadece “ateş” kuralı iki farklı poliklinikte çıkmıştır. “ateş” kuralı rasgele orman algoritmasında kulak burun boğaz polikliniğine atanırken diğer algoritmelerde enfeksiyon polikliniğine atanmıştır. Dört algoritmada enfeksiyon polikliniği belirlendiği için chatbot tanımlamasında enfeksiyon polikliniği olarak belirtilmiştir.

Karar ağaçları bölümünde anlatılan hata tahminlemesine göre bir sınıflandırma algoritmasında göre yaprak düğümlerin hata değeri ara düğümün hata değerinden

büyükse budama yapılmalıdır. Şekil 3.11.'de üçüncü dallanmada biten bir yapı söz konusudur. Buna göre işlemler %80 için güvenilirlik için yapılmış ve 0,80'e karşılık gelen Z puanı 0,845 olarak bulunmuştur.



Şekil 3.11. Hata tahminlemesi için karar ağacı örneği

Dallanmanın başlangıcından dolayı işlemler Dermatoloji polikliniği üzerinden yapılmıştır. f değeri eğitim verisi hata değeri olup polikliniğe atanan değer daldaki bulunan tüm örnek sayısına bölünmesi ile elde edilir. Bir, iki ve üç için sırasıyla f değerleri 0,840, 0,880 ve 0,200'dür. e değeri ise; hata tahmini değerini verip formül 2.7'de verilen şekilde hesaplanmıştır. Bir, iki ve üç için sırasıyla e değerleri 0,870, 0,474 ve 0,907'dir. e ve f değerlerinin hesaplanması Tablo 3.23.'te verilmiştir.

Tablo 3.23. Hata tahminlemesi hesaplama

f (Eğitim verisi hata değeri)	e (Hata tahmini)
$f_1 = \frac{72}{85} = 0,847$	$e_1 = \frac{(0,847 + \frac{0,845^2}{2 \times 85} + 0,845 \sqrt{\frac{0,847}{85} - \frac{0,847^2}{85} + \frac{0,845^2}{4 \times 85^2}})}{(1 + \frac{0,845^2}{85})} = 0,877$
$f_2 = \frac{71}{80} = 0,887$	$e_2 = \frac{(0,887 + \frac{0,845^2}{2 \times 80} + 0,845 \sqrt{\frac{0,887}{80} - \frac{0,887^2}{80} + \frac{0,845^2}{4 \times 80^2}})}{(1 + \frac{0,845^2}{80})} = 0,913$
$f_3 = \frac{1}{5} = 0,200$	$e_3 = \frac{(0,200 + \frac{0,845^2}{2 \times 5} + 0,845 \sqrt{\frac{0,200}{5} - \frac{0,200^2}{5} + \frac{0,845^2}{4 \times 5^2}})}{(1 + \frac{0,845^2}{5})} = 0,383$

Yaprak düğümlere olan değerlerden tek bir e değerinin hesaplanması ve ara düğümlerle karşılaştırılması gerekmektedir. Bunun için düğümlerde bulunan örnek sayısı, ara düğümlerdeki tüm veri sayısına bölünür ve ilgili düğüm için bulunan e değeri ile çarpılır. Yaprak düğümler için ortak e değeri formül 3.1’de görüldüğü gibi hesaplanmış ve 0,881 olarak bulunmuştur. Bulunan değer ara düğümün hata tahmin değeri olan 0,877’den büyük olduğu için budama yapılmasına karar verilmiştir.

$$e_{2,3} = \frac{80}{85} \times 0,913 + \frac{5}{85} \times 0,383 = 0,881 \quad (3.1)$$

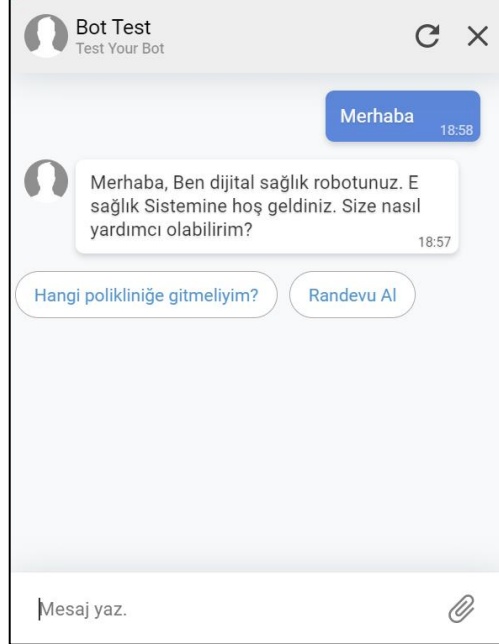
Diğer dallanmalarda da frekans değeri 10’un altında olan kısımlarda aynı durum söz konusu olduğundan dolayı chatbot için frekans değeri 10 ve üzerinde olanların kural olarak alınmasına karar verilmiştir.

Bulunan anahtar kelimeler chatbota tanımlanmıştır. Şekil 3.12.’deki gibi anahtar kelimelere verilecek olan cevaplar tanımlanmıştır.



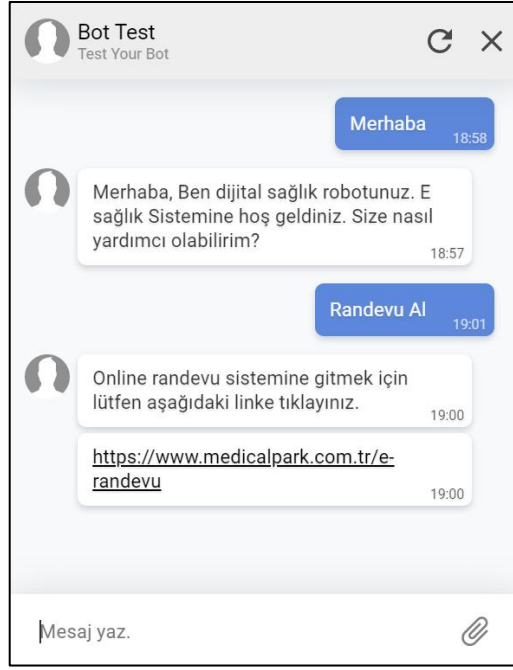
Şekil 3.12. Chatbot anahtar kelime tanımlama

Chatbot ilk açıldığında boş bir ekran kullanıcıyı karşılamaktadır. Kullanıcının başlangıç komutu olarak selam vermesinin ardından kullanıcı – chatbot iletişimi başlamış olur. İlk etapta Chatbot kullanıcıya otomatik bir mesaj gönderir. Şekil3.13.’de görüldüğü gibi “Hangi polikliniğe gitmeliyim?” ve “Randevu Al” şeklinde iki seçenek mevcuttur.



Şekil 3.13. Chatbot başlangıç komutları

Randevu Al butonuna tıklandığı zaman chatbot herhangi bir sorgulama yapmadan kullanıcıyı Şekil 3.14.’teki gibi hastanenin randevu sayfasına göndermektedir.



řekil 3.14. Chatbot randevu al butonu

řekil 3.15.’de görölen “Hangi polikliniđe gitmeliyim?” butonunun arka planında ise bu alıřmada bulunan anahtar kelimeler kullanılmıřtır. Chatbotun kullanıcıların řikayetlerine göre gitmeleri gereken polikliniđin ne olduđunu söylediđi kısımdır.



řekil 3.15. Chatbot hangi polikliniđe gitmeliyim butonu

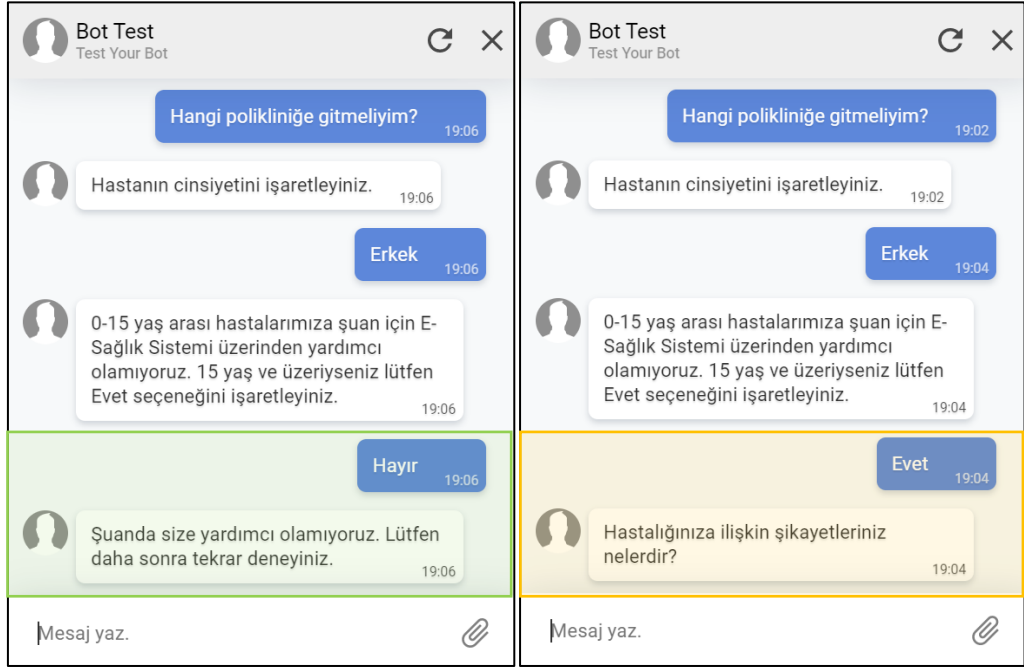
Kullanıcıya sorulan ilk soru řekil 3.16.’da göröldüđu gibi cinsiyettir. Bu alıřmada kadın ve erkek bireylerin anahtar kelimeleri aynı alınmıřtır. Ancak erkek

bireyler kadın doğum polikliniğine gidememektedir. Bu durumu bir kısıt olarak belirtmek gerekmektedir. Bu nedenle öncelikle kullanıcının cinsiyeti sorulmaktadır. Kullanıcıya sorulan ikinci soru Şekil 3.16.'daki gibi yaştır. 0-15 yaş arası bireylerde toplanan veriler Çocuk Sağlığı ve Hastalıkları polikliniğinde yoğunlaşmış olduğundan dolayı veri ezberlemesi söz konusuydu. Verinin bu şekilde olmasının nedeni bu yaş grubundaki bireylerin genellikle bu polikliniğe gidiyor olmasıydı. Bu nedenle 0-15 yaş arası bireyler metin madenciliği çalışmasında analiz dışı tutuldu. Kullanıcı cinsiyetini seçtiğinde “0-15 yaş arası hastalarımıza şu an için E-Sağlık Sistemi üzerinden yardımcı olamıyoruz. 15 yaş ve üzeriyse lütfen evet seçeneğini işaretleyiniz.” şeklinde uyarı gelmektedir.



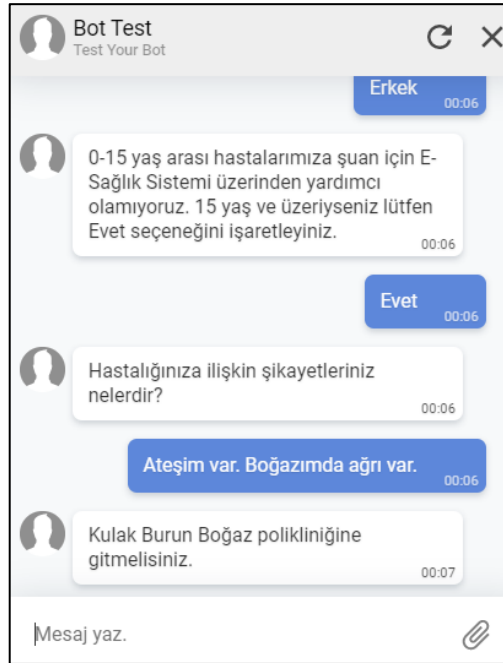
Şekil 3.16.Chatbot cinsiyet sorusu

Şekil 3.17.'de görüldüğü gibi eğer “Hayır” seçeneği seçilirse “Şu anda size yardımcı olamıyoruz. Lütfen daha sonra tekrar deneyiniz.” yanıtı kullanıcıya verilmektedir. “Evet” seçildiğinde ise hastanın şikayetleri sorulmaktadır.



Şekil 3.17. Chatbot yaş sorusu

Hastanın girmiş olduğu şikayetlere göre hangi polikliniğe gitmesi gerektiğini Şekil 3.18.'de gösterilmiştir.



Şekil 3.18. Chatbot hastalığa ilişkin şikayetler sorusu

Anahtar kelimelerin içerisinde bulunduğu poliklinikler için chatbot kullanılabilir. Verilen cevap kullanıcı için yeterli olmamışsa ve kullanıcı ek şikayetler

belirtmek isterse tekrar yazabilmektedir. Chatbot yeni yazılan kelimeleri analiz ederek kullanıcının şikayetine uygun poliklinik önerisini verecektir. Eđer kullanıcı şikayetleri doğrultusunda öğrendiđi poliklinikten randevu almak isterse “Randevu Al” demesi yeterli olur. Böylece hastanenin online randevu sistemine bağlanarak randevu oluşturabilir.

4. TARTIŞMA VE SONUÇ

Bu çalışmada Samsun'da bulunan özel bir hastanenin Ocak-Aralık 2019 yılları arasında acil servisine başvuran yeşil alan muayenesine sahip hastalarının verileri kullanılmıştır. Yapılan literatür çalışması sonucunda sağlık alanında metin madenciliği tekniklerinin çok sık kullanıldığı görülmüştür. Hastalık veya ilaç tanısı gibi tanı koymaya yönelik yapılan çalışmaların spesifik olarak tek branşı ya da hastalığı ele aldığı görülmüştür. Bu çalışmayı benzerlerinden ayıran yönü poliklinik sayısının fazla olması ve özel bir alana inmeden yeşil alan muayenesindeki şikayetlere geniş açıdan bakmasıdır. Çalışmanın ana amacı ise acil serviste oluşan yoğunluğun azaltılarak hastaların tek seferde doğru polikliniğe gitmelerini sağlayacak bir sistem tasarlamaktır. Böylece acil servisin hizmet kalitesi artırılabilir ve müşteri memnuniyetini yükseltilebilir.

Literatürde sıklıkla kullanılan Apriori algoritması, ikili ağırlıklandırma yöntemi çalışmada anahtar kelime bulurken kullanılmış ve hasta şikayetlerinin Türkçe olması nedeniyle metin analizinde Zemberek Doğal Dil Kütüphanesi kullanılmıştır. 0,80 güven ve 0,01 minimum destek değerine göre 32 adet birliktelik kuralı bulunmuştur. Bu kurallardan 5 tanesi üçlü, 16 tanesi ikili ve 11 tanesi tekli kelimelerden oluşmaktadır. Apriori algoritmasına benzerlik gösteren FP-Growth algoritmasında 0,01 minimum destek değeri ve 2 minimum dize boyutuna göre toplam 37 kural bulunmuştur. Üçlü ve tekli birliktelik kuralları FP-Growth algoritmasında, ikili birliktelik kuralları ise Apriori algoritmasında sayısal olarak daha fazladır. Ancak kurallarda bulunan anahtar kelime grupları aynıdır.

