



**MASTİTİS HASTALIĞININ TEŞHİSİ İÇİN GUGUK KUŞU ALGORİTMASI
KULLANILARAK KURAL TABANLI SINIFLANDIRMA ALGORİTMASI
GELİŞTİRİLMESİ**

FADİME DEMİRTAŞ

YÜKSEK LİSANS TEZİ

MEKATRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

Dr. Öğr. Üyesi Kenan ZENGİN

Haziran - 2018

Her hakkı saklıdır

**T.C.
GAZIOSMANPAŞA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MEKATRONİK MÜHENDİSLİĞİ ANA BİLİM DALI**

YÜKSEK LİSANS TEZİ

**MASTİTİS HASTALIĞININ TEŞHİSİ İÇİN GUGUK KUŞU ALGORİTMASI
KULLANILARAK KURAL TABANLI SINIFLANDIRMA ALGORİTMASI
GELİŞTİRİLMESİ**

FADİME DEMİRTAŞ

**TOKAT
Haziran - 2018**

Her hakkı saklıdır



Bu tez çalışması;

**Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından
116O332 nolu proje ile desteklenmiştir.**

Fadime DEMİRTAŞ tarafından hazırlanan “**Guguk Kuşu Algoritması ile Bulanık Mantığa Dayalı Kural Tabanlı Bir Sınıflandırma Algoritması ile Mastitis Hastalığı Tespiti**”adlı tez çalışmasının savunma sınavı 28 HAZİRAN 2018 tarihinde yapılmış olup aşağıda verilen Jüri tarafından Oy Birliği / Oy Çokluğu ile Gaziosmanpaşa Üniversitesi Fen Bilimleri Enstitüsü MEKATRONİK MÜHENDİSLİĞİ ANA BİLİM DALI'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

İmza

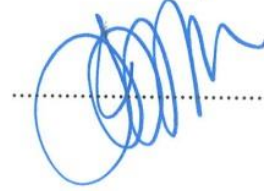
Danışman
Dr. Öğr. Üyesi Kenan ZENGİN
Gaziosmanpaşa Üniversitesi


.....

Üye
Prof. Dr. Sefa TARHAN
Gaziosmanpaşa Üniversitesi


.....

Üye
Dr. Öğr. Üyesi Hidayet TAKÇI
Cumhuriyet Üniversitesi


.....

ONAY


Prof. Dr. Ebubekir ALTUNTAŞ
Fen Bilimleri Enstitüsü Müdürü
26/07/2018


TEZ BEYANI

Tez yazım kurallarına uygun olarak hazırlanan bu tezin yazılmasında bilimsel ahlak kurallarına uyulduđunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduđunu, tezin içerdđi yenilik ve sonuçların başka bir yerden alınmadıđını, kullanılan verilerde herhangi bir tahrifat yapılmadıđını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadıđını beyan ederim.

FADİME DEMİRTAŞ

28 Haziran 2018

ÖZET

YÜKSEK LİSANS TEZİ

MASTITİS HASTALIĞININ TEŞHİSİ İÇİN GUGUK KUŞU ALGORİTMASI KULLANILARAK KURAL TABANLI SINIFLANDIRMA ALGORİTMASI GELİŞTİRİLMESİ

FADİME DEMİRTAŞ

GAZİOSMANPAŞA ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

MEKATRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

(TEZ DANIŞMANI:DR. ÖĞR. ÜYESİ KENAN ZENGİN)

Son yıllarda doğadan esinlenme yoluyla, problem uzayı içerisinde çözüm arayan ve yerel optimumlara takılma oranı azaltılmış başarılı sürü zekalı optimizasyon algoritmaları geliştirilmiştir. Guguk kuşu algoritması bu anlamda bu tür algortimalar için oldukça iyi bir örnektir. Optimizasyon algoritmaları genellikle analitik çözümü zor olan problemler üzerinde uygulandığından sınıflandırma gibi makine öğrenmesi yöntemlerinde de en iyileme amaçlı kullanılmaktadır. Bu çalışmada Guguk Kuşu Algoritması kullanılarak kural tabanlı bir sınıflandırma algoritması geliştirilmiştir. Geliştirilen algoritma süt sığırlarında görülen, süt üretiminde büyük ekonomik zararlara yola açan ve tesbit edilmesi klasik yöntemlerle oldukça zor olan mastitis hastalığının erken dönemde teşhisi amaçlı kullanılmıştır. Oluşturulan algortimanın C#.NET programlama dili kullanılarak yazılımı gerçekleştirilmiştir. Yazılım mastitis hastalığı teşhisi için bilgisayarlı ortamda süt üretimi yapılan bir işletmeden elde edilen gerçek veri seti üzerine uygulanmıştır. Elde edilen deneysel sonuçlarda geliştirilen sınıflandırma algoritması kullanılarak mastitis hastalığının erken dönemde teşhisi için oldukça başarılı ve güvenilir bir tahminleyici model elde edilmesi sağlanmıştır.

2018, 46 SAYFA

ANAHTAR KELİMELER: Veri Madenciliği, Kural Tabanlı Sınıflandırma, Optimizasyon, Guguk Kuşu Optimizasyon Algoritması

ABSTRACT

MASTER THESIS

DEVELOPMENT OF RULE-BASED CLASSIFICATION ALGORITHMS BY USING CUCKOO OPTIMIZATION ALGORITHM FOR DIAGNOSING MASTITIS PATIENT'S DISEASE FADİME DEMİRTAŞ

**GAZIOSMANPASA UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

DEPARTMENT OF MECHATRONICS ENGINEERING

SUPERVISOR: ASST. PROF. DR. KENAN ZENGİN

In recent years, successful swarm optimization algorithms have been developed, which have reduced of getting stuck to local optimums rate by searching from within the problem space through natural inspiration. The Cuckoo Bird Algorithm is a good example for this kind of algorithms. Optimization algorithms are also used for optimization in machine learning methods such as classification because they are applied on problems that are difficult to solve analytically. In this study, a rule-based classification algorithm has been developed using the Cuckoo Bird Algorithm. The algorithm developed was used for the early detection of mastitis disease in dairy cattle, which leads to large economic losses in milk production and is difficult to detect with conventional methods. The generated algorithm was implemented by using the C # .NET programming language. The software has been applied to the actual data set obtained from a computer operated dairy farm for the diagnosis of mastitis disease. By using the classification algorithm developed in the experimental results obtained, it is possible to obtain a highly successful and reliable predictor model for the early detection of mastitis disease.

2018, 46 PAGE

KEYWORDS: Data Mining, Rule Based Classification, Optimisation, Cuckoo Optimization Algorithm

ÖNSÖZ

Bu tez çalışmasında literatürde sıkça kullanılan kural tabanlı sınıflandırma algoritmalarına bir yenisini eklemek amaçlandı. Optimizasyon algoritmalarından biri olan guguk kuşu optimizasyon algoritması kullanılarak yeni bir sınıflandırma algoritması geliştirildi.

Tez konusunu seçerken ve tez aşamasında yardımını esirgemeyen tez danışmanım Dr. Öğr. Üyesi Kenan Zengin' e, TÜBİTAK projesine beni dahil ederek bir proje sürecinin nasıl işlediğine dair fikir sahibi olmamı sağlayan Prof. Dr. Sefa Tarhan'a ve tüm eğitim hayatım boyunca maddi ve manevi yardımlarını benden esirgemeyen aileme teşekkürlerimi sunarım.

FADİME DEMİRTAŞ

28 Haziran 2018

İÇİNDEKİLER

Sayfa

ÖZET	i
ABSTRACT.....	ii
ÖNSÖZ	iii
İÇİNDEKİLER	iv
SİMGE VE KISALTMALAR	v
ŞEKİL LİSTESİ.....	vi
ÇİZELGE LİSTESİ.....	vii
1. GİRİŞ	1
2. KAYNAK ÖZETLERİ	4
3. SINIFLANDIRMA ve YAPAY ZEKA OPTİMİZASYON ALGORİTMALARI.....	8
3.1. Veri Madenciliği.....	8
3.2. Sınıflandırma.....	9
3.2.1. Kural tabanlı sınıflandırma algoritmaları	12
3.2.2. Metasezgisel yöntemler.....	15
3.3. Kümeleme	17
3.4. Birliktelik Kuralları.....	18
3.5. Yapay Zeka Optimizasyon Algoritmaları	19
3.5.1. Guguk kuşu optimizasyon algoritması	20
4. METARYEL ve YÖNTEM.....	25
4.1. Hayvan Metaryeli	25
4.2. Süt Örnekleri.....	25
4.3. Laboratuvar İşlemleri	26
4.4. Kullanılan Veri Setleri.....	31
4.5. Geliştirilen Algoritma.....	35
5.BULGULAR veya BULGULAR VE TARTIŞMA	37
6. SONUÇ.....	41
7. KAYNAKLAR	43
8. ÖZGEÇMİŞ.....	46

SİMGELER ve KISALTMALAR

Simgeler

Açıklama

Kısaltmalar

Açıklama

CMT

California Mastitis Testi

GA

Genetik Algoritma

GOA

Guguk Kuşu Optimizasyon Algoritması

PSO

Parçacık Sürü Optimizasyon Algoritması

YY

Yumurta Yarışması

ŞEKİL LİSTESİ

<u>Şekil</u>	<u>Sayfa</u>
Şekil 3.1. Veri madenciliği süreci.....	9
Şekil 3.2. Metasezgisel yöntemler	16
Şekil 3.3. Guguk kuşlarının yaşam alanları	22
Şekil 3.4. GOA akış diyagramı.....	24
Şekil 4.1. İşletmenin bilgisayarlı sürü yönetim sistemi	25
Şekil 4.2. Süt numuneleri.....	25
Şekil 4.3. Hasta hayvanın elektriksel iletkenlik ölçüm değerleri	28
Şekil 4.4. Sağlıklı hayvanın elektriksel iletkenlik ölçüm değerleri	29
Şekil 4.5. Hazırlanan mastitis veri seti	32
Şekil 4.6. Iris veri setinden bir görüntü.	33
Şekil 4.7. Zoo veri setinden bir görüntü	34
Şekil 4.8. Önerilen yöntemin akış diyagramı	36
Şekil 5.1. Iris veri setinin geliştirilen algoritmaya uygulanışı.	37
Şekil 5.2. Zoo veri setinin geliştirilen algoritmaya uygulanışı	38
Şekil 5.3. Mastitis veri setinin geliştirilen algoritmaya uygulanışı.....	39
Şekil 5.4. Geliştirilen yöntem ve diğer yöntemlerin karşılaştırılması.....	40

ÇİZELGE LİSTESİ

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 1.1. Mastitis erken teşhis durumunda iyileşme oranları	1
Çizelge 3.1. Karışıklık matrisi	13
Çizelge 5.1. Weka programında Iris veri seti için elde edilen sonuçlar	38
Çizelge 5.2. Weka programında Zoo veri seti için elde edilen sonuçlar	39
Çizelge 5.3. Weka programında Mastitis veri seti için elde edilen sonuçlar.....	40

1. GİRİŞ

Hayvanlarda çok çeşitli meme hastalıkları görünmektedir. Bunlardan meme bezi yangısı olarak bilenen mastitis, süt hayvancılığında en çok rastlanan ve hayati öneme sahip bir meme hastalığıdır. Mastitis çok etmenli bir hastalık olup özellikle bakterilerin etkisi büyüktür (Philpot ve Nickerson, 2000). Bakteriler yanında mikoplazma, mantar, maya ve klamidyalar gibi etiyolojik etkenlerde hastalık oluşumuna neden olmaktadır. Hayvan yaş ve ırkı, hava koşulları ve mevsimsel etkileri, sağım ortamının hijyeni, stres, beslenme, memenin anatomik ve fizyolojik özellikleri gibi faktörler de hastalık oluşumunda etkili olmaktadır (Şeker ve ark., 2000; Cedden ve ark., 2002; Rişvanlı ve Kalkan, 2002).

Hayvancılıkta çok sık karşılaşılan mastitis hastalığı süt üretiminde azalmaya, tedavi masraflarındaki ciddi artışa, hayvanların kesilme yüzdesinin artmasına ve ölümlere neden olduğu için sığır yetiştiriciliğinde büyük oranda ekonomik zararlara neden olmaktadır (Sabuncuoğlu ve Çoban, 2006). Türkiye’de mastitisten kaynaklanan, hayvancılık işletmelerinin yıllık ekonomik kayıplar milyonlarla ifade ediliyor. Türkiye’de bu hastalık nedeniyle ortaya çıkan ekonomik zararın yılda 41,5 milyon TL olduğu öngörülmektedir (Atasever ve Erdem, 2008). Mastitis, etkilediği her hayvanda süt üretimini olumsuz yönde etkilemektedir. Hastalık sütün kimyasal ve fiziksel yapısında önemli değişiklikler meydana getirir. Mastitis hastalığı nedeniyle süt veriminde gerçekleşen azalma, sürünün yenilenmesi, tedavi ve veteriner hekim giderleri, süt kalitesinde düşüş ve ölüm nedeniyle kayıplar meydana gelmektedir (Memmedova, 2012).

Mastitis hastalığı klinik ve subklinik mastitis olmak üzere ikiye ayrılmaktadır. Memenin yapısında ve sütte değişikliklerin gözle görülecek şekilde olduğu klinik mastitis, memede fiziksel değişikliklerin (şişme, kızarıklık vb.) yanında halsizlik, kilo kaybı gibi hastalık belirtileri ile kendini göstermektedir. Subklinik mastitis meme dokusunda gözle görülür belirtiler oluşturmadığı için saptanması daha zordur (Özyurtlu, 2011). Süt sığırcılığı işletmelerinde sıklıkla rastlanan bu hastalık yüksek ekonomik kayıplara neden olmaktadır. Genellikle mastitislerin %70’ni subklinik mastitis oluşturmakta ve süt oranında %3 – 26 oranında süt kaybına neden olduğu belirtilmektedir (Yeşilmen ve

ark., 2012). Bu bağlamda subklinik mastitisin erken dönem de tespiti sayesinde süt üretiminde kalite artırılabilir ve sütçülük işletmelerinde ekonomik kayıpların mümkün oldukça en aza indirilmesi sağlanabilir. Mastitise erken müdahale etme durumunda elde edilen iyileşme oranları Çizelge 1.1' de gösterilmiştir.

Çizelge 1.1. Mastitise erken müdahalenin süresine göre elde edilen iyileşme oranları (Anonim, 2016)

6 saatten az	% 86.5
7-12 saat arası	% 81
13-24 saat arası	% 69
24 saatten fazla	% 47

Büyük hayvan işletmelerin de mastitis hastalığına karşı erken önlemler almak amacıyla bilgisayar ortamında veriler doğrultusunda çok hızlı bir şekilde tanı koyabilmek oldukça önemlidir.

Veri kümeleri birden fazla değişkenin olduğu ve her değişkenin birden fazla değer aldığı yapılardır. Veri kümelerinin karmaşıklığı nedeniyle veriden bilgi edinmek zor bir süreçtir. Veriden bilgi edinmek veriyi anlamlandırmak için ondan kural çıkarımı yapmak oldukça önemlidir. Sınıflandırma kural madenciliği ile kolayca anlaşılabilir kapsamlı kurallar çıkarılmaya çalışılmaktadır.

Son yıllarda doğadan esinlenerek oluşturulan sürü zekalı optimizasyon algoritmalar oldukça popülerdir. Bu tez çalışması içerisinde sınıflandırma kural madenciliği daha önce bu problem kapsamında kullanılmamış sürü zekalı yapay zekâ optimizasyon algoritmalarından olan guguk kuşu optimizasyon algoritması ile gerçekleştirilmiştir. Veri seti içerisinde kapsayıcı kurallar bulunmaya çalışılmıştır.

