



**T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

**PATOLOJİK TANI VE BULGULARIN DERİN
ÖĞRENME YÖNTEMLERİ İLE ANALİZİ VE
TAHMİNİ**

Ahmet Anıl ŞAKIR

BURDUR, 2022

**T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

**PATOLOJİK TANI VE BULGULARIN DERİN
ÖĞRENME YÖNTEMLERİ İLE ANALİZİ VE
TAHMİNİ**

Ahmet Anıl ŞAKIR

Danışman: Doç. Dr. Ali Hakan IŞIK

BURDUR, 2022

ÖNSÖZ ve TEŞEKKÜR

Bu araştırma için beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan değerli Danışman Hocam Doç. Dr. Ali Hakan IŞIK'a teşekkürlerimi sunarım.

Araştırmalarım sırasında yardımlarını gördüğüm değerli hocalarım Prof. Dr. Özlem ÖZMEN'e ile Dr. Öğr. Üyesi Volkan İPEK'e teşekkür ederim.

0671-YL-20 No`lu Proje ile tezimi maddi olarak destekleyen Burdur Mehmet Akif Ersoy Üniversitesi Bilimsel Araştırma Projeleri Koordinatörlüğü'ne teşekkür ederim.

Eğitim hayatımın her aşamasında beni her anlamda destekleyen aileme sonsuz sevgi ve saygılarımı sunarım.

Ocak, 2022

Ahmet Anıl ŞAKIR

İÇİNDEKİLER

	Sayfa
ÖNSÖZ ve TEŞEKKÜR	i
İÇİNDEKİLER.....	ii
ŞEKİL DİZİNİ.....	iii
ÇİZELGE DİZİNİ	v
SİMGELER VE KISALTMALAR DİZİNİ	vi
ÖZET	vii
SUMMARY	viii
1. GİRİŞ	1
2. GENEL BİLGİLER	3
3. MATERYAL VE YÖNTEM.....	5
3.1. Veri Tamamlama.....	5
3.1.1. Yanlış Veri Düzeltme	6
3.1.2. Eksik Veri	9
3.1.3. Eksik Veri Tamamlama	11
3.2. Karar Ağacı Metodu İle Tahmin	17
3.3. KNN Metodu İle Tahmin	21
3.4. RMSE, MSE ve MAE Yöntemleri ile Hata Payı Hesaplama.....	22
4. ARAŞTIRMA BULGULARI VE TARTIŞMA	24
4.1. Grafik Bulguları	24
4.1.1. Analiz Sonucu Elde Edilen Grafikler	24
4.1.2. İstatistik Analizi Sonucu Elde Edilen Grafikler	31
4.2. Karar Ağacı Metodu Sonuçları.....	39
4.3. KNN Metodu Sonuçları.....	41
4.4. Tartışma.....	43
5. SONUÇ.....	45
KAYNAKLAR.....	46

ŞEKİL DİZİNİ

	Sayfa
Şekil 1.1. Rutin defterde bulunan verilerin örnek hali	1
Şekil 3.1. Çalışmada izlenen akış şeması.....	5
Şekil 3.2. Veri tabanındaki eksik veriler	6
Şekil 3.3. Tür kategorisine ait değerler	7
Şekil 3.4. Yaş kategorisine ait değerler	7
Şekil 3.5. Düzenlemeden sonraki yaş kategorisi	8
Şekil 3.6. Veri türleri.....	9
Şekil 3.7. Eksik verilerin bar grafiği ile gösterimi	10
Şekil 3.8. Eksik verilerin matrix grafiği ile gösterimi.....	10
Şekil 3.9. Eksik verilerin heatmap grafiği ile gösterimi.....	11
Şekil 3.10. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setindeki eksik veri sayısının yapısı	12
Şekil 3.11. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setinin matrix grafiği ile gösterimi	12
Şekil 3.12. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setinin heatmap grafiği ile gösterimi	13
Şekil 3.13. Şehirlerin kategori sayısı grafiği	14
Şekil 3.14. İnceleme türü kategori sayısı grafiği.....	15
Şekil 3.15. Hayvan türleri kategori sayısı grafiği.....	15
Şekil 3.16. Hayvan yaşları kategori sayısı grafiği.....	16
Şekil 3.17. Patolojik-anatomik tanı kategori sayısı grafiği	17
Şekil 3.18. Hastalık türleri kategori sayısı grafiği.....	17
Şekil 3.19. Veri tamamlama sonucu kategorilerin eksik veri sayısı	18
Şekil 3.20. Örnek bir karar ağacı şeması.....	19
Şekil 4.1. Burdur ili ilçelerin kategori sayısı grafiği	24
Şekil 4.2. Şehirlerin kategori oranını gösteren Pie grafiği	25
Şekil 4.3. Hayvan cinsiyet oranını gösteren Pie grafiği	25
Şekil 4.4. İnceleme türü kategori oranını gösteren Pie grafiği.....	26
Şekil 4.5. Hayvan türleri kategori oranını gösteren Pie grafiği.....	26

Şekil 4.6. Hayvan yaşları kategori oranını gösteren Pie grafiği.....	27
Şekil 4.7. Hastalık türleri kategori oranını gösteren Pie grafiği.....	28
Şekil 4.8. Şehirlerden gelen numunelerin hastalık türleri oranı grafiği	28
Şekil 4.9. Şehirlerden gelen numunelerin hayvan türleri oranı grafiği	29
Şekil 4.10. Hayvan cinsiyetinin inceleme türüne oranı grafiği	29
Şekil 4.11. Hayvan türlerinin hastalık türüne oranı grafiği.....	30
Şekil 4.12. Hayvan cinsiyetinin yıllara oranı grafiği.....	30
Şekil 4.13. 2000-2020 yılları arası gelen numune sayısı grafiği.....	31
Şekil 4.14. 2000-2020 yılları arası hayvan yaş dağılımı grafiği	32
Şekil 4.15. Değişik illerden gelen hayvan yaş dağılımı grafiği.....	33
Şekil 4.16. Hayvan yaş gruplarında uygulanan inceleme türleri grafiği.....	34
Şekil 4.17. Hayvan cinsiyetinin hayvan yaşına dağılımı grafiği.....	35
Şekil 4.18. Hayvan türlerinin hayvan yaşına dağılımı grafiği.....	36
Şekil 4.19. Hastalık türlerinin hayvan yaşına dağılımı grafiği.....	37
Şekil 4.20. Hayvan türlerinin hayvan cinsiyetine oranı grafiği.....	38
Şekil 4.21. Hastalık türlerinin hayvan cinsiyetine oranı grafiği.....	39
Şekil 4.22. Karar ağacı modeli tahmin ve hata payı grafiği	40
Şekil 4.23. Derinlik sayısı 15 olan model grafiği.....	41
Şekil 4.24. KNN modeli tahmin ve hata payı grafiği.....	42
Şekil 4.25. Komşu sayısı 8 olan model grafiği	42
Şekil 4.26. Model sonuç karşılaştırma grafiği.....	43

ÇİZELGE DİZİNİ

	Sayfa
Tablo 2.1. Kanserin fethi ile sonuçlanan belirli yapay zekâ yeniliklerinin bir zaman çizelgesi	4
Tablo 4.1. Karar ağacı modeli sonucu çıkan değerler tablosu.....	40
Tablo 4.2. KNN modeli sonucu çıkan değerler tablosu.....	41



SİMGELER VE KISALTMALAR DİZİNİ

A	: Absorbans
CV	: Döngüsel Voltametri
FBE	: Fen Bilimleri Enstitüsü
MAKÜ	: Mehmet Akif Ersoy Üniversitesi
T	: Transmittans
TDK	: Türk Dil Kurumu
λ	: Dalga boyu
η	: Verim



ÖZET

Yüksek Lisans Tezi

Patolojik Tanı ve Bulguların Derin Öğrenme Yöntemleri ile Analizi ve Tahmini

Ahmet Anıl ŞAKIR

**Burdur Mehmet Akif Ersoy Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı**

Danışman: Doç. Dr. Ali Hakan IŞIK

Ocak, 2022

İnsan hastalıklarında olduğu gibi, hayvan hastalıklarında da hızlı teşhisin önemi büyüktür. Hastalık tedavilerinin doğru yapılabilmesi açısından teşhisin hızlı konulması kadar, konulan teşhisin yüksek doğruluk oranında olması da gerekmektedir. Bu çalışmada Burdur Mehmet Akif Ersoy Üniversitesi Veteriner Fakültesi Patoloji Anabilim Dalına ait 2000-2020 yılları arasında incelenen verilerden oluşan veri seti içerisinde bulunan hastalık türlerinin, karar ağacı sınıflandırma modeli ile KNN sınıflandırma modeli kullanılarak tahmini yapılmıştır. Veri seti içerisinde bulunan yaş, tür, şehir, cinsiyet gibi kategoriler ise grafikler halinde analiz edilmiştir. Tahmin ve analiz işlemlerinin doğru sonuçlar verebilmesi açısından veri seti bazı ön işlemlerden geçirilerek düzeltilip, eksik veriler tamamlanmıştır. Tahmin ve analizlerden elde edilen sonuçların hayvan hastalık teşhislerinde hızlı ve doğru teşhis koyulmasına olanak sunacağı düşünülmektedir.

Anahtar Kelimeler: patoloji, sınıflandırma, karar ağacı, KNN, derin öğrenme

Hazırlanan bu Yüksek Lisans tezi Burdur Mehmet Akif Ersoy Üniversitesi Bilimsel Araştırma Projeleri tarafından 0671-YL-20 proje numarası ile desteklenmiştir.

SUMMARY

M. Sc. Thesis

Analysis and Estimation of Pathological Diagnosis and Findings with Deep Learning Methods

Ahmet Anıl ŞAKIR

**Burdur Mehmet Akif Ersoy University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering**

Supervisor: Assoc. Prof. Ali Hakan IŞIK

January, 2022

As in human diseases, rapid diagnosis of animal diseases is of great importance. In order for the disease treatments to be carried out properly, the diagnosis must be of high accuracy, as well as the rapid diagnosis. In this study, the disease types in the data set consisting of the data examined between the years 2000-2020 belonging to the Department of Pathology of the Faculty of Veterinary Medicine of Burdur Mehmet Akif Ersoy University were estimated by using the decision tree classification model and the KNN classification model. Categories such as age, type, city, and gender in the data set were analyzed in graphics. For the estimation and analysis processes to give accurate results, the data set was corrected by going through some pre-processes and the missing data in the data set was completed. It is thought that the results obtained from the estimation and analysis will allow rapid and accurate diagnosis in animal disease diagnoses.

Keywords: pathology, classification, decision tree, KNN, deep learning

The present M.Sc. Thesis was supported by Burdur Mehmet Akif Ersoy University Scientific Research Projects Under the Project number of 0671-YL-20.

1. GİRİŞ

Patoloji kelimesi Latince kaynaklı olup, kelime anlamı hastalık olan *pathos* ve kelime anlamı bilim olan *logos* kelimelerinin birleşmesi ile meydana gelen bir tanımdır (Slauson ve Cooper, 1990; Carlton ve McGavin, 1995; Cheville, 1999). Patolojinin alt dallarından birisi olan *Medikal Patoloji*, insan hastalıklarıyla ilgilenirken, *Veteriner Patoloji* ise hayvan hastalıkları ile ilgilenmektedir (Slauson ve Cooper, 1990; Kahraman, 1996; Cheville, 1999). Bu tezin konusu olan patolojik veriler, Burdur Mehmet Akif Ersoy Üniversitesi Veteriner Fakültesi Patoloji Anabilim Dalının 2000-2020 yılları arasında incelenen verilerden oluşmaktadır (Şekil 1.1).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	Protokol	İnceleme	Numunenin	Tarih	Telefon N	Şehir/İlçe	Tür	İrk	Cinsiyet	Yaş	Anamnez/Makrosko	Mikrosko	Mikrosko	Hastalık	TI	Patolojik-	Makrosko	Mikrosko	Rapor	Bakteriyel/Virolojik	Genetik	A	Parazitolo	Toksikolojik	Analiz	Sonucu
2	18/058	Nekropsi		2/27/2018			Oğlak			1 günlük	Hepatit, enterit, deri altı ödem	Diğer	Hepatit, pnömoni, mekonyum aspirasyonu													
3	18/059	Nekropsi		2/28/2018			Karamanlı Oğlak			1 hafta	Fibrinli pnömoni	1 hafta	Bakteriyel Fibrinopurulent pnömoni													Pasteurella ve E. Coli pozitif
4	18/060	Nekropsi		2/28/2018			Karamanlı Oğlak			1 hafta	Fibrinli pnömoni	1 hafta	Bakteriyel Fibrinopurulent pnömoni													Pasteurella ve E. Coli pozitif
5	18/152	Nekropsi		4/19/2018			Çanköy/C Oğlak	Honamlı	Erkek	1,5 ay	15 kadar o	Bağırsakla	Bağırsak	İçerisinde	koksidiyoz	çiddetli m	WhatsApp	Image 20	Myokarditis	tanısının	ardından	hasta	sağlık	koksidiyozis		
6	18/061	Nekropsi		2/28/2018			Oğlak			1 hafta			Bakteriyel Fibrinopurulent bronkopnömoni													Pasteurella ve E. Coli pozitif
7	18/062	Nekropsi		2/28/2018			Oğlak			1 hafta			Bakteriyel Fibrinopurulent bronkopnömoni													Pasteurella ve E. Coli pozitif
8	18/063	Nekropsi		3/1/2018			Tavas/DEP Oğlak	Kıl Keçisi	Erkek	10 günlük			Hemorajik abomazitis													Hemorajik gastroenteritis, sepsisemi
9	18/064	Nekropsi		3/1/2018			Yelalan/D Kuzu	Sakız		1 ay			Beyaz kas													Metabolik Beyaz kas, purulent pnömoni
10	18/065	Nekropsi		3/1/2018			Buzağı																			Atrezia ani ve rekti, sakrokoksigeal agenezis
11	18/066	Nekropsi		3/2/2018			Barkaraa Koyun	Pırıt	Dişi	5 yaş	Sinirsel	bulgular														Bakteriyel Listeriozis, sarkosistozis, sepsisemi
12	18/067	Nekropsi		3/2/2018			9 Oğlak			10 günlük																Paraziter Kriptosporidiozis, sepsisemi
13	18/068	Nekropsi		3/2/2018			Beşkavak Oğlak			1 ay			Hemorajik enteritis													Enterotoksemi ?
14	18/069	Nekropsi		3/2/2018			Oğlak			1 günlük																Neonatal sepsisemi
15	18/070	Nekropsi		3/6/2018			Serik Kuzu			15 günlük			Karaciğerde	kanama,	enteritis,	akciğer	od	Enteritis,	sepsisemi							
16	18/071	Organ nu		3/7/2018			Alyon Sığır			Yetişkin			Formalde	Multifokal	kanama	alanları	gözlemlendi.									Multifokal miyozitis

Şekil 1.1. Rutin defterde bulunan verilerin örnek hali (Kişisel bilgi içerdiği için *numune sahibi* ve *telefon numarası* kategorileri gösterilmemiştir)

İnsan hastalıklarında olduğu gibi hayvan hastalıklarında da hızlı teşhisin önemi büyüktür. Teşhisin hızlı olmasının yanında doğruluk değerinin de yüksek olması gerekmektedir. Günümüzde bilgisayar uygulamaları sağlık alanında yaygın olarak kullanılmaktadır. Bunun en güzel örneklerinden biri yapay zekâ uygulamalarının *kanser tedavisi aracı* olarak kullanılmasıdır (Sütçü ve AYTEKİN, 2018).