Bu çalışmada literatürde sıklıkla kullanılan karar ağacı, destek vektör makineleri, k-en yakın komşu ve rasgele orman algoritmaları ile sınıflandırma analizi yapılmıştır. Bulunan en başarılı sınıflandırma algoritması %80,4 ile destek vektör makineleri algoritmasına aittir. Destek vektör makineleri algoritmasında test verisinde bulunan 4.756 verinin 3.822 tanesi doğru sınıflandırılmıştır. %90,2 ile kulak burun boğaz polikliniği en yüksek doğruluğu verirken %52,2 ile psikiyatri polikliniği en düşük doğruluğu vermiştir. Karar ağacı algoritması budama yapılmış ve yapılmamış olarak iki halde analiz edilmiştir. MDL yöntemi ile budama yapılan karar ağacının doğruluğu %77 iken budama yapılmamış karar ağacının doğruluğu %74,3 olarak bulunmuştur. Budama yapılmış karar ağacında 4.756 verinin 3.660 tanesi doğru sınıflandırılmıştır. %89,9 ile kulak burun boğaz polikliniği en yüksek doğruluğa

sahipken %38,3 ile onkoloji polikliniği en düşük doğruluğa sahiptir. Budama yapılmamış karar ağacında ise 4.756 verinin 3.533 tanesi doğru sınıflandırılmıştır. %89 ile kulak burun boğaz polikliniği en yüksek doğruluğa sahipken %15,2 ile fizik tedavi polikliniği en düşük doğruluğa sahiptir. K-en yakın komşu algoritmasında %73,4 doğruluk bulunmuş olup 4.756 veriden 3.492 tanesi doğru sınıflandırılmıştır. %89,3 ile kulak burun boğaz polikliniği en yüksek doğruluğa sahipken %15,4 ile fizik tedavi polikliniği en düşük doğruluğa sahiptir. Rasgele orman algoritmasında %71,8 doğruluk bulunmuş olup 4.756 verinin 3.415 tanesi doğru sınıflandırılmıştır. %87 ile kulak burun boğaz polikliniği en yüksek doğruluğa sahipken fizik tedavi ve enfeksiyon polikliniklerinde doğru pozitif ve yanlış pozitif değerleri sıfır olduğu için algoritma sonuç vermemiştir.

Karar Ağacı, Destek Vektör Makineleri, K-En Yakın Komşu ve rasgele Orman algoritmalarına ait sınıflandırma matrislerinden doğru atama yapılabilmemiş hasta epikriz notları tespit edilmiştir. Bu epikriz notlarının atama matrislerindeki kelimeler incelenerek hangi kelimelerin hangi polikliniğe atandığı bulunmuştur. Frekansı 10'un üzerinde olan kelimeler chatbot için kullanılmıştır. Frekans değeri 10'un altında olan verilerin hata tahmini yüksek olduğu için böyle bir sınıra ihtiyaç duyulmuştur. Frekans sayısı onun üzerinde olan veri sayısı budanmamış karar ağacında 2.749, budanmış karar ağacında 2.615, destek vektör makinelerinde 1.465, rasgele orman algoritmasında 2.666 ve k-en yakın komşu algoritmasında 2.725 olarak tespit edilmiştir. En çok kelime atamasının Kulak Burun Boğaz polikliniğinde olduğu görülmüştür. Bu durumun en büyük nedeni Kulak Burun Boğaz polikliniği veri sayısının diğer polikliniklerden fazla olmasıdır.

Bulunan kelimeler Dahi.ai'de oluşturulan çevrimiçi chatbota tanımlanmıştır. Böylece hastaların acil servisi gereksiz meşgul etmeyerek gitmeleri gereken polikliniği öğrenebilecekleri alternatif bir sistem geliştirilmiştir. İlerleyen zamanlarda chatbotun arka planında kullanıcıların şikayetleri ve poliklinik verileri toplanmaya devam edilecektir. Böylece poliklinikler arasındaki eşit olmayan dağılım azalacak ve veri sayısı artacaktır. Bu sayede doğruluk oranı artabilecek ve yapılan sınıflandırmalar daha sağlıklı olacaktır. Ayrıca hastane yönetimi ile görüşülecek ve oluşturulan chatbotun hastanenin web sitesinde hastaların kullanımına sunulması için gerekli işlemler yapılacaktır.

KAYNAKÇA

- Abdullahi, A. H. (2018). *An Intrusion Detection Approach Based On Binary Particle Swarm Optimization And Naive Bayes*. Yüksek Lisans Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümü, 61, Konya.
- Akgül, G., Çelik, A. A., Aydın, Z. E. ve Öztürk, Z. K. (2020). Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı. *Bilişim Teknolojileri Dergisi*, 13:3, 255-268.
- Akis, M. S. (2019). *Using Text Mining Algorithms Of Health Data For Analysis Purposes*. Master of Science Thesis. Dokuz Eylül University Graduate School of Natural and Applied Sciences, 40, İzmir.
- Akşehirli, Ö. Y., Ankaral, H., Aydın, D. ve Saraçlı, Ö. (2013). Tıbbi Tahminde Alternatif Bir Yaklaşım: Destek Vektör Makineleri. *Türkiye Klinikleri Journal of Biostatistics*, 5, 19-28.
- Altıntaş, Y. Y. (2010). *Veri Madenciliğinin Tıpta Kullanımı Ve Bir Uygulama : Hemodiyaliz Hastaları İçin Risk Seviyelerine Göre Risk Faktörlerinin Etkileşimlerinin İncelemesi*. Yüksek Lisans Tezi. Gazi Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Bölümü, 222, Ankara.
- Altunkaynak, B. (2017). *Veri Madenciliği Yöntemleri ve R Uygulamaları*. Ankara: Seçkin Yayıncılık.
- Amanet, H. (2017). *Türkçe Sosyal Medya Metinlerinde Duygu Analizi*. Yüksek Lisans Tezi. Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, 94, Trabzon.
- Amasyalı, M. F., Balcı, S., Varlı, E. N. ve Mete, E. (2012). Türkçe Metinlerin Sınıflandırılmasında Metin Temsil Yöntemlerinin Performans Karşılaştırılması. *EMO Bilimsel Dergi*, 2:4.
- Ay, H. (2019). *Üretim Sürecinin Doğal Dil İşleme İle Modellenmesi* Yüksek Lisans Tezi. Eskişehir Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 51, Eskişehir.
- Ayatollahi, H., Gholamhosseini, L. ve Salehi, M. (2019). Predicting coronary artery disease: a comparison between two data mining algorithms. *BMC Public Health*, 19:1, 448. doi:10.1186/s12889-019-6721-5.
- Aydemir, E. (2018). *Weka İle Yapay Zekâ: Makine Öğrenmesi - Yapay Sinir Ağları - Derin Öğrenme*. Ankara: Seçkin Yayıncılık.
- Bayazıt, D. A., Çebi, D. A., Aydın, F., Akçapınar, D. G., Yurdugül, D. H., Aydın, M., Kokoç, D. M., Şahin, D. M., Yürüm, O. R., Metin, D. S. K., Keskin, D. S., Yıldırım, D. S., Aydoğdu, D. Ş., Kışla, D. T. ve Güyer, D. T. (2020). *Eğitsel Veri Madenciliği ve Öğrenme Analitikleri*. Anı Yayıncılık.
- Budak, H. (2018). Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22:Özel. doi:10.19113/sdufbed.01653.
- Chakaravarthy, V. T., Pandit, V. ve Sabharwal, Y. (2009). Analysis of sampling techniques for association rule mining. Proceedings of the 12th International Conference on Database Theory - ICDT '09,
- Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, 37:1, 51-89.
- Cihan, Ş., Karabulut, B., Arslan, G. ve Cihan, G. (2018). Koroner Arter Hastalığı Riskinin Veri Madenciliği Yöntemleri İle İncelenmesi. *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 10:1, 85-93.
- Cohen, S., Jannot, A. S., Iserin, L., Bonnet, D., Burgun, A. ve Escudie, J. B. (2019). Accuracy of claim data in the identification and classification of adults with congenital heart diseases in electronic medical records. *Arch Cardiovasc Dis*, 112:1, 31-43. doi:10.1016/j.acvd.2018.07.002.
- Çalış, K., Gazdağı, O. ve Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. *Bilişim Teknolojileri Dergisi*, 6:1, 1-7.

- Çamlıca, H.ve Dişçi, R. (2008). Tanı testlerinde sınır değerlerinin belirlenmesi. *Türk Onkoloji Dergisi*, 23:1, 26-33.
- Çelik, B. (2019). *Mekanik Arızaların Veri Madenciliği Apriori Algoritması ile Analiz Edilmesi*. Yüksek Lisans Tezi. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 76, Kocaeli.
- Çomak, E. (2004). *Destek vektör makineleri çoklu sınıf problemleri için çözüm önerileri*. Yüksek Lisans Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 49, Konya.
- Dahi.ai. (2021). Dahi.ai Kullanım Kılavuzu. Erişim: 9 Mart 2021, <http://dahi.ai/dahi/bot>
- Deliismail, İ. (2019). *Veri Madenciliği Teknikleri Kullanılarak İnternette Alışveriş Yapan Tüketicilerin Analizi*. Yüksek Lisans Tezi. Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 85, Sakarya.
- Demircoğlu, H. Z.ve Bilge, H. Ş. (2015). Yumurtalık Kanseri Veri Kümesindeki Gen İfadelerinin Veri Madenciliği ile Analizi. *Marmara Journal of Pure & Applied Sciences*, 27:4. doi:10.7240/mufbed.89154.
- Demirel, Ş. (2019). *Karar Ağacı Algoritmaları Ve Çocuk İşçiliği Üzerine Bir Uygulama*. Yüksek Lisans Tezi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, 115, İstanbul.
- Diaz-Uriarte, R.ve Alvarez de Andres, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7, 3. doi:10.1186/1471-2105-7-3.
- Dilbaz, B., Kaplanoğlu, M.ve Kaplanoğlu, D. K. (2020). Teletıp ve Telesağlık: Geçmiş, Bugün ve Gelecek. *Eurasian Journal of Health Technology Assessment: EHTA*, 4:1, 40-56.
- Durnagöl, F. (2020). *Uluslararası Haber Raporlarının Rapor İçeriklerinde Kullanılan İfadelere Göre Makine Öğrenmesi Yöntemiyle Sınıflandırılması Ve Denetlenmesi*. Yüksek Lisans Tezi. İstanbul Aydın Üniversitesi Lisansüstü Eğitim Programı Bilgisayar Mühendisliği Anabilim Dalı, 74, İstanbul.
- Elmas, Ç. (2018). *Yapay Zeka Uygulamaları*. Ankara: Seçkin Yayıncılık.
- Elmas, M. (2012). *Destek Vektör Makineleri İle Fiyat Tahminleri Ve Kuyumculuk Sektöründe Bir Uygulama*. Yüksek Lisans Tezi. İstanbul Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 187, İstanbul.
- Erpolat, S. (2012). Otomobil Yetkili Servislerinde Birliktelik Kurallarının Belirlenmesinde Apriori ve FP-Growth Algoritmalarının Karşılaştırılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 12:2, 137-146.
- Ersöz, F. (2019). *Veri Madenciliği Teknikleri ve Uygulamaları*. Ankara: Seçkin Yayıncılık.
- Ertek, S. (2011). Endokrinolojide tele-sağlık ve tele-tıp uygulamaları. *Acıbadem Üniversitesi Sağlık Bilimleri Dergisi*, 2:3, 126-130.
- Erten, F. (2015). *Metin Madenciliği Tabanlı Bir Web Sitesi Sınıflandırma Aracı Tasarımı*. Yüksek Lisans Tezi. Maltepe Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 37, İstanbul.
- Espino, J.ve Wagner, M. (2001). Accuracy of ICD-9-coded Chief Complaints and Diagnoses for the Detection of Acute Respiratory Illness. *Proceedings. AMIA Symposium*, 164-168, Pittsburgh.
- Fedchenko, M., Mandalenakis, Z., Hultsberg-Olsson, G., Dellborg, H., Eriksson, P.ve Dellborg, M. (2020). Validation of myocardial infarction diagnosis in patients with congenital heart disease in Sweden. *BMC Cardiovasc Disord*, 20:1, 460. doi:10.1186/s12872-020-01737-1.
- Ford, E., Carroll, J. A., Smith, H. E., Scott, D.ve Cassell, J. A. (2016). Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*, 23:5, 1007-1015. doi:10.1093/jamia/ocv180.
- Frohnert, B. K., Lussky, R. C., Alms, M. A., Mendelsohn, N. J., Symonik, D. M.ve Falken, M. C. (2005). Validity of hospital discharge data for identifying infants with cardiac defects. *J Perinatol*, 25:11, 737-742. doi:10.1038/sj.jp.7211382.