Bu tez çalışmasının amacı, süt sığırcılığında hayati öneme sahip ve süt verimini etkileyen mastitis hastalığının erken dönemde tespiti için, sezgisel algoritmalarından biri olan guguk kuşu optimizasyon algoritması kullanılarak kural tabanlı bir sınıflandırma algoritması geliştirmek ve bu sayede bir karar destek mekanizmasını oluşturmaktır. Dolayısı ile ilk önce mastitis hastalığının erken dönem tespitinde kullanılabilecek

değişkenler belirlenmiş ve bunların etki oranları tespit edilmiştir. Mastitisin tespitiyle ilgili çalışmalarda, literatürde, çoğunlukla sütün iletkenliği, sütteki somatik hücre sayısı ve sütün akış hızı girdi değişkenleri olarak kullanıldı. Bu çalışmada ise bu değişkenlerden sadece süt iletkenliği kullanılmış olup bununla beraber hayvanların günlük su tüketimleri ve iklim değerleri de girdi değişkeni olarak kullanıldı.

Parametrelerin tesbitinin ardından ikinci hedef olarak guguk kuşu optimizasyon algoritması kullanılarak kural tabanlı bir sınıflandırma algoritması geliştirildi, mastitis teşhisi için karar destek mekanizmasının alt yapısı oluşturuldu.

Üçüncü hedef olarak geliştirilen karar destek mekanizmasının tahmin başarısı test edilerek eğitilmiş bir model kural seti elde edildi.

2. KAYNAK ÖZETLERİ

Bulanık mantık yöntemi literatürde bir çok alanda kullanılmıştır. Bu alanlardan biri olan süt sığırcılığı alanında da sınıflandırma amacıyla kullanılabilir. Mastitisin sınıflandırılması amacıyla; somatik hücre sayısı, elektriksel iletkenlik ve süt verimi gibi parametreler kullanılarak bulanık mantık yöntemi ile bir sistem geliştirilmiştir. Araştırma sonucunda yanlış uyarıların sayısında azalma olduğu ve dolayısıyla oluşturulan modelin pratik uygulamalarda kullanımının uygun olduğu görülmüştür. Çalışmanın başarı sonucu %99 'dur (De mol ve Ouweltjes, 2001). Cavero ve ark. (2006), bulanık mantık tabanlı bir uzman sistem modeli ile mastitis teşhisi çalışmasını sütün elektriksel iletkenliği, süt sağım hızı gibi parametreleri ile gerçekleştirmişlerdir. Çalışma hassaslık, belirlilik ve hata oranlarına göre değerlendirilmiştir. Değerlendirme sonucunda başarı performansı ise %41.9 ile % 95.5 arasında değişkenlik gösterdiği saptanmıştır. Çiğ süt kalitesinin tespiti için bulanık mantık modeli yöntemiyle mikrobiyolojik açıdan değerlendirildiği bir başka çalışmada girdi değişkenleri olarak başta somatik hücre sayısı olmak üzere yağlı ve yağsız kuru madde yüzdesi de kullanılmıştır. Bulanık model sonuçları bu konuda uzman olan bir kişinin verdiği hükümler ile karşılaştırıldığında %82.5 oranında bir başarı görülmüştür (Mahreban ve ark., 2012).

Yapay sinir ağları kullanılarak mastitisin erken dönemde teşhisi konusunda çeşitli çalışmalar yapılmıştır. Yapay sinir ağları kullanılan başka bir çalışmada mastitise etki eden faktörler arasında hayvanların laktasyon sayıları, günlük süt verimleri ile beraber somatik hücre sayısının da önemli faktörler arasında yer aldığı tespit edilmiştir (Yang ve ark., 2000). Krieter ve ark. (2007), yapay sinir ağlarını kullanarak kurdukları modelde, parametre olarak süt sağımı ile ilgili parametreler, elektrik iletkenliği ve laktasyondaki gün sayısı kullanılmıştır. Araştırma sonucunda mastitis teşhisinde ve başarı oranı ise %51.3 ile % 80.5 arasında değişkenlik göstermiştir.

Veri madenciliği yöntemlerinden olan sınıflandırılma, kural keşfi için en çok kullanılan yöntemlerdendir. Sınıflandırma kural madenciliğinde kullanılan optimizasyon algoritmalarından biri genetik algoritmadır. Kuralların kapsama özelliği dikkate alınarak uygunluk fonksiyonu oluşturulup üç veri seti üzerinde uygulanmış ve optimal bir süre

içerisinde kural keşfi yapılmıştır (Al-Maqaleh, 2012). Genetik algoritma kullanılarak yapılan bir diğer çalışmada ise uygunluk fonksiyonunun kapsam özelliğine ek olarak anlaşılabilirlik ve doğruluk gibi özelliklere sahip olabilmesi için çalışılmıştır. Genetik algoritma bileşenleri olan kodlama, genetik operatörler ve uygunluk fonksiyonu üzerinde durulmuştur. Önerilen algoritma dört veri seti üzerinde uygulanmış ve karar ağaçları yöntemi ile kıyaslanıp nispeten daha kapsayıcı kurallar bulunmuştur (Al-Maqaleh ve Shahbazkia, 2012).

Genetik algoritma kullanılarak yapılan bir kural keşfi çalışmasında kuralların rastgele dağılıp doğruluk oranında ki sapmaların önüne geçmek hedeflenmiştir. Başlangıç popülasyonun rastgele oluşturulması durumunda bireyler çözüme uzakta olabilir veya yerel noktalardan kurtulamayabilir. Gerçekleştirilen çalışmada başlangıç popülasyonu genelleştirilmiş tek-düze popülasyon yöntemi ile oluşturulup rassallık yöntemi olarak ortadan kaldırılmıştır. Sapmaların önüne geçen bu yöntem ile kromozomların gen aralıkları bilinmekte ve o aralığın alt ve üst değerlerine sahip iki kromozom oluşturulup popülasyon bu iki değer arasında oluşturulmaktadır (Gündoğan ve ark., 2004).

Veri madenciliği alanında karşılaşılan en önemli problemlerden biri olan sınıflandırma kurallarının keşfi işlemi için metasezgisel algoritmaların kullanımına sıkça rastlanmaktadır. Bunlardan sürü tabanlı algoritmalarından biri olan karınca kolonisi algoritması kullanılarak yapılan bir çalışmada “ayır-ve-yönet” mantığı kullanılarak her iterasyonda tek bir karınca koşturulması yöntemi ile iki veri seti üzerinde başarılı sonuçlar elde edilmiştir (Alataş ve Akın, 2004). Bir diğer sürü tabanlı algoritma olan kedi sürüsü optimizasyon algoritması kurallar üzerinde budama işlemi yapılmadan uygulanmıştır. Nümerik veritabanlarının da otomatik sınıflandırma kuralları keşfi için yapılan çalışma dört veri tabanında uygulanmış ve Weka programından elde edilen sonuçlar ile kıyaslanmıştır (Akyol ve Alataş, 2016).

Sınıflandırma kural madenciliği problemi kapsamında uygulanan algoritmalarda ki uygunluk fonksiyonun doğru seçimi çok önemlidir. Farklı kriterler ve parametreler eklenerek uygunluk fonksiyonu üzerinde iyileştirmeler yapmak maliyet üzerinde pozitif etki oluşturacaktır. Yapay kimyasal reaksiyon optimizasyon algoritması kullanılarak kural madenciliği yapılan bir çalışmada uygunluk fonksiyonuna anlaşılabilirlik ve

kapsam özelliğinin yanı sıra ilginçlik kriteri de eklenmiştir. Bu kriter sayesinde beklenmedik ve ilginç kurallar bulunmaya çalışılmıştır. Çalışma iki veritabanı üzerinde uygulanmış ve bulunan sonuçlar genetik algoritma, parçacık sürü ve karınca kolonisi optimizasyon algoritmaları ile kıyaslanarak nispeten daha yüksek tahmin oranları elde edilmiştir (Alataş ve Özer, 2017).

Kültürel algoritma kullanılarak kural keşfi problemi üzerine bir optimizasyon çalışmasında popülasyon rastgele belirlenmiştir. Çaprazlama oranı %80 ve mutasyon oranı %20 alınarak bulunan sonuçların eğitim setinde %50 ile %78 arasında değiştiği saptanmıştır (Srinivasan ve Ramakrishnan, 2012). Parçacık sürü optimizasyon algoritması kullanılarak yapılan bir çalışmada ise veri kümesi üzerinde başlangıçta genetik görelî indirgeme algoritması kullanılarak veri ön işleme uğramıştır. Çalışma az sayıda veri içeren data setlere uygulanarak doğruluk oranı olarak %85.19 değerine ulaşılmıştır (Sahebi ve Ebrahimi, 2015).

Sınıflandırma kurallarının keşfi insan sağlığı açısından önemli birçok hastalığın tespiti üzerinde kullanılmıştır. Bu hastalıklardan biri olan diyabetin erken dönemde tespiti üzerine yapılan bir çalışmada sosyal tabanlı sezgisel optimizasyon algoritması kullanılmıştır. Bu çalışmada kurallar üzerinde herhangi bir ön işlem yapılmadan değişken sayısı dikkate alınarak bir başlangıç popülasyonu oluşturulmuştur. Diyabet veri seti üzerinde algoritma %79.06 doğruluk oranı göstermiştir (Kızılloluk ve Alataş, 2015). Diyabet üzerine yapılan bir başka çalışmada ise kural madenciliği örümcek maymunu optimizasyon algoritması kullanılarak yapılmıştır. Örümcek maymunlarının doğada parçalara ayrılıp yiyecek bulma davranışlarından esinlenilerek oluşturulan bu algoritma da grup liderleri vardır ve grubu yönlendirip gıda kaynakları bulmakla yükümlüdür. Bulduğu gıda kaynaklarına göre konum değişikliği yapabilir. Grubun diğer üyeleri de sınırları korumak ve liderleri ile sosyal bağı korumak zorundadırlar. Bu algoritmanın gruplara ayrılma özelliği kullanılarak veri seti parçalara bölünmüş ve en iyi kurallar bu gruplar içerisinde algoritmanın grup lideri özelliği kullanılarak bulunmuştur. Yapılan çalışma, sosyal tabanlı sezgisel optimizasyon algoritma kullanılarak yapılan çalışmaya nazaran daha iyi performans göstermiş ve diyabetin erken dönemde teşhisinde doğruluk oranı olarak %89.84 oranına ulaşmıştır (Cheruku ve ark., 2017).

Bu alıřmada st iřletmelerinde nemli kayıpların nne gemek iin mastitis hastalıęının erken dnemde teřhisinin yapılması amalanmıřtır. alıřmada, veriler zerinde etkili hesaplama yntemleri ile anlamlı iliřkiler oluřturmayı saęlayan veri madencilięi ve yapay zek-optimizasyon algoritmaları kullanıldı. Gerekleřtirilen model ile veri madencilięi sınıflandırma kurallarının otomatikleřtirilmesi ve performansının ykseltilmesi hedeflenmiřtir. Bu baęlamda literatrde sınıflandırma kural madencilięinin optimize edilmesinde daha nce kullanılmayan yeni bir algoritma olan guguk kuřu optimizasyon algoritması kullanılmıřtır.



3. SINIFLANDIRMA ve YAPAY ZEKÂ OPTİMİZASYON ALGORİTMALARI

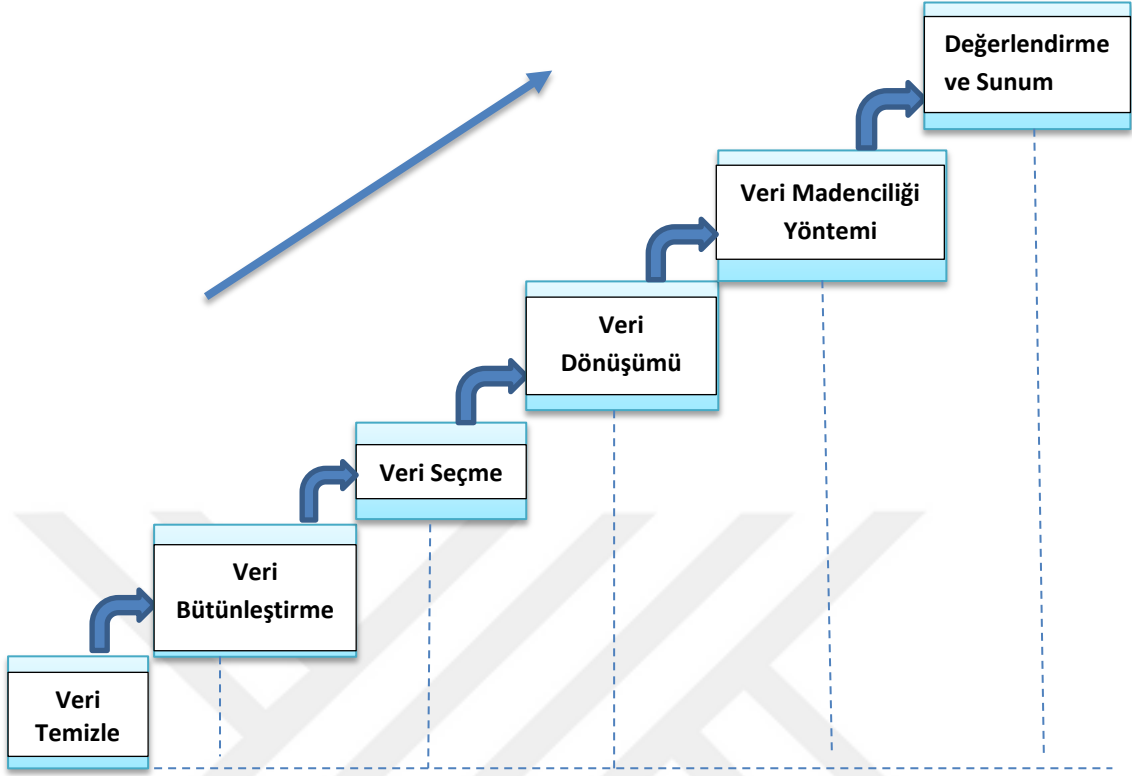
3.1. Veri Madenciliği

Veriler üzerinde anlamlı sonuçlar elde edebilmek için çeşitli istatistiksel ve matematiksel yöntemler kullanılabilir. Ancak veri boyutu arttıkça problemler oluşur. Bu büyük boyutlardaki verileri çözümleyebilmek için veri madenciliği kavramı ortaya çıkmıştır. Veri madenciliği büyük hacimli veri setlerinde istatistik, makine öğrenmesi ve yapay zekâ yöntemleriyle bilginin keşfi sürecidir (Han ve ark., 2011). Teknik olarak veri madenciliği ise veri setleri içerisindeki gizli ve anlamlı örüntüleri bulma işlemidir.

Veri madenciliği günümüzde oldukça yaygın bir kullanım alanına sahiptir. Örneğin; sağlık, pazarlama, banka ve sigorta, web madenciliği, yüzey analizi ve coğrafi bilgi sistemleri gibi birçok alanda kullanılmaktadır.