Rutin defterdeki veriler hayvan türlerine göre değerlendirildiğinde incelenen numunelerin çoğunluğu ruminantlara ait olup bunu köpek, kedi gibi hayvanlar takip etmektedir. Bu numunelerin ölüm sebepleri arasında ergin ruminantlarda sindirim ve solunum sistemi hastalıklarının başta geldiği, yavru ruminantlarda neonatal sepsisemilerin önemli ölçüde ölüme sebebiyet verdiği gözlemlenmiştir. Kedi ve köpeklerde ise en sıklıkla tümörlerle karşılaşmıştır (ÖZMEN, 2006).

Veri işleme ve analizi için *Pandas* kütüphanesi kullanılmıştır (URL-1). *Pandas* kütüphanesinin kullanılmasındaki ana amaç *dataframe* olarak veri iskeleti yapılarını

kullanmasıdır (URL-2). Aynı zamanda farklı formatlardan içe veri aktarma özelliği sayesinde de orijinalde excel içerisinde bulunan rutin defter verilerini basit bir şekilde içe aktarabilmektedir (URL-3). Pandas kütüphanesi aynı zamanda *NumPy*, *Matplotlib* gibi kütüphanelerde birlikte çalışabildiği için bu kütüphaneleri de kullanma olanağı sağlamaktadır. NumPy kütüphanesi matris ve dizi kullanarak, bu diziler üzerinde matematiksel işlemler yapılabilmesini sağlayan bir kütüphanedir (URL-7, URL-8). Pandas kütüphanesi object olarak dataframe kullandığı için bu dataframeleri dizi olarak NumPy ile işlemek mümkündür. Bir diğer kütüphane olan Matplotlib ise, grafik oluşturmayı sağlayan bir kütüphanedir (URL-9). Line, bar, pie plot gibi çeşitli grafikleri dataframe üzerinden kolaylıkla oluşturabilmektedir.

Eksik veri tamamlamada kullanılan metot, *sklearn.impute* kütüphanesinde bulunan *SimpleImputer* metodudur. SimpleImputer metodu verinin biçimine göre (sayısal veya kategorik) farklı *stratejiler* kullanmaktadır (URL-4). Rutin defter içinde bulunan veriler kategorik veri oldukları için, kategorik verilerde kullanılan stratejiler eksik veri tamamlama işlemlerinde kullanılmıştır. Hastalık türü tahmininde kullanılan metotlar ise, *sklearn.tree* kütüphanesinde bulunan *DecisionTreeClassifier* metodu ve *sklearn.neighbors* kütüphanesinde bulunan *KNeighborsClassifier* metodudur (URL-5; URL-13).

2. GENEL BİLGİLER

Hayvanlarda görülen hastalıkların tedavi edilebilmesi ve kontrol altına alınabilmesi için teşhisin önemi büyüktür. Teşhisin hızlı ve doğru koyulabilmesi için ölü veya canlı hayvan numunelerindeki lezyonların değerlendirilmesi önemlidir. Teşhisin konulabilmesi genel ve özel patoloji ve bu kategorilerin teşhis üzerinde uygulanmasının anlaşılabilmesi temeline dayanmaktadır (Jones ve Hunt, 1993; Özmen, 2006). Hastalıkların doğru tedavisi için teşhisin kesin olarak ve doğru yapılması gerekmektedir. Teşhis için numuneler hayvanların ölümünden sonra alınabileceği gibi canlı hayvanlardan da cerrahi işlemler uygulanarak patolojik numuneler elde edilebilir. Bu cerrahi işlemlerde dokuların doğru teknik kullanarak elde edilememesi veya hatalı doku elde edilmesi gibi nedenlerle teşhislerin yanlış koyulmasında, dolayısıyla tedavilerin düzgün yapılamamasına neden olabilir (Brown, 2015; Özmen, 2021). Alınan patolojik numunelerin doğru bir şekilde işlenmesi ve patolojik bulguların düzgün bir şekilde değerlendirilmesi teşhisteki başarı oranını artırmakla kalmayıp, tedavi sürecinde de başarılı olma oranını artırır (Özmen, 2021).

Tablo 2.1’de görüleceği gibi yapay zekâ ve derin öğrenme uygulamaları geçmişten günümüze epey fazla yol kaydetmiş, sadece sağlık alanı değil çoğu alanda yaygın olarak kullanılabilir. Örneğin yapay zekâ uygulamaları geliştiren 100 şirketten sekiz adeti sağlık alanında uygulama geliştirmektedir (Sütçü ve AYTEKİN, 2018). Teknoloji geliştikçe bu sayının ve üretilen uygulamaların artması hem hızlı hem doğruluk oranı yüksek teşhisler geliştirme konusunda insanlığa büyük yarar sağlayabilir. Bu çalışmada da hayvan hastalıklarının analiz edilmesi, tahmin sonucuna bağlı olarak hızlı teşhis konulması ve bunun yanında da doğruluk oranının teşhis hızı ile bağlantılı olarak yüksek olması hedeflenmiştir.

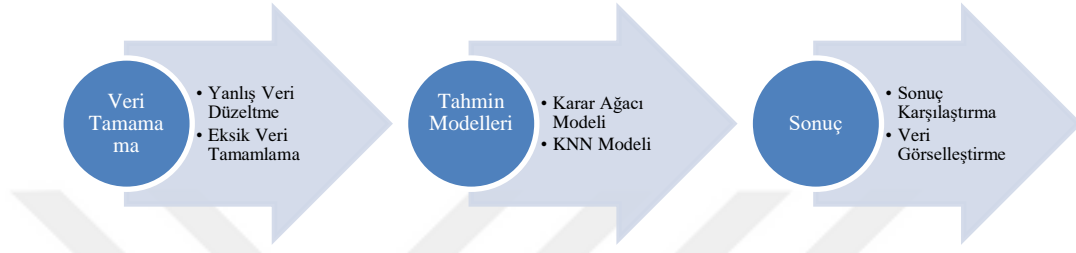
Bu çalışmada hayvan hastalık türlerinin, hayvan türleri ve patolojik-anatomik tanımlar ile tahmini ile rutin defterde mevcut olan birkaç kategori kullanılarak analiz edilmesi amaçlanmıştır.

Tablo 2.1.Kanserin fethi ile sonuçlanan belirli yapay zekâ yeniliklerinin bir zaman çizelgesi (Sütçü ve Aytekin, 2018)

Tarih	Açıklama
1952	Marvin Minsky, ilk bağlantı nöral ağ öğrenme makinesi olan Stochastic Sinirsel Analog Takviye Hesaplayıcısı'nı (SNARC) ve muhtemelen ilk kendi kendine öğrenen makineyi tanıttı.
1975	Bilgisayar destekli makinelerde zorlukları çözen, çok-katmanlı sinir ağlarını eğitimi ve 1980'lerde sinir ağlarının yaygın kullanımını sağlayan geri yayılım algoritması geliştirdi.
2000 civarı	"Derin Öğrenme" ifadesi ilk defa, bir makine öğrenimini, denetlenmemiş bir şekilde yapılandırılmamış verilerden öğrenebilen ağların oluşturulmasını tanımlamak için kullanıldı.
2011-2012	Evrişimli (konvolüsyonel) sinir ağı AlexNet, ana tanıma derinlemesine girmenin yolunu açan, görsel tanımda görülmemiş düzeyde doğruluk elde etti.
Ocak 2017	Stanford Üniversitesi'ndeki araştırmacılar, kanserli deriyi ve lezyonu görsel olarak insan dermatologla aynı hassasiyet düzeyinde tanımlayabilen derin bir öğrenme teknolojisi geliştirdi.
Şubat 2017	Microsoft, yapay zekâ ve makine öğrenme teknolojilerini kanser tedavisi de dahil olmak üzere sağlık sorunlarına uygulamak için tasarlanmış bir girişim olan Healthcare NExT'i kurdu.
Mart 2017	Google,ın GoogleNet derin öğrenme teknolojisi, kanserli tümörleri insan klinisyenlerden daha yüksek doğrulukla tespit etti.
Ekim 2017	Intel, kanser teşhisi dahil olmak üzere derin öğrenme görevlerini hızlandırabilen Nervana Sinir Ağı İşlemcisi (NNP) çipi duyurdu.
2021-2026 Civarı	Microsoft, kanserli hücreleri tespit etmek ve yeniden programlamak ve onları zararsız hale getirmek için insan vücudunun içinde yapay zekâ ile çalışan bir bilgisayarı piyasaya sürecek.

3. MATERYAL VE YÖNTEM

Bu çalışmada izlenecek yolu gösteren akış şeması Şekil 3.1’de verilmiştir. Öncelikle veri tamamlama işlemleri gerçekleştirilmiş, tamamlanmış verileri kullanarak 2 ayrı sınıflandırma modeli oluşturulmuştur. Çıkan sonuçlar karşılaştırılıp daha sonrasında ise veri görselleştirme işlemleri uygulanmıştır.



Şekil 3.1. Çalışmada izlenen akış şeması

3.1. Veri Tamamlama

Analiz ve tahmin işlemlerine geçmeden önce veri setinde eksik veya hatalı olan veriler düzeltilmelidir. Bu sayede daha doğru analizlerin yapılması sağlanırken, tahmin sonucu çıkan başarı oranının artması amaçlanmaktadır. Veri tabanında toplamda 24 farklı kategoriden 5500 yakın inceleme bulunurken, bazı verilerin eksik veya hatalı girildiği görülmektedir. Şekil 3.2’de eksik verilerin kategori bazında değerleri verilmiştir.

Eksik Veriler	
Protokol Numarası	0
İnceleme Türü	2329
Numunenin Sahibi / Geldiği Yer	28
Tarih	120
Telefon Numarası	2893
Şehir/İlçe/Mahalle	2109
Tür	137
Irk	2275
Cinsiyet	1961
Yaş	1421
Anamnez/Klinik Bulgular/Numune Bilgileri	921
Makroskobik Bulgular	5297
Mikroskobik Bulgular	5297
Mikroskobik Tanı	5505
Hastalık Türü	4516
Patolojik-Anatomik Tanı	103
Makroskobik Resimler	5404
Mikroskobik Resimler	5502
Rapor	5485
Bakteriyolojik Analiz Sonucu	5422
Virolojik Analiz Sonucu	5478
Genetik Analiz Sonucu	5505
Parazitolojik Analiz Sonucu	5452
Toksikolojik Analiz Sonucu	5504
dtype: int64	

Şekil 3.2. Veri tabanındaki eksik veriler

3.1.1. Yanlış Veri Düzeltme

Genellikle yanlış harf kullanımı veya küçük büyük harf kullanımı gibi nedenlerden dolayı yanlış verilerin veri tabanında bulunması analiz sonucunu olumsuz etkileyebilmektedir. Şekil 3.3'te tür kategorisine ait değerlerin bir kısmı ile bu değerlerden kaçar adet olduğunu gösterilmektedir. Şekil 3.3'te görüleceği üzere "Oğlak" değeri 3 farklı şekilde yazılmış olup, aynı olan bu değeri 3 ayrı veri kümesine bölünmüştür. Bir diğer örnekte "Sığır" olarak girilmiş olan veride görülebilir. Aynı şekilde "Deve kuşu" ile "devekuşu" örneği yanlış veriye örnek verilebilir. Yani aynı türler birden fazla kez gözüktüğü gibi, bazı türlerin farklı kelimelerle ifade edildiği görülmektedir (sığır ile inek vb.). Tür kategorisinde 301 küme bulunuyor. Yanlış verilerin düzeltilmesinden sonra bu küme sayısı 69'a kadar düşmektedir. Küme sayısındaki bu düşüş ile daha genel ve toplu

kategorize edilmiş bir kategori oluşmuş olup, analiz sonuçlarında da daha net ve kesin sonuçlar çıkmasına olanak sağlamaktadır.