- Göker, H.ve Tekedere, H. (2017). Fatih Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10:3, 291-299.
- Gümüşçü, A., Taşaltın, R.ve Aydılek, İ. B. (2016). C4.5 Karar ağaçlarında genetik algoritma ile budama. *Dicle Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 5:2, 77-80.
- Gümüşsu, E. (2019). *Metin Madenciliği Teknikleri İle Sosyal Medya Gönderilerinin Analiz Edilmesi: Ankilozan Spondilit Hastalığı Örneği* Yüksek Lisans Tezi. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Akıllı Sistemler Mühendisliği Anabilim Dalı, 232, Samsun.
- Haltaş, A., Alkan, A.ve Karabulut, M. (2015). Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 30:3, 417-427.
- Hürriyet Gazetesi. (2019). Hürriyet Gazetesi - Sağlıkta Neyim Var Projesi. Erişim: 17 Nisan 2021, <https://www.hurriyet.com.tr/gundem/saglikta-neyim-var-donemi-41373673>
- Işık, A. H.ve Güler, İ. (2010). Teletipte Mobil Uygulama Çalışması ve Mobil İletişim Teknolojilerinin Analizi. *International Journal of InformaticsTechnologies*, 3:1.
- Kahya, A. N. (2020). Wikipedia'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. *ESTUDAM Bilişim Dergisi*, 2:1, 11-14.
- Karamaşa, Ç. (2018). *Bayramlarda Gerçekleşen Trafik Kazalarının Birlikte Kuralları İle Analiz Edilmesi*. Doktora Tezi. Anadolu Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 207, Eskişehir.
- Karasu, E. (2019). *Metin Madenciliği Tekniklerini Kullanarak Faaliyet Bazlı Proje Yönetiminde Bilgi Yönetimi Süreçleri İçin Derin Öğrenme Modeli*. Yüksek Lisans Tezi. İstanbul Üniversitesi Cerrahpaşa Lisansüstü Eğitim Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 89, İstanbul.
- Kaşıkcı, T.ve Gökçen, H. (2014). Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7:1, 25-32. doi:10.12973/bid.2014.
- Kaya, A. (2018a). *Sağlıkla İlgili Sosyal Medya Forum Verilerinde İçeriğinin Bilgilendirici Özellikleri: İlaç Yan Etkilerinin Araştırması Üzerine Bir Çalışma*. Yüksek Lisans Tezi. Okan Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Programı, 119, İstanbul.
- Kaya, S. (2018b). *Doğal Dil İşleme Teknikleriyle Yazar-Kitap Tanıma*. Yüksek Lisans Tezi. İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 65, İstanbul Üniversitesi.
- Kılıç, E., Tavus, M. R.ve Karhan, Z. (2015). Classification of breaking news taken from the online news sites. 23rd Signal Processing and Communications Applications Conference, IEEE, Malatya, Türkiye.
- KNIME. (2021). KNIME Kullanım Kılavuzu. Erişim: 19 Ocak 2021, <https://www.knime.com/>
- Kocatekin, T. (2013). *Text Mining In Turkish Radiology Reports*. Master's Thesis. Bahcesehir University Graduate School of Natural and Applied Sciences Computer Engineering, 47, İstanbul.
- Korkmaz, S.ve Hoşman, İ. (2018). Sağlık Sektöründe Tele-Tıp Uygulamaları: Tele-Tıp Uygulama Boyutlarını İçeren Bir Araştırma. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 4:3, 251-263.
- Kökver, Y., Barışçı, N., Çiftçi, A.ve Ekmekçi, Y. (2014). Hipertansiyona Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi. *Engineering Sciences*, 9:2, 15-25.
- Kurban, H. (2021). Metin Madenciliği ile Tıbbi Tedavi Alanlarının Yakınlıklarının Ölçülmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 21, 518-526. doi:10.31590/ejosat.833199.
- Kuş, O. (2021). COVID-19 Salgını ve Mültecilere Yönelik Dijital Nefret Söylemi: Büyük Veri Perspektifinden Metin Madenciliği Tekniği ile Kullanıcı Kaynaklı İçeriklerin Analizi. *TRT Akademi*, 6:11, 108-131. doi:10.37679/trta.830736.
- Kuzey, C. (2012). *Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin*

- Ölçülmesi ve Bir Uygulama*. Doktora Tezi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 318, İstanbul.
- Küçüköner, E. M.ve Yavuz, E. (2016). Yaşlı Ve Engelli Hastalar İçin Uzaktan Takip Sistemi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 4:2. doi:10.21923/jesd.17324.
- Liddy, E. D. (2001). *Natural language processing*. In Encyclopedia of Library and Information Science: 2nd Edition. NY. Marcel Decker, Inc.
- Lyalina, S., Percha, B., LePendu, P., Iyer, S. V., Altman, R. B.ve Shah, N. H. (2013). Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records. *Journal of the American Medical Informatics Association*, 20:2, 297-305. doi:10.1136/amiajnl-2013-001933.
- Malkoçoğlu, A. B. V.ve Malkoçoğlu, Ş. U. (2020). Rasgele Orman ve Lojistik Regresyon Algoritmalarının Karşılaştırmalı Performans Analizi. 5th International Conference on Computer Science and Engineering, Sanya, Çin.
- Melek, C. (2012). *Metin Madenciliği Teknikleri İle Şirketlerin Vizyon İfadelerinin Analizi*. Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı, 101, İzmir.
- Mortazavi, B. J., Downing, N. S., Bucholz, E. M., Dharmarajan, K., Manhapra, A., Li, S. X., Negahban, S. N.ve Krumholz, H. M. (2016). Analysis of Machine Learning Techniques for Heart Failure Readmissions. *Circ Cardiovasc Qual Outcomes*, 9:6, 629-640. doi:10.1161/CIRCOUTCOMES.116.003039.
- Mosayebi, A., Mojaradi, B., Naeini, A. B.ve Hosseini, S. H. K. (2020). Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer. *PLoS One*, 15:10, 1-23. doi:10.1371/journal.pone.0237658.
- Nabiyev, V.ve Erümit, A. K. (2020). *Eğitimde Yapay Zeka Kuramdan Uygulamaya*. Ankara: Pegem Akademi Yayıncılık.
- Nabiyev, V. V. (2016). *Yapay Zeka*. Ankara: Seçkin Yayıncılık.
- Oğuzlar, A. (2011). *Temel Metin Madenciliği*. Bursa: Dora Basım Yayın.
- Paksoy, V. (2017). Telemedicine applications from the perspective of health economics. *Health Care Academician Journal*, 4:1. doi:10.5455/sad.13-1481481194.
- Pawlak, Z. a. (2003). A rough set view on Bayes' theorem. *International Journal of Intelligent Systems*, 18:5, 487-498. doi:10.1002/int.10099.
- Rençber, Ö. F. (2018). *Sınıflandırma Problemlerinde Çoklu Lojistik Regresyon, Yapay Sinir Ağı ve ANFIS Yöntemlerinin Karşılaştırılması: İnsani Gelişmişlik Endeksi Üzerine Uygulama*. Ankara: Gazi Kitabevi.
- Saedsayad. (2021). Karar Ağacı Budama Kriterleri. Erişim: 24 Mart 2021, http://www.saedsayad.com/decision_tree_overfitting.htm
- Sağlık Bakanlığı. (2019). Sağlıkta Neyim Var Projesi. Erişim: 17 Nisan 2021, <https://www.saglik.gov.tr/TR,58587/fahrettin-koca-tbmmde-parlamento-muhabirleri-ile-bulustu.html>
- Savasere, A., Omiecinski, E.ve Navathe, S. (1995). An Efficient Algorithm For Mining Association Rules In Large Databases. Proceedings of the 21st VLDB Conference, Zurich, Swizerland.
- Seçkin, K. (2011). *Metin Madenciliğinde Kullanılan Yöntemlerin Karşılaştırılması: Siyasi Parti Liderlerinin Grup Genel Toplantı Konuşmaları İle Bir Uygulama*. Yüksek Lisans Tezi. Sakarya Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 122, Sakarya.
- Silahtaroglu, G., Baykal, E.ve Canbolat, Z. N. (2020). Weekly Emotional Changes Amidst Covid-19: Turkish Experience. *Ekonomi, İşletme ve Maliye Araştırmaları Dergisi*, 2:3, 280-304.
- Singh, P., Singh, S.ve Pandi-Jain, G. S. (2018). Effective heart disease prediction system using data mining techniques. *Int J Nanomedicine*, 13:T-NANO 2014 Abstracts, 121-124. doi:10.2147/IJN.S124998.
- Sukanya, M.ve Biruntha, S. (2012). Techniques on Text Mining. International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

- Şahin, İ. (2019). *Metin Madenciliği Ve Makine Öğrenmesi İle İnternet Sayfalarının Sınıflandırılması*. Hacettepe Üniversitesi Endüstri Mühendisliği Anabilim Dalı, Ankara.
- Şeker, S. E. (2015). Metin Madenciliği (Text Mining). *YBS Ansiklopedi*, 2:3, 30-32.
- Taş, Y. (2018). *Birliktelik Kuralları Madenciliği Ve Bir Uygulama*. Yüksek Lisans Tezi. Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Bilişim Sistemleri Anabilim Dalı, 95, Sivas.
- Taşçı, M. E.ve Şamli, R. (2020). Veri Madenciliği İle Kalp Hastalığı Teşhisi. *European Journal of Science and Technology*, 88-95. doi:10.31590/ejosat.araconf12.
- Tokalakçioğlu, E. (2020). *Karar Ağacı Temelli Geri Bildirim Sistemi Tasarımı: Türkiye Mutluluk Ağacı Üzerine Bir Uygulama*. Yüksek Lisans Tezi. İstanbul Teknik Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 90, İstanbul.
- Tokunaga, T.ve Iwayama, M. (1994). *Text categorization based on weighted inverse document frequency*. Tokyo Institute of Technology Department of Computer Science. Advanced Research Laboratory, Hitachi Ltd.
- Tokyürek, E. (2019). *Birliktelik Kural Çıkarım Algoritmaları Kullanılarak Market Sepet Analizi*. Yüksek Lisans Tezi. Şeyh Edebalı Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı, 62, Bilecik.
- Toplu, S.ve Cangür, Ş. (2020). Text Mining Method in the Field of Health. *Konuralp Tıp Dergisi*, 236-246. doi:10.18521/ktd.700789.
- Uzer, M. S. (2014). *Örüntü Tanıma Uygulamalarında Yapay Zekâ Ve Öznitelik Dönüşüm Metotları Kullanılarak Geliştirilen Öznitelik Seçme Algoritmaları*. Doktora Tezi. Selçuk Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı, 103, Konya.
- Vyawahare, R., Mandape, P.ve Chawhan, S. (2015). Implementation Of DIC Using Association Rule Mining Algorithm. *International Journal of Engineering Sciences & Research Technology*, 4:3, 679-684.
- Wikipedia. (2021). Zemberek Doğal Dil Kütüphanesi. Erişim: 13 Mart 2021, [https://tr.wikipedia.org/wiki/Zemberek_\(yaz%C4%B1%C4%B1m\)](https://tr.wikipedia.org/wiki/Zemberek_(yaz%C4%B1%C4%B1m))
- Wu, C. S., Kuo, C. J., Su, C. H., Wang, S. H.ve Dai, H. J. (2020). Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records. *J Affect Disord*, 260, 617-623. doi:10.1016/j.jad.2019.09.044.
- Wu, Y., Dang, M., Li, H., Jin, X.ve Yang, W. (2019). Identification of genes related to mental disorders by text mining. *Medicine (Baltimore)*, 98:42, 1-7. doi:10.1097/MD.0000000000017504.
- Yang, Y. (1998). An Evaluation of Statistical Approaches to Text Categorization. *Information Retrieval I*, 69-90.
- Yelmen, İ. (2016). *Doğal Dil İşleme Yöntemleriyle Türkçe Sosyal Medya Verileri Üzerinde Duygu Analizi*. Yüksek Lisans Tezi. İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Ana Bilim Dalı, 71, İstanbul.
- Yıldız, O. (2016). Metin Madenciliğinde Anahtar Kelime Seçimi Bir Üniversite Örneği. *Yönetim Bilişim Sistemleri Dergisi*, 2:3, 29-50.
- Yolcular, B. O., Bilge, U.ve Samur, M. K. (2018). Kulak Burun Boğaz Taburcu Notlarından Birliktelik Kurallarının Çıkartılması. *Bilişim Teknolojileri Dergisi*, 11:1, 35-42. doi:10.17671/gazibtd.319690.
- Yücebaş, S. C. (2018). Karmaşık Hastalıkların Teşhisinde Veri Madenciliği Yöntemlerinin Başarım Karşılaştırması. *Çanakkale Onsekiz Mart Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 4:1, 14-27.
- Zor, T. (2021). *Denim Tasarımı İçin Hissel Metin Madenciliği Yaklaşımı*. Yüksek Lisans Tezi. Erciyes Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 91, Kayseri.

ÖZ GEÇMİŞ

Müberra Terzi 2014 yılında Samsun Anafartalar Anadolu Lisesi'ni bitirdikten sonra Ondokuz Mayıs Üniversitesi Mühendislik Fakültesi Endüstri Mühendisliği'nden 2018 yılında birincilikle mezun oldu. 2018 yılında OMÜ LEE Akıllı Sistemler Mühendisliği Yüksek Lisans programını girdi. Mezuniyetinden sonra 2018-2020 yılları arasında VM Medicalpark Samsun Hastanesinde Endüstri Mühendisi olarak çalıştı. 2020 yılı sonrasında OMÜ Endüstri Mühendisliği Bölümünde Araştırma Görevlisi olarak çalışmaya başladı.

İletişim Bilgileri

E mail : muberraterzii@gmail.com

ORCID ID: <https://orcid.org/0000-0002-3939-4268>

Yayınlanmış Çalışmalar:

1. TERZİ, M., “**Türkiye’de Sağlık Sektöründe Veri Madenciliği Kullanım Alanları**” Black Sea Journal of Health Science, 2019. (Derleme)
2. TERZİ, M., MURAT, N. (2020) “**Metin Madenciliği Teknikleri Kullanılarak Özet Epikrizlere Dayalı Poliklinik Sınıflandırılması**” Uluslararası Marmara Fen ve Sosyal Bilimler Kongresi, 2020. (Özet Bildiri / Sözlü Sunum)

EKLER

EK-1

Veri gizliliğinden dolayı alınan diğer parametreler eklenmemiştir. Sadece metin madenciliği analizinde kullanılan parametreler eklenmiştir.