Veri madenciliğinde bilginin keşfi aşaması oldukça zordur ve bu süreci kolaylaştırmak için aşağıdaki adımlar kullanılmıştır:

- Veri temizleme (Gürültülü ve tutarsız verileri silme)
- Veri bütünleştirme (Farklı türdeki verilerin tek türe dönüştürülmesi)
- Veri seçme (Çözümleme işleminin çok uzun sürmesi durumunda eğer sonucun değişmeyeceğine inanılıyorsa analiz ile ilgili olan verileri belirlemek)
- Veri dönüşümü (Değişkenlerin ortalama ve varyanslarının birbirinden önemli derece de farklı olduğu ve büyük varyanslı değişkenlerin diğerleri üzerinde baskısının fazla olduğu veri setlerinde uygulanır. Bu aşamada veriyi veri madenciliği tekniklerinde kullanılabilir hale dönüştürmek için normalleştirilmesi veya standartlaştırılması gerekir.)
- Veri madenciliği yöntemi (Veri hazır hale geldikten sonra veri modellerini yakalayabilmek için akıllı yöntemlerin uygulandığı önemli bir adım)
- Modeli değerlendirme (Ölçümlere göre elde edilen örüntüleri değerlendirme bu aşamada yapılır)
- Bilgi sunumu (Elde edilen bilginin görselleştirme ve raporlama araçları kullanarak sunumunu gerçekleştirmek) (Han ve ark., 2011).



Şekil 3.1. Veri madenciliği süreci (Han ve ark., 2011)

Veri madenciliği aşaması, veriler üzerinde anlamlı örüntüler elde edebilmek için gerekli olan algoritmanın belirlendiği çok önemli bir aşamadır. Bu bölümde kullanılan akıllı yöntemler üç kısımda incelenmektedir (Özkan, 2008).

- 1- Sınıflandırma
- 2- Kümeleme
- 3- Birliktelik kuralları

3.2. Sınıflandırma

Veri madenciliği alanının önemli konularından olan sınıflandırma; veri seti içerisindeki ortak özelliklere göre verinin dağılımı işlemidir (Özkan,2008). Sınıflandırma da sınıflar ve hangi değerlerin hangi sınıfa ait olduğu bilinmektedir. Yeni verilerin sahip olduğu ortak özellikler kullanılarak hangi sınıfa ait olduğu tespit edilmektedir.

Sınıflandırma işlemini gerçekleştirebilmek için belirli bir süreç takip edilmektedir:

1. Sınıflandırma işleminin en temel aşaması model oluşturma işlemidir. Model oluşturulurken veri setinin öznitelikleri dikkate alınır. Model oluşturmak için kullanılan veri kümesi, öğrenme ve test kümesi olarak ikiye ayrılmaktadır. Bu ayrım çeşitli yöntemler ile yapılabilmektedir. Örneğin; holdout, k-katlı çapraz doğrulama, bootstrap ve birini dışarda bırakan çapraz doğrulama kullanılan yöntemler arasındadır. Model oluşturma işleminde farklı sınıflandırma yöntemleri kullanılabilir. Bunlar algoritmik tanımlama kuralları olan “if-then-else”, karar ağaçları veya matematiksel formüller kullanılarak oluşturulan modeller olabilir.

2. Model değerlendirme: aşaması en az model oluşturma aşaması kadar önemlidir. Modelin başarısı(doğruluğu) test kümesi örnekleri kullanılarak belirlenir. Modelin doğruluğunu, doğru sınıflandırılmış verilerin kümesinin test kümesine oranı olarak belirlenir. (Han, 2011).

Literatürde birçok sınıflandırma algoritması bulunmaktadır. Bunlardan bazıları:

Karar ağaçları: Sınıflandırma kural madenciliği bir hastalığın erken teşhisinde en çok başvurulan ve insan düşünce yapısına en uygun yöntemlerden biridir. Kural madenciliği sayesinde kullanıcıların rahat bir şekilde anlayabilecekleri kurallar çıkarılmaktadır ve bu kurallar sayesinde yeni gelen örnek hakkında fikir üretilebilmektedir. Karar ağaçları var olan sınıflandırma yöntemleri arasında kolay yorumlanabilmesi ve anlaşılabilirliği açısından en bilinen modellerdendir. Genel hatları ile bilginin bir ağaç yapısı üzerinde modellenmesidir. En üst yapı “kök” en son yapı ise yaprak olarak adlandırılmaktadır. Yapraklar genelde sonuç kısmının yer aldığı şekillerdir. Sınıflandırma ağaçlarında genelde entropi hesabı yapılmaktadır (Özkan, 2008). Karar ağaçlarının ID3, C4.5, C5.0 ve CART gibi çeşitleri bulunmaktadır. Bunlar akış şemalarına benzeyen algoritmalarlardır. Üzerinde çalışılan veri setinden kurallar çıkarmayı amaçlamaktadırlar (Alataş ve Özer, 2017).

Karar ağaçları yorumlanması basit yapılardır. Düşük hesaplama karmaşıklığına sahiptirler. Bu nedenle veri setini oldukça kısa bir sürede işleyebilmektedir. Ancak veri setindeki kümelerin saf alt kümelerini belirleme çabası içerisinde bulduklarından

gürültülü veya tutarsız verilere karşı hassas hale gelebilirler. Bu yöntemin bir diğer dezavantajı ise çok fazla dallanma yapması durumunda(overfitting) başarı kriteri düşecektir. Burada karar ağaçları yerine kural tabanlı bir sınıflandırma algoritması kullanımı soruna çare olabilmektedir.

C4.5 ve C5.0 algoritmaları: Quinlan (1993) tarafından oluşturulan ID3 algoritmasının daha ileri bir sürümü olan C4.5 algoritması en yaygın olarak bilinenler arasındadır. ID3 algoritması sadece kategorik öznitelikler için uygulanabilir. C4.5'in geliştirilmesi ile sayısal özniteliklerin de sınıflandırma ağacına dahil olması sağlanmaktadır. C5.0 algoritması ise C4.5 algoritmasından türetilmiştir ve genellikle büyük veri setleri için tercih edilmektedir. (Çalış ve ark., 2014).

Karar listeleri de karar ağaçlarında olduğu gibi akış şeması gibi veri setinden elde edilen veriyi açık bir şekilde belirtmektedir. Ancak karar ağaçlarından farklı olarak “böl ve yönet” mantığını değil de “ayır ve yönet” mantığını kullanmaktadırlar. Bu mantıkta çalışan en bilindik algoritmalar ise AQ ailesi ve CN2'dir. Bu algoritmalara eğitim verileri uygulandığında EĞER–O ZAMAN (IF–THEN) kural listesi elde edilir. Yeni bir veri geldiğinde bu liste de o yeni veriye uygun kural sırası ile aranır. Bir veri bir kural tarafından sınıflandırılırsa daha fazla kural denenmez bu da esneklik açısından bir dezavantaj oluşturabilmektedir. Eğer veri hiçbir kural tarafından kapsanmıyorsa o zaman listenin sonunda varsayılan bir kurala uygulanır (Hacıbeyoğlu ve ark., 2011).

One-R algoritması: R. C. Holt tarafından geliştirilen genel olarak basit bir algoritmadır. Bu algoritmada basit olarak çalışma mantığı hangi sınıf kullanılırsa daha iyi sonuç elde edileceğine bakılmaktadır. Her bir özellik için bir kural üretilir bunlar içerisinde en küçük hata oranına sahip olan seçilmektedir (Holt, 1993).

Part algoritması: Böl ve yönet mantığı ile çalışan bu algoritma karar listeleri oluşturmaktadır. Bu karar listeleri kurallar kümesini sıralamaktadır. Çalışması sırasında bir C4.5 karar ağacına benzer yapı oluşturularak en iyi kural bir yaprakta yer almaktadır (Witten ve Frank, 2000).

JRip algoritması: W. Cohen tarafından 1995 yılında önerilmiş basit yapılı bir algoritmadır. Veri seti içerisindeki tüm pozitif kurallar boş bir kural kümesine dahil

edilir. Kural kümesini oluşturduktan sonra kural kümesinin veri seti ile uyumunu artırabilmek için optimizasyon uygulanır ve boyut azaltılır (Cohen, 1995).

En yakın k-komşu algoritması: Mevcut olan veri setinde ki sınıfları belirli olan değerlerinden faydalanarak, yeni bir gözlemin hangi sınıfa ait olduğunu saptamak amacıyla kullanılır. Algoritma da ilk olarak en önemli parametrelerden olan “k” değeri belirlenir. Bu değer belirtilen bir noktaya en yakın komşuların sayısıdır. Belirtilen nokta ile diğer tüm noktalar arasında hesaplama yapılır. Hesaplaması yapılan bütün noktalar arasından en küçük değere sahip olan k tane değer alınır. Bu k tane değer içerisinde en yakın komşu kategorileri içerisinde en çok tekrarlanan seçilir ve yeni değer kategorisi olarak kabul edilir (Taşçı ve Onan, 2016). Bu algoritma gürültülü verilere karşı daha dirençli, analitik olarak izlenebilir ancak veri setinin boyutunun artması durumunda ki işlem ağırlığı açısından oldukça maliyetlidir.

Bayes sınıflandırıcılar: Sınıflandırma işleminde kestirimin istatistiksel yöntemlerle yapıldığı bir yöntemdir. Bir örneğin hangi sınıfa dahil olduğunun olasılığını kestirir. Bu algoritmanın kodlaması ve sonucunda ne çıkacağını kestirimi kolaydır ancak değişkenler arasında ki ilişkiyi bu algoritma ile gösteremeyiz (Jensen, 2001).

Destek vektör makinası: Temelleri istatistiksel yöntemlere dayanmaktadır. Veride bir ayırım yapabilmek için en uygun fonksiyonu bulmaya çalışır. Algoritma içerisinde temel olarak sınıflandırma doğrusal veya doğrusal olmayan bir fonksiyonun kullanılması ile gerçekleştirilebilmektedir. Bu algoritmanın kullanımı sürecinde çekirdek fonksiyonun seçilmesi ve parametre optimizasyonunun yapılması oldukça önemlidir. (Ayhan ve Erdoğan, 2014).

3.2.1. Kural tabanlı sınıflandırma algoritmaları

Sınıflandırma metodlarından biri olan kural tabanlı sınıflandırma yöntemi ile veri kümesine yeni veri girişinde başarı oranını artırmak için kullanılır.

Kural tabanlı sistemler If-then gibi koşullu durumlar ile oluşturulan sınıflandırma yöntemleridir. Kurallar içerisinde “ve yada veya” bağlaçları ile bağlanmış koşullu cümleler yer alabilir. Sisteme verilen koşullar sağlandığı durumdan verinin

sınıflandırıldığı sistemlerdir (Şeker, 2017). Bu yöntem içerisinde tanım ve sonuçları barındırır. Bu tanım ve sonuçlar sayesinde sisteme yeni veri girişi yapıldığında bu veriyi doğru bir şekilde sınıflandırır.

Kural tabanlı sınıflandırma algoritmaları oluşturulurken baz alınan kriter kazanç(başarı) durumlarıdır. Oluşturulan algoritma kendi kendine kurallar üretebilmelidir ve her kural için bir başarı değeri hesaplanmalıdır ve yazılan kuralın başarısının en fazla olması istenmektedir. Mesala bir veri seti üzerine kural tabanlı bir sınıflandırma algoritması uygulanmak isteniyorsa yazılan kural istenen durumları en çok kapsayan olmakla birlikte, istenmeyen durumları da en aza indirecek başarıda olmalıdır. Bir algoritmanın başarı kriterleri bir çok yöntem ile belirlenmektedir. Bunlar;

- Kural karmaşıklığı
- Anma
- Duyarlılık
- Kural kapsamasıdır (Freitas ve ark., 2000).

Sınıflandırma başarısı gösteriminde kullanılan bir yapı olan karışıklık (confusion) matrisi yardımı ile sınıflandırma algoritmalarının doğruluk değerleri üzerinde kıyaslama yapılmaktadır. Karışıklık matrisi Çizelge 3.1' de gösterilmektedir.

Çizelge 3.1. Karışıklık matrisi

Gerçek/Tahmini Sınıf	(Tahmini) Sınıf1	(Tahmini) Sınıf2
(Gerçek) Sınıf1	TP	FN
(Gerçek) Sınıf2	FP	TN

True Positive(TP) değeri: Kuralın uygulanması ile tahmin sonucu sınıf değeri evet(positive) ve gerçekte de sınıfının evet(positive) olduğu kayıt sayısını göstermektedir.

False Positive(FP) değeri: Kuralın uygulanması ile tahmin sonucu sınıf değeri evet(positive) ve gerçekte de sınıfının hayır(negatif) olduğu kayıt sayısını göstermektedir.

True Negative(TN) değeri: Kuralın uygulanması ile tahmin sonucu sınıf değeri hayır(negatif) ve gerçekte de sınıfının hayır(negatif) olduğu kayıt sayısını göstermektedir.

False Negative(FN) değeri: Kuralın uygulanması ile tahmin sonucu sınıf değeri hayır(negatif) ve gerçekte de sınıfının evet(positive) olduğu kayıt sayısını göstermektedir (Aydilek, 2017).

Kuralların keşfi işlemi için kullanılan kriterler olan Denklem 3.1 ve 3.2 de sırasıyla anma ve özgüllük kriterleri verilmiştir.

$$anma = \frac{TP}{TP+FN} \quad (3.1)$$

$$özgüllük = \frac{TN}{TN+FP} \quad (3.2)$$

Bir kural tabanlı sınıflandırma algoritmasında verilen eğitim kümeleri(training set) ya da test kümeleri(test set) başarı performansını etkileyen kriterler arasındadır. Bu nedenle eğitim ve test kümelerinin kullanım oranları önemlidir. Burada kullanılan metodlardan bazıları şunlardır:

Yüzdesel bölmek(Holdout method): En basit metodlardan olup verilen veri setinin %66'sı eğitim %33'ü ise test işlemi için ayrılmaktadır. Burada verilerin rastgele seçimi sonucunda başarı oranı biraz şansa bağlı kalmaktadır. Burada şansın etkisini biraz daha azaltmak için cross-validation method kullanılmaktadır.

Cross-validation: Veri setini parçalara bölerek başarı oranını arttırmayı hedefleyen bir yöntemdir. Burada parçalardan ilki test, kalan parçalar ise eğitim amacıyla kullanılabilir. Örneğin; 10 parçalık bir veri seti için 10 kere üzerinde çalışılmış olunur. Neticede çıkan sonucun ortalaması alınmaktadır. Ortalama değer sonucuna göre başarı oranı belirlenmiş olur. Cross-validation metodunun uygulanabilmesi için en az 10 veriye ihtiyaç vardır.

Boostrap: Genelde kullanım alanı olarak küçük veri setleri için tercih edilmektedir. Verilerin yetersiz olduğu durumlarda eğitim ve test verilerinin ayrılması

zorlaşacağından ve başarı oranının düşeceğinden bu yönetime başvurulmaktadır. Verilerin yerlerinde değişiklik yapılması ile yeni eğitim verisi ortaya çıkarılır. Her seferinde tek bir satır seçilerek işleme devam edilir. Seçilen her bir veri daha sonra dışarıda bırakılmaktadır. Cross validation metoda benzerlikler göstermektedir. Her bir verinin hem test hem de eğitim verisinin içinde bulunması olarak da düşünülebilir (Şeker, 2017).