Oğlak	558	iguana	2
Köpek	476	Baykuş	2
Kuzu	360	kanatlı	2
Keçi	315	SÜLÜN	2
Buzağı	254	Lama	2
Sığır	218	Japon balığı	2
Kedi	176	gümüş sülün	2
kuzu	173	Penguen	2
oğlak	172	TAVŞAN YAVRUSU	2
Koyun	162	Sığır	2
OĞLAK	145	ala geyik	2
keçi	139	Deve kuşu	2
köpek	137	Sığırcık	2
NaN	137	devekuşu	2
KÖPEK	133	Yunus	2
Muhabbet kuşu	120	VAŞAK	2

Şekil 3.3. Tür kategorisine ait değerler

Çoğu kategoride harf ve benzer kelime yazımı olarak eksik ve yanlış veri gözüktüğü de yaş kategorisi en düzensiz olan kategori olabilir (Şekil 3.4). Bunun nedeni ise standart tek değerde değil de birden çok değerde veri girilmesidir (günlük, haftalık, aylık vb.). Bunu düzeltmek için ise her bir veri tek değer üzerinden belirtilmelidir. Bunu en iyi ve düzgün olarak günlük şekilde ifade edebiliriz.

NaN	1421
1 aylık	143
2 aylık	127
2 yaşlı	96
3 aylık	91
3 yaşlı	81
10 günlük	80
1 yaşlı	71
1 haftalık	70
1,5 aylık	66
1 günlük	64
4 yaşlı	63
15 günlük	62
2 yaş	57
20 günlük	54

Şekil 3.4. Yaş kategorisine ait değerler

Yaş kategorisinin düzeltmeden sonraki hali Şekil 3.5’te gözükebilir. Düzeltmeden önce 647 farklı kategori bulunurken düzeltmeden sonra ise 149 farklı kategori bulunmaktadır. Bu da daha standart ve etkili bir sonuç elde etmek için düzeltilmesi gerektiğini göstermektedir.

NaN	1421
0	293
10	268
30	219
1	209
365	207
60	204
730	191
15	177
7	170
1460	137
90	133
45	123
3	112
2	109

Şekil 3.5. Düzenlemeden sonraki yaş kategorisi

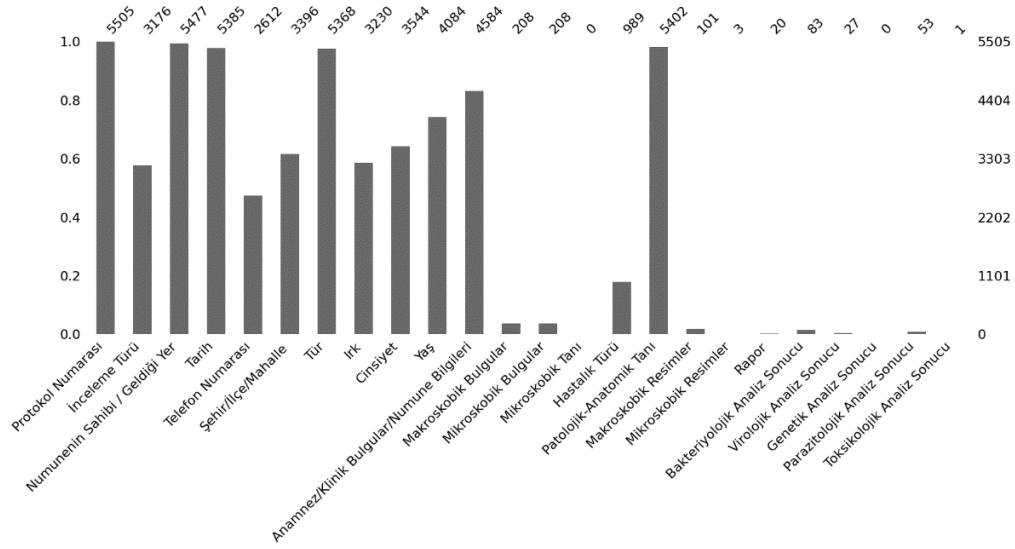
Yaş kategorisinde gözüken “NaN” değeri not a number yani sayı olmayan veri anlamına gelir. Bazı kategoriler orijinal halde *object* olarak geçtiği ve bunların düzenlenmesi için *string* değerinde olması gerekmektedir. Bu kategorileri *string* değerine dönüştüğünde, orijinal halde “null” olarak geçen boş değerler *string* halinde boş olarak gözükmemektedir. Bu nedenle “null” değeri yerine “NaN” değeri kullanılmaktadır (Şekil 3.6).

Protokol Numarası	object
İnceleme Türü	object
Numunenin Sahibi / Geldiği Yer	object
Tarih	datetime64[ns]
Telefon Numarası	object
Şehir/İlçe/Mahalle	object
Tür	object
İrk	object
Cinsiyet	object
Yaş	object
Anamnez/Klinik Bulgular/Numune Bilgileri	object
Makroskopik Bulgular	object
Mikroskopik Bulgular	object
Mikroskopik Tanı	float64
Hastalık Türü	object
Patolojik-Anatomik Tanı	object
Makroskopik Resimler	object
Mikroskopik Resimler	object
Rapor	object
Bakteriyolojik Analiz Sonucu	object
Virolojik Analiz Sonucu	object
Genetik Analiz Sonucu	float64
Parazitolojik Analiz Sonucu	object
Toksikolojik Analiz Sonucu	object
dtype:	object

Şekil 3.6. Veri türleri

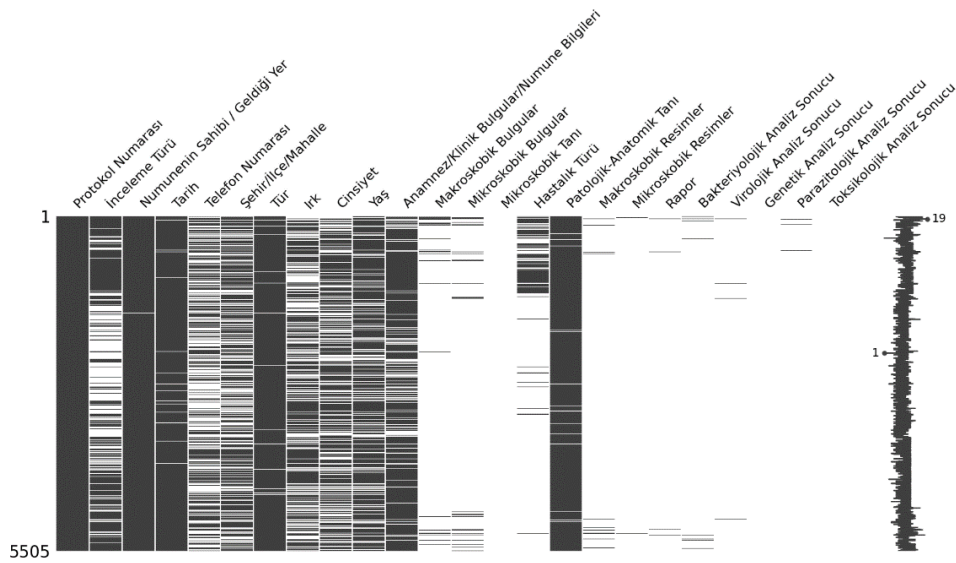
3.1.2. Eksik Veri

Eksik veri tamamlanması yapılmadan önce veri görselleştirilmesi yapılarak hangi verilerde eksiklik olduğu belirlenmiştir. Bu sayede eksik veri tamamlanması daha basit bir şekilde gerçekleşmesi amaçlanmıştır. Eksik verilerin görselleştirilmesinde *missingno* kütüphanesine ait bar, matrix ve heatmap (ısı haritası) grafikleri kullanılmıştır (URL-6).



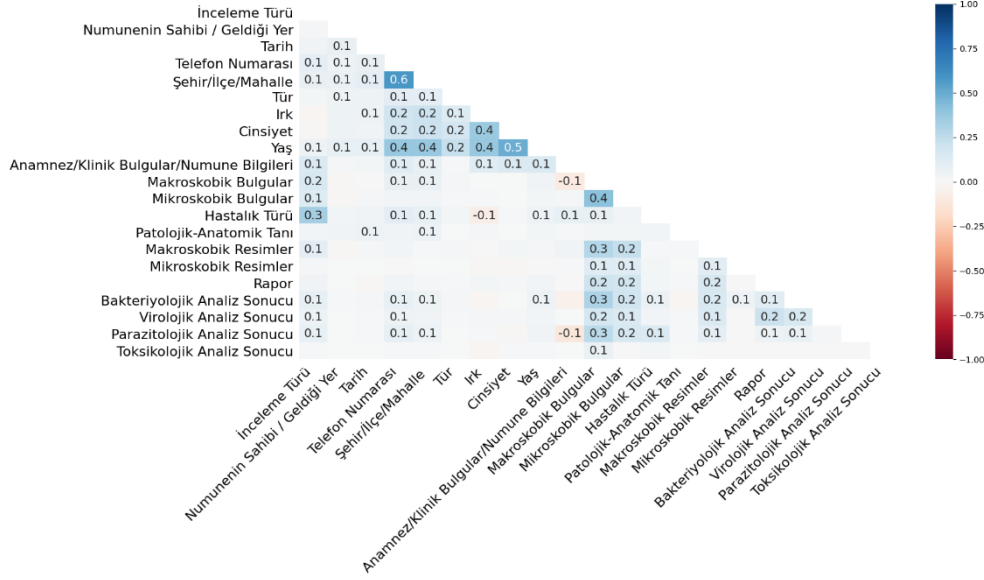
Şekil 3.7. Eksik verilerin bar grafiği ile gösterimi

Şekil 3.7’de gösterilen grafik eksik verilerin kategorilerine göre dağılımını vermektedir. Protokol numarası her girdi için bir tanımlayıcı biçimde olduğu için eksiksiz olduğu görülmektedir. Çok eksik olan kategorilerde ise hastalık türü kategorisi dışında kalanlar sonuca pek etki etmeyeceği için inceleme ve analizlerde kullanılmayacaktır. Şekil 3.8’de ise eksik verilerin matrix grafiği verilmiştir. Matrix grafiği eksik verileri konumlarına göre gösterdiği için bar grafiğine göre daha detaylı bir görünüm vermektedir. Matrix grafiğine göre sütun bazında fazla eksik veri içeren kategorilerin analizi veya tamamlanması zor olacaktır.



Şekil 3.8. Eksik verilerin matrix grafiği ile gösterimi

Eksik verileri göstermek için son olarak heatmap yani ısı haritası grafiği Şekil 3.9’da verilmiştir. Diğer iki grafiğe göre daha karmaşık bir yapıda olan ısı haritası grafiğinde, kategoriler arasındaki eksik veri ilişkilerini görmek mümkündür.



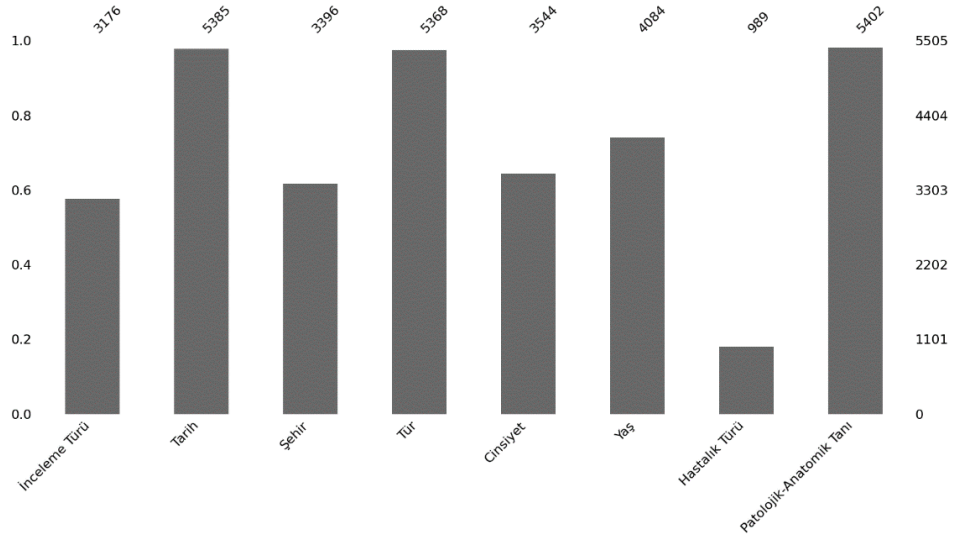
Şekil 3.9. Eksik verilerin heatmap grafiği ile gösterimi

Isı haritası grafiği -1 ile 1 değer aralığında kategoriler arası ilişkileri gösterir. -1 değeri, bir kategoride görülen eksik verinin diğer kategoride kesinlikle eksik olmadığını gösterir. 0 değeri kategorilerdeki eksik verilerin birbirleriyle bir alakası bulunmadığını gösterirken, 1 değeri ise bir kategoride görülen eksik veri diğer kategoride de kesinlikle eksik olduğu anlamını taşır. Tamamen dolu veya tamamen boş kategorileri ısı haritasında göstermek anlamsız olacağından çıkarılması gerekmektedir. Bu sebepten ötürü *Protokol Numarası* kategorisi eksiksiz olduğu için ısı haritası grafiğine dahil edilmemiştir. -1 ile 1 arası kalan değerlerde ise kesin olmamakla beraber yakın olduğu değer sonucuna bağlı olarak kategoriler arası ilişkiyi gösterir.