Cinsiyet	Yaş	Özet Epikriz	Yönlendirilen Poliklinik
Kadın	33	Bulantı Kusma	Dahiliye
Kadın	28	İshal Karın Ağrısı	Dahiliye
Kadın	28	Bulantı Kusma Halsizlik ve Dalma Nöbetleri Aşırı Bulantı Kusma Sonrası Boş ve Anlamsız Bakışları Olan Hasta Yakınlarınca Getirildi	Dahiliye
Kadın	29	Bulantı Kusma Karın Ağrısı Ateş Sabah İshali Başlamış Giderek Artmış Sulu İshali Giderek Artmış	Dahiliye
Kadın	40	İshal Kusma	Dahiliye
Kadın	27	Bulantı Kusma Halsizlik Şikâyeti ile Gelen Hasta	Dahiliye
Kadın	28	Ateş Üşüme Titreme Bulantı Sabah Başlamış	Dahiliye
Kadın	18	Karın Ağrısı Bulantı Kusma	Dahiliye
Kadın	47	Baş Ağrısı Bulantı Kusma	Dahiliye
Kadın	31	Baş Ağrısı Bulantı Kusma	Dahiliye
Erkek	27	İshal Kusma	Dahiliye
Erkek	27	Bulantı Kusma İshal Karın Ağrısı Gece Başlamış	Dahiliye
Erkek	29	Bulantı Kusma Baş Dönmesi Baş Ağrısı Şikâyeti ile Geldi	Dahiliye
Erkek	17	İshal Kusma Karın Ağrısı	Dahiliye
Erkek	36	Kusma İshal	Dahiliye
Erkek	29	Bulantı Kusma	Dahiliye
Erkek	18	Bulantı Kusma İshal	Dahiliye
Erkek	23	Bayılma ve Halsizlik Şikâyeti ile Özel Araç ile Getirilen Hasta	Dahiliye
Erkek	41	Bulantı Kusma Addison ve Hasimato Hastası Tek Böbrek	Dahiliye
Erkek	57	Karın Ağrısı	Dahiliye
Kadın	39	Baş Ağrısı Bulantı Kusma	Dermatoloji
Kadın	76	Dudaklarda Kaşıntı ve Kızarıklık Şikâyeti ile Gelen Hasta	Dermatoloji
Kadın	80	Sol Ayakta Şişlik Kızarıklık	Dermatoloji
Kadın	42	Yüzüne Krem Sürdükten Sonra Alerjik Reaksiyon Gelişmiş	Dermatoloji
Kadın	49	Alerji	Dermatoloji
Kadın	34	Alerji	Dermatoloji
Kadın	21	Dudakları Şişmiş	Dermatoloji
Kadın	17	Alerji	Dermatoloji

Kadın	63	Tüm Vücutta Kızarıklık ve Kaşıntı Şikâyeti ile Gelen Hasta	Dermatoloji
Kadın	23	Ürtiker	Dermatoloji
Erkek	32	Vücutta Kızarıklık ve Kaşıntı	Dermatoloji
Erkek	23	Yaygın Ürtiker	Dermatoloji
Erkek	47	Sağ Bacakta Kızarıklık	Dermatoloji
Erkek	49	Alerji	Dermatoloji
Erkek	47	Kafada Küçük Yaralar	Dermatoloji
Erkek	27	Vücutta Kızarıklık ve Kaşıntı	Dermatoloji
Erkek	33	Boğaz Ağrısı ve Ciltte Döküntü	Dermatoloji
Erkek	41	Ateş Alında Şişlik	Dermatoloji
Erkek	64	Ellerde Şişlik Döküntü Ateş	Dermatoloji
Erkek	25	Vücutta Kaşıntı	Dermatoloji
Kadın	33	Ateş	Enfeksiyon
Kadın	38	Ateş	Enfeksiyon
Kadın	38	Ateş	Enfeksiyon
Kadın	21	Ateş	Enfeksiyon
Kadın	29	Ateş	Enfeksiyon
Kadın	23	Ateş	Enfeksiyon
Kadın	60	Ateş	Enfeksiyon
Kadın	59	Ateş	Enfeksiyon
Kadın	30	Ateş	Enfeksiyon
Kadın	67	Ateş	Enfeksiyon
Erkek	31	Ateş Boğaz Ağrısı	Enfeksiyon
Erkek	61	Ateş	Enfeksiyon
Erkek	58	Ateş	Enfeksiyon
Erkek	58	Ateş	Enfeksiyon
Erkek	32	Ateş	Enfeksiyon
Erkek	66	Ateş	Enfeksiyon
Erkek	59	Ateş	Enfeksiyon
Erkek	32	Ateş	Enfeksiyon
Erkek	29	Ateş	Enfeksiyon
Erkek	24	Ateş	Enfeksiyon
Kadın	38	Bel Ağrısı	Fizik Tedavi
Kadın	48	Ftr Polikliniğinden Tedavi Amaçlı Gönderilen Hasta	Fizik Tedavi
Kadın	50	Bel Ağrısı	Fizik Tedavi
Kadın	61	Bel Ağrısı	Fizik Tedavi
Kadın	31	Bel Ağrısı	Fizik Tedavi
Kadın	55	Bel Ağrısı	Fizik Tedavi
Kadın	59	Bel Ağrısı	Fizik Tedavi
Kadın	29	Bel Ağrısı	Fizik Tedavi
Kadın	97	Aygın Eklem Ağrısı Şikâyeti ile Gelen Hasta	Fizik Tedavi
Kadın	37	Boyun Ağrısı Gece Başlamış	Fizik Tedavi
Erkek	34	Bel Ağrısı	Fizik Tedavi
Erkek	37	Bel Ağrısı	Fizik Tedavi
Erkek	44	Bel Ağrısı	Fizik Tedavi

Erkek	39	Bel Ağrısı	Fizik Tedavi
Erkek	39	Bel Ağrısı	Fizik Tedavi
Erkek	56	Bel Ağrısı	Fizik Tedavi
Erkek	34	Bel Ağrısı	Fizik Tedavi
Erkek	43	Bel Ağrısı	Fizik Tedavi
Erkek	42	Boyun Ağrısı	Fizik Tedavi
Erkek	25	Bel Ağrısı	Fizik Tedavi
Kadın	92	Acile Getirilen Hasta Solunum Sıkıntısı	Göğüs Hastalıkları ve Cerrahisi
Kadın	26	Ateş Öksürük Balgam	Göğüs Hastalıkları ve Cerrahisi
Kadın	47	Sırt Ağrısı Nefes Darlığı	Göğüs Hastalıkları ve Cerrahisi
Kadın	18	Nefes Darlığı Boğaz Ağrısı	Göğüs Hastalıkları ve Cerrahisi
Kadın	38	Ateş Öksürük	Göğüs Hastalıkları ve Cerrahisi
Kadın	48	Öksürük Boğaz Ağrısı Baş Ağrısı	Göğüs Hastalıkları ve Cerrahisi
Kadın	75	Nefes Darlığı Öksürük Şikâyeti ile Gelen Hasta Siroz Hastası	Göğüs Hastalıkları ve Cerrahisi
Kadın	57	Kas Distrofi Hastası	Göğüs Hastalıkları ve Cerrahisi
Kadın	52	Göğüs Ağrısı Yeni Başlamış	Göğüs Hastalıkları ve Cerrahisi
Kadın	22	Ateş Öksürük	Göğüs Hastalıkları ve Cerrahisi
Erkek	68	Öksürük Nefes Darlığı Şikâyeti ile Gelen Hasta	Göğüs Hastalıkları ve Cerrahisi
Erkek	83	112 İle Sevkli Olarak Hemothoraks Sebebi ile Entübe ve Göğüs Tüpü Olan Hasta	Göğüs Hastalıkları ve Cerrahisi
Erkek	74	Kulak Ağrısı Nefes Darlığı	Göğüs Hastalıkları ve Cerrahisi
Erkek	76	112 Ambulansı ile Getirildi Nefes Darlığı Ateş	Göğüs Hastalıkları ve Cerrahisi
Erkek	25	Öksürük ve Nefes Darlığı Şikâyeti ile Gelen Hasta	Göğüs Hastalıkları ve Cerrahisi
Erkek	86	Ateş Balgamlı Öksürük	Göğüs Hastalıkları ve Cerrahisi
Erkek	79	Nefes Darlığı	Göğüs Hastalıkları ve Cerrahisi
Erkek	29	Göğüs Ağrısı Ateş Yürürken Düşmüş Ateşi Gece Çıkmış	Göğüs Hastalıkları ve Cerrahisi
Erkek	75	Öksürük Nefes Darlığı Tansiyon Yüksekliği	Göğüs Hastalıkları ve Cerrahisi
Erkek	38	Göğüs Ağrısı	Göğüs Hastalıkları ve Cerrahisi
Kadın	25	Adet Sancısı Bulantı Hastamız Adetinin 1 Günündedir	Kadın Doğum ve Hastalıkları
Kadın	27	Vajinal Kanaması Olmuş	Kadın Doğum ve Hastalıkları
Kadın	45	Karın Ağrısı Adet Dönemindedir	Kadın Doğum ve Hastalıkları
Kadın	34	23 Hafta 4 Günlük Gebe Olan ve Kasılma Şikâyeti ile Geldi Ciproleks 5 Mg Kullanmaya Başlamış	Kadın Doğum ve Hastalıkları
Kadın	34	Kasık Ağrısı	Kadın Doğum ve Hastalıkları
Kadın	34	İdrar Yaparken Yanma Gebelik	Kadın Doğum ve Hastalıkları
Kadın	28	Ateş Memelerinde Ağrı Yeni Doğum Yapmış	Kadın Doğum ve Hastalıkları
Kadın	26	Vajinal Kanama Gebe Hasta	Kadın Doğum ve Hastalıkları

Kadın	31	Denize Atlama	Kadın Doğum ve Hastalıkları
Kadın	22	C/S Sonrası Ağrı	Kadın Doğum ve Hastalıkları
Kadın	26	Gebe ve İshal	Kadın Doğum ve Hastalıkları
Kadın	40	Adet Sancısı	Kadın Doğum ve Hastalıkları
Kadın	21	Bulantı Adet Sancısı Alkolü Biraz Fazla Kaçırılmış	Kadın Doğum ve Hastalıkları
Kadın	22	Adet Sancısı	Kadın Doğum ve Hastalıkları
Kadın	25	İshal Kusma Ateş 26 Hf Gebe Belinde Sırtlarında Ağrı Varmış	Kadın Doğum ve Hastalıkları
Kadın	38	Adet Döneminin Bitmesine Yakın Böyle Ağrıları Oluyormuş Dün Karnı Ağrımış	Kadın Doğum ve Hastalıkları
Kadın	19	Kasık Ağrısı	Kadın Doğum ve Hastalıkları
Kadın	29	Baş Dönmesi Bulantı Kusma Gece Başlamış Gebe Hasta	Kadın Doğum ve Hastalıkları
Kadın	18	Kasık Ağrısı	Kadın Doğum ve Hastalıkları
Kadın	37	Gebelik YB İhtiyacı	Kadın Doğum ve Hastalıkları
Kadın	26	Boğaz Ağrısı Göğüs Ağrısı Havzadan Sevкли Olarak Entübe	Kardiyoloji
Kadın	86	Olarak Mı Ön Tanısı ile Gelen Hasta	Kardiyoloji
Kadın	56	Tansiyonu Çıkmış	Kardiyoloji
Kadın	31	Çarpıntı	Kardiyoloji
Kadın	33	Göğüs Ağrısı	Kardiyoloji
Kadın	28	Sol Omuz ve Sırt Ağrısı Şikâyeti ile Gelen Hasta	Kardiyoloji
Kadın	44	Hipertansiyon	Kardiyoloji
Kadın	77	Göğüs Ağrısı Başında Yanma Hissi ve Mide Bulantısı Şikâyeti ile Gelen Hasta	Kardiyoloji
Kadın	62	Bulantı Kusma Boyun Ağrısı Baş Ağrısı Evde Tansiyonu Yükselmiş ve İlaç Almış	Kardiyoloji
Kadın	63	Tansiyon Yüksekliği	Kardiyoloji
Erkek	40	Tansiyon Yüksekliği Baş Ağrısı Sinirlendikten Sonra Olmuş	Kardiyoloji
Erkek	29	Boğaz Ağrısı Göğüs Ağrısı	Kardiyoloji
Erkek	56	Göğüste Ağrı Sırtta Vuruyor 1 Sene Önce Uzdr Emre Beyce Anjiyo Yapılmış Stent Takılmış	Kardiyoloji
Erkek	38	Göğüs Ağrısı Şikâyeti ile Gelen Hasta	Kardiyoloji
Erkek	36	Göğüs Ağrısı	Kardiyoloji
Erkek	77	Göğüs Ağrısı	Kardiyoloji
Erkek	70	Nefes Ağrısı Göğüs Ağrısı Dış Merkezden Sevкли Geldi 112 ile Getirildi	Kardiyoloji
Erkek	37	Tansiyon Yüksekliği	Kardiyoloji
Erkek	22	Halsizlik	Kardiyoloji

Erkek	28	Sol Yan Ağrısı	Kardiyoloji
Kadın	28	Boğaz Ağrısı	Kulak Burun Boğaz
Kadın	33	Ateş Boğaz Ağrısı Nefes Darlığı Öksürük	Kulak Burun Boğaz
Kadın	17	Boğaz Ağrısı ve Yutma Güçlüğü Şikâyeti ile Gelen Hasta	Kulak Burun Boğaz
Kadın	35	Boğaz Ağrısı	Kulak Burun Boğaz
Kadın	33	Ateş Boğaz Ağrısı Nefes Darlığı Öksürük	Kulak Burun Boğaz
Kadın	34	Ateş Boğaz Ağrısı Şikâyeti ile Gelen Hasta	Kulak Burun Boğaz
Kadın	35	Boğaz Ağrısı	Kulak Burun Boğaz
Kadın	28	Boğaz Ağrısı	Kulak Burun Boğaz
Kadın	30	Boğaz Ağrısı Ateş	Kulak Burun Boğaz
Kadın	34	Nefes Darlığı Bulantı	Kulak Burun Boğaz
Erkek	29	Baş Ağrısı Halsizlik Ateş	Kulak Burun Boğaz
Erkek	36	Ateş Geniz Akıntısı Şikâyeti Sebebi ile Gelen Hasta	Kulak Burun Boğaz
Erkek	23	Burun Tıkanıklığı Grip	Kulak Burun Boğaz
Erkek	31	Boğaz Ağrısı	Kulak Burun Boğaz
Erkek	29	Boğaz Ağrısı Ateş Halsizlik	Kulak Burun Boğaz
Erkek	44	Ateş Boğaz Ağrısı	Kulak Burun Boğaz
Erkek	29	Ateş Boğaz Ağrısı	Kulak Burun Boğaz
Erkek	31	Ateş	Kulak Burun Boğaz
Erkek	21	Boğaz Ağrısı	Kulak Burun Boğaz
Erkek	31	Ateş Boğaz Ağrısı Öksürük	Kulak Burun Boğaz
Kadın	25	Baş Dönmesi	Nöroloji
Kadın	49	Bacağında Çekme Varmış Birkaç Gündür Ağrı Varmış Ancak Şimdi Çekme Varmış	Nöroloji
Kadın	35	Baş Ağrısı	Nöroloji
Kadın	58	İnr Takibi İnr Takibi İstiyor 4 Günde Bir	Nöroloji
Kadın	33	Bel Ağrısı	Nöroloji
Kadın	53	Diz Ağrısı Halsizlik Bacakta Uyuşukluk	Nöroloji
Kadın	34	Bulantı Kusma Baş Ağrısı	Nöroloji
Kadın	69	Bulantı Kusma Sabah Başlamış 26 Aralık Çarşamba Öğleden Sonra Ayancık Devlet Hastanesine Şiddetli Baş Ağrısı Nedeniyle	Nöroloji
Kadın	53	Yatırılmış Gerekli Müdahale ve Takibe Rağmen Ağrısı Geçmeyince Sinop Devlet Hastanesine Sevk Edilmiş	Nöroloji
Kadın	79	Baş Dönmesi ve Mide Bulantısı Şikâyeti ile Gelen Hasta	Nöroloji
Erkek	43	Baş Ağrısı Şikâyeti ile Gelen Hasta	Nöroloji
Erkek	29	Baş Ağrısı Şikâyeti ile Gelen Hasta	Nöroloji