3.2.2. Metasezgisel yöntemler

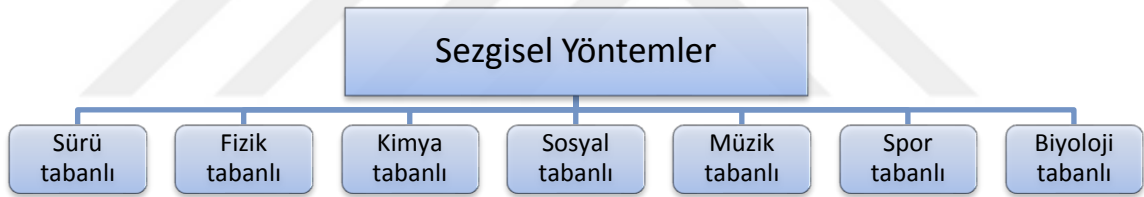
Metasezgisel yöntemler, karmaşık optimizasyon problemlerine etkili çözümler üretebilmek için geliştirilmiş algoritmalarıdır. Arama uzayını etkin ve verimli bir şekilde inceleyerek karmaşık ve büyük ölçekli problemlere çözüm üretmek amaçlanmıştır. Bu yöntemler, her zaman küresel en iyi çözümün bulunmasını garantileyemeseler de, etkili çözümler üretebildikleri için tercih edilmektedirler (Onan, 2013).

Sınıflandırma işlemi için karar ağaçları, Bayes sınıflandırma, en yakın komşu algoritması, karar destek makineleri vb. gibi birçok yöntem bulunmaktadır. Ancak metasezgisel algoritmalar açıklayıcı, anlaşılabilir olma ve sınırlayıcılara takılmama özellikleri ile daha iyi performanslar sergileyebilmektedir. Metasezgisel yöntemler son yıllarda popüler olmuştur ve devam eden çalışmalar ile de gittikçe büyük bir çalışma alanı haline gelmiştir. Bunun nedenleri ise şu şekilde ifade edilebilir:

- a. Metasezgisel yöntemler çalışma prensibi olarak çözüm uzayı çeşidine, karar değişkenleri ve sınırlayıcı sayılarından bağımsız çalışabilmektedirler.
- b. Bir problemde aynı anda farklı tipte öznitelikler, maliyet fonksiyonları ve kısıtlayıcıların olması halinde diğer algoritmalara göre daha yüksek performans göstererek genel çözüm stratejileri sunabilmektedir.
- c. Kurulması ve zor olan veya kurulunca da yüksek maliyetler içeren matematiksel modellere gerek duymamaktadır.
- d. Hesaplama zamanı açısından az maliyetlidir.
- e. Üzerlerinde değişiklik yapma ve uyarılama işlemleri oldukça kolaydır.
- f. Karmaşık ve büyük ölçekli problemlerde yüksek performans göstermektedir (Kar, 2016).

Bu gibi avantajları nedeniyle metasezgisel algoritmalar mühendislik, işletme, endüstri gibi farklı birçok alanda rağbet görmektedir.

Genel olarak sezgisel algoritmalar; sürü zekasına dayalı (parçacık sürü optimizasyonu, kurt koloni algoritması, karınca koloni optimizasyonu), fizik tabanlı (parçacık çarpışma algoritması, ısı işlem algoritması, yerçekimsel arama algoritması, merkez kuvvet optimizasyonu), kimya tabanlı (yapay olarak kimyasal reaksiyon algoritması), sosyal tabanlı (parlamentar optimizasyon algoritması, tabu arama algoritması), müzik tabanlı (armoni arama), spor tabanlı (lig şampiyonası algoritması) ve bu tez çalışması kapsamında kullanılan guguk kuşu optimizasyon algoritmasının dahil olduğu biyolojik tabanlı (ateşböceği algoritması, enzim algoritması, arı koloni algoritması) olmak üzere Şekil 3.2’ de görüldüğü üzere yedi kısımda incelenebilir(Karaboğa, 2011; Akyol ve Alataş, 2012).



Şekil 3.2. Metasezgisel yöntemler (Akyol ve Alataş, 2011)

Literatürde bu algoritmaların melez olarak kullanımı üzerine de çalışmalar bulunmaktadır.

Gerçekleştirilen başarılı algoritmalar olmasına rağmen, geniş kapsamlı olarak en iyi sonucu veren literatürde bir algoritmaya rastlanmıştır. Bu bağlamda son yıllarda ki çalışmalar yeni metasezgisel algoritmalar üretme veya var olanlar üzerinde iyileştirmeler yaparak devam etmektedir.

3.3. Kümeleme

Veri madenciliğinin temel konularından biri olan kümeleme; mevcut olan veriler arasında benzerlik ve farklılıklar göz önünde bulundurularak yapılan gruplandırma işlemidir (Çelik, 2013). Veriler birbirine benzeyen yanları ile alt gruplara ayrılırlar. Ayırma işlemi yapılırken veriler üzerinde uygulanan yöntemde benzerlik veya farklılık dikkate alınmaktadır. Elde edilen gruplar sayesinde araştırmacı uygun ve özetleyici bir bilgi çıkarımı yapabilir. Kümeleme işlemi ile bulunan her bir grubun kendi içerisinde homojen olması, birbirleri arasında ise heterojen olmaları gerekmektedir.

Kümelemeye dayalı modellerde, sınıflandırma modellerinden farklı olarak veriler belirli olmayan bir sınıfa göre gruplandırılır. Sınıflandırma da verilerin ait olduğu sınıf bilinmemektedir ancak kümelemede bu bilinmemektedir.

Kümeleme yöntemlerinde gruplandırma benzerlik veya uzaklık dikkate alınarak yapılmaktadır (Özkan, 2008). Yaygın olarak kullanılan uzaklık ölçüleri; Öklid, Manhattan ve Minkowski 'dir. Benzerlik kavramı uzaklığın tersi bir kavramdır. Benzerlik hesabında kullanılan yöntemlerden bazıları ise; Dice, Overlap ve Jaccard vb. (Silahtaroglu, 2013).

Birçok kümeleme yöntemi literatürde var olmaktadır. En çok tercih edilen kümeleme algoritmaları hiyerarşik ve hiyerarşik olmayan kümeleme algoritmaları olarak ikiye ayrılmaktadır. Hiyerarşik kümelemede toplama veya ayırma tekniği kullanılarak yapılmaktadır. Toplama tekniği ile başlangıçta veri setinde ki tüm değerler ayrı bir küme olarak düşünülür ve daha sonra aşamalı olarak birleştirilip alt kümeler oluşturulur. Ayırma tekniğinde ise başlangıçta veri setinde ki tüm değerler tek bir küme içine dahil edilir ve daha sonra aşamalı bir şekilde birbirine benzemeyen noktalar arasında ayrışım yapılır. Hiyerarşik durumda bulunmayan kümeleme tekniklerde ise küme sayısı başlangıçta bilinmemekte veya makul olacak şekilde bir küme adedi belirlenmektedir. Hiyerarşik olmayan metodlar, hiyerarşik metodlara göre daha nispeten daha büyük veri setine uygulanmaktadır. Bu yöntemde küme içi ortalama hata minimize edilmeye çalışılır (Silahtaroglu, 2013).

3.4. Birliktelik Kuralları

Büyük veri setleri içerisinde hemen göze çarpmayan birtakım ilişkiler olabilir. Bu ilişkilerin saptanması büyük veri setleri içerisinde zor bir süreç olabilmektedir. Bu süreç birliktelik kuralları yöntemi kullanılarak yapılır. Birliktelik kuralları yönteminde olayların birlikte gerçekleşme olasılıkları geçmiş veriler analiz edilerek yapılır. Bu yöntem ile gerçekleştirilen çalışmalarda geleceğe yönelik tahminler elde edilmektedir (Özçakır ve Çamurcu, 2007).

Birliktelik kuralları kullanılarak yapılan en yaygın uygulamalardan biri perakende satışlarda müşterilerin satın alma eğilimlerini saptayıdır. Buradaki amaç, müşteriler tarafından satın alınan ürünler arasında bir ilişki bulmaktır. Belirlenen bu ilişkilere dayanarak alışveriş alışkanlıkları belirlenir ve kazançlı bir pazarlama ortamı oluşur. Müşterilerin bu şekilde yaptıkları tüm ürünleri ele alarak satın alma eğilimlerini ortaya koyan uygulamalara pazar sepeti uygulaması denilmektedir.

Birliktelik kurallarında kullanılan temel kavramlar;

Destek sayısı: Destek ve güven ölçütlerinin hesaplanmasında kullanılan bir değerdir.

Destek: Veride ilişkinin ne kadar sık tekrarlandığını tanımlar. X ve Y 'yi ürün grubu olarak ve N ise tüm alışverişlerin sayısı olarak kabul edersek destek hesaplaması şu şekilde ifade edilir:

$$destek(X \rightarrow Y) = \frac{sayı(X,Y)}{N} \quad (3.3)$$

Güven: X malını almış bir müşterinin Y malını alma ihtimalini vermektedir ve şu şekilde ifade edilmektedir:

$$güven(X \rightarrow Y) = \frac{sayı(X,Y)}{sayı(X)} \quad (3.4)$$

Hesaplanan güven ve destek değerlerinin büyük olması ile birliktelik kurallarının güçlü olduğu kanısına varılmaktadır.

3.5. Yapay Zekâ Optimizasyon Algoritmaları

Optimizasyon genel olarak bir problemin en iyi çözümünü elde etme işlemidir. Belirli sınırlamalar içerisinde, belirli parametre değerleri bulmayı amaçlayan problem türlerine optimizasyon problemleri denir. Optimizasyon işleminde ilk basamak olarak karar değişkenleri olarak adlandırılan parametreler setinin belirlenmesi gerekir. Daha sonra bu parametreler göz önünde bulundurularak probleme göre istenilen minimize veya maksimize işlemini gerçekleştirecek uygunluk fonksiyonu belirlenir. Optimum sayılacak çözüm en küçük yapılacak problem türü için en düşük maliyet değerine sahip çözüm iken en büyük yapılacak problem için ise maliyet değerinin en büyük olması şartı aranır (Karaboğa, 2011).

Uygunluk fonksiyonu matematiksel formda aşağıdaki gibi tanımlanabilir:

n değişkenli $x = (x_1, x_2, \dots, x_n)$ vektörü tanımlanır ise burada x_i , i. parametrenin değerini göstermektedir. Maliyet fonksiyonu $f(x)$,

$$f(x) = f(x_1, x_2, \dots, x_n) \quad (3.5)$$

Karmaşık sistemlerde matematiksel bir model oluşturmak oldukça zordur. Modelin kurulum aşaması başarılı bir şekilde gerçekleşse bile işlem maliyeti oldukça fazla gelmektedir.

3.5.1. Guguk Kuşu Optimizasyon Algoritması

Guguk kuşlarının yaşam tarzından esinlenerek Ramin Rajabioun (2011) tarafından geliştirilen yeni bir sezgisel algoritma olan guguk kuşu optimizasyon algoritmasının temelini doğadaki kuşların yerleşimi ve üremesi oluşturmaktadır. Guguk kuşları doğada kendilerine bir konak aile seçerler. Yumurtalarını o konak yuvanın yumurtalarına benzetir. Guguk kuşları konak yuvanın gerçek sahibi kuş uzaklaştığında gelir ve kendi yumurtasını yuvadaki kuşun yumurtaları arasına bırakır. Bu olay sadece 10 sn sürmektedir. Ev sahibinin yumurtalarından bir tanesini de yuvadan atar. Yuvaya bırakılan guguk kuşu yumurtaları ev sahibi yumurtalardan genelde bir gün önce yumurtadan çıkıyor. Yumurtadan çıktıktan sekiz saat sonra yuvadaki diğer yumurtaları yuvadan atıyor. Bu sayede yuvanın tek sahibi olup yuvayı işgal ediyorlar ve üvey annenin getirdiği bütün yiyecekleri yiyorlar (Rajabioun, 2011).

Popülasyonu başlatma

Evrimsel algoritmalar da olduğu gibi guguk kuşu optimizasyon algoritmasında da bir başlangıç popülasyonu ile başlar. Bir optimizasyon problemini çözebilmek için problem değişkenlerini bir popülasyon halinde göstermek gerekir. Bu popülasyon içerisindeki diziler algoritmalara göre farklı isimler alabilmektedir. Örneğin; GA ve PSO algoritmalarında çözüm uzayındaki her bir nokta “Kromozom” ve “Parçacık Konumu” şeklinde kullanılmaktadır. GOA ’da ise kuşların bulunduğu habitat alanı olarak ifade edilen “ortam” kelimesiyle adlandırılmaktadır (Rajabioun, 2011).

N boyutlu bir problem için :

$$ortam = [x_1, x_2, x_3, \dots, x_N] \quad (3.6)$$

Ortamdaki bireyler kullanılarak maliyet hesabı yapılır:

$$maliyet = f_0(ortam) = f_0(x_1, x_2, x_3, \dots, x_N) \quad (3.7)$$

GOA ’yı başlatabilmek için N problem boyutuna ve N_{pop} adet ortam sayısına bağlı olarak $N \times N_{pop}$ matrisi oluşturulur. Daha sonra ortama yerleştirilecek olan yumurtalar için y_{min} ve y_{max} değerleri belirlenir ve bu değerler arasında yumurta sayıları rastgele

belirlenir. Algoritma içerisinde genel geçer olarak kabul edilen aralık 5 ila 20 arasındadır (Rajabioun, 2011).

Guguk kuşları doğa da yumurtalarını doğal ortamlarının maksimum uzaklığını dikkate alarak bırakırlar. Bu uzaklık değeri algoritma içerisinde Yumurta Yarıçapı (YY) olarak adlandırılmaktadır. Bu değer hesaplanırken Denklem 3.11'de olduğu gibi guguk kuşunun o anki yumurta sayısı ile toplam yumurta sayının limiti alınır (Rajabioun, 2011).

$$YY = \alpha \times \frac{\text{guguk kuşunun yumurta sayısı}}{\text{toplam yumurta sayısı}} \times (y_{max} - y_{min}) \quad (3.8)$$

α parametresi YY 'nin değerini ayarlayabilmek için kullanılan bir tamsayı değeridir.

Yumurtlama süreci

Yumurtalar, yumurtlama yarıçapı dikkate alınarak yumurta merkezleri etrafında rastgele yerleştirilir. Yumurtalar yerleştirildikten sonra ev sahibi kuş tarafından fark edilen yumurtalar yuvadan atılır. O anki yumurtalar arasında istenmeyen %p (genelde %10) kadarı silinir (Demirtaş ve Zengin, 2016). Algoritmada tanımlanan N_{max} parametresi ortamda yaşamını sürdüren guguk kuşu sayısını kontrol altında tutmak için kullanılır. Bu nedenle her iterasyonda en iyi maliyet değerlerine bakılarak içlerinde N_{max} sayısı kadar guguk kuşu hayatta kalacaktır (Mahmoudi ve Lutfi, 2015).

Göç etme süreci

Guguk kuşları yavruları büyüdüğünde kendileri ve yavruları için daha çok yiyecek kaynağı olan daha iyi bir ortama göç ederler. Guguk kuşları farklı ortamlarda grup oluştururlar ve bu gruplar içerisinde en iyi olan grup seçilerek oraya göç işlemi gerçekleştirirler. Guguk kuşu gruplarının belirlenmesinde ise K-Ortalama Algoritması kullanılır. Bu algoritma MacQueen (1967) tarafından geliştirilen bir veri seti üzerinde kümeleme yapabilmek için kullanılabilir en sade ve basit algoritmalardandır.