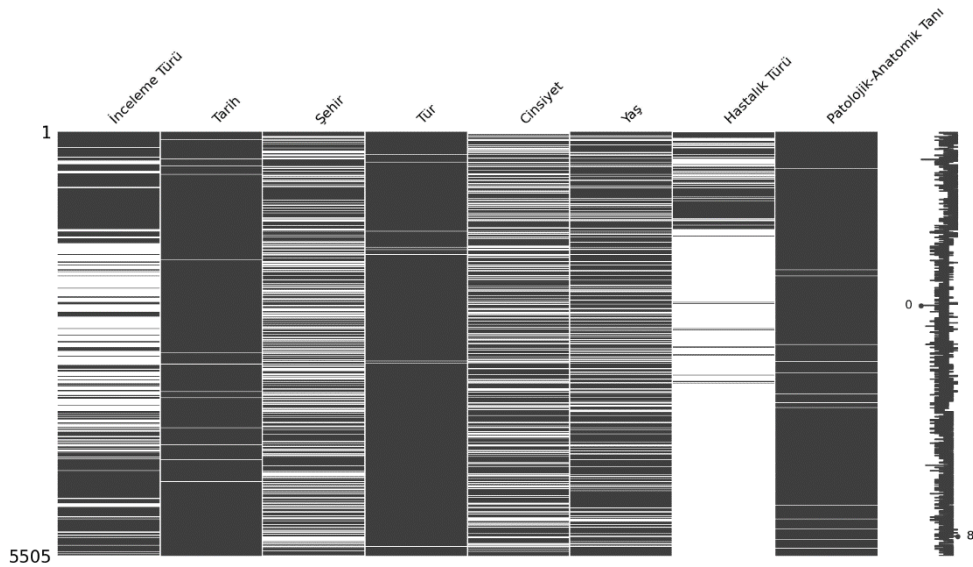
3.1.3. Eksik Veri Tamamlama

Veri setindeki hatalı veriler düzeltildikten sonra, eksik veri tamamlama işlemleri uygulanabilir. Öncelikle eksik kategorilerin çıkarılması ve hatalı verilerin düzeltilmesinden sonraki eksik veri sayısını gösteren bar grafiği Şekil 3.10’da verilmiştir. Eksik verilerin matrix grafiği gösterimi ise Şekil 3.11’de verilmiştir. Eksik verilerin, bütün veri sayısına

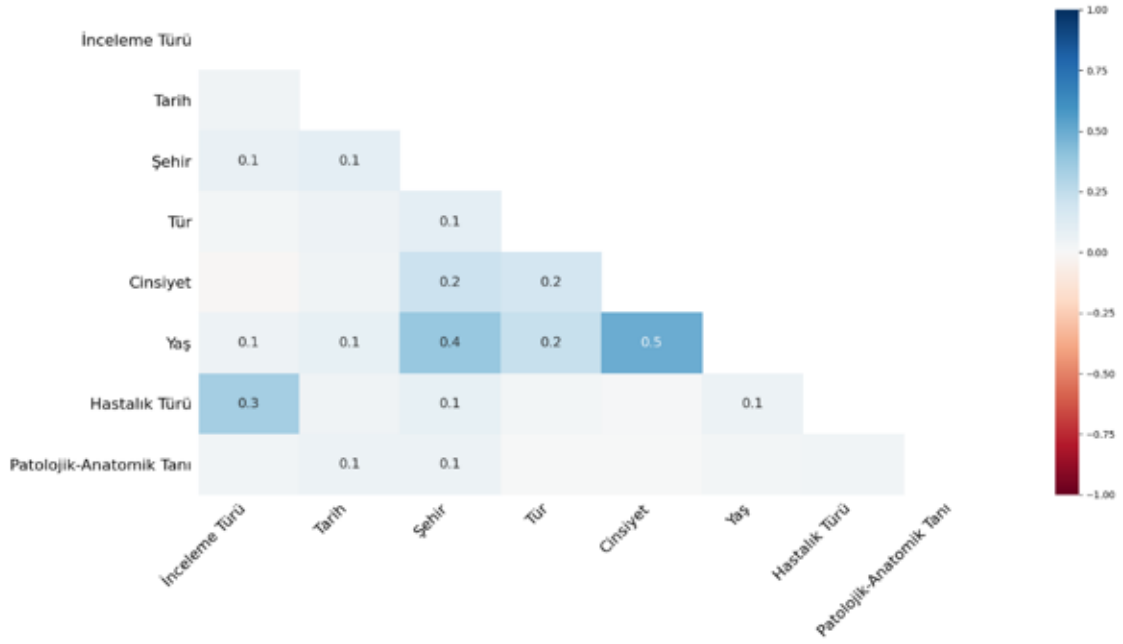
bölünmesiyle veri setindeki eksik veri oranını öğrenebiliriz. Yüzdeler olarak öğrenmemiz içinde elde ettiğimiz sonucu 100 ile çarpmamız yeterli olacaktır. Bu işlemlerin yapılmasından sonra veri setinde %59,54 oranında eksik veri olduğu hesaplanmıştır. Kategori çıkarılma işlemi yapıldıktan sonraki veri setindeki, analiz ve tahmin işlemlerinde kullanılacak olan eksik veri miktarı ise %28,89 oranındadır. Çıkan bu sonuçta, neredeyse her 4 veriden 1 tanesinin eksik olduğu göstermektedir.



Şekil 3.10. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setindeki eksik veri sayısı



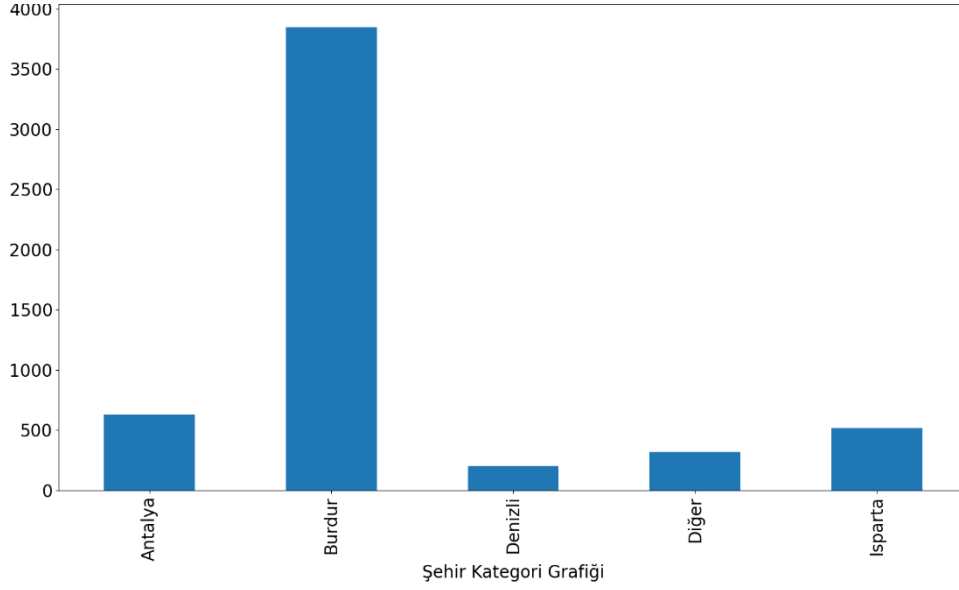
Şekil 3.11. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setinin matrix grafiği ile gösterimi



Şekil 3.12. Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra veri setinin heatmap grafiği ile gösterimi

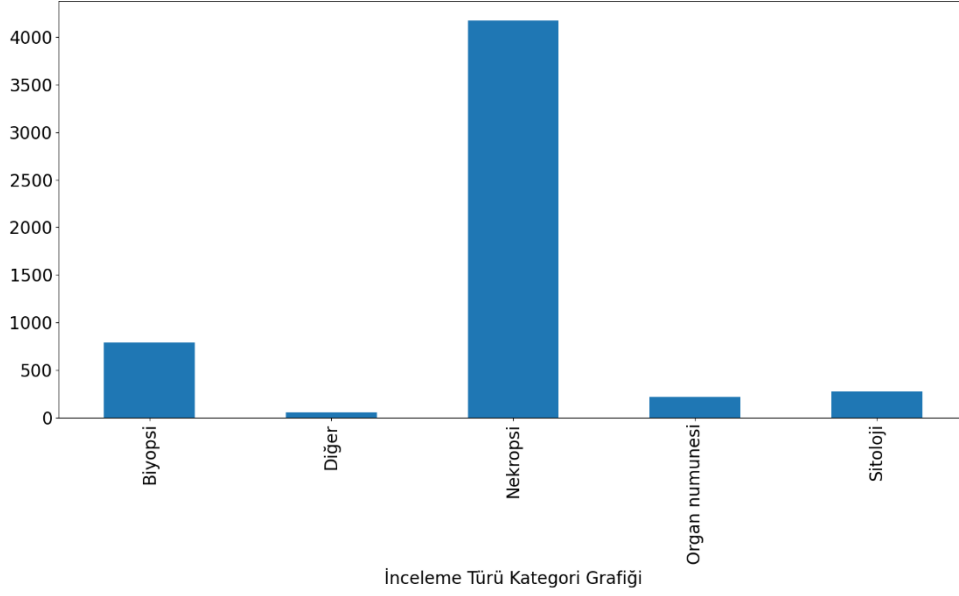
Kategori çıkarma ve yanlış veri düzeltme işlemlerinden sonraki ısı haritası Şekil 3.12’de görülmektedir. Daha önceden de belirtildiği gibi ısı haritası -1 ile 1 arası değerlerden oluşmaktadır. -1 ile 1 arasında kalan değerlerde ise kategorilerden herhangi biri veya diğerini içeren ancak her ikisini birden olmayan verileri göstermektedir. Isı haritası kesin değerli kategorilerin tamamlanmasında kullanılabilir önemli bir görselleştirme tekniğidir.

Eksik kategoriler çıkarıldıktan ve yanlış veriler düzeltildikten sonra çıkan veri grafiklerine göre, *Hastalık Türü* kategorisi dışında kalan kategorilerin yarısından fazlasının eksik olmadığı görülmektedir. Bu kategoriler arasında *Tarih* kategorisi eksik veri tamamlama olarak en basit kategoridir. Veri setine girilen tarih aralığına bakarak eksik verilerin tamamlanması mümkündür. Bu işlemi de eksik olan veriden önce gelen verinin tarihine bakılarak yapılabilir. Bu yöntem ile *Tarih* kategorisi eksiksiz olarak doldurulmuştur.

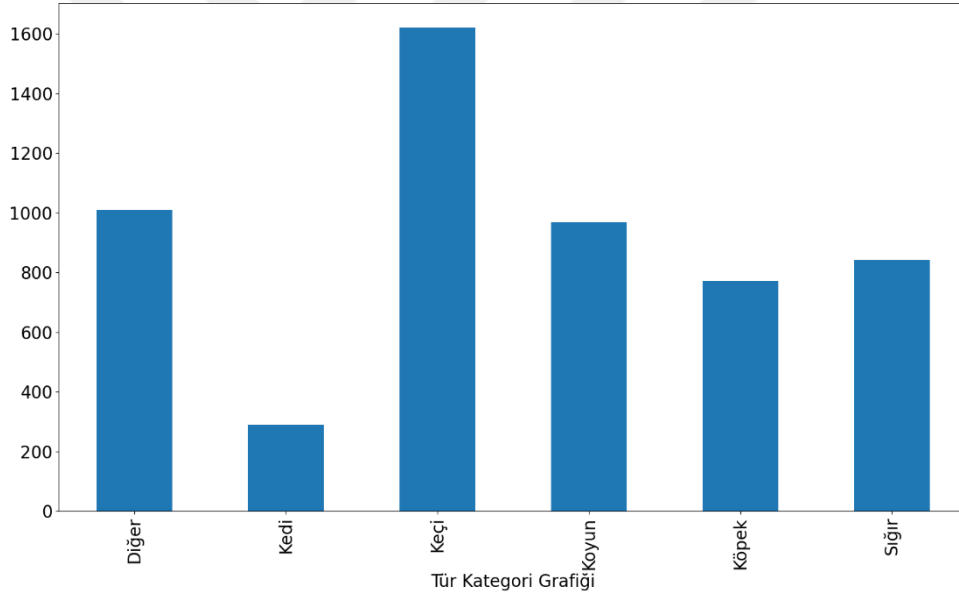


Şekil 3.13. Şehirlerin kategori sayısı grafiği

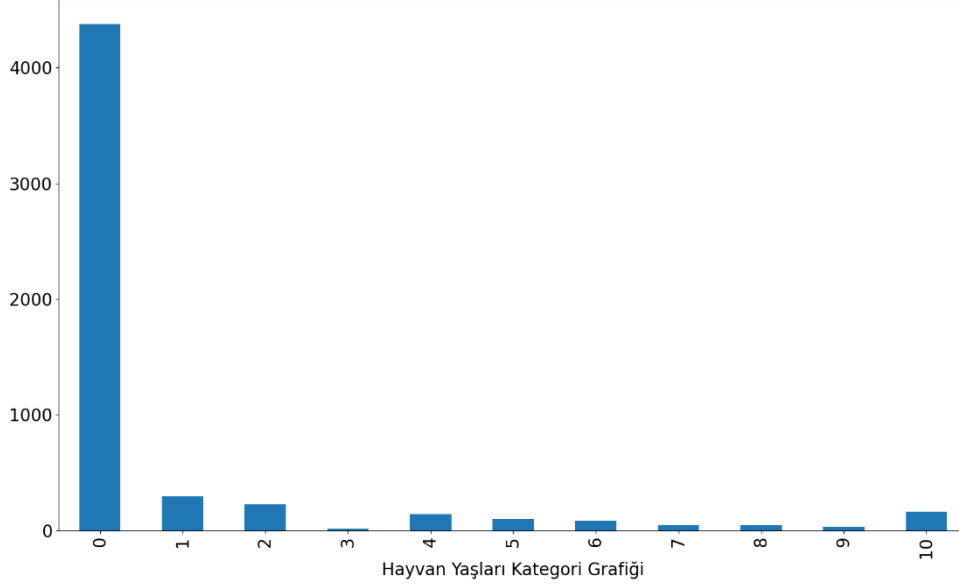
El ile tamamlanabilecek bir diğer kategori ise numunelerin geldiği şehir kategorisidir (Şekil 3.13). Numune sahipleri ile şehir kategorileri karşılaştırılarak eksik veriler tamamlanmıştır. Geri kalan kategorileri doldurabilmek için *sklearn.impute* kütüphanesine ait *SimpleImputation* metodu kullanılabilir. Bu metot hem sayısal hem kategorik verilere uygun bir şekilde veri tamamlama işlemi uygulamaktadır. Kategorik veriler için iki ayrı strateji bulunmaktadır. Bunlar: *most frequent*, yani veri tabanında en sık bulunan veriler ve *constant* yani sabit bir değer girilmesidir (URL-4). Geriye kalan eksik kategorilerde *most frequent* metodu uygulanmış olup *Hastalık Türü* harici kategorilerdeki eksik veriler tamamlanmıştır. İnceleme türü, hayvan türleri ve hayvan yaşları kategori grafikleri Şekil 3.14-3.16’da verilmiştir.