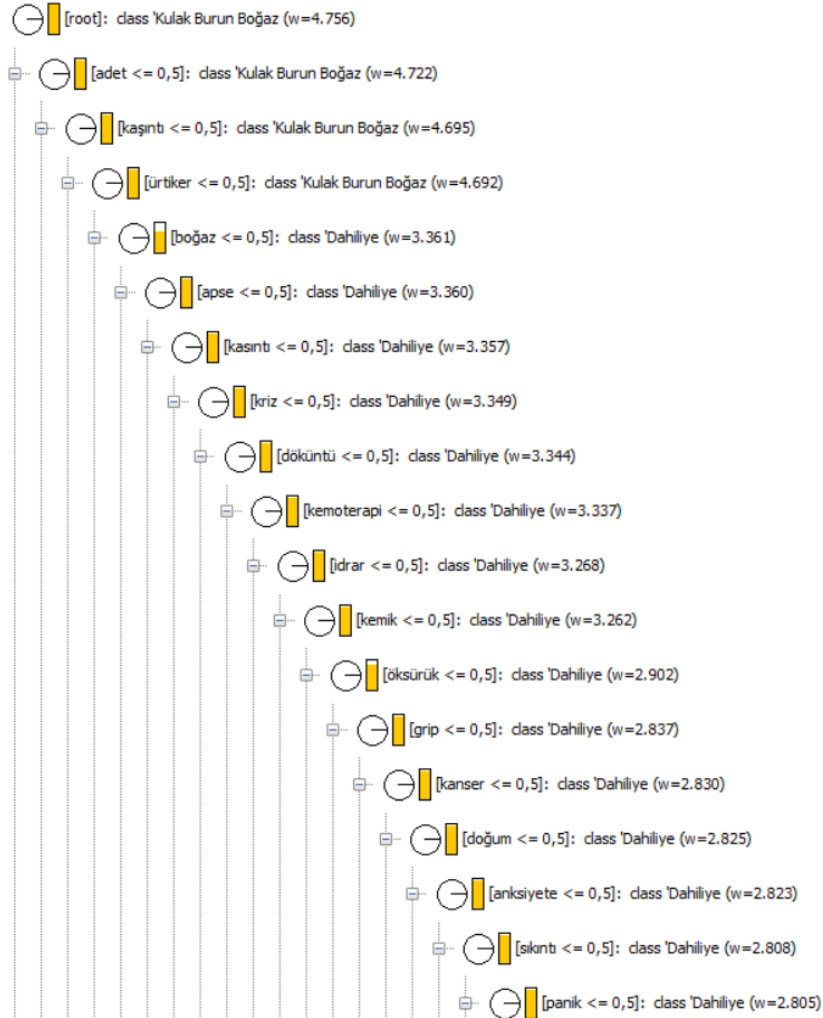
Erkek	45	Baş Dönmesi	Nöroloji
Erkek	63	Baş Ağrısı	Nöroloji
Erkek	21	Baş Ağrısı ve Baş Dönmesi Şikâyeti ile Gelen Hasta	Nöroloji
Erkek	30	Baş Ağrısı	Nöroloji
Erkek	43	Kasılma	Nöroloji
Erkek	35	Baş Ağrısı	Nöroloji
Erkek	37	Baş Ağrısı	Nöroloji
Erkek	30	Baş Dönmesi	Nöroloji
Kadın	67	Kemik Metastatik Ca İshali Var	Onkoloji
Kadın	38	Bulantı Kusma Halsizlik Lenfoma Tedavisi Görüyor Omu Fakültede	Onkoloji
Kadın	33	Ateş Kolon Ca	Onkoloji
Kadın	74	Yanık Radyoterapi Alıyor	Onkoloji
Kadın	44	Ateş Balgamlı Öksürük Bulantı Sırt Ağrıları Meme Ca	Onkoloji
Kadın	63	Bulantı Kusma İshal Kemoterapi Sonrası Karaciğer Ca	Onkoloji
Kadın	63	Kc Ca Bulantı Kusma Öksürük Eklem Ağrıları Kc Ca Onkoloji Hastası	Onkoloji
Kadın	64	Halsizlik Bulantı İştahsızlık Ca Hastası Karın Ağrısı Bulantı	Onkoloji
Kadın	27	Öksürük Eklem Ağrıları Over Ca Opere	Onkoloji
Kadın	55	Meme Ca Bulantı Kusma Ağrıları Nefes Darlığı Öksürük	Onkoloji
Erkek	73	Kolon Ca Hastası Ateş Yüksekliği	Onkoloji
Erkek	50	Nefes Darlığı Öksürük	Onkoloji
Erkek	62	Bacakta Şişlik Ağrı Varmış Işın Tedavisi Alıyormuş Bacağından Ameliyat Olmuş	Onkoloji
Erkek	78	Şiddetli Bel Ağrısı Bel Ağrısı Nedeni ile Yapılan Tetkikerde Lomber Vertebralarda Metastatik Lezyonlar Saptanan ve Primeri Araştırılan Hasta Ağrı Palyasyonu İçin Geldi	Onkoloji
Erkek	50	Huzursuzluk İç Daralması Akıllı İlaç Kullanıyormuş Kemik Tm	Onkoloji
Erkek	63	Ateş Sırtta Ağrı Akc Ca	Onkoloji
Erkek	60	Öksürük Nefes Darlığı	Onkoloji
Erkek	70	Ateş Öksürük Dün Başlayan Şikayetleri Giderek Artmış Ac Ca Sebebi ile Tedavi Sırasında	Onkoloji
Erkek	55	Göğüs Ağrısı Olan ve Kaşıntısı Olan Hasta	Onkoloji
Erkek	76	Bulantı Kusma Kolon Ca Dr Dilek Hanım Takibinde Kemoterapi Alıyor	Onkoloji

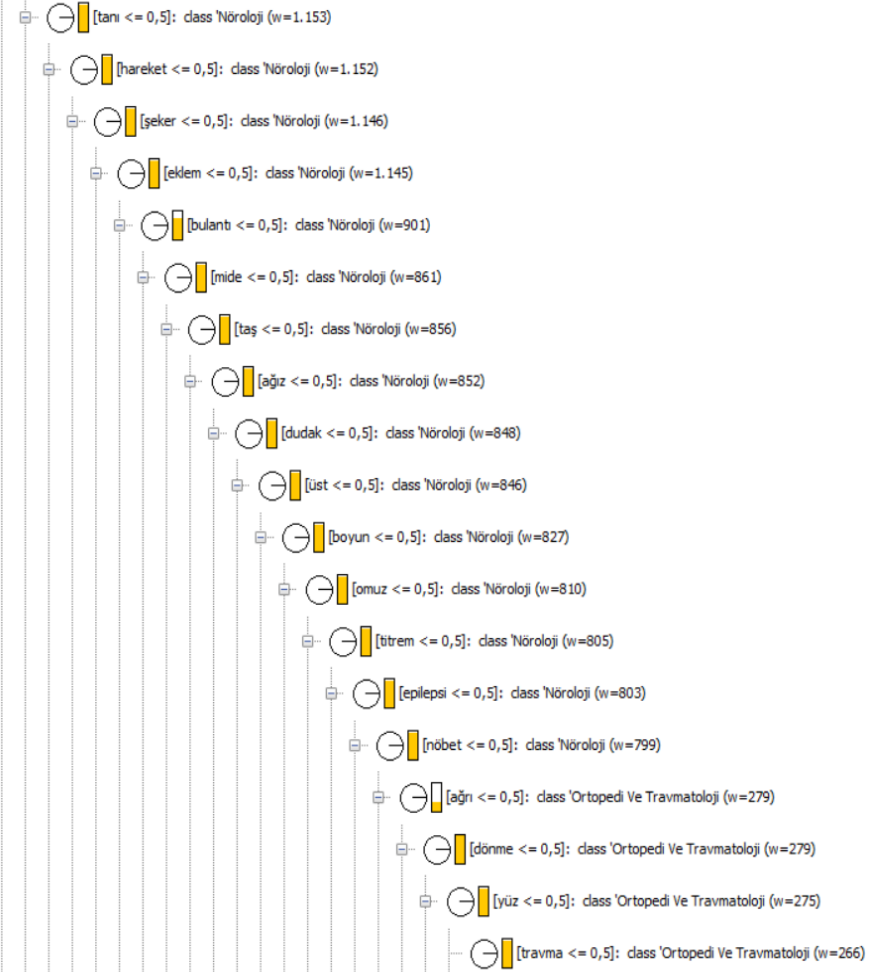
Kadın	70	Düşme Sonucu Sol Dizde Ağrı Şikâyeti ile 112 ile Gelen Hasta Dün Pazar Yerinde Kayarak Düşmüş	Ortopedi ve Travmatoloji
Kadın	17	Düşme Sonucu Sağ Ayak Bileğinde Ağrı Şişlik Şikâyeti ile Geldi	Ortopedi ve Travmatoloji
Kadın	41	Ateş Sağ Ayakta Ağrı	Ortopedi ve Travmatoloji
Kadın	71	Bacak Ağrısı	Ortopedi ve Travmatoloji
Kadın	39	Düşme Ayakta Burkulma	Ortopedi ve Travmatoloji
Kadın	34	Sol Ayak Ağrısı Düşme	Ortopedi ve Travmatoloji
Kadın	70	Dün Parmağı İş Yaparken Ağrımış Şimdi Düzeltmemiş	Ortopedi ve Travmatoloji
Kadın	56	Yürürken Ayağı Kaymış Dizi Burkulmuş	Ortopedi ve Travmatoloji
Kadın	33	Diz ve Bacak Ağrısı	Ortopedi ve Travmatoloji
Kadın	69	Sol Eli Üzerine Düşmüş	Ortopedi ve Travmatoloji
Erkek	34	Sağ Ayakta Ağrı Şikâyeti ile Gelen Hasta	Ortopedi ve Travmatoloji
Erkek	41	Sol Ayağında Ağrı Hastamız Merdiven İnerken Ayağını Burkmuş	Ortopedi ve Travmatoloji
Erkek	34	Sol El Başparmakta Cam Kesisi	Ortopedi ve Travmatoloji
Erkek	45	Omuz Çıkığı	Ortopedi ve Travmatoloji
Erkek	29	Bacak Ağrısı	Ortopedi ve Travmatoloji
Erkek	30	Diz Ağrısı Düşmüş	Ortopedi ve Travmatoloji
Erkek	38	Düşme	Ortopedi ve Travmatoloji
Erkek	36	Burkulma Sonucu Sol Ayak Bileğinde Şişme	Ortopedi ve Travmatoloji
Erkek	57	Omuzda Ağrı	Ortopedi ve Travmatoloji
Erkek	42	Evde Ayağını Ezmiş Sol Ayak 2 Nci Parmağının Üzerine Odun Düşmüş	Ortopedi ve Travmatoloji
Kadın	33	Bayılma ve Kasılma Şikâyeti ile Gelen Hasta Annesi Yoğun Bakımda Olan Hasta	Psikiyatri
Kadın	25	Dişlerde Kenetlenme ve Kasılma Sebebi Le Acil Servise Getirilen Hasta	Psikiyatri
Kadın	40	Huzursuzluk Sinirlilik Ellerde Titreme Annesi Fenalaşmış Yoğun Bakıma Yatmış Sonra Olmuş	Psikiyatri
Kadın	66	Başından Başlayıp Sirtına Bütün Vücuduna Yayılan Yanma Varmış Dün Aynı Şikayetle Başka Hasta haneye Gitmiş Bir şey Çıkmamış Mayıs Ayından Beri Şikâyeti devam Ediyormuş	Psikiyatri
Kadın	23	Kasılma ve Ağlama Atağı	Psikiyatri
Kadın	52	Bayılma	Psikiyatri
Kadın	47	Baş Ağrısı ve Ağlama	Psikiyatri

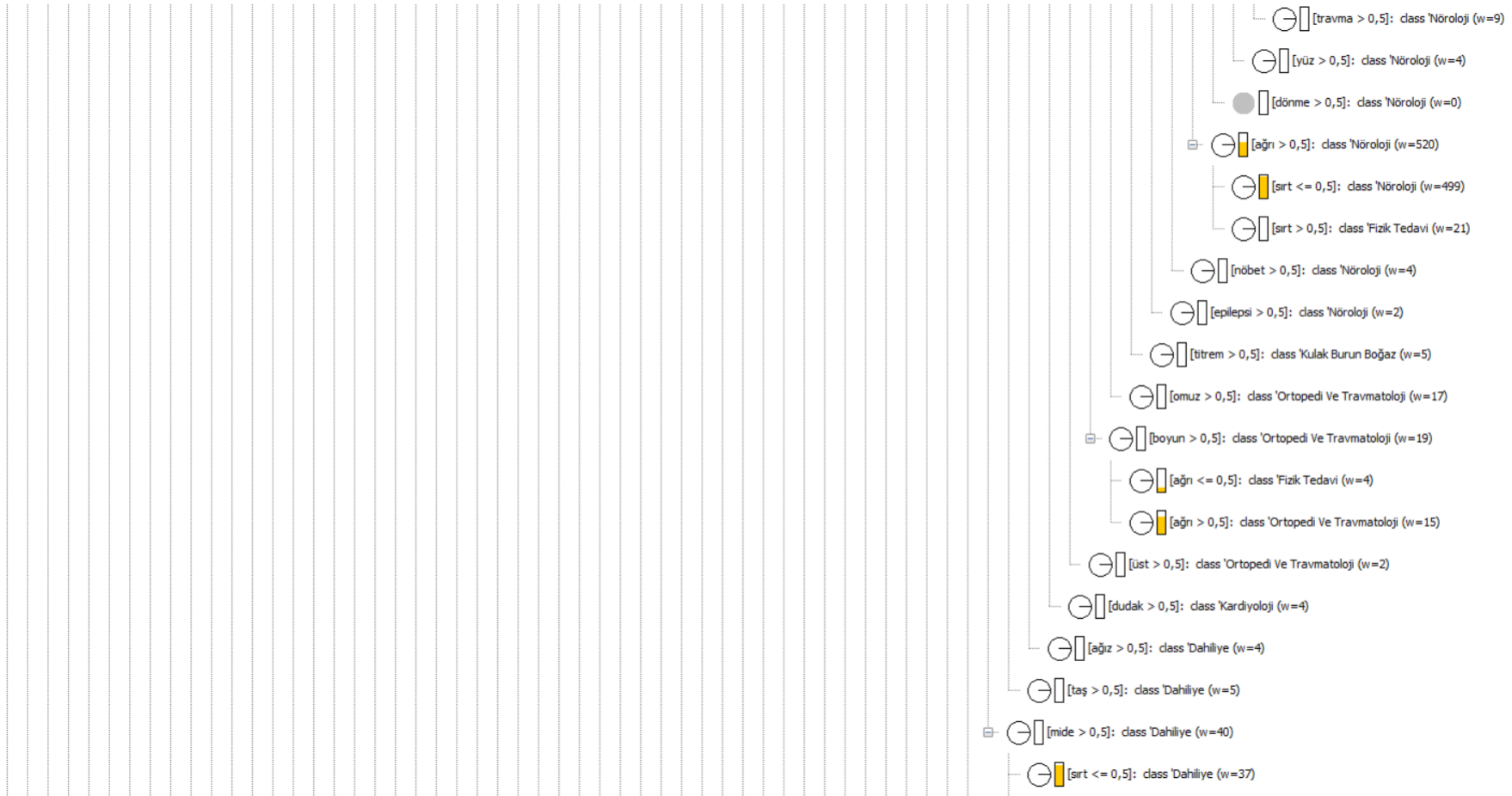
		Bayılma ve Kasılma Sebebi ile	
Kadın	33	Gelen Hasta Annesi Yoğun Bakımda Vefat Etmiş	Psikiyatri
Kadın	22	Çarpıntı	Psikiyatri
Kadın	21	İç Sıkıntısı	Psikiyatri
Erkek	28	Nefes Darlığı	Psikiyatri
Erkek	37	Babasını Kaybetmiş	Psikiyatri
Erkek	41	Oyunda Kasılma Göğüste Yanma	Psikiyatri
Erkek	30	Bayılma Sebebi ile Patoloji Polikliniğinden Getirilen Hasta	Psikiyatri
Erkek	48	Ellerinde Kasılma ve Baygınlık Şikâyeti ile Gelen Hasta	Psikiyatri
Erkek	30	Huzursuzluk Sinirlilik	Psikiyatri
Erkek	17	Sinir Krizi	Psikiyatri
Erkek	23	Nefes Alamama	Psikiyatri
Erkek	22	İç Sıkıntısı	Psikiyatri
Erkek	43	Huzursuzluk Sinirlilik	Psikiyatri
Kadın	37	Yan Ağrısı İdrar Yaparken Yanma	Üroloji
Kadın	50	Yan Ağrısı	Üroloji
Kadın	60	İdrar Yaparken Yanma	Üroloji
Kadın	31	Sol Yan Ağrısı	Üroloji
Kadın	40	İdrar Yaparken Yanma	Üroloji
Kadın	75	İdrarda Yanma	Üroloji
Kadın	44	Yan Ağrısı	Üroloji
Kadın	44	İdrarda Yanma	Üroloji
Kadın	29	Sol Yan Ağrısı Şikâyeti ile Gelen Hasta	Üroloji
Kadın	50	Ateş Sık İdrar Gitme	Üroloji
Erkek	29	Yan Ağrısı	Üroloji
Erkek	27	Sol Yan Ağrısı Daha Öncede Taşı Varmış	Üroloji
Erkek	51	Sol Yan Ağrısı Daha Önce Taşı Varmış Yeni Tomografi Çekilmiş	Üroloji
Erkek	52	Yan Ağrısı	Üroloji
Erkek	27	Yan Ağrısı	Üroloji
Erkek	39	Yan Ağrısı	Üroloji
Erkek	37	Yan Ağrısı	Üroloji
Erkek	32	Yan Ağrısı İdrar Yaparken Yanma	Üroloji
Erkek	88	Ateş İdrar Yapmada Zorluk	Üroloji
Erkek	27	Yan Ağrısı	Üroloji