Algoritma temel olarak dört aşamadan oluşmaktadır ve guguk kuşu kümelerini bölmek için uygulanan aşamalar aşağıdaki gibidir:

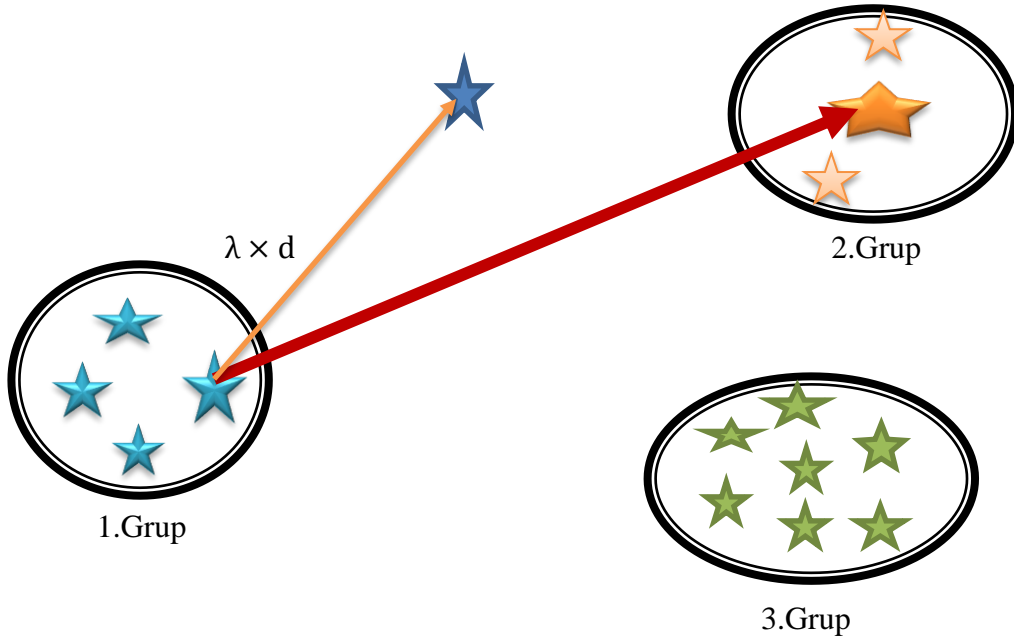
1. Küme merkezleri belirlenir. Guguk konumları arasında k adet rastgele yer belirlenerek veya tüm konumların ortalaması alınarak saptanır.
2. Guguk konumlarının birinci aşamada belirlenen küme merkezlerine olan uzaklığı hesaplanır. Sonuçlar dahilinde k adet küme içerisinde kendilerine konumca en yakın olana yerleştirilirler.
3. Oluşan k adet kümenin yeni merkez noktaları o küme içerisindeki guguk konumlarına göre belirlenir.
4. Merkez noktalarında herhangi bir değişiklik olmayıncaya kadar ikinci ve üçüncü adımlar tekrarlanır.

Her hangi bir guguk kuşu için, hedef ortama olan uzaklığının λ katı kadar hedefe yaklaşmış olur ve φ radyan kadar bir sapma eğilimindedir. λ ve φ parametreleri şu şekilde tanımlanmaktadır:

$$\lambda \sim \text{Random}(0,1), \varphi \sim (-\omega, \omega) \quad (3.9)$$

ω parametresi en iyi maliyete yaklaşmak amacıyla $\frac{\pi}{6}$ (rad) seçilir.

Guguk kuşlarının yaşam alanları içerisinde gerçekleşen göç işlemi Şekil 3.3'de gösterilmektedir.

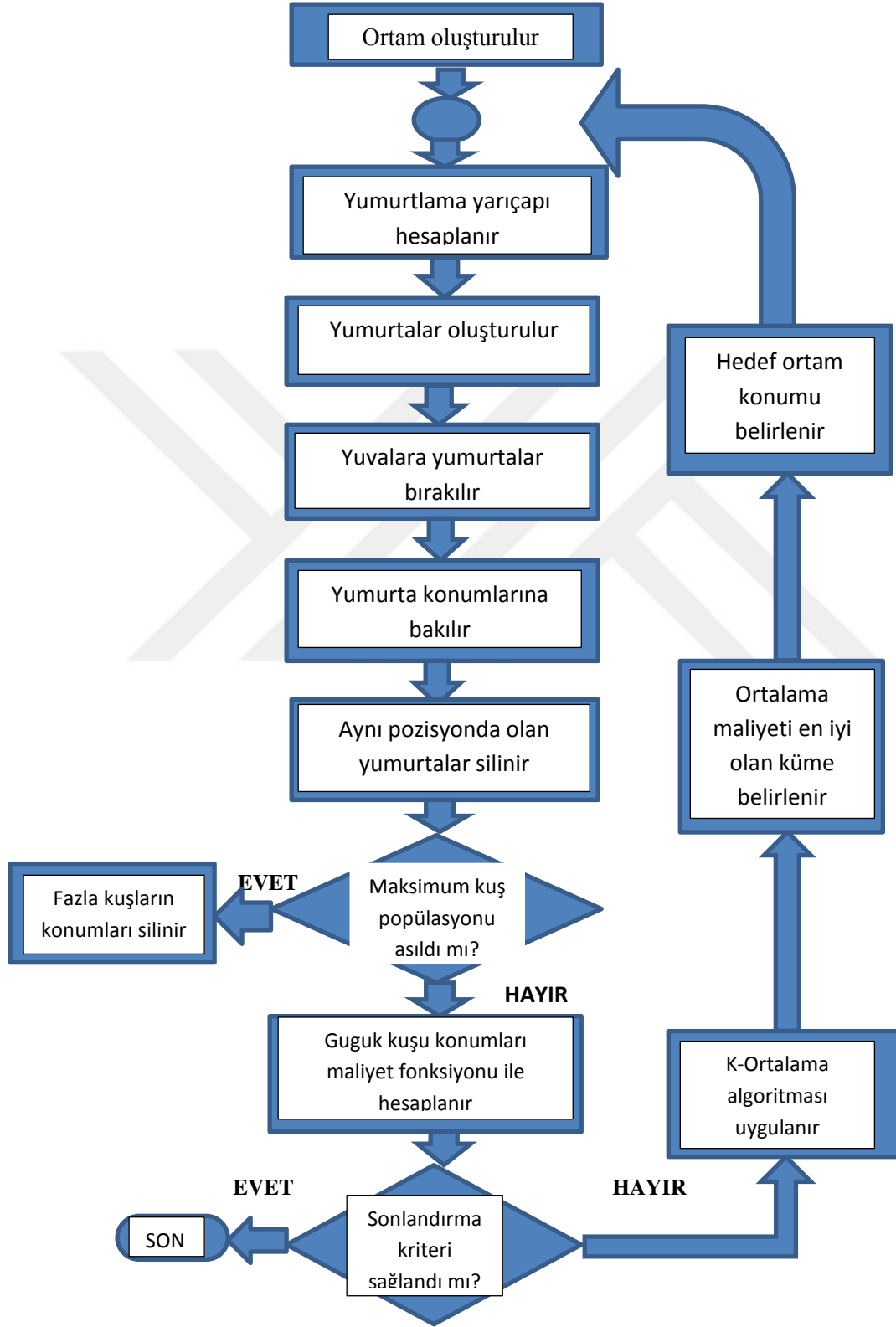


Şekil 3.3. Guguk kuşlarının yaşam alanları (Rajabioun, 2011).

Guguk kuşu optimizasyon algoritmasının sözde kodu:

1. İlk başlangıç ortamı oluşturulur.
2. Her kuşa rastgele sayıda yumurtalar tahsis edilir.
3. Her kuş için maksimum yumurtlama yarıçapı belirlenir.
4. Belirlenen yumurtlama yarıçapı alanı içine yumurta bırakılır.
5. Ev sahibi kuşlar tarafından farkedilen yumurtalar yok edilir.
6. Cıvcıvlerin yumurtadan çıkmasına ve büyümesine izin verilir.
7. Büyüyen her kuşun yaşam ortamı değerlendirilir.
8. Alan içinde yaşayabilecek kuş sayısı limitlendirilir ve istenmeyen alandakiler yok edilir.
9. En iyi kuş grubu tespit edilir ve hedef yaşam alanı seçilir.
10. Yeni guguk kuşu nüfusunun hedef yaşam alanına göç etmesi sağlanır.
11. İstenilen şart sağlanmış ise optimizasyon durdurulur, sağlanmamış ise 2. adımdan devam edilir (Rajabioun, 2011).

GOA' nın akış diyagramı Şekil 3.4'deki gibidir:



Şekil 3.4. GOA akış diyagramı

4. MATERYAL VE YÖNTEM

4.1. Hayvan Materyali

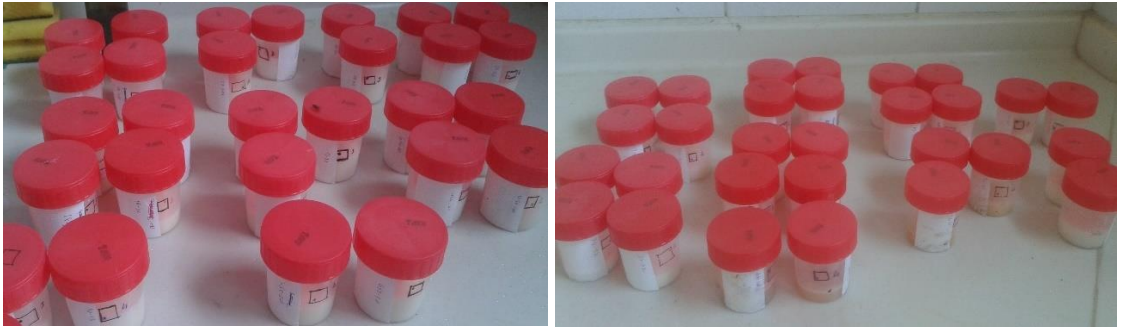
Çalışmanın hayvan materyalini Tokat ili içerisinde bulunan bir tarım işletmesinde yetiştirilen siyah alaca(holstein) türü sığırlar oluşturdu. İşletmede hayvanlar otomatik sağım makineleri kullanılarak sağılmaktadır.



Şekil 4.1. İşletmenin bilgisayarlı sürü yönetim sistemi

4.2. Süt Örnekleri

İşletmede sağımlar 03:00–06:00 10:00-13:00 15:00–18:00 arasında olmak üzere günde üç kez yapılmaktadır. Hayvanın memesinin dört lobundan da süt örnekleri alınmıştır. Süt örneklerinin hijyenik bir şekilde toplanabilmesi için Şekil 4.2’de gösterilen 50 cc’lik tüpler kullanılmıştır. Süt örneklerinin düzgün şekilde alınması, verilerin düzenli ve sağlıklı olması için iki ay boyunca işletmeye gidilmiştir.



Şekil 4.2. Süt numuneleri

4.3. Laboratuvar İşlemleri

Mastitis hastalığının subklinik evresinde süt ve etkilenen meme lobu normal görünmektedir. Subklinik hızlı bir şekilde klinik mastitise dönüşebilir yada subklinik olarak hastalık seyrine devam edip süt üretimini ve süt kalitesini düşürerek hem hayvan bazında hem de işletme bazında sorunlara neden olabilir. Uzun yıllar bu hastalık üzerinde çalışmalar devam etmiştir ve hastalık döneminde sütteki tuz oranının artması nedeniyle sütün elektriksel direncinin düştüğü görülmüştür (Atasever ve Erdem, 2008). Mastitisin teşhisinde kullanılan yöntemlerden biri olan elektriksel iletkenlik 1940'ların başında, mastitisin belirlenmesinde kullanılmaya başlayan bir yöntemdir. Sütte bulunan Na ve Cl iyonlarının artışı nedeniyle sütteki elektriksel iletkenlik değerleri değişiklik göstermektedir.

Sütün elektriksel iletkenliği ve meme sağlığı arasındaki bağlantının fark edilmesinden sonra J. Draminski 1989 yılında Polonya ziraat üniversiteleri ile yürüttüğü ortak çalışmaların neticesinde Draminski elektronik mastitis dedektörü geliştirilmiştir.

Bu çalışma bünyesinde işletmede toplanan numunelerin elektriksel iletkenliği, dört meme lobundan alınan sonuçları birlikte gösterebilen özel ekran uygulamalı bir cihaz olan Draminski elektronik mastitis detektörü ile ölçüldü. Ölçümlerde ki güvenilirliği artırmak için, ineğin yaşı, cinsi ve fizyolojik özelliğinin yanında hayvanların nasıl beslendiği ve metabolik düzensizlikler de dikkate alınmalıdır.

Gözle görülür değişimler olan süt örneklerini test etmek önerilmemektedir. Bu süt örneklerinin ölçüm değerleri yanıltıcı olabilmektedir. Bir ineğe ait dört meme lobunun elektriksel iletkenliği ölçümü sırasında dört değer arasında ciddi farklılıklar olması da subklinik mastitis olduğu anlamına gelebilmektedir ve bu yüzden hayvanın dört meme lobunun ölçülmesi subklinik mastitisin teşhisi için önemli bir unsurdur.

Genç ve sağlıklı bir inekteki okuma değerlerinin çoğu 400'e yakındır. Hayvan test edildikten sonra elde edilen 300 değerinin altındaki okumalar mastitis riskine işaret edebilir. Hayvan dokuz yaş üzeri ise sağlıklı olsalar dahi bu hayvanlarda gözlemlenen okuma değeri genellikle 290 civarındadır. Örneğin yaşlı bir hayvandan ölçüm sonucu okunan değerler üç lobda 290, bir lobda 250 ise bu düşük lobda enfeksiyon riski olabileceği anlamına gelmektedir.

Draminski mastitis dedektörü kullanılarak alınan ölçüm değerleri şu kriterler dikkate alınarak yorumlanmıştır:

1. Okunan değer 250'nin altında ise hayvanda ki enfeksiyonun şiddetinde ani bir artış olduğu ve subklinik mastitis den ileri klinik mastitis seviyelerine geçiş olabileceğine dair riskin bulunduğunu gösterir.

2. Okunan değer 300'ün üzerinde ise incelenen süt örneğinin kalitesinin iyi düzeylerde olduğu ve hayvan sağlığında bir tehlike olmadığı şeklinde yorumlanabilir. Bu kısımda okuma değerleri genelde 300-360 aralığındadır ve bu aralık değerlerinde subklinik mastitis görülme riski düşüktür.

3. Okunan değer 300 – 250 arasındaki değerler subklinik mastitis için dikkat etmek gerektiğini gösterir. Değerler düştükçe subklinik mastitis enfeksiyonunun görülme olasılığı artar. Bir inekteki alınan ölçüm değerlerinden üç meme lobunun 300 üzerinde ve eğer tek bir meme lobunda ki ölçüm değeri 250 ila 300 arasında ise burada ki ani düşüş meme de subklinik mastitis olma ihtimalini göstermektedir.

Meme loblarından elde edilen ölçüm değerleri arasındaki en yüksek ve en düşük ölçüm değerleri arasındaki farkın 40-50 olması subklinik mastitis başlangıcını gösterir. Bu nedenle hayvanların süt değerlerinin takibi hayati önem taşımaktadır. Her inek tek tek test edilmeli ve hastalık teşhisi için yapılan değerlendirme mümkün olduğu kadar inek bazında yapılmalıdır. Verilen değerler genel geçer kabul edilen rakamlar olmaktadır ve bu değerler ineklerde değişim gösterebilmektedir. Hayvan sürekli olarak takip edilmeli ve elektriksel iletkenlik değerindeki iniş çıkışlara dikkat edilmelidir.