Şekil 3.14. İnceleme türü kategori sayısı grafiği

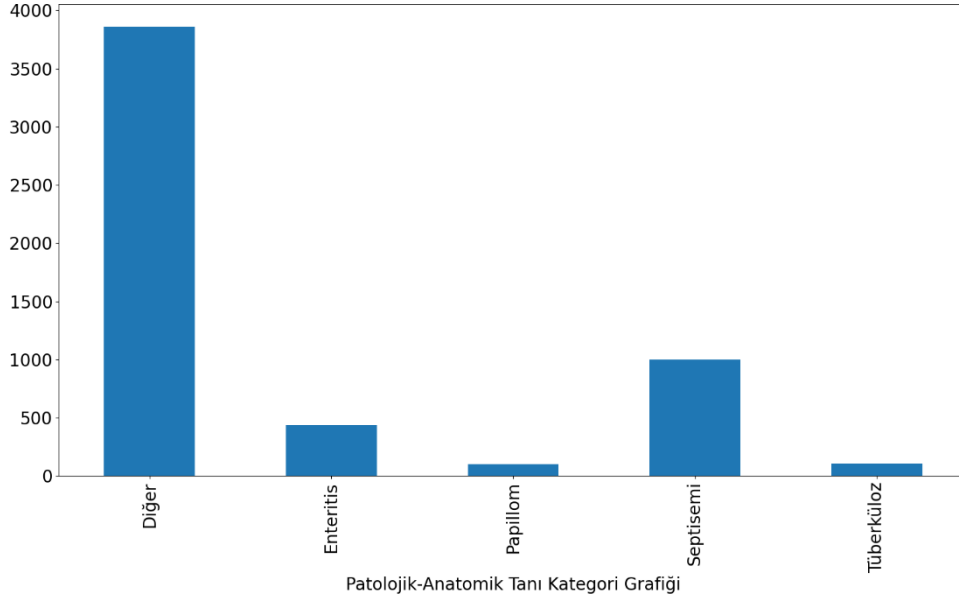


Şekil 3.15. Hayvan türleri kategori sayısı grafiği

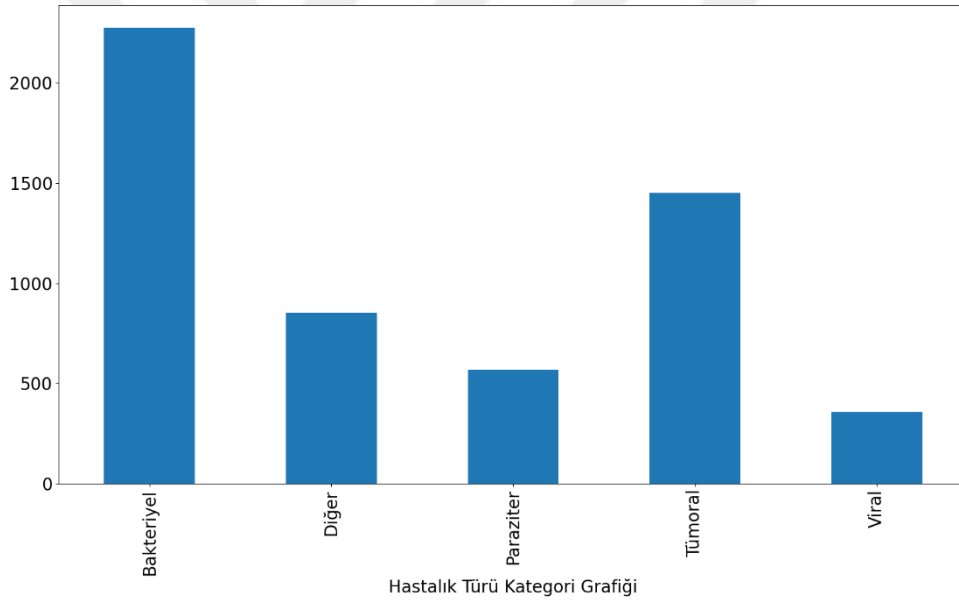


Şekil 3.16. Hayvan yaşları kategori sayısı grafiği (10 yaş ve üzeri hayvanlar 10 kategorisine dahil edilmiştir)

Hastalık Türü kategorisinin *imputation* metodu ile tamamlanmamasının nedeni, kategorideki veri sayısının çok fazla eksik olması ve bu metot ile tamamlanırsa çıkacak sonucun kategorideki en çok bulunan veri ile aynı olacağıdır. *Hastalık Türü* kategorisinin tamamlanabilmesi için *Patolojik-Anatomik Tanı* kategorisiyle karşılaştırılıp, çıkan sonuçların ortak değerlerine göre alınması gerekmektedir (Şekil 3.17-3.18). Bu işlem sonucunda 5 ana kategori olarak sayılabilecek olan, bakteriyel, tümoral, paraziter viral ve diğer olarak ayrılmıştır. Diğer kategorisi geriye kalan dört kategorinin birlikte olduğu durumları içerdiği gibi anomaliler ve travmatik lezyonlarda içinde barındırmaktadır.



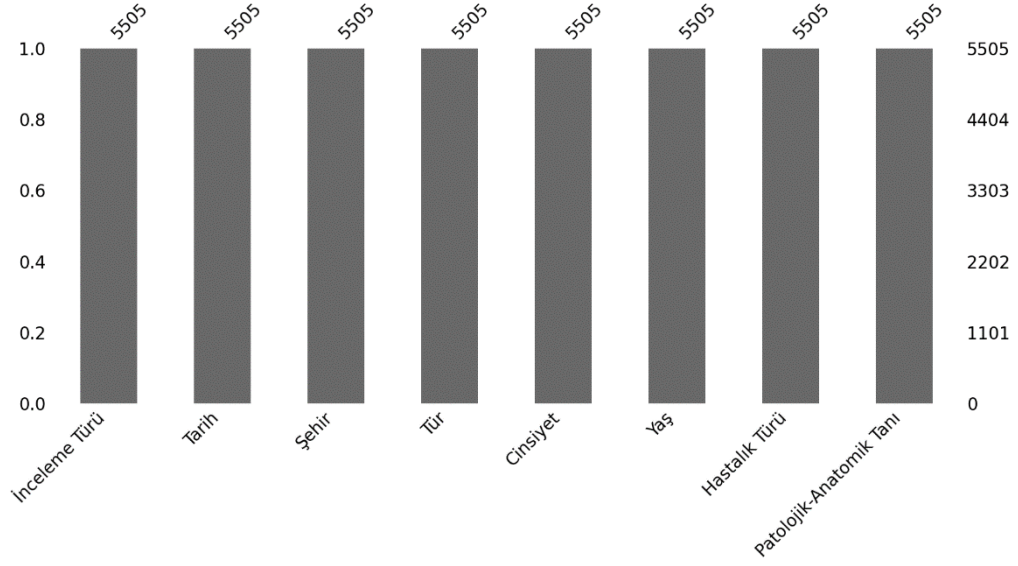
Şekil 3.17. Patolojik-anatomik tanı kategori sayısı grafiği



Şekil 3.18. Hastalık türleri kategori sayısı grafiği

3.2. Karar Ağacı Metodu ile Tahmin

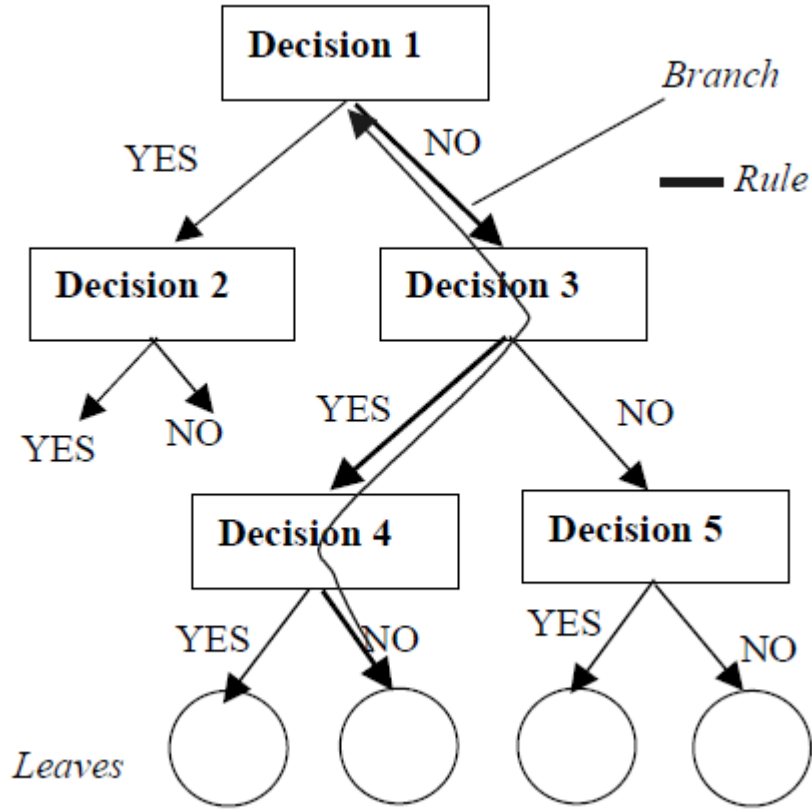
Veri tamamlama işlemleri tamamlandıktan sonra model tahmini işlemleri eksiksiz veri seti üzerinde gerçekleştirilmiştir (Şekil 3.19).



Şekil 3.19. Veri tamamlama sonucu kategorilerin eksik veri sayısı

Tahmin modelleri sınıflandırma (classification) ve regresyon (regression) olmak üzere iki farklı şekilde ifade edilebilir. Karar ağacı metodu ile elde edilen tahmin sınıflandırma olarak sonuçlanmaktadır. Sınıflandırma ise verilerin ortak özelliklerine göre önceden belirlenmiş sınıflara tanımlanmasıdır. Bu tezde kullanılan veri seti içerisinde bulunan Hastalık Türü kategorisi sınıflandırma için kullanılmıştır. Bu nedenle kategori bakteriyel, tümöral, paraziter, viral ve diğer olmak üzere 5 ayrı sınıfa ayrılmıştır.

Sınıflandırma için kullanılan *Karar Ağacı* metodu akış şeması grafiğine benzer yapıda olup verilerin belirlenen sınıflara ayrılmasını sağlar (Şekil 3.20). Sınıflandırma yapılırken *tümevarım* yöntemi kullanılır. Karar ağacı metodunda sınıflandırma sırasında çok sayıda test gerçekleştirerek, sınıflandırma sonucunun en iyi şekilde çıkması sağlanır. Bu işlem sırasında her bir test, karar ağacı dallarından birini oluşturur. Test işleminin sona ermesi için yaprak düğümüne (leaf node) ulaşması gerekir. Ağaç kökünden sınıflandırma sonucunu içeren yaprak düğümüne kadar olan sıralama, eğer-sonra (if-then) yapısında olan *kural* olarak adlandırılır (Emel ve Taşkın, 2005).



Şekil 3.20. Örnek bir karar ağacı şeması (Bounsaythip ve Esa, 2001)

Karar ağacında kullanılan tümevarım yöntemine ağaç tümevarımı veya ağaç indüksiyonu (tree induction) denir. Bu yöntem boş ağaç ile başlanır. Bu ağacı doldurmak için kullanılan ağaç indüksiyonu yinelemeli bir yöntem olup (recursive), dört adımdan oluşur (Emel ve Taşkın, 2005; Zorman vd., 2001):

- 1.Adım: Eğitim seti (training set) içindeki eğitim nesnelerinin sonucu aynı olması durumunda, bu sonuç kullanılarak bir yaprak düğümü (leaf node) oluşturulur ve 4. adımdan devam edilir.
- 2.Adım: Sezgisel veya bulgusal değerlendirme fonksiyonu (heuristic evaluation function) kullanılarak ağaç kökünden başlayarak mevcut düğüm giden yolda daha önce kullanılmamış nitelikler içinden en iyi nitelik seçilir ve bu niteliğe ait olarak bölünmüş bir düğüm oluşturulur. Eğitim seti alt kümelere ayrılır.
- 3.Adım: Oluşturulan her bir alt eğitim seti için 1.adımdan devam edilir.
- 4.Adım: Bir üst seviyeye çıkılarak yineleme işlemi (recursion) uygulanır.

Bu dört adım sonucunda temel olarak iki işlem gerçekleşmektedir. Bunlar, bölme (splitting) işlemi ve budama (pruning) işlemidir. Bu işlemlerden sonra yineleme yönteminin bitmesi için durdurma kriteri gelmektedir (Emel ve Taşkın, 2005).

Bölme işlemi, eğitim setinin daha küçük alt kümelerine ayrılmasını sağlayan bir yöntemdir. Yinelemeli bir yöntem olan bölme işleminde ilk yineleme, ağaç kökünü de içeren bütün eğitim setini kapsar. Geriye kalan yinelemeler ise eğitim setinin alt kümelerini de içine alan türev düğümler kullanılarak işlenir. Her bir bölme adımında, veriler analiz edilerek en iyi sınıflandırma seçilir. Bölme işleminin en önemli karakteristik özelliği açgözlü (greddy) olmasıdır, bu nedenle bu yöntemde algoritma en iyi sonucu elde edip edilmediğini bulmak için ağaç üzerinde ilerleyen adımlara bakmaz (Emel ve Taşkın, 2005; Bounsaythip ve Esa, 2001).