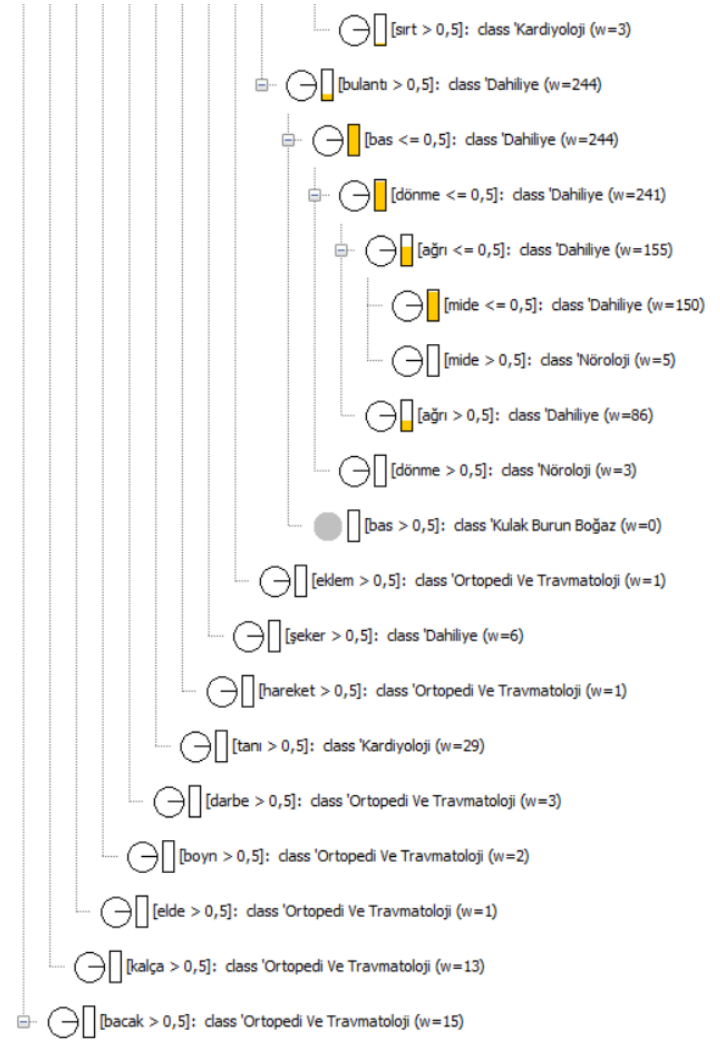
EK- 2

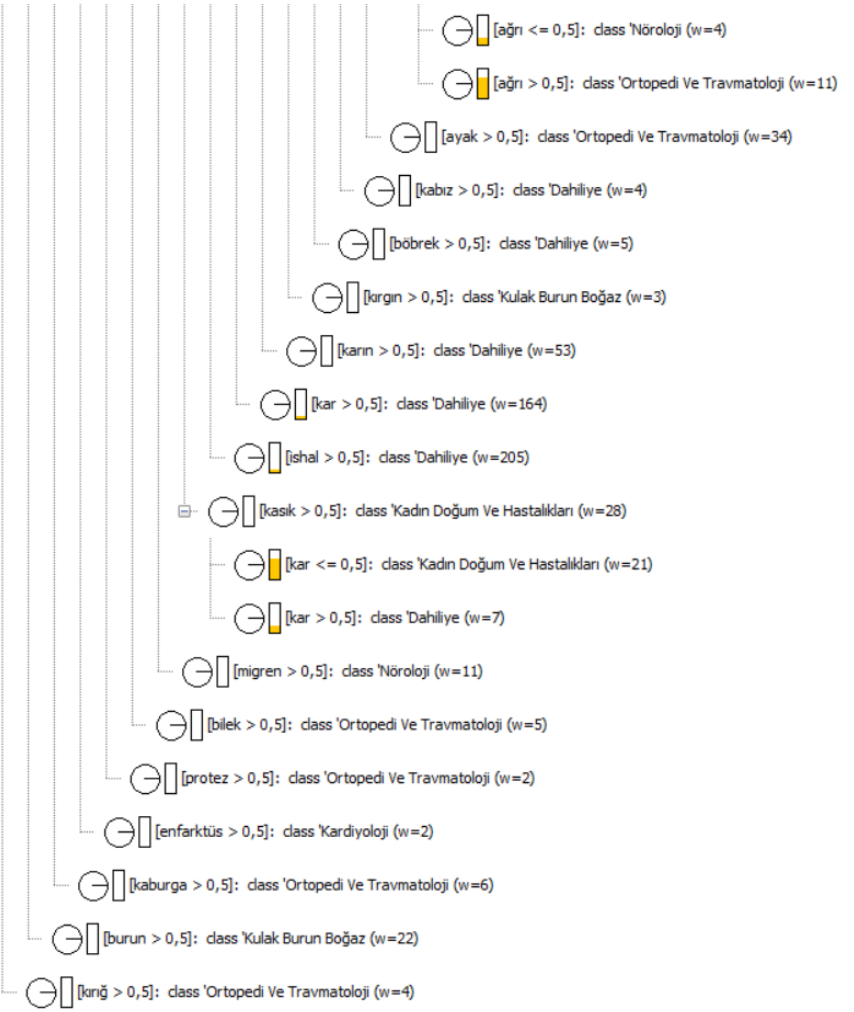
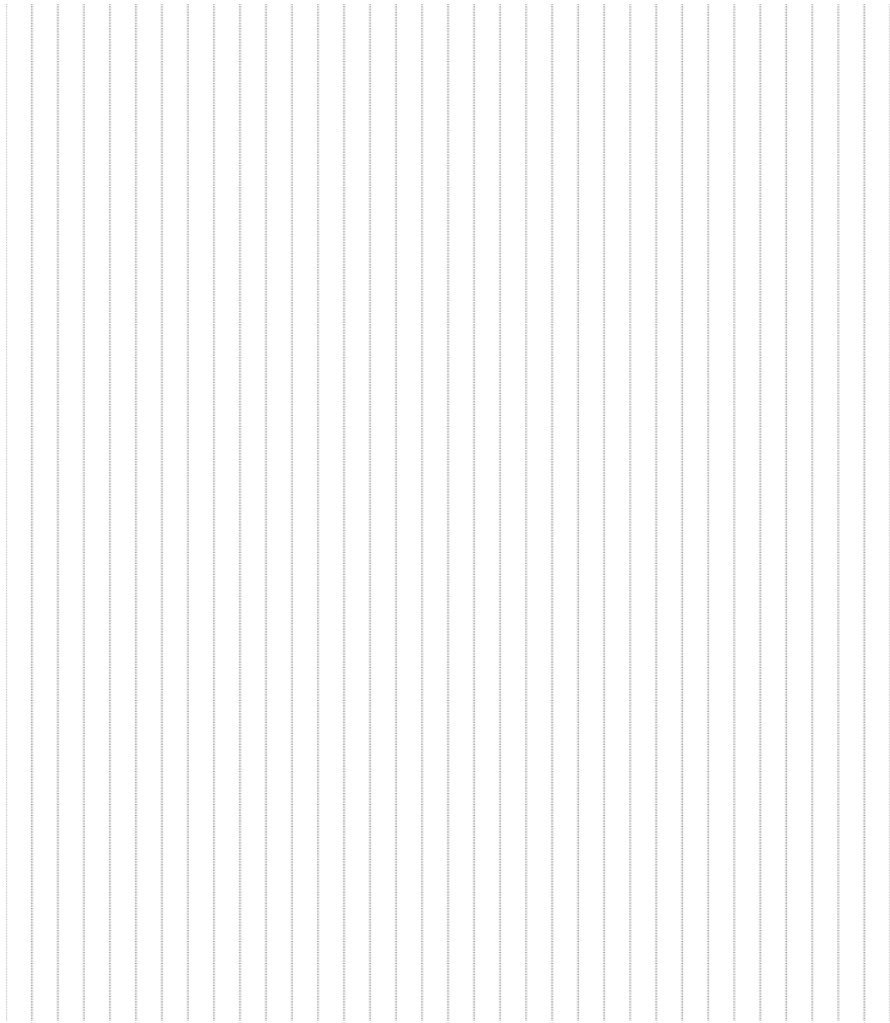
Budanmamış karar ağacına aittir.

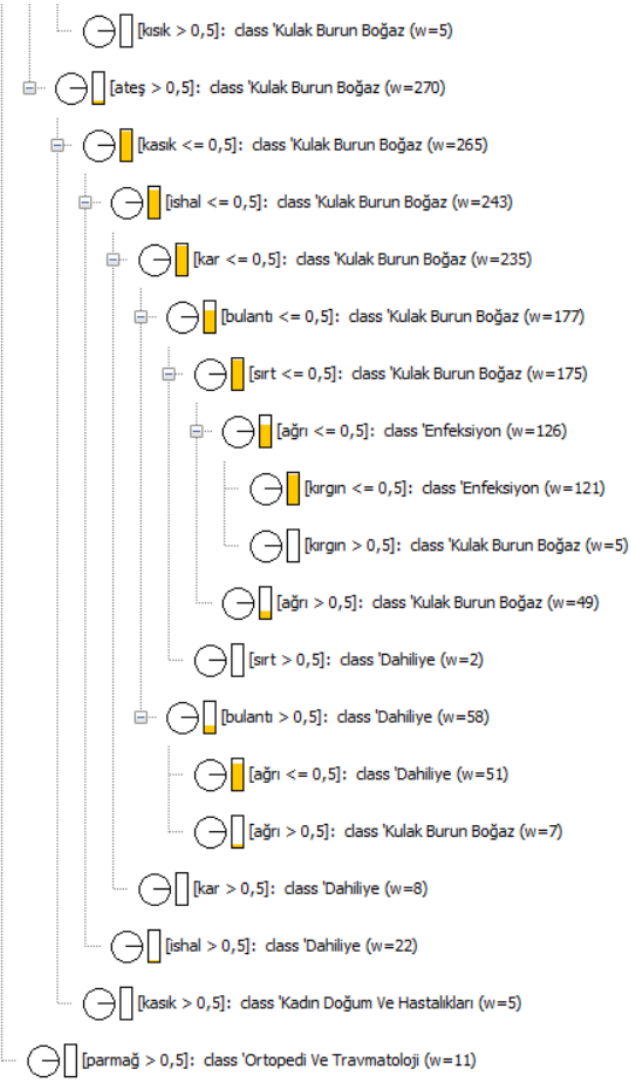
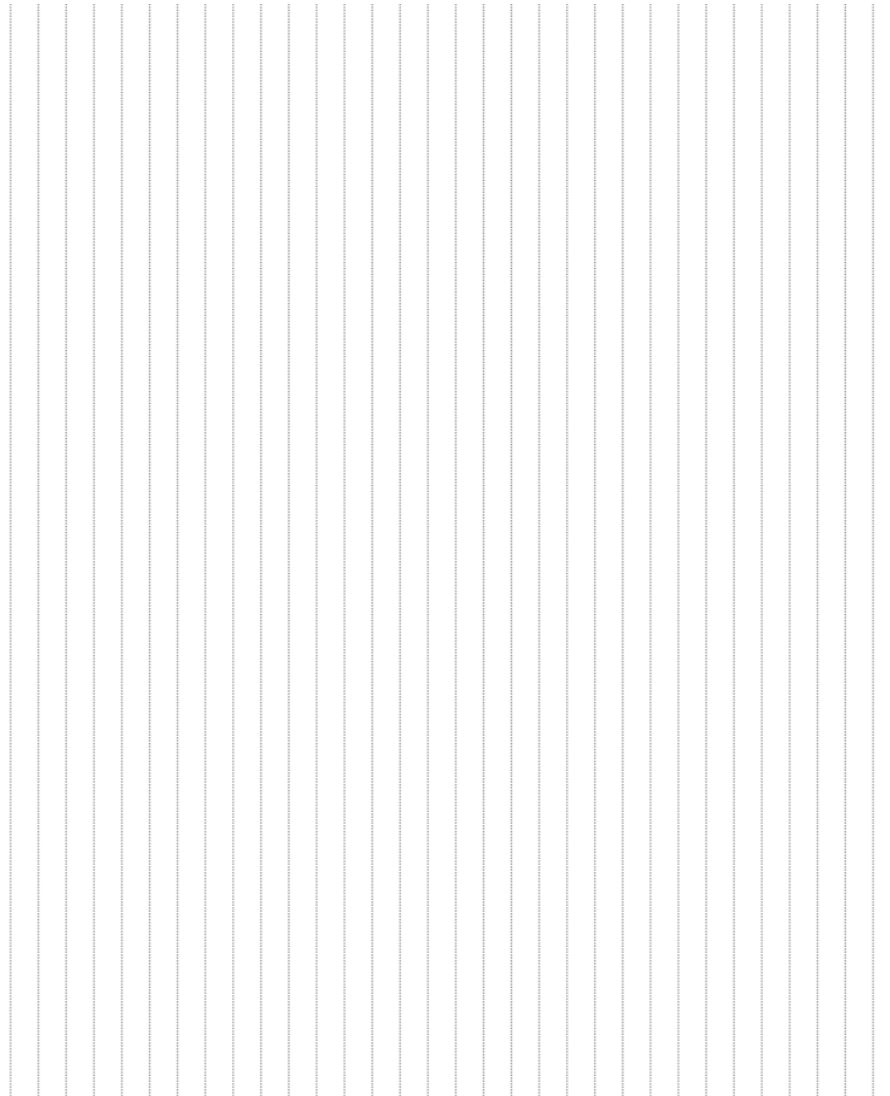