Şekil 4.3' de Draminski mastitis dedektörü kullanılarak yapılan bir hayvanın dört meme lobuna ait elektriksel iletkenlik değerleri gösterilmektedir. Şekilde gösterilen ölçüm değerleri genç bir ineğe ait ölçüm değerleridir. Ölçüm değerleri sol-ön meme lobu için: 410, sağ ön: 200, sol arka: 400 ve sağ arka ise 420 olarak gözlemlenmiştir. Burada özellikle sağ ön meme lobunda var olan ani düşüş dikkat çekmektedir Üç meme lobundan elde edilen değerlerin 400 civarında olması bu meme loblarının sağlıklı olduğunu göstermektedir. Sağ ön meme lobundan ölçülen değerinin düşüklüğü ve aynı zamanda diğer meme lobları ile arasındaki fark bu meme lobunda enfeksiyon olduğunu göstermektedir.



Şekil 4.3. Hasta hayvanın elektiksel iletkenlik ölçüm değerleri

Sağlıklı bir hayvanda ölçülen değerler arasında ki rakamsal farklılıklar 50'nin altında olmakta ve eğer değerleri ölçülen inek 9 yaş üzeri yaşlı bir inek değilse 300 üzerinde bir elektiksel iletkenlik değeri gözlemlenmektedir. Şekil 4.4'de meme loblarında enfeksiyon bulunmayan sağlıklı bir ineğe ait elektiksel iletkenlik ölçüm değerleri bulunmaktadır. Bu hayvanda dört meme lobunun elektiksel iletkenlik değerleri 300'ün üzerindedir ve meme lobları arasındaki farklılıklar normal değerlere sahiptir.



Şekil 4.4. Sağlıklı hayvanın elektriksel ölçüm değerleri

Toplanan verilerin hasta olup olmadığının belirlenmesi için California mastitis testi(CMT) kullanılmıştır.

California mastitis testinin uygulanışı: California mastitis testi kabına memenin her bir lobundan ayrı ayrı süt örnekleri alınır. Sütlere eşit miktarlarda ayraç eklenir ve kap yatay düzlem içinde yavaşça çevrilerek sütteki değişimlere göre testin sonucu okunur.

Değerlendirilmesi: Test kabı hafifçe eğildiğinde kolayca akan karışımın altından daha yavaş akan yapışkan bir tabaka görülmez ise süt sağlıklı hastalık bulunmamakta demektir.

Test kabı eğildiğinde hafif bir jelöz kıvam oluşmakla birlikte çevirme hareketleri sırasında kaybolur veya kolay akan süt karışımının altında daha yavaş akan ince bir tabaka gözlemlendiğinde sütün şüpheli olduğu kanısına varılır.

Kap çevrildiğinde jelöz bir tabaka oluşur ise bu sütün sağlıklı olmadığı ve mastitis hastalığının varlığı anlaşılır. (Busato ve ark.,2000; Furumura ve ark.,1995).



4.4. Kullanılan veri setleri

Bu tez çalışmasında GOA mantığı ile bir kural tabanlı sınıflandırma algoritması geliştirmek amacıyla C# programlama dili kullanılarak bir program kodlanmıştır. Kodlaması yapılan bu program süt ineklerinin mastitis hastalığı teşhisini erken dönemde farkedebilmek amacıyla hazırlanmış mastitis veri setinde uygulandı. Hazırlanan veri seti için düzenli olarak süt işletmesine gidilip veriler toplandı. Veri seti içerisinde yer alan özniteliklerden biri olan su miktarı da aynı proje kapsamında geliştirilen otomasyon sistemi ile elde edilmiştir. Hazırlanan veri seti Şekil 4.5' de görüldüğü üzere 5 özellikten, 540 kadar kayıt sayısından ve 2 sınıftan oluşmaktadır. Sığırdaki var olan 4 meme lobu içerisinde farklılıklar olmaktadır. Aynı hayvana ait bir meme lobu ile diğer bir meme lobu arasında farklılıklar bulunmaktadır. Örneğin aynı hayvanın bir meme lobunda mastitis hastalığı gözlemlenirken diğer bir meme lobu sağlıklı olabilir. Bu nedenle veri seti içerisinde meme lobları teker teker dikkate alınmıştır. Aslında her bir meme lobu bir hayvan olarak değerlendirilip hastalığın hassas bir şekilde tespiti sağlandı. Eğer değerlendirme de meme lobları dikkate alınmadan hayvan bazında bir değerlendirme ile dört lobunda elektriksel iletkenliği alınıp ortalama bir değer hesaplanmış olsaydı hasta bir meme lobu var olsa bile diğer lobları sağlıklı olduğu için alınan ortalama değerde hasta meme lobu fark edilmecekti.

Veri seti içerisinde bulunan su miktarına ait öznitelikte ki veriler ise bu tez çalışmasının bağlı olduğu TÜBİTAK destekli proje kapsamında gerçekleştirilen otomatik suluk sayesinde elde edilmiştir. Hayvanlardaki su tüketiminin hastalıklar ile arasında bir ilişki olup olmadığı üzerine bilgi sahibi olmak amacıyla toplanmış verilerdir.

Sıcaklık özneliği ise mastitis hastalığının mevsimsel olarak değişimini belirlemek amacıyla data setine alınmıştır. Mastitis hastalığında ki bazı bakterilerin mevsim sıcaklıklarına göre artış veya azalış gösterdiği bilinmektedir(Alpay ve Yeşilbağ, 2009).

Sınıf kısmında yer alan 1 değeri hastalığın var olduğunu, 0 değeri ise hastalığın olmadığını göstermektedir.

1	ID	quarter	EI	waterintake	temp	class
2	TR601045230	2	270	92,00	34,8	1
3	TR601351564	2	310	81,60	32	1
4	TR601596548	4	410	54,19	34,3	0
5	TR601421597	3	380	82,37	30	0
6	TR601553564	4	370	100,40	34,3	1
7	TR601551769	4	390	83,00	36,5	0
8	TR601421737	1	340	74,22	36,5	0
9	TR601421737	2	260	74,22	31,7	1
10	TR601421737	3	280	74,22	36,9	1
11	TR601553564	1	240	65,00	33,6	1
12	TR601553826	3	390	86,50	42,9	0
13	TR601553798	1	250	79,85	29,5	1
14	TR601351926	2	300	94,15	33,4	1
15	TR601351769	1	380	74,85	34,5	0
16	TR601421737	4	400	52,21	39,6	0
17	TR601171815	1	370	85,46	29,7	0
18	TR601277264	4	260	69,75	42,9	1
19	TR601553664	3	380	58,20	32,1	0
20	TR601351659	2	220	74,49	32	1
21	TR601351564	2	300	73,00	36,5	0
22	TR601277264	2	340	85,60	29,5	0
23	TR601277264	4	290	85,60	33,6	1

Şekil 4.5. Hazırlanan mastitis veri seti

Kodlaması yapılan bu programda literatürde sıklıkla kullanılan UCI veri ambarında yer alan Iris ve Zoo veri setleri de kullanıldı. Bu veri kümeleri literatürde geliştirilen bir çok kural tabanlı sınıflandırma algoritması için kullanılmıştır. Bu bağlamda geliştirilen algoritmanın literatürde var olan algoritmalar ile olan başarı oranını karşılaştırma fırsatı elde edildi.

Kullanılan veri setlerinden biri olan Iris' de 4 öznitelik, 3 sınıf ve 150 kayıt sayısı bulunmaktadır. Iris veri setine ait bir kesit Şekil 4.6' de gösterilmektedir.

1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Sınıf
2	5.1	3.5	1.4	0.2	Iris-setosa
3	4.9	3.0	1.4	0.2	Iris-setosa
4	4.7	3.2	1.3	0.2	Iris-setosa
5	4.6	3.1	1.5	0.2	Iris-setosa
6	5.0	3.6	1.4	0.2	Iris-setosa
7	5.4	3.9	1.7	0.4	Iris-setosa
8	4.6	3.4	1.4	0.3	Iris-setosa
9	5.0	3.4	1.5	0.2	Iris-setosa
10	4.4	2.9	1.4	0.2	Iris-setosa
11	4.9	3.1	1.5	0.1	Iris-setosa
12	5.4	3.7	1.5	0.2	Iris-setosa
13	4.8	3.4	1.6	0.2	Iris-setosa
14	4.8	3.0	1.4	0.1	Iris-setosa
15	4.3	3.0	1.1	0.1	Iris-setosa
16	5.8	4.0	1.2	0.2	Iris-setosa
17	5.7	4.4	1.5	0.4	Iris-setosa
18	5.4	3.9	1.3	0.4	Iris-setosa
19	5.1	3.5	1.4	0.3	Iris-setosa
20	5.7	3.8	1.7	0.3	Iris-setosa
21	5.1	3.8	1.5	0.3	Iris-setosa

Şekil 4.6. Iris veri setinden bir görüntü

Kullanılan veri setlerinden bir diğeri olan Zoo veri setinde 17 öznitelik, 7 sınıf ve 101 tane kayıt bulunmaktadır. Zoo veri setine ait örnek bir görüntü Şekil 4.7’ de bulunmaktadır.

Animal	Hair	Feathers	Eggs	Milk	Airborne	Aquatic	Predator	Toothed	Backbone	Breathes	Venomous	Fins	Legs	Tail	Domestic	Catsize	Snif
aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1	1	1
carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	4
catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1	0	1
cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2
chub	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
clam	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	7
crab	0	0	1	0	0	1	1	0	0	0	0	0	4	0	0	0	7
crayfish	0	0	1	0	0	1	1	0	0	0	0	0	6	0	0	0	7
crow	0	1	1	0	1	0	1	0	1	1	0	0	2	1	0	0	2
deer	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
dogfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	1	4
dolphin	0	0	0	1	0	1	1	1	1	1	0	1	0	1	0	1	1
dove	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2
duck	0	1	1	0	1	1	0	0	1	1	0	0	2	1	0	0	2

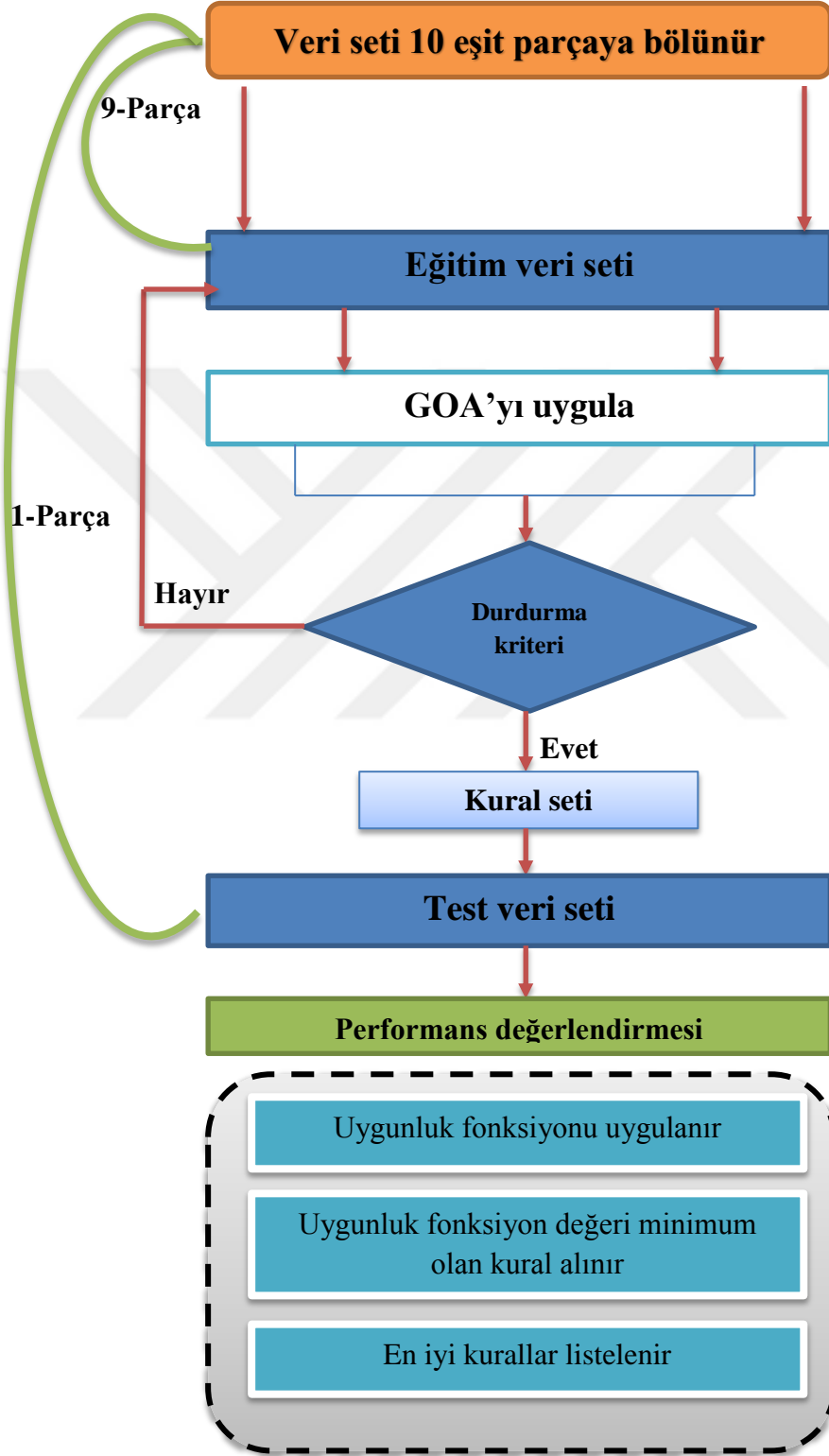
Şekil 4.7. Zoo veri setinden bir görüntü

4.5. Geliştirilen algoritma

GOA algoritması kullanılarak gerçekleştirilen bu uygulamada her bir sınıf etiketini en doğru şekilde tanımlayan kuralı bulmak amaçlanmaktadır. Böylece her bir sınıf etiketine göre kurallar listesi oluşmaktadır. Şekil 4.8’de önerilen yöntemin akış şeması yer almaktadır.

Geliştirilen algoritma ilk önce veri seti içerisindeki her bir öznitelik için özniteliğin aldığı minimum ve maksimum değerlerini tespit etmekte ve aralık listesi oluşturmaktadır. Aralık listesi başlangıçta kullanıcıdan girilmesi istenen aralık sayısına göre oluşturulmaktadır. Varsayılan olarak aralık sayısı 50 olarak belirlenmiştir. Yani bir öznitelik 0 ile 1000 arasında değer alıyor ise bu değeri aralık sayısına bölerek aralıklar hesaplanmaktadır. Bu aralıklar oluşturulacak kuralın koşul gövdesini oluşturmaktadır. Örneğin “A” özniteliği 0 ile 20 arasında değer almış ise sınıfı “1” dir şeklinde kurallar oluşturulmaktadır. Kurallar, içerisinde sadece bir adet öznitelik olduğunda “Birli Kural”, öznitelikler içinde herhangi iki adet özniteliğin kombinasyonu olduğunda da “İkili Kural” olarak adlandırılmaktadır. Algoritma birli ve ikili olmak üzere olası tüm kuralları oluşturarak veri seti üzerindeki kural uzayını oluşturmaktadır. Guguk kuşu optimizasyon algoritması bu kural uzayı içerisinde en iyi kuralları aramak için kullanılmaktadır. En iyi kuralları tesbit etmek için kural uzayı habitatlara ayrılmaktadır. Habitatlar kural uzayının alt kümeleridir. Bu noktada herbir habitat içerisindeki yumurta sayısı yani kural sayısı hesaplanmaktadır. Kural uzayındaki her bir kural bir habitata yerleştirilir. Guguk kuşu algoritmasına göre öncelikle kuralların kapsama sayıları elde edilerek kapsama sayısı sıfır ya da varsayılan değerinin altında olan kurallar listeden silinerek yuvadan atılmış olur. Kuralların kapsama sayısı, herbir kuralın koşul gövdesindeki şartı sağlayan kayıt sayısı olarak hesaplanmaktadır. Guguk Kuşu algoritmasındaki herbir iterasyonda kuralların kapsama sayısına göre habitatlar arasında geçiş sağlanmaktadır. Geçişler uygunluk fonksiyonu olan kural kapsama sayısına göre gerçekleşmektedir. Bu tez çalışmasında uygunluk fonksiyonu olarak TP+TN değeri alınmıştır. Herbir iterasyonda kurallar habitatlar arasında geçiş yapmakta ve en iyi kurallar bir habitat içerisinde toplanmaktadır. İtereasyonlar habitatlar arası kural geçişi kalmayınca kadar devam etmektedir. Bu sayede sınıflandırma için en iyi kurallar elde

edilmektedir. Sınıflandırmada ezberleme olmaması için 10 katlı çapraz doğrulama yapılarak üzerinde çalışılan veri seti için en uygun kurallar listesi elde edilmektedir.