Karar ağacının oluşabilmesi için yinelenen yöntemin durması gerekmektedir. Bu durdurma işlemi için durdurma kriterleri kullanılır. Durdurma kriterleri genellikle, maksimum ağaç derinliği, bölme işlemi için ele alınan düğüm içindeki minimum öge sayısı veya yeni düğüm içerisinde yer alması gereken minimum öge sayısı gibi birkaç kural içermektedirler. Kullanılan veri türüne veya kullanıcının isteğine göre bu kurallarla ilişkili parametreleri değiştirebilir. Genellikle bu yöntemi kullanan uygulamalar ağaçları maksimum derinlikte inşa ederler. Maksimum derinlikteki böyle bir ağaç eğitim setindeki tüm nesnelere kesin bir olasılıkla tahmin ederken, büyük ihtimalle verilere aşırı öğrenme (overfitting) gösterirler (Bounsaythip ve Esa, 2001).

Karar ağacı oluştuktan sonra, fazla uyum nedeniyle istenmeyen düğümleri veya alt kümeleri kaldırmak için budama yöntemi kullanılır. Budama yöntemi, bölmeleri ve onların oluşturduğu alt kümeleri kaldırır ve karar ağacını daha kararlı bir hale getirir. Maksimum derinlikle ağaç oluşturan uygulamalar otomatik olarak budama yöntemi içerirler. Karar ağacı eğitimi tamamlandıktan sonra, ağacın tepe kısmından başlanarak, herhangi bir sonuç düğümüne erişene kadar oluşan yol kullanılarak yeni verileri için tahmin işlemi uygulanabilir (Bounsaythip ve Esa, 2001).

Sklearn kütüphanesine ait DecisionTreeClassifier, sınıflandırma yapmak için kullanılan bir kütüphanedir. Sınıflandırma yapmak için öncelikle sınıflandırmada kullanılacak olan kategoriler veya sütunlar ile sınıflandırma sonucu içeren sütun ayrı dataframelere bölünmelidir. *Hastalık Türü* sınıflandırması yapabilmek için *Tür* ile *Patolojik-Anatomik Tanı* sütunları ayrılmıştır. Bazı durumlarda ayrılan kategoriler sınıflandırılmadan

önce bir takım ön işlemlerden geçmeleri gerekmektedir. Burada metin içeren kategorilerdeki verilerin numerik değere dönüştürülmesi için *one-hot* metodu kullanılmaktadır. Pandas kütüphanesinde *one-hot* metodu tek satır kod olarak *pandas.get_dummies* metodu ile kolay bir şekilde işlenebilmektedir (URL-10). Sınıflandırmada kullanılacak olan kategoriler hazır olduğunda model eğitime başlanılabilir. Modelin ne kadar başarılı olduğunu belirlenmesi için model eğitiminden önce veri setinin eğitim seti ve test seti olarak ikiye ayrılması gerekmektedir. Bu ayırma işlemi için *Sklearn.model_selection* kütüphanesine ait *train_test_split* metodu kullanılmıştır (URL-11). Bu metoda ait parametrelerden birisi olan *test_size* ile ayırma oranı belirlenebilir. Bu veri setini girilen parametre boyutu kadar test seti, geri kalanı eğitim seti olarak ayırmamıza olanak sağlamıştır. Veri seti, eğitim ve test olmak üzere ikiye ayrıldıktan sonra sınıflandırma için model eğitime başlanmadan önce sınıflandırma için kullanılacak olan modelin tanımlanması gerekir. Bu tanımlama işlemini *Sklearn.tree* kütüphanesine ait olan *DecisionTreeClassifier* fonksiyonu çağırılarak yapılabilir. Tanımlama işleminin ardından *fit* fonksiyonu kullanılarak eğitim tamamlanır (URL-5). Eğitimin tamamlanmasının ardından başarı oranı ölçülebilir hale gelir ve model yeni girilecek veriler için tahmin yapmak için hazır duruma gelir. Başarı oranını ölçmek için *DecisionTreeClassifier* fonksiyonuna ait olan *predict* parametresi kullanılarak test setindeki değerler eğitilmiş model üzerinden tahmin edilir. Daha sonra bu tahminler ile gerçek sonuç değerleri karşılaştırılması için *Sklearn.metrics* kütüphanesine ait *accuracy_score* metodu kullanılmıştır (URL-12).

3.3. KNN Metodu ile Tahmin

Bu çalışmada kullanılacak diğer sınıflandırma yöntemi olan KNN (k-nearest neighbors), *K en yakın komşu yöntemi*, ise eğitim kümesindeki en yakın komşuların ortalama değerine göre sınıflandırma yapmaktadır. Buradaki *k* değeri seçilecek komşu sayısını göstermektedir. En yakın komşu bulmak için, bulunacak noktaya en yakın olan örnek ile aralarındaki mesafenin hesaplanması gerekmektedir. Aradaki bu mesafenin ölçülmesi için, Öklid, Manhattan, Minkowski gibi farklı uzaklık metrik ölçü yöntemleri kullanılabilir. Yaygın olarak Öklid (Euclidean) uzaklığı bu yöntemler arasındaki en çok tercih edilen yöntemdir (URL-14; URL-15).

İki nokta arası (Örn. $P = (p_1, p_2, \dots, p_n)$ ve $Q = (q_1, q_2, \dots, q_n)$ noktaları) doğrusal uzaklığın hesaplanmasında kullanılan Öklid uzaklığı formül 3.1'de verilmiştir (URL-14).

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (3.1)$$

KNN için tasarlanan *Sklearn.neighbors* kütüphanesine ait *KNeighborsClassifier* metodu mesafe ölçümü gibi ön adımları bünyesinde barındıran bir metottur (URL-13). Karar ağacı yönteminde de olduğu gibi model eğitime başlanmadan önce veri seti ön işlemlerden geçip, eğitim seti ve test seti olarak ikiye ayrılması gerekir. Modelin tanımlanması için ise *KNeighborsClassifier* metodu, komşu sayısını gösteren k parametresi ile kullanılmıştır (URL-13). Bu işlemin ardından *fit* metodu ile veri setleri modele aktarılır ve eğitim tamamlanmış olur. *DecisionTreeClassifier* kütüphanesinde bulunan *predict* parametresi *KNeighborsClassifier* metodunda da bulunmaktadır. Bu parametre ile tahmin işlemi gerçekleştirilir. Başarı oranı için ise yine *sklearn.metrics* kütüphanesine ait *accuracy_score* metodu kullanılır (URL-12).

3.4. RMSE, MSE ve MAE Yöntemleri ile Hata Payı Hesaplama

Tahmin edilen sonuç ile gerçek değerler arasındaki farkın hesaplanması için kullanılan yöntemler arasında MSE (mean squared error), RMSE (root mean squared error) ve MAE (mean absolute error) gibi yöntemler bulunmaktadır. MSE yani ortalama mutlak hata payı, veri setindeki her iki değişken arasındaki mutlak farkın ortalama değeridir. MAE sonucu çıkan değer 0'a yakın olması, değer en iyi sonucu verdiğini gösterir. MAE hesaplanmasını sağlayan formül 3.2'de verilmiştir (URL-18).

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}. \quad (3.2)$$

Ortalama kare hata olarak geçen MSE yöntemi ise, tahmin edilen sonuçların veri setindeki gerçek sonuçlardan ne kadar farklı olduğunu belirten bir mutlak sayı vermektedir. MSE değeri de 0'a yaklaştıkça en iyi sonuç ortaya çıkmaktadır. MSE formül 3.3'te verilmiştir (URL-19).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2. \quad (3.3)$$

RMSE (veya RMSD), kök ortalama kare hata yöntemi ise MSE yöntemi sonucunun karekökü olup, MSE yöntemine göre daha küçük sonuçlar vermektedir. Bu nedenle genellikle MSE yöntemine göre daha çok tercih edilir. RMSE formül 3.4'te verilmiştir (URL-20).

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}. \quad (3.4)$$

Bu çalışmadaki tahminler ile gerçek sonuç arasında oluşan hata payını ölçmek için *sklearn.metrics* kütüphanesine ait *mean_squared_error* ile *mean_absolute_error* metotları kullanılmıştır. *mean_squared_error* metodunda bulunan *squared* parametresinin *true* veya *false* şeklinde değiştirilmesi ile MSE ile RMSE yöntemleri arasında geçiş yapılabilir (URL-16).

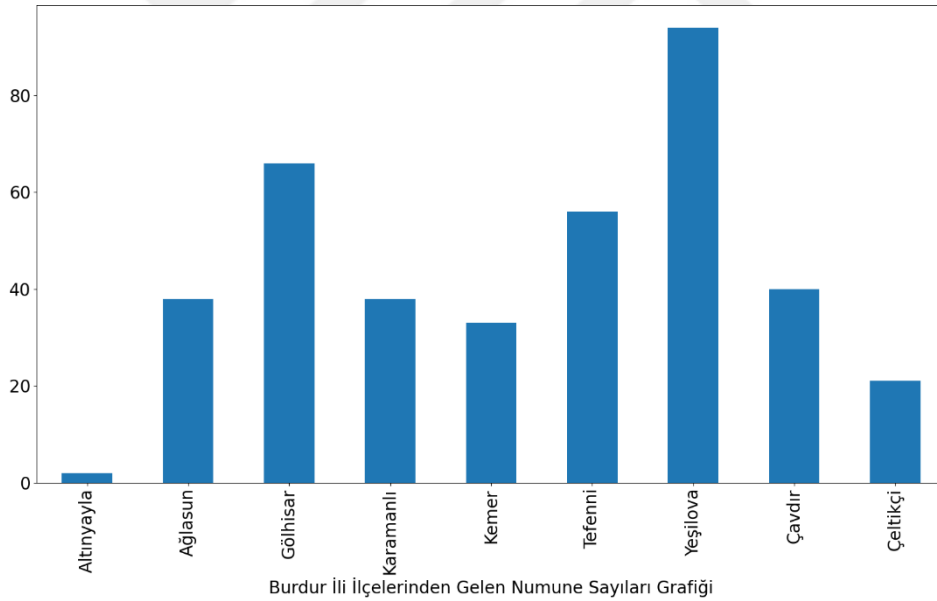
4. ARAŞTIRMA BULGULARI VE TARTIŞMA

4.1. Grafik Bulguları

Rutin deftere ait kategoriler üzerinde yapılan analizler sonucu bazı grafikler çıkarılmıştır. Bu grafiklerin bir kısmı *Pandas* ve *Matplotlib* kütüphaneleri kullanılarak direk *dataframe*ler üzerinden çıkarılırken, bir kısmı içinse SPSS 22.00 versiyonlu program kullanılmıştır. Çıkan grafiklerin okunabilir ve düzgün olması için kategoride en çok bulunan değerler gösterilmiş, geri kalan değerler ise diğer olarak toplu bir şekilde düzenlenmiştir.

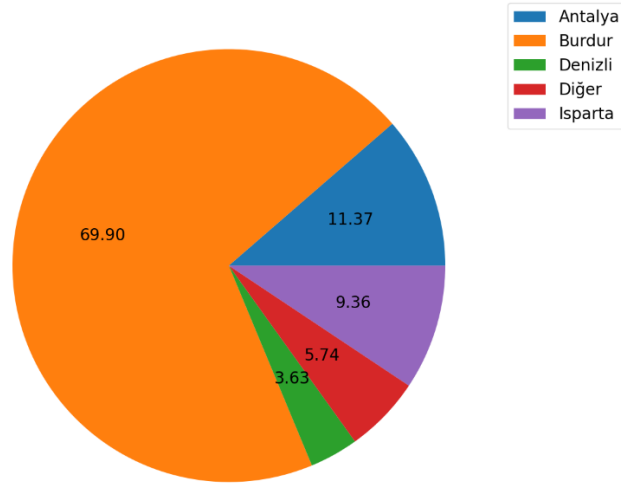
4.1.1. Analiz Sonucu Elde Edilen Grafikler

Rutin defter içerisindeki şehir kategorisinde 24 ayrı değer bulunmaktadır. Bunlar numunelerin geldiği il sayısıdır. İller haricinde Burdur ili ilçelerinden gelen numunelerde bulunmaktadır. Bunlar il kategorisinde Burdur içinde gösterilmiştir. Burdur iline ait ilçelerden gelen numune sayısı grafiği şekil 4.1’de verilmiştir.



Şekil 4.1. Burdur ili ilçelerin kategori sayısı grafiği

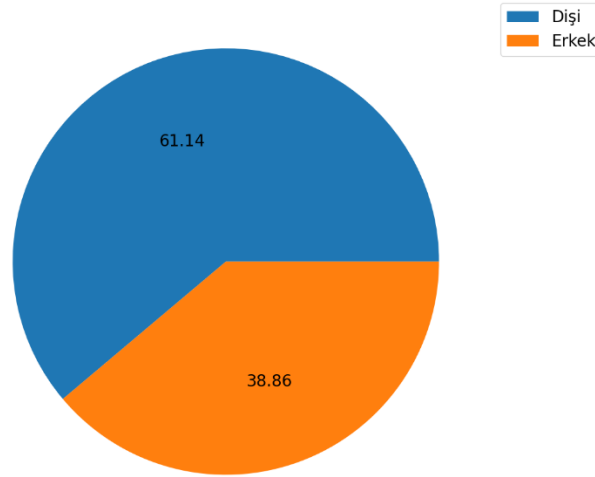
Bu grafiğe göre en çok numune Burdur ili içerisinde Yeşilova ilçesinden gelmiştir. İl bazında ise en çok numune Burdur ilinden gelmiştir (Şekil 4.2).