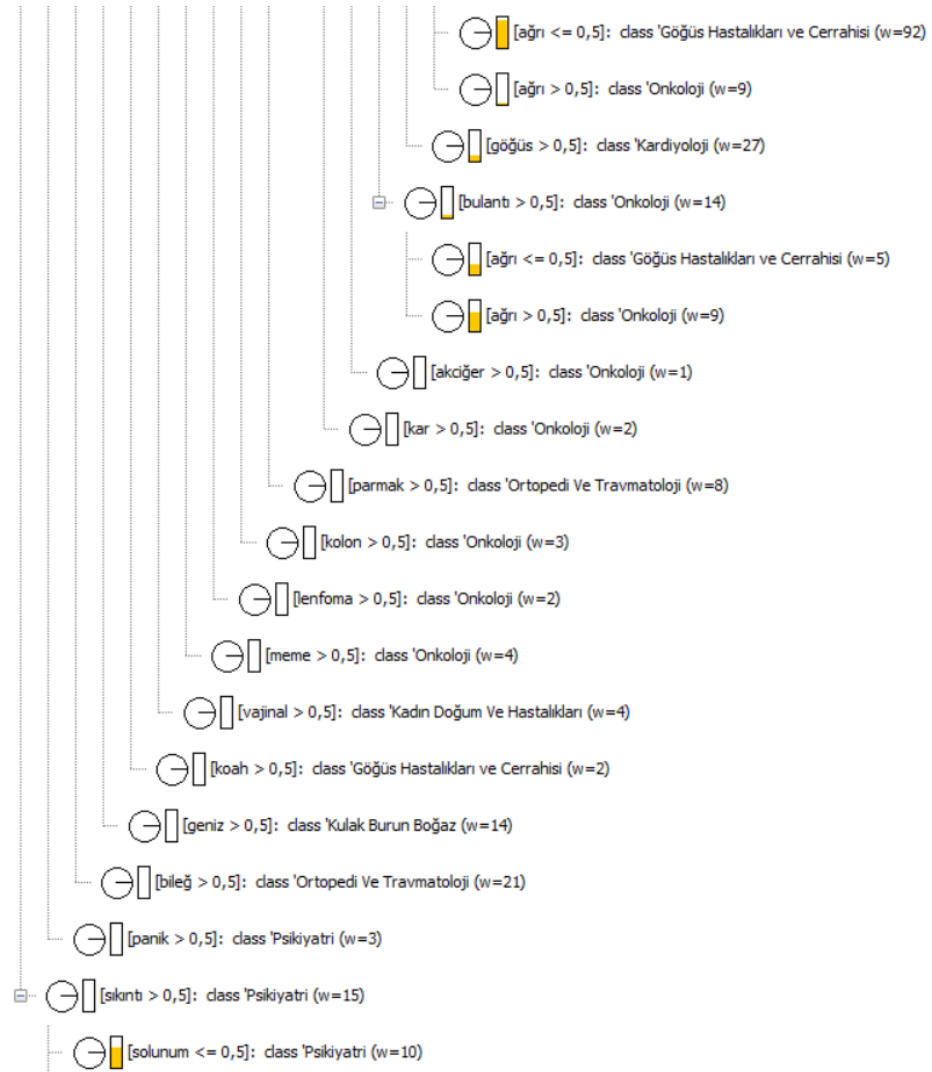


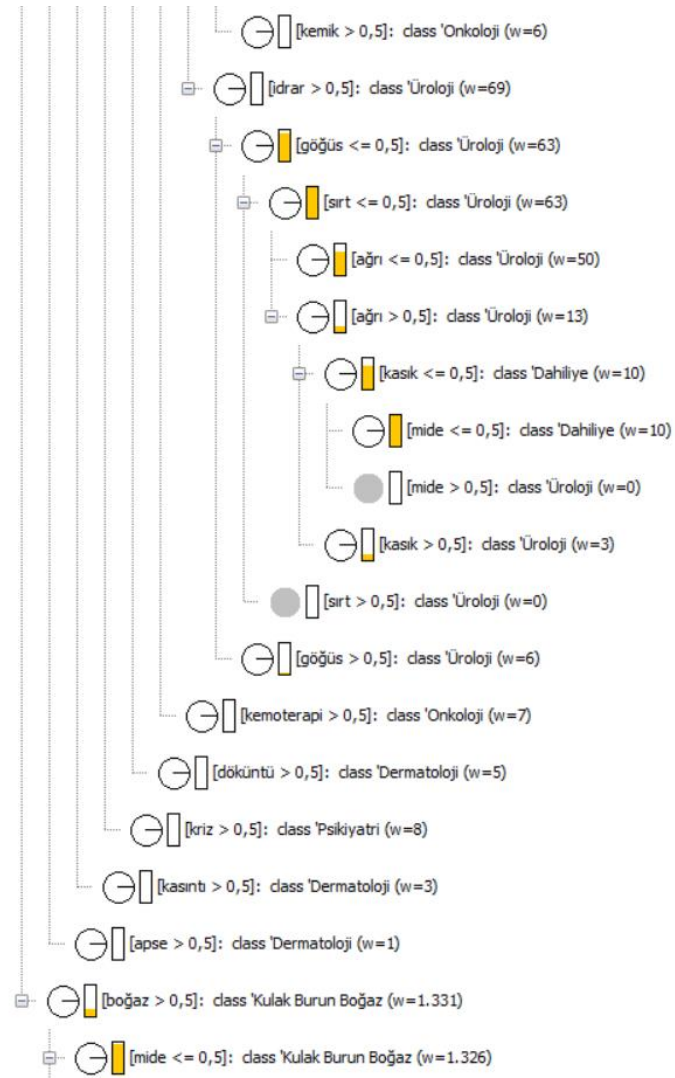


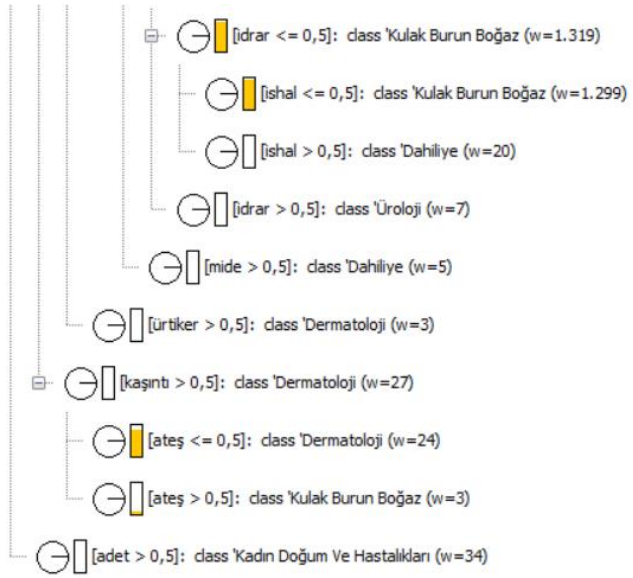


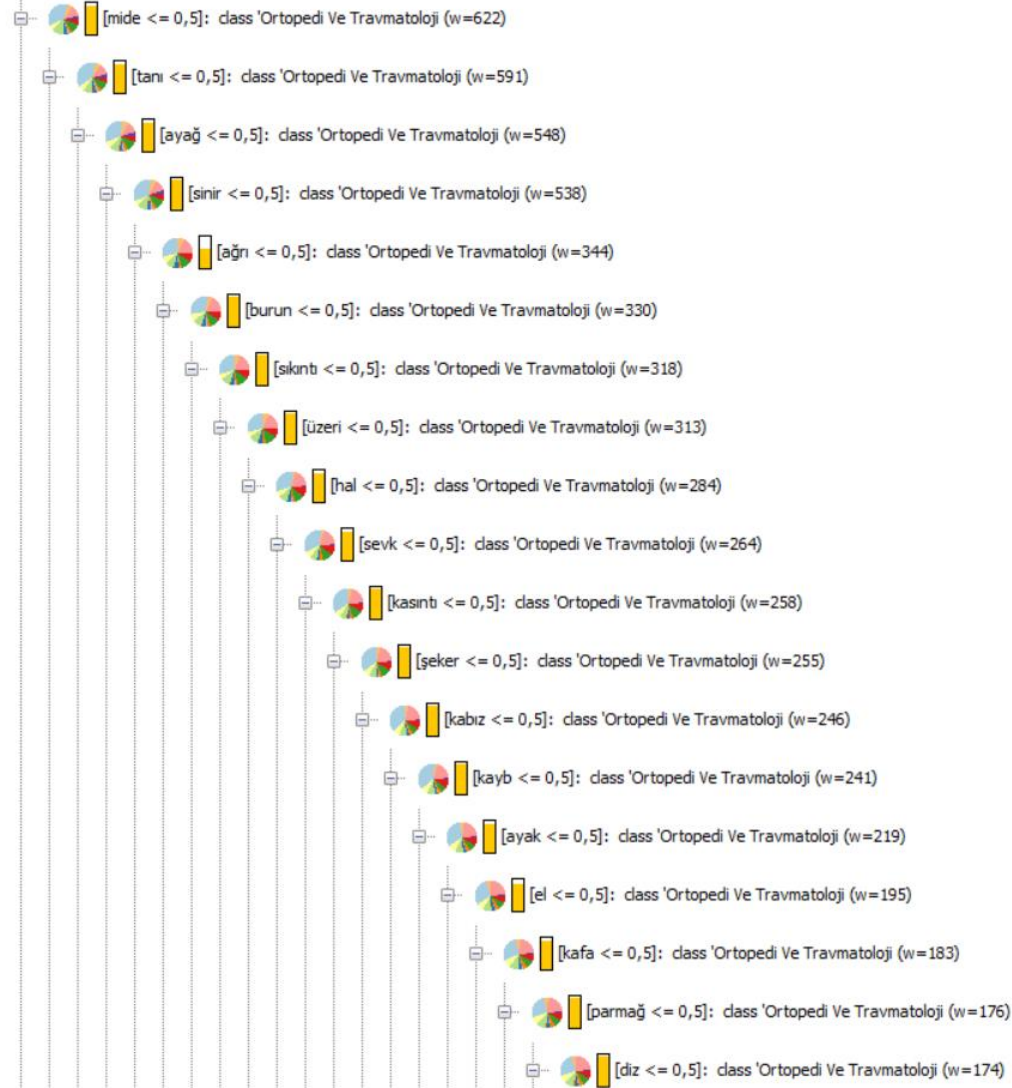


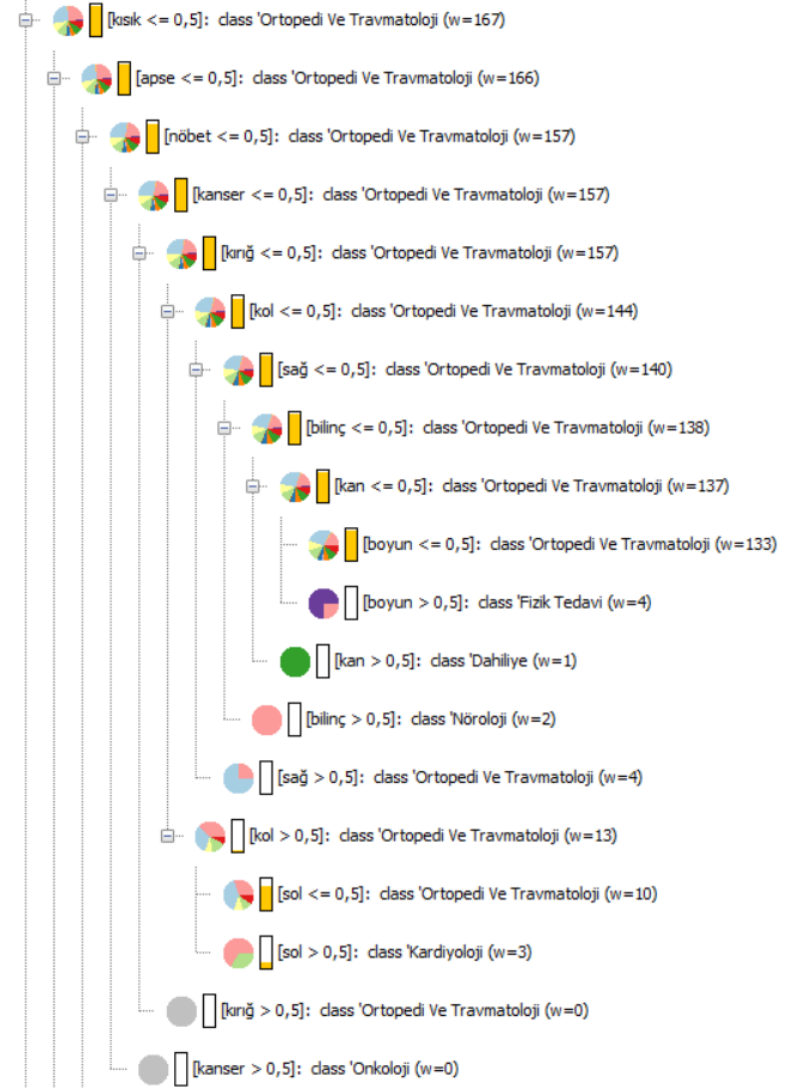
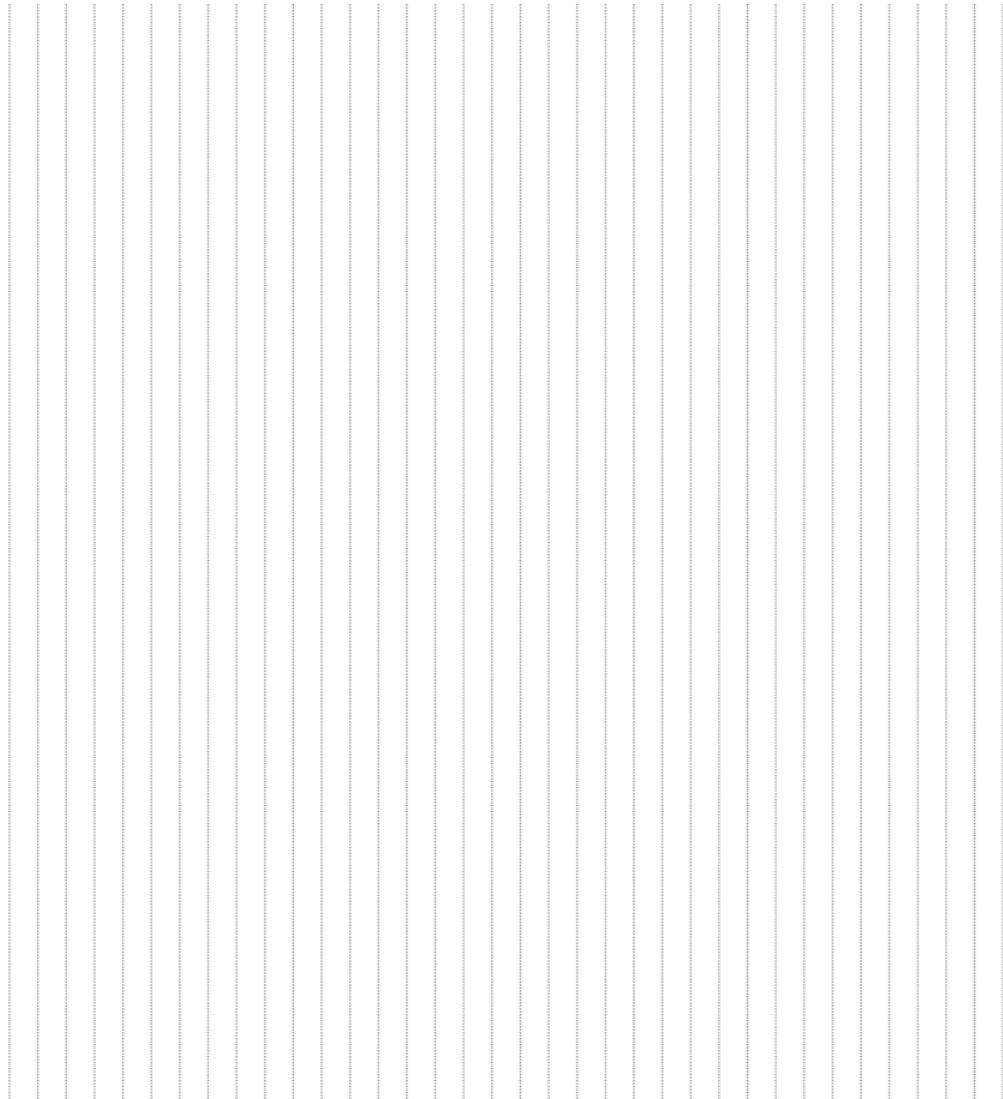




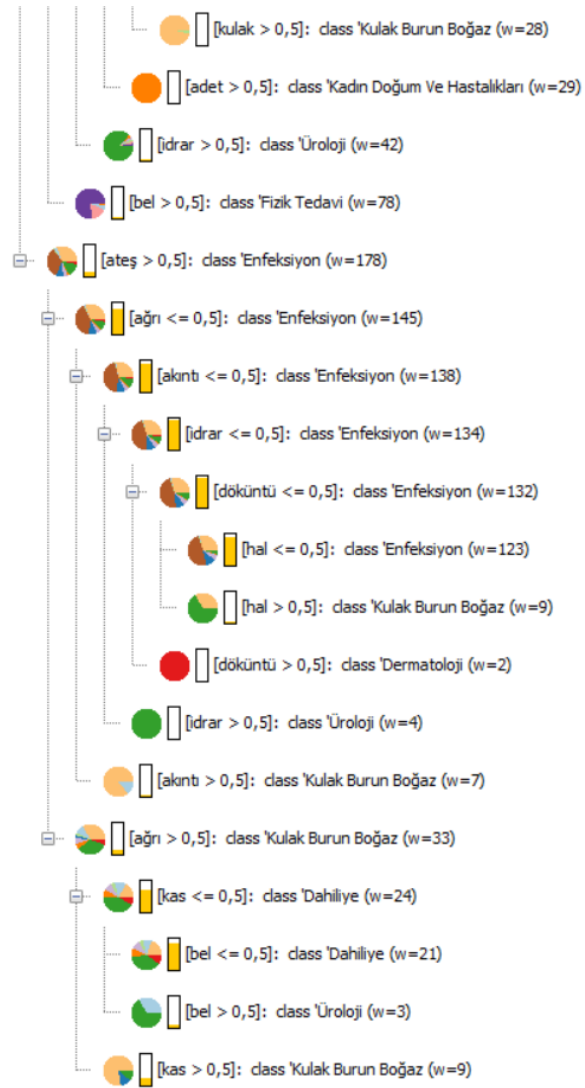


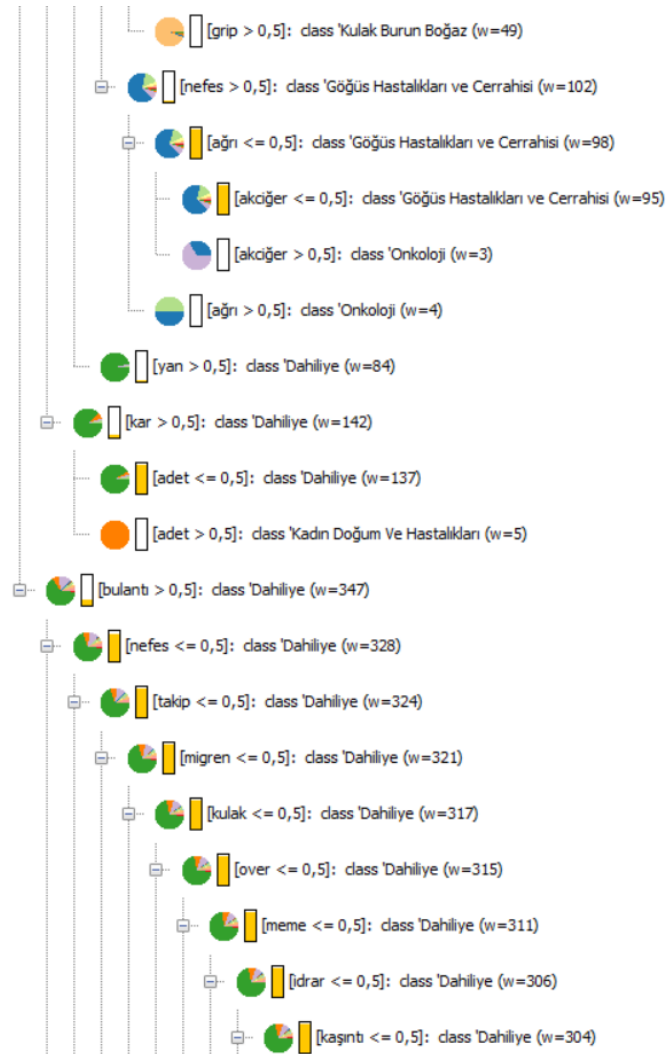


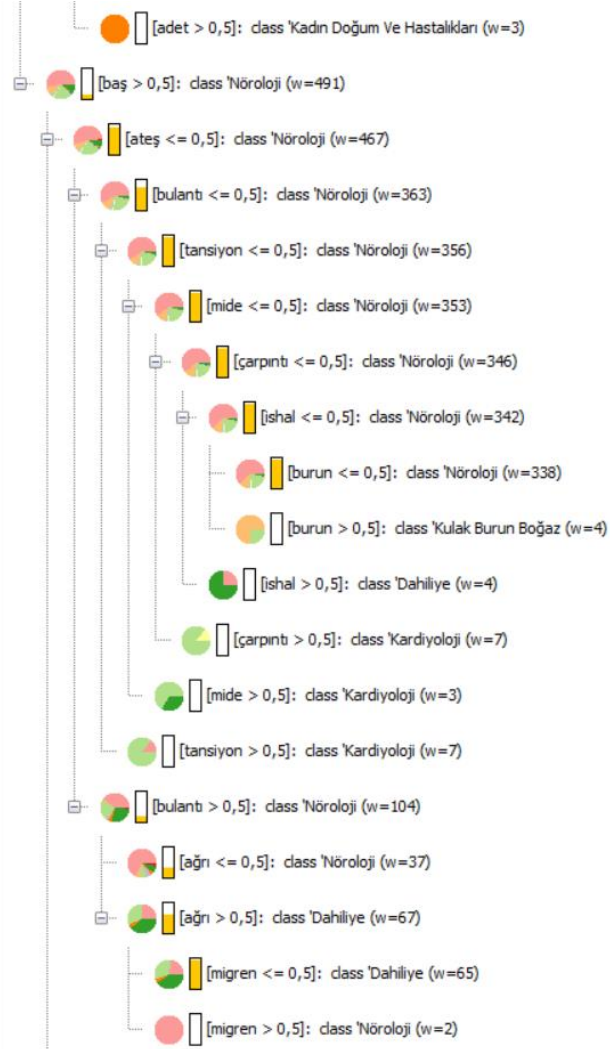


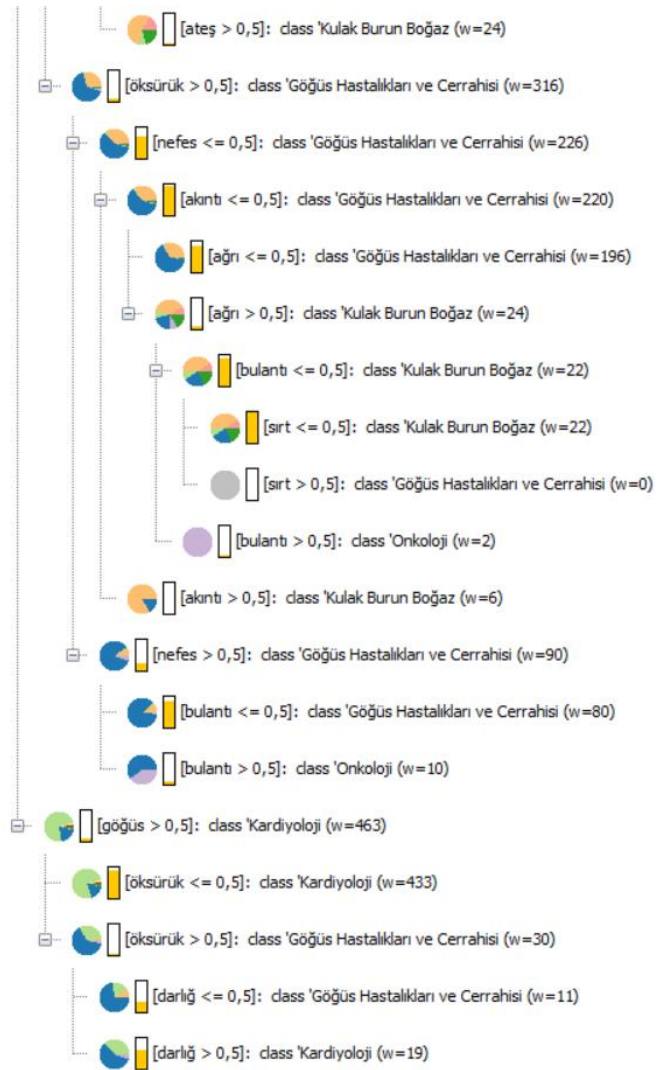