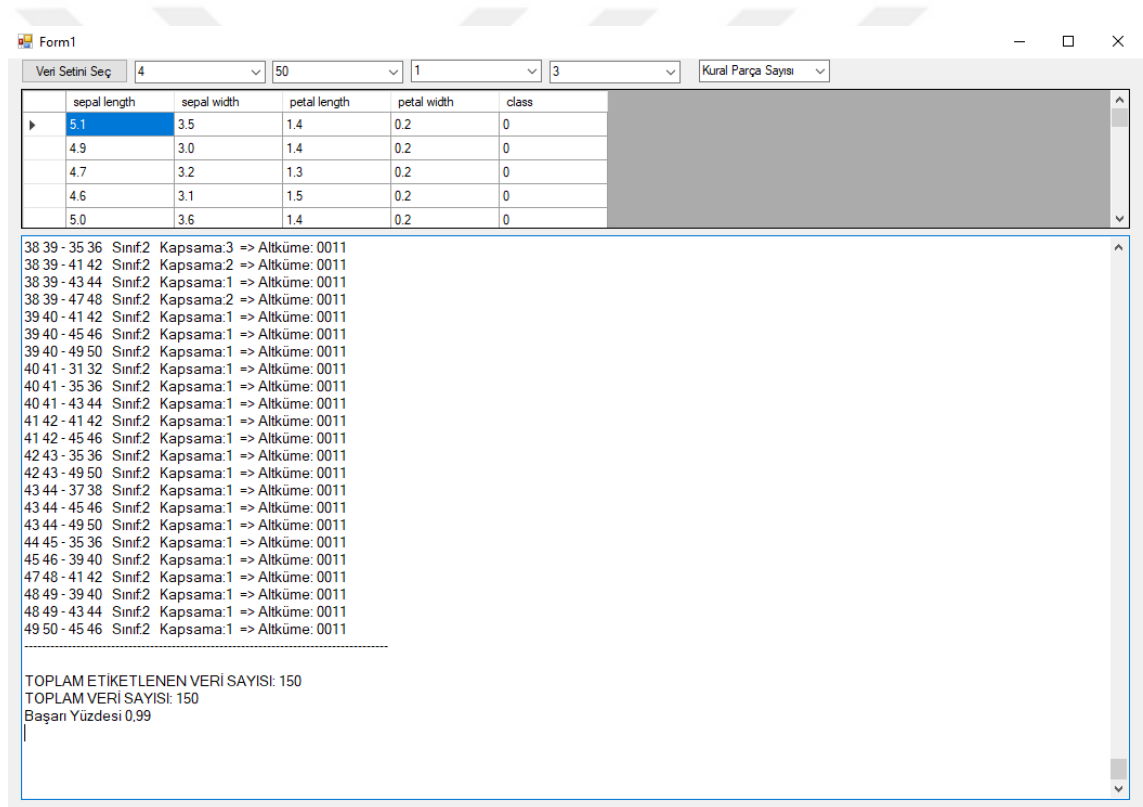


Şekil 4.8. Önerilen yöntemin akış diyagramı

5. BULGULAR VE TARTIŞMA

C# programlama dilinde kodlanmış olan programın başarı performansını test edebilmek için Waikato Üniversitesi tarafından geliştirilen WEKA (Waikato Environment for Knowledge Analysis) kullanılmıştır. WEKA da bulunan One-R, Part ve JRip algoritmaları kullanılarak başarı performansı karşılaştırılmıştır.

İris veri setinin geliştirilen algoritma olan guguk kuşu optimizasyon algoritmasına uygulanarak elde edilen sonuçları Şekil 5.1’ de görülmektedir. Elde edilen doğruluk yüzdesi %99 şeklindedir. Programda oluşturulan kural sayısı ise 766 tanedir.



The screenshot shows the WEKA software interface. At the top, there are several dropdown menus for selecting data sets and parameters. Below these, a table displays the Iris dataset with columns for sepal length, sepal width, petal length, petal width, and class. The first row is highlighted in blue. Below the table, a large text area displays the results of the optimization algorithm, including a list of rules and their associated accuracy values. At the bottom, a summary of the results is provided.

sepal length	sepal width	petal length	petal width	class
5.1	3.5	1.4	0.2	0
4.9	3.0	1.4	0.2	0
4.7	3.2	1.3	0.2	0
4.6	3.1	1.5	0.2	0
5.0	3.6	1.4	0.2	0

38 39 - 35 36 Sınıf2 Kapsama:3 => Altküme: 0011
38 39 - 41 42 Sınıf2 Kapsama:2 => Altküme: 0011
38 39 - 43 44 Sınıf2 Kapsama:1 => Altküme: 0011
38 39 - 47 48 Sınıf2 Kapsama:2 => Altküme: 0011
39 40 - 41 42 Sınıf2 Kapsama:1 => Altküme: 0011
39 40 - 45 46 Sınıf2 Kapsama:1 => Altküme: 0011
39 40 - 49 50 Sınıf2 Kapsama:1 => Altküme: 0011
40 41 - 31 32 Sınıf2 Kapsama:1 => Altküme: 0011
40 41 - 35 36 Sınıf2 Kapsama:1 => Altküme: 0011
40 41 - 43 44 Sınıf2 Kapsama:1 => Altküme: 0011
41 42 - 41 42 Sınıf2 Kapsama:1 => Altküme: 0011
41 42 - 45 46 Sınıf2 Kapsama:1 => Altküme: 0011
42 43 - 35 36 Sınıf2 Kapsama:1 => Altküme: 0011
42 43 - 49 50 Sınıf2 Kapsama:1 => Altküme: 0011
43 44 - 37 38 Sınıf2 Kapsama:1 => Altküme: 0011
43 44 - 45 46 Sınıf2 Kapsama:1 => Altküme: 0011
43 44 - 49 50 Sınıf2 Kapsama:1 => Altküme: 0011
44 45 - 35 36 Sınıf2 Kapsama:1 => Altküme: 0011
45 46 - 39 40 Sınıf2 Kapsama:1 => Altküme: 0011
47 48 - 41 42 Sınıf2 Kapsama:1 => Altküme: 0011
48 49 - 39 40 Sınıf2 Kapsama:1 => Altküme: 0011
48 49 - 43 44 Sınıf2 Kapsama:1 => Altküme: 0011
49 50 - 45 46 Sınıf2 Kapsama:1 => Altküme: 0011

TOPLAM ETİKETLENEN VERİ SAYISI: 150
TOPLAM VERİ SAYISI: 150
Başarı Yüzdesi 0.99

Şekil 5.1. Iris veri setinin geliştirilen algoritmaya uygulanışı

Part, One-R ve JRip algoritmaları uygulanarak Iris veri seti üzerinde elde edilen kurallar Çizelge 5.1’ de gösterilmektedir. Tablo da gösterildiği üzere Weka programında uygulanan One-R, Part ve JRip algoritması kullanılarak elde edilen kural sayısı 3 tanedir. Veri seti içerisinde bulunan 3 sınıf için de birer kural bulunmuştur.

Algoritmaların veri seti üzerindeki doğru sınıflandırma değerleri ise %92, %94 ve %95.3 şeklindedir.

Çizelge 5.1. Weka programında Iris veri seti için elde edilen sonuçlar

Kullanılan Algoritmalar	Bulunan Kural Sayısı	Doğruluk Yüzdesi
One-R	3	%92
Part	3	%94
JRip	3	%95.3

Zoo veri setinin, geliştirilen algoritma olan guguk kuşu optimizasyon algoritmasına uygulanarak elde edilen sonuçları Şekil 5.2' de görmektedir. Elde edilen doğruluk yüzdesi %90 şeklindedir. Programda oluşturulan kural sayısı ise 836 tanedir.

The screenshot shows the Weka software interface for the Zoo dataset. The top part displays a table of features with their values for four different instances. Below this, a list of generated rules is shown, each with its corresponding accuracy percentage. The rules are as follows:

Rule	Accuracy
36 37 - 49 50 Sınıf:1 Kapsama:1 => Altküme: 10000000000000001	100%
37 38 - 0 1 Sınıf:3 Kapsama:1 => Altküme: 10000000000000001	100%
37 38 - 49 50 Sınıf:1 Kapsama:2 => Altküme: 10000000000000001	100%
42 43 - 49 50 Sınıf:4 Kapsama:1 => Altküme: 10000000000000001	100%
43 44 - 0 1 Sınıf:6 Kapsama:1 => Altküme: 10000000000000001	100%
44 45 - 49 50 Sınıf:3 Kapsama:1 => Altküme: 10000000000000001	100%
45 46 - 0 1 Sınıf:3 Kapsama:1 => Altküme: 10000000000000001	100%
46 47 - 0 1 Sınıf:1 Kapsama:2 => Altküme: 10000000000000001	100%
46 47 - 49 50 Sınıf:4 Kapsama:1 => Altküme: 10000000000000001	100%
47 48 - 0 1 Sınıf:1 Kapsama:1 => Altküme: 10000000000000001	100%
48 49 - 0 1 Sınıf:6 Kapsama:1 => Altküme: 10000000000000001	100%
48 49 - 49 50 Sınıf:1 Kapsama:2 => Altküme: 10000000000000001	100%
49 50 - 49 50 Sınıf:1 Kapsama:1 => Altküme: 10000000000000001	100%
49 50 - 49 50 Sınıf:1 Kapsama:30 => Altküme: 01000000000000001	100%
49 50 - 0 1 Sınıf:2 Kapsama:14 => Altküme: 00100000000000001	100%
49 50 - 49 50 Sınıf:2 Kapsama:6 => Altküme: 00100000000000001	100%
0 1 - 49 50 Sınıf:1 Kapsama:31 => Altküme: 00010000000000001	100%
49 50 - 0 1 Sınıf:1 Kapsama:9 => Altküme: 00001000000000001	100%
49 50 - 49 50 Sınıf:1 Kapsama:32 => Altküme: 00001000000000001	100%
49 50 - 49 50 Sınıf:2 Kapsama:3 => Altküme: 00000100000000001	100%
49 50 - 49 50 Sınıf:4 Kapsama:1 => Altküme: 00000000000100001	100%
49 50 - 0 1 Sınıf:4 Kapsama:9 => Altküme: 00000000000010001	100%
49 50 - 49 50 Sınıf:1 Kapsama:6 => Altküme: 00000000000000011	100%

Below the rules, the following summary statistics are displayed:

TOPLAM VERİ SAYISI: 101
Başarı Yüzdesi 0.9
Toplam Süre 201.8990877 saniye

Şekil 5.2. Zoo veri setinin geliştirilen algoritmaya uygulanışı

Zoo veri seti için Weka programında bulunan programların uygulanması elde edilen sonuçlar Çizelge 5.2' de görünmektedir.

Çizelge 5.2. Weka programında Zoo veri seti için elde edilen sonuçlar

Kullanılan Algoritmalar	Bulunan Kural Sayısı	Doğruluk Yüzdesi
One-R	101	%42.5
Part	8	%92
JRip	6	%81

Mastitis veri setinin, geliştirilen algoritma olan guguk kuşu optimizasyon algoritmasına uygulanarak elde edilen sonuçları Şekil 5.3' de görünmektedir. Elde edilen doğruluk yüzdesi %99 şeklindedir. Programda oluşturulan kural sayısı ise 1072 tanedir.

The screenshot shows the Weka software interface with the Mastitis dataset loaded. The data table has the following columns: EI, waterintake, temp, and class. The data points are:

EI	waterintake	temp	class
270	92	34,8	1
310	81,6	32	1
410	54,19	34,3	0
380	82,37	30	0
370	100,4	34,3	1

Below the data table, a list of generated rules is shown, each with a unique ID and a structure: `Sınıf:0 Kapsama:2 => Altküme: 011`. The rules are numbered from 42 to 49. At the bottom of the interface, the following summary information is displayed:

TOPLAM VERİ SAYISI: 537
Başarı Yüzdesi 0,99
Toplam Süre 46,4274696 saniye

Şekil 5.3. Mastitis veri setinin geliştirilen algoritmaya uygulanışı

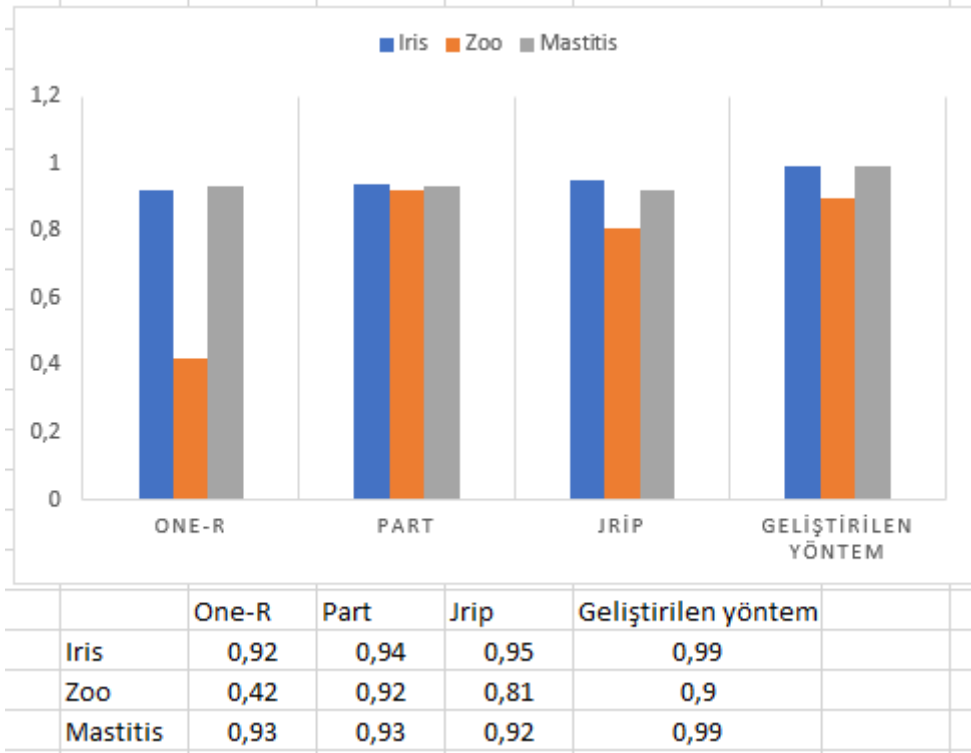
Mastitis hastalığının erken dönemde teşhisi için toplanan veriler ile oluşturulan mastitis data setinin weka programında bulunan One-R, Part ve Jrip algoritmalarına

uygulanması sonucunda elde edilen kural sayısı ve doğruluk yüzdesi Çizelge 5.3’ de görünmektedir.