Şehir Oran Grafiği

Şekil 4.2. Şehirlerin kategori oranını gösteren *Pie* grafiği

Veri setindeki bütün hayvan türlerinin cinsiyet grafiği şekil 4.3'te verilmiştir. Bu grafiğe göre 2000-2020 yılları arasında gelen numunelerin çoğunluğunu dişi hayvanlar oluşturmaktadır.

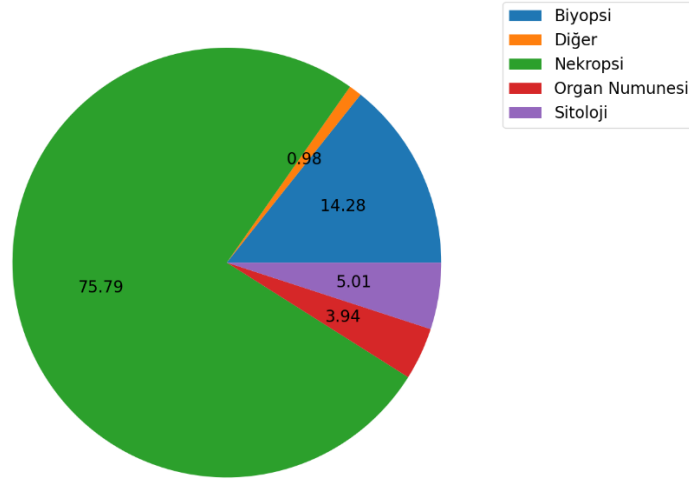


Cinsiyet Oran Grafiği

Şekil 4.3. Hayvan cinsiyet oranını gösteren *Pie* grafiği

Gelen numunelere uygulanan İnceleme Türü kategorisi ise dört ana alana ayrılmıştır. Bunlar; Biyopsi, Nekropsi, Sitoloji, Organ Numunesidir. Bu dört alanın birleşimi ise diğer kategorisi içerisinde gösterilmiştir (Örn. Biyopsi;Organ Numunesi). Bu alanların dağılımı Şekil 4.4'te gösterilmektedir. Numunelerde uygulanan inceleme türleri arasında en çok

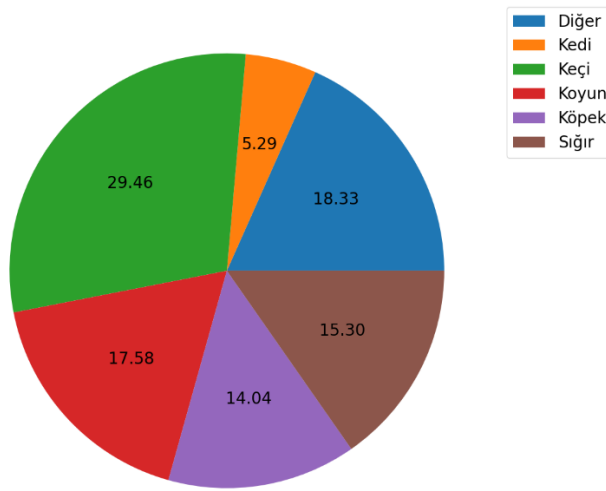
Nekropsi uygulaması yapılmıştır. Bunu sırası ile *Biyopsi*, *Sitoloji*, *Organ Numunesi* ve diğer kategorisi izlemektedir.



İnceleme Türü Oran Grafiği

Şekil 4.4. İnceleme türü kategori oranını gösteren *Pie* grafiği

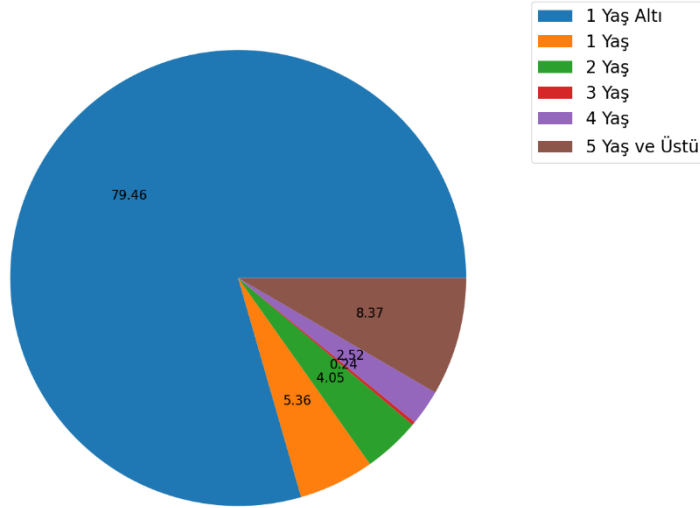
Gelen numunelerin türlere göre ayırım grafiği Şekil 4.5'te verilmiştir. Rutin defterinde gözüken 68 ayrı kategoride tür bulunmaktadır. Bunların çoğunluğunu oluşturan ilk beş kategori Şekil 4.5'te verilmiştir. Bu beş kategori dışındaki değerler ise diğer olarak gösterilmiştir. Bu değerlere göre en çok numune gelen hayvan türü keçi olurken bunu sırasıyla; koyun, sığır, köpek ve kedi takip etmektedir.



Tür Oran Grafiği

Şekil 4.5. Hayvan türleri kategori oranını gösteren *Pie* grafiği

Gelen numunelerin yaş oranını gösteren *pie* grafiği Şekil 4.6'da gösterilmiştir. Grafikten çıkan sonuçlara göre en çok numune 1 yaş altı olarak geldiği gözükmektedir. 5 yaş ve üzeri değerler ise 5 yaş üzeri kategorisi olarak gösterilmiştir.



Yaş Oran Grafiği

Şekil 4.6. Hayvan yaşları kategori oranını gösteren *Pie* grafiği (5 yaş ve üzeri hayvanlar 5 yaş kategorisine dahil edilmiştir)

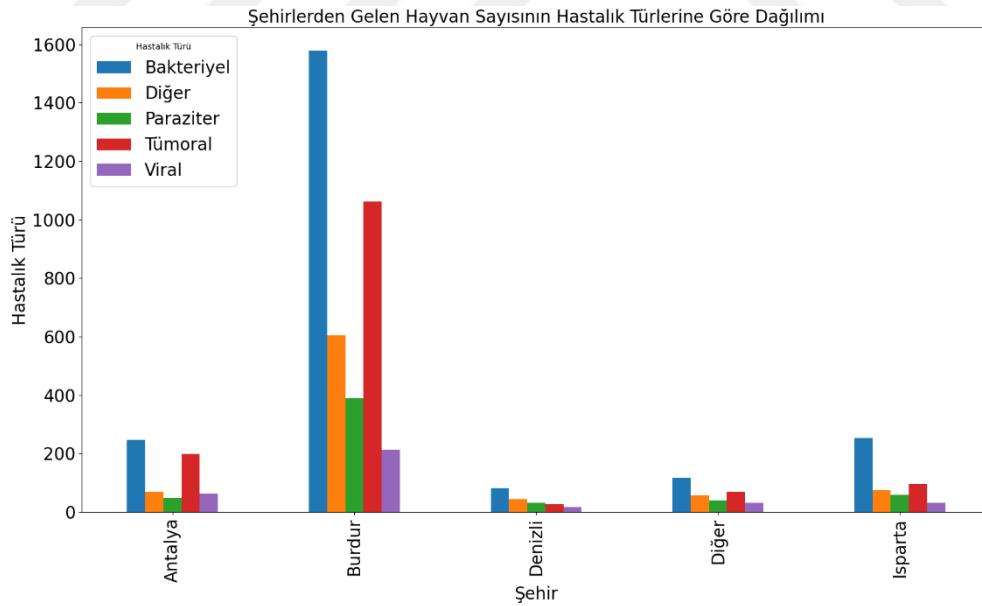
Sınıflandırma sonucu kategorisi de olan Hastalık Türü kategorisi dört ayrı sonuç vermektedir (Şekil 4.7). Bunlar sırasıyla; Bakteriyel, Tümöral, Paraziter ve Viral olarak sıralanmaktadır. Bu değerler dışında kalan sonuçlar ise diğer kategorisi içerisinde verilmektedir. Bu dört sonucun birleşimini de içeren diğer kategorisi anomali ve travmatik değerlerini de içermektedir.



Hastalık Türü Oran Grafięi

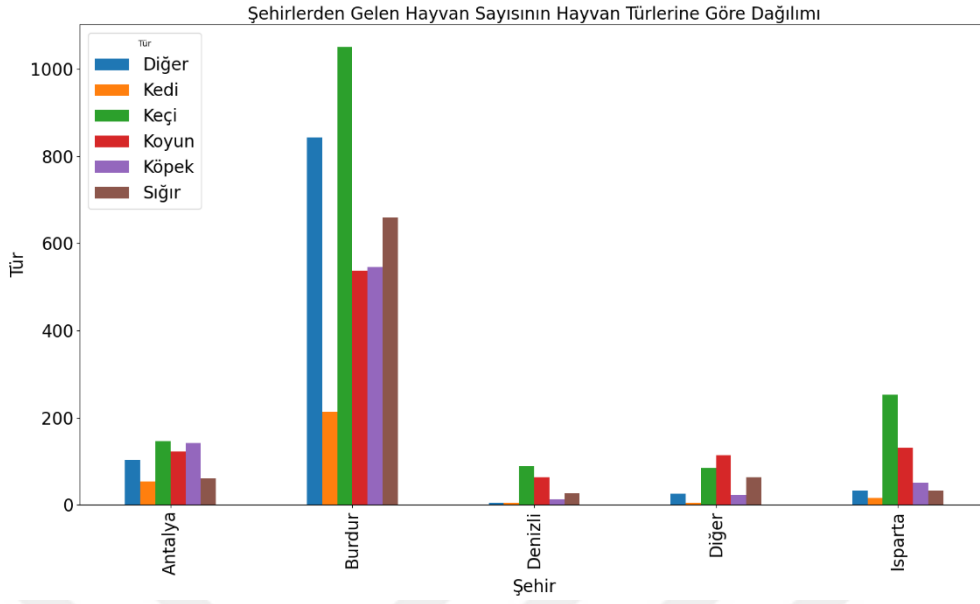
Şekil 4.7. Hastalık türleri kategori oranını gösteren Pie grafięi

Şehirlerden gelen numunelerin hastalık türünü gösteren grafik Şekil 4.8’de verilmiştir. Grafięe göre en çok numunenin geldięi Burdur ilinde en çok gözükten hastalık bakteriyel olarak gözükmemektedir. Bakteriyel hastalıktan sonra ise tümör (26.38%) hastalık gelmektedir.



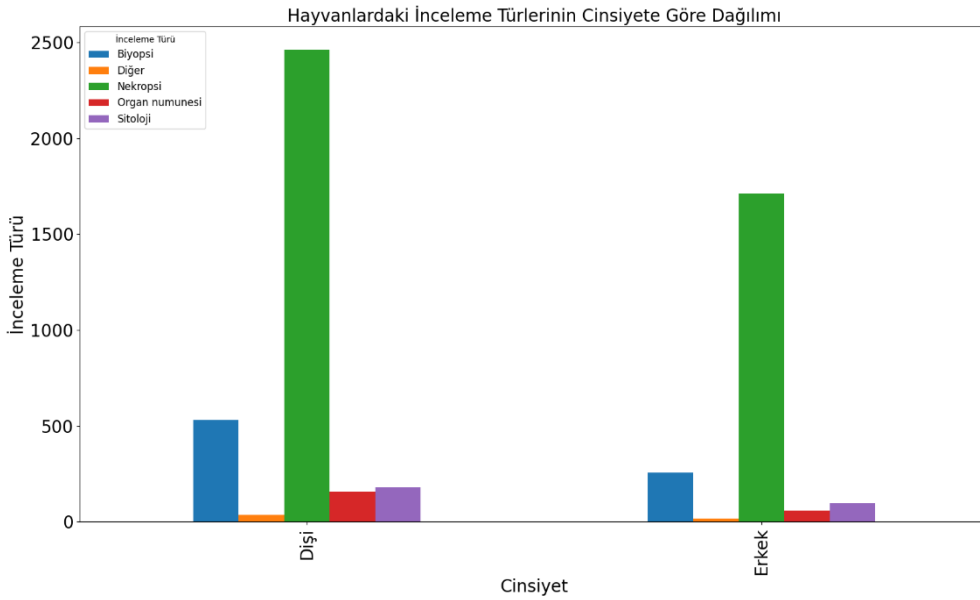
Şekil 4.8. Şehirlerden gelen numunelerin hastalık türleri oranı grafięi

Şehirlerden gelen numunelerin Tür kategorisi Şekil 4.9’da verilmiştir. Yine en çok numunenin geldięi il olan Burdur ilinden en çok gelen tür keçi olmuştur. Keçi türünü sırasıyla; sığır, köpek, koyun ve kedi izlemektedir.



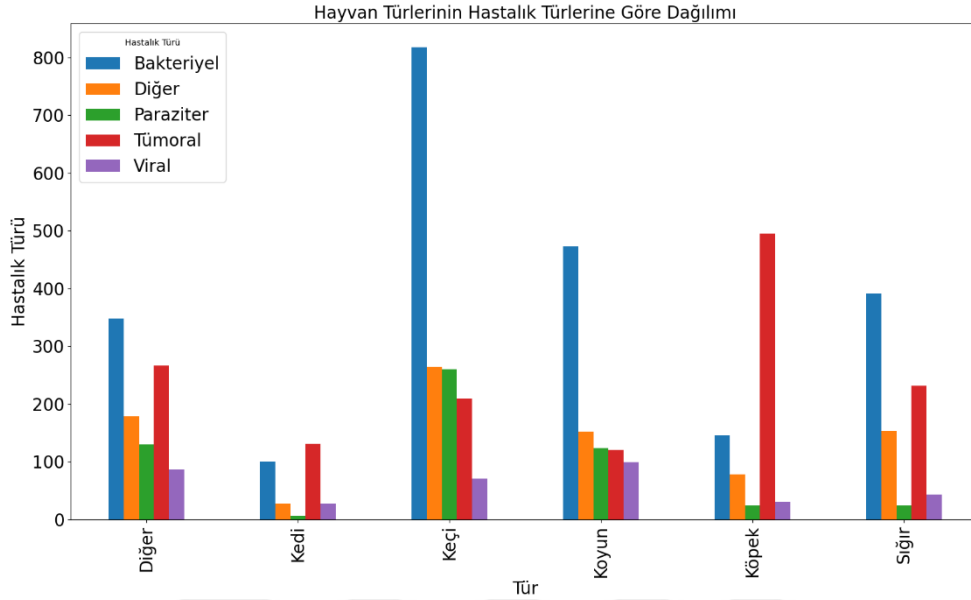
Şekil 4.9. Şehirlerden gelen numunelerin hayvan türleri oranı grafiği

Hayvan cinsiyetinin inceleme türüne göre oranlarını gösteren grafik Şekil 4.10'da verilmiştir. Bu grafiğe göre en çok inceleme türü olan nekropsinin dişi hayvanlarda daha çok uygulandığı görülmektedir.