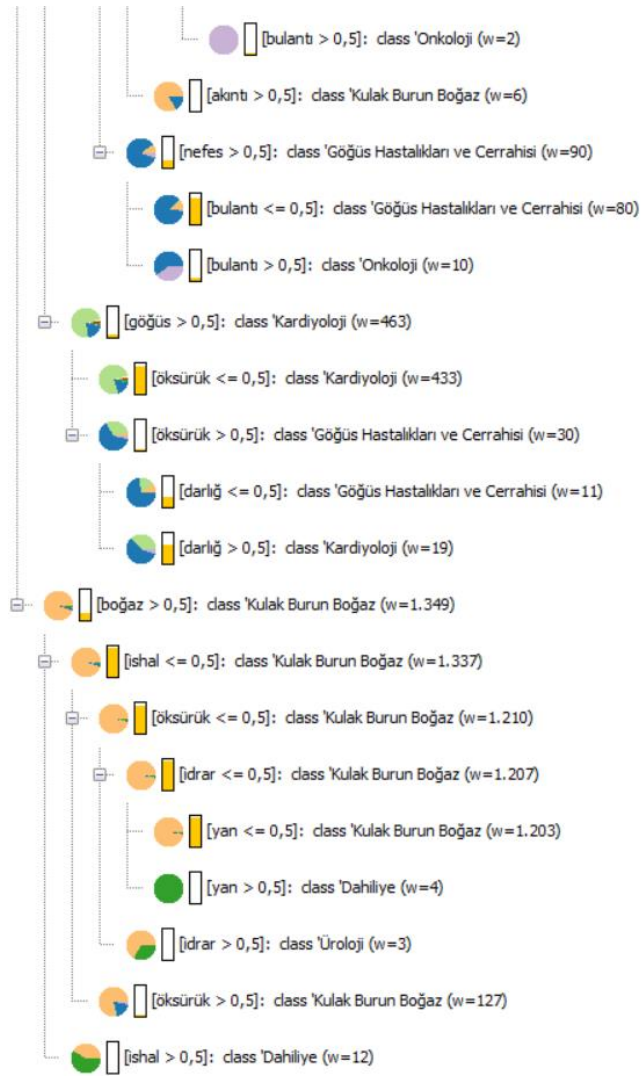












Ek-4

7.10.2019

VM MEDICALPARK HASTANESİ BAŞHEKİMLİĞİNE

Hastanenede Yatan Misafir Hizmetleri Yetkilisi olarak görev yapmaktayım. Yüksek lisans tezim için hastaların ICD kodları ve gittikleri polikliniklerin verilerini son 5 yıllık olarak, kimlik bilgileri veya hasta mahremiyetini ihlal edecek olan veriler olmadan temin etmek istiyorum.

Gereğinin yapılmasını arz ederim.

Müberra TERZİ

Yatan Misafir Hizmetleri Yetkilisi