Çizelge 5.3. Weka programında Mastitis veri seti için elde edilen sonuçlar

Kullanılan Algoritmalar	Bulunan Kural Sayısı	Doğruluk Yüzdesi
One-R	2	%93.5
Part	9	%93.1
JRip	2	%92.5

GOA kullanılarak kural tabanlı sınıflandırma algoritmasının diğer yöntemlere oranla daha iyi sonuçlar verdiği Şekil 5.4’ de görünmektedir. Kullanılan veri setlerine göre Zoo veri setinde Part algoritmasının gerisinde kalmıştır ancak diğer veri setlerinde yüksek performans göstermektedir.



Şekil 5.4. Geliştirilen yöntem ve diğer yöntemlerin karşılaştırılması

6. SONUÇ

Mastitis hastalığı süt sığırı işletmelerinde görülen bir hastalık olup klinik ve subklinik olarak ikiye ayrılmaktadır. Subklinik mastitisin gözle görülebilen belirtileri olmaması nedeniyle teşhisi oldukça zordur. Bu nedenle hastalığa neden olan etkenlerin belirlenip, bulunan bu etkenler doğrultusunda teşhisin erken dönemde yapılabilmesi için bir sistemin geliştirilmesi oldukça önemlidir. Bu çalışma da hastalığın erken dönemde teşhisini gerçekleştirebilecek bir algoritma geliştirildi. Geliştirilen bu algoritma süt sağım işletmesine düzenli aralıklarla gidilip hazırlanan veri seti üzerinde uygulandı.

Optimizasyon problemlerinin çözümü için literatürde birçok çalışma bulunmaktadır. Bunlar içerisinde metasezgisel yöntemler kullanılarak geliştirilen algoritmalar, büyük boyutlardaki problemlerde süre açısından sağladığı avantaj nedeniyle tercih edilmektedir.

Bu tez çalışmasında, guguk kuşu algoritması kullanılarak kural tabanlı yeni bir sınıflandırma algoritması geliştirildi. Geliştirilen bu algoritma mastitis hastalığının teşhisi için oluşturulan veri seti ile UCI veri ambarından alınan 2 veri seti üzerine uygulandı. Elde edilen sonuçlar ile Weka programında var olan One-R, Part ve Jrip algoritmaları ile elde edilen sonuçlar karşılaştırılmıştır. Karşılaştırma sonucunda mastitis veri seti üzerinde GOA kullanılarak geliştirilen kural tabanlı sınıflandırma algoritmasının doğruluk yüzdesi %99 şeklindedir. Weka programında bulunan One-R algoritması ile %93.5; Part algoritması ile %93.1 ve JRip ile %92.5 doğruluk yüzdesi elde edilmiştir. Karşılaştırma sonuçlarına bakıldığında ise Weka' da bulunan algoritmalaradan daha iyi sonuçlar elde edildiği görülmektedir. İris veri seti için geliştirilen algoritmanın doğruluk yüzdesi %99 One-R algoritması %92, Part %94 ve JRip %95.3 değerlerini vermektedir. Elde edilen sonuçlar incelendiğinde One-R, Part ve JRip algoritmalarından daha iyi sonuçlar elde edildiği görülmektedir. Geliştirilen algoritma zoo veri seti üzerinde uygulandığında %90 doğruluk yüzdesi elde edilmiştir. Weka programında ise One-R için %42.5, Part %92 ve JRip için de %81 değerleri bulunmuştur. Sonuçlar doğrultusunda Part algoritmasına göre daha düşük bir performans gösterdiği görülmektedir. Ancak One-R ve JRip'e göre daha iyi sonuçlar elde edilmektedir.

Mastitis hastalığının erken dönemde teşhisi konusunda geliştirilen algoritmanın sonuçları başarılıdır. Bu bağlamda bu algoritma kullanılarak mastitis hastalığı da dahil olmak üzere bir çok hastalığın erken dönemde teşhisi konusunda bu algoritma kullanılabilir.

Elde edilen sonuçlar karşılaştırılarak guguk kuşu optimizasyon algoritması kullanılarak geliştirilen kural tabanlı sınıflandırma algoritmasının etkili bir yöntem olduğu görülmektedir. Algoritma üzerinde iyileştirmeler yapılarak daha yüksek bir başarıya ulaşılabilir. Daha iyi sonuçların elde edilmesi amacıyla algoritmanın ilk aşaması olan popülasyon üretme kısmında değişiklikler yapılabilir. Gelişigüzel bir popülasyon üretimi yerine arama uzayına sistematik olarak düzenli bir şekilde dağıtılabilir. Ayrıca Guguk kuşu algoritmasının yumurta yarıçapı gibi değerleri üzerinde de iyileştirmeler yapılabilir. Algoritmanın sonlanma şartı üzerinde de çalışmalar yapılabilir.

7. KAYNAKLAR

- Anonim, 2016. İneklerde mastitis ve korunma yöntemleri. http://www.ndsyb.com/images/stories/yazilar/kenan_bayoglu/mastitis.pdf
- Akyol, S. ve Alataş, B., 2012. Güncel Sürü Zekası Optimizasyon Algoritmaları. Nevşehir Bilim ve Teknoloji Dergisi, 1(1).
- Akyol, S. ve Alataş, B., 2016. Kedi sürüsü optimizasyon algoritmasıyla doğru ve anlaşılabilir nümerik sınıflandırma kurallarının otomatik keşfi. Journal of the Faculty of Engineering and Architecture of Gazi University, 31(4), 839-857.
- Alataş, B. ve Akın, E., 2004. Sınıflandırma Kurallarının Karınca Koloni Algoritmasıyla Keşfi, ELECO'2004, Bursa, 357-361, Aralık 2004.
- Alataş, B. ve Özer, A. B., 2017. Genelleş tirilmiş ilginç sınıflandırma kurallarının yapay kimyasal reaksiyon optimizasyon algoritması ile keşfi. Journal of the Faculty of Engineering and Architecture of Gazi University, 32(1), 109-126.
- Al-Maqaheh, B. M. A., 2012. Genetic algorithm approach to automated discovery of comprehensible production rules. In Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference on (pp. 69-71). IEEE.
- Al-Maqaheh, B. M., ve Shahbazkia, H., 2012. A genetic algorithm for discovering classification rules in data mining. International Journal of Computer Applications, 41(18).
- Atasever, S. ve Erdem, H., 2008. Süt sığırlarında mastitis ile sütün elektriksel iletkenliği arasındaki ilişkiler, OMÜ Zir. Fak. Derg., 23 (2): 131-136.
- Alpay, G. ve Yeşilbağ, K. (2009). Mastitis Olgularında Virusların Rolü" The roles of Viruses in Mastitis". Uludağ Üniversitesi Veteriner Fakültesi Dergisi, 28(1).
- Aydilek, İ. B. (2017). Değiştirilmiş ateşböceği optimizasyon algoritması ile kural tabanlı çoklu sınıflama yapılması. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 32(4).
- Ayhan, S. ve Erdoğan, Ş., 2014. Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi, 9(1).
- Aytuğ, O., 2013. Metasezgisel Yöntemler ve Uygulama Alanları. Çukurova Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 17(2).
- Busato A., Trachsel P., Schallibaum M., Blum, JW., 2000. Udder health and risk factors for subclinical mastitis in organic dairy farms in Switzerland. Prev. Vet. Med., 28, 44: 205-220.
- Cavero, D., Tölle, K-H., Rave, G., Buxade, C., Krieter, J., 2007. Analysing serial data for mastitis detection by means of local regression, Livestock Sci., 110, 101 – 110.
- Cedden, F. Kor, A. ve Keskin, S., 2002, Laktasyonun geç döneminde keçi sütünde somatik hücre sayımı; yaş, süt verimi ve bazı meme özellikleri ile olan ilişkileri, YYÜ Ziraat Fak Tarım Bilimleri Derg., 12 (2): 63-67.
- Cheruku, R. Edla, D. R. ve Kuppili, V., 2017. SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification. Computers in biology and medicine, 81, 79-92.
- Cohen W.W., 1995. Fast Effective Rule Induction, Twelfth International Conference on Machine Learning, Morgan Kaufmann, 115-123.
- Çalış, A., Kayapınar, S., ve Çetinyokuş, T., 2014. Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir

- Uygulama. Journal of Industrial Engineering (Turkish Chamber of Mechanical Engineers), 25.
- Çelik, Ş., 2013. Kümeleme analizi ile sağlık göstergelerine göre Türkiye'deki illerin sınıflandırılması.
- Demirtaş, F. ve Zengin, K., 2016. "Guguk Kuşu Algoritmasının Gezgin Satıcı Problemine Uygulanması ve Simülasyonu " EEB 2016 Elektrik-Elektronik ve Bilgisayar Sempozyumu.
- De Mol, R. M. and Ouweltjes, W., 2001. Detection model for mastitis in cows milked in an automatic milking system, Preventive Veterinary Med., 49: 71-82.
- Furumura K., Imanishi M., Kashiwamura F., Shinde Y., Kawabata S., Hayashi M., 1995. On-line image processing prototype for the detection of mastitis in cows. Anim. Sci. Tec., 66, 10: 882-888.
- Fidelis, M. V., Lopes, H. S., & Freitas, A. A. (2000). Discovering comprehensible classification rules with a genetic algorithm. In Evolutionary Computation, 2000. Proceedings of the 2000 Congress on (Vol. 1, pp. 805-810). IEEE.
- Gündoğan, K. K. Alataş, B. ve Karci, A., 2004. Mining classification rules by using genetic algorithms with non-random initial population and uniform operator. Turkish Journal of Electrical Engineering & Computer Sciences, 12(1), 43-52.
- Han, J., Pei, J., ve Kamber, M., 2011. Data mining: concepts and techniques. Elsevier.
- Hacibeyoglu, M., Arslan, A., ve Kahramanli, S., 2011. Improving Classification Accuracy with Discretization on Data Sets Including Continuous Valued Features. Ionosphere, 34(351), 2.
- Holt R. C., 1993. Very Simple classification rules perform well on most commonly used datasets, Machine Learning, 11, 69-90.
- Jensen, F.V., 2001. Bayesian Networks and Decision Graphs, Springer.
- Kar, A. K., 2016. Bio inspired computing–A review of algorithms and scope of applications. *Expert Systems with Applications*, 59, 20-32.
- Kiziloluk, S. ve Alatas, B., 2015. Automatic mining of numerical classification rules with parliamentary optimization algorithm. Advances in Electrical and Computer Engineering, 15(4), 17-24.
- Karaboğa, D., 2011. Yapay Zeka Optimizasyon Algoritmaları. Nobel Yayın Evi, 245 s, İstanbul.
- Krieter, J., Caverio, D., ve Henze, C., 2007. Mastitis detection in dairy cows using neural Networks, GIL Jahrestagung, volume101of LNI, page 123-126.
- Mahmoudi, S. ve Lotfi, S., 2015. Modified cuckoo optimization algorithm (MCOA) to solve graph coloring problem. Applied soft computing, 33, 48-64.
- Mehraban, S. M., Mohebbi, M., Shahidi, F., Vahidian, K. A., ve Qhods, R. M., 2012. Application of fuzzy logic to classify raw milk based on qualitative properties. Int. J. Agri. Sci. 2(12): 1168-1178.
- Memmedova, N., 2012. Süt sığırlarında mastitisin bazı yapay zekâ yöntemleri kullanılarak erken dönemde tespiti (Doctoral dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü).
- Özçakır, F. C. ve Çamurcu, A. Y., 2007. Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 6(12), 21-37.
- Özkan, Y., 2008. Veri Madenciliği Yöntemleri. Papatya Yayıncılık Eğitim, 240 s, İstanbul.

- Özyurtlu, N., 2011. İneklerde Mastitisin Ekonomik ve Sağlık Açısından Önemi. Dicle Üniversitesi Veteriner Fakültesi Dergisi, (1), 36-38.
- Philpot WN., ve Nickerson SC., 2000. Winning the fight against mastitis. Naperville, Westfalia Surge, Inc.
- Quinlan, J. R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, USA.
- Rajabioun, R., 2011. Cuckoo optimization algorithm. Applied soft computing, 11(8), 5508-5518.
- Rişvanlı, A. ve Kalkan, C., 2002. Sütçü ineklerde yaş ve irkin subklinik mastitisli memelerin sütlerindeki somatik hücre sayıları ile mikrobiyolojik izolasyon oranlarına etkisi, YYÜ, Vet. Fak. Derg., 13(1-2): 84-87.
- Sabuncuoğlu, N. ve Çoban, Ö., 2006. Mastitis ekonomisi, AÜ Vet. Bil. Derg., 1 (1-2) 1-5.
- Sahebi, H. R. ve Ebrahimi, S., 2015. A Fuzzy Classifier Based on Modified Particle Swarm Optimization for Diabetes Disease Diagnosis. Advances in Computer Science: an International Journal, 4(3), 11-17.
- Sezer, E. A. Bozkır, A. S., Yağız, S., Gökçeoğlu C. 2010. "Karar Ağacı Derinliğinin CART Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama," Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Kayseri.
- Silahtaroglu, G., 2013. Veri Madenciliği: Kavram ve Algoritmaları. Papatya Yayıncılık Eğitim, 304 s, İstanbul.
- Srinivasan, S. ve Ramakrishnan, S., 2012. Multi objective optimization of classification rules using cultural algorithms. Procedia Engineering, 30, 457-465.
- Şeker, İ., Rişvanlı, A., Kul, S., Bayraktar, M., ve Kaygusuzoğlu, E., 2000. İsviçre esmeri ineklerde meme özellikleri ve süt verimi ile CMT skoru arasındaki ilişkiler, Lalahan Hay. Araş. Ens. Derg., 40 (1): 29-38.
- Şeker, Ş. E., 2017. Weka ile Veri Madenciliği. Bilgisayar Kavramları Yayınları, 376 s, İstanbul.
- Taşcı, E. ve Onan, A., 2016. K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. Akademik Bilişim.
- Witten I.H. ve Frank E., 2000. Weka machine learning algorithms in java, in Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann , 265-320.
- Yang, X. Z., Lacroix, R., ve Wade, K. M., 2000. Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks. Can. J. Anim. Sci. 80: 415-426.
- Yeşilmen, S., Özyurtlu, N., ve Bademkiran, S., 2012. Diyarbakır yöresinde Sublinik mastitisli İneklerde etken izolasyonu ve duyarlı antibiyotiklerin belirlenmesi, Dicle Univ. Vet. Fak. Derg, 1, 24-29.

9. ÖZGEÇMİŞ

Kişisel Bilgiler:

Adı-Soyadı: Fadime DEMİRTAŞ

Uyruğu: TC

E-posta: demirtas.fadime@gmail.com

EĞİTİM:

Lisans:	Fırat Üniversitesi - Mühendislik Fakültesi	
	Bilgisayar Mühendisliği	2010
Yüksek Lisans:	Gaziosmanpaşa Üniversitesi Fen Bilimler Enstitüsü	
	Mekatronik Mühendisliği	2018

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler

DEMİRTAŞ F., ZENGİN K. ve NASİP Ö. F., (2017). A Rule-Based Classification Algorithm Based On Coa. 1 st International Symposium on Multidisciplinary Studies and Innovative Technologies

Ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler

Demirtaş, F. ve Zengin, K. "Guguk Kuşu Algoritmasının Gezgin Satıcı Problemine Uygulanması ve Simülasyonu" EEB 2016 Elektrik-Elektronik ve Bilgisayar Sempozyumu 2016.

YABANCI DİL

İngilizce