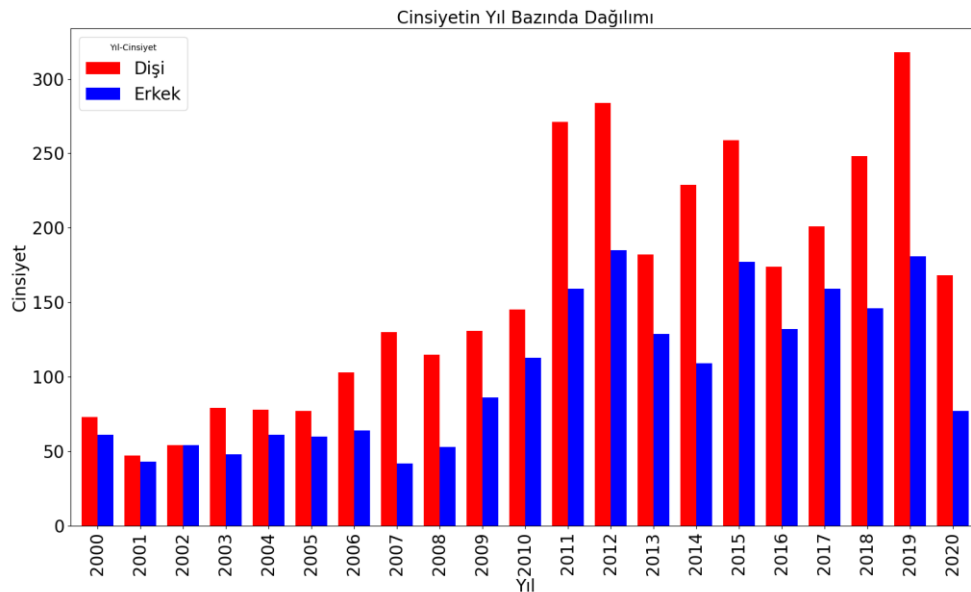
Şekil 4.10. Hayvan cinsiyetinin inceleme türüne oranı grafiği

Hayvan türleri arasında en çok gözükten hastalık türü grafiği Şekil 4.11’de verilmiştir. Bakteriyel hastalık türü en çok keçi türünde gözükürken, köpek türünde ise en çok tümöral hastalık türü görülmektedir.



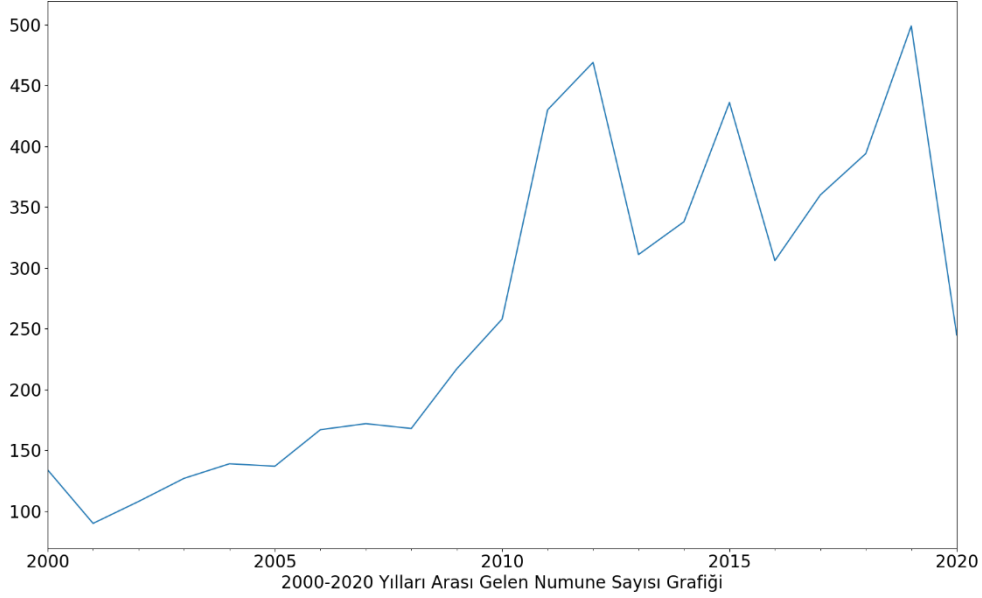
Şekil 4.11. Hayvan türlerinin hastalık türüne oranı grafiği

2000-2020 yılları arası gelen numunelerin cinsiyet oranını gösteren grafik Şekil 4.12’de görülmektedir. Bu grafiğe göre dişi hayvanlar, erkek hayvanlara göre neredeyse her yıl daha fazla gelmiştir.



Şekil 4.12. Hayvan cinsiyetinin yıllara oranı grafiği

2000-2020 yılları arasında gelen numune sayısını gösteren çizgi grafiği Şekil 4.13'te verilmiştir. Grafiğe göre 2011, 2012, 2014, 2015, 2018, 2019 yıllarında numune sayılarının arttığı gözükmemektedir. 2020 yılındaki düşüşün sebebi rutin defter verilerinin 2020 yılı Ağustos ayına kadar olmasından kaynaklanmaktadır.

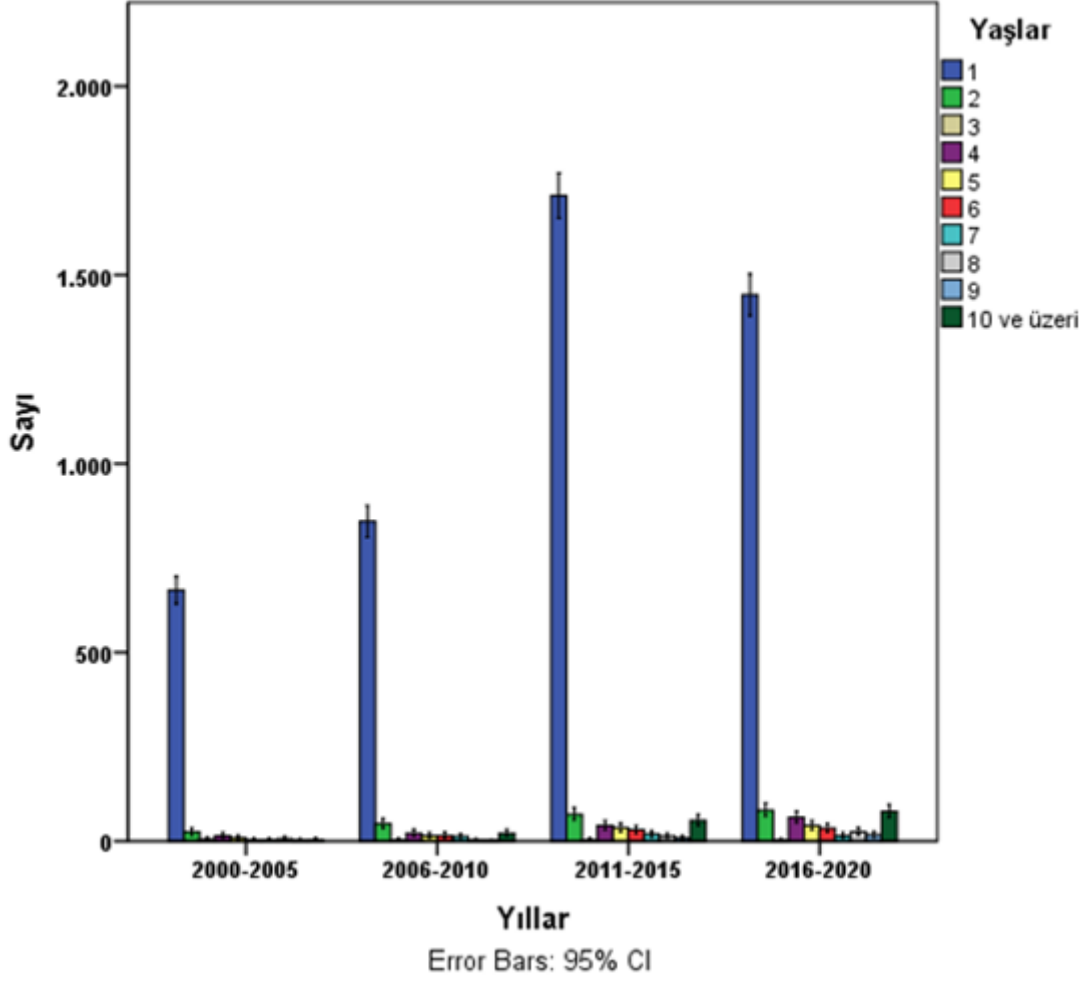


Şekil 4.13. 2000-2020 yılları arası gelen numune sayısı grafiği

4.1.2. İstatistik Analizi Sonucu Elde Edilen Grafikler

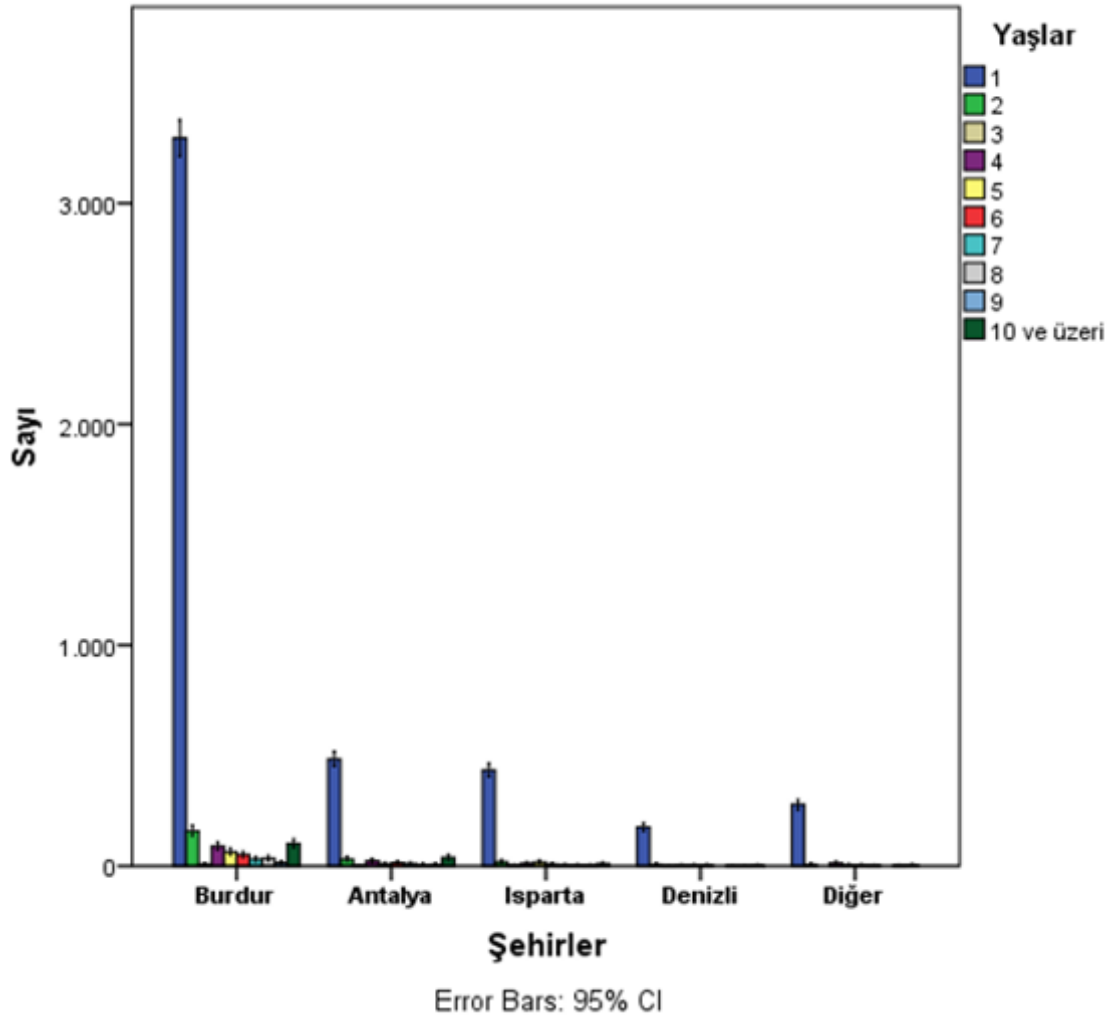
Bu çalışmanın verilerinin istatistik analizi SPSS (v.22.00) paket programda yapılmıştır. Verilerin grafik gösterimlerinin yapılmasında da aynı program kullanılmıştır.

Şekil 4.14'te hayvan yaşlarının yıllara göre dağılım grafiği verilmiştir. En çok bulunan 1 yaş ve altı kategorisi 2011-2015 yılları arasında gelmiş olup, 10 yaş ve üzeri hayvanların ise genellikle 2015 ve sonraki yıllarda geldiği görülmektedir.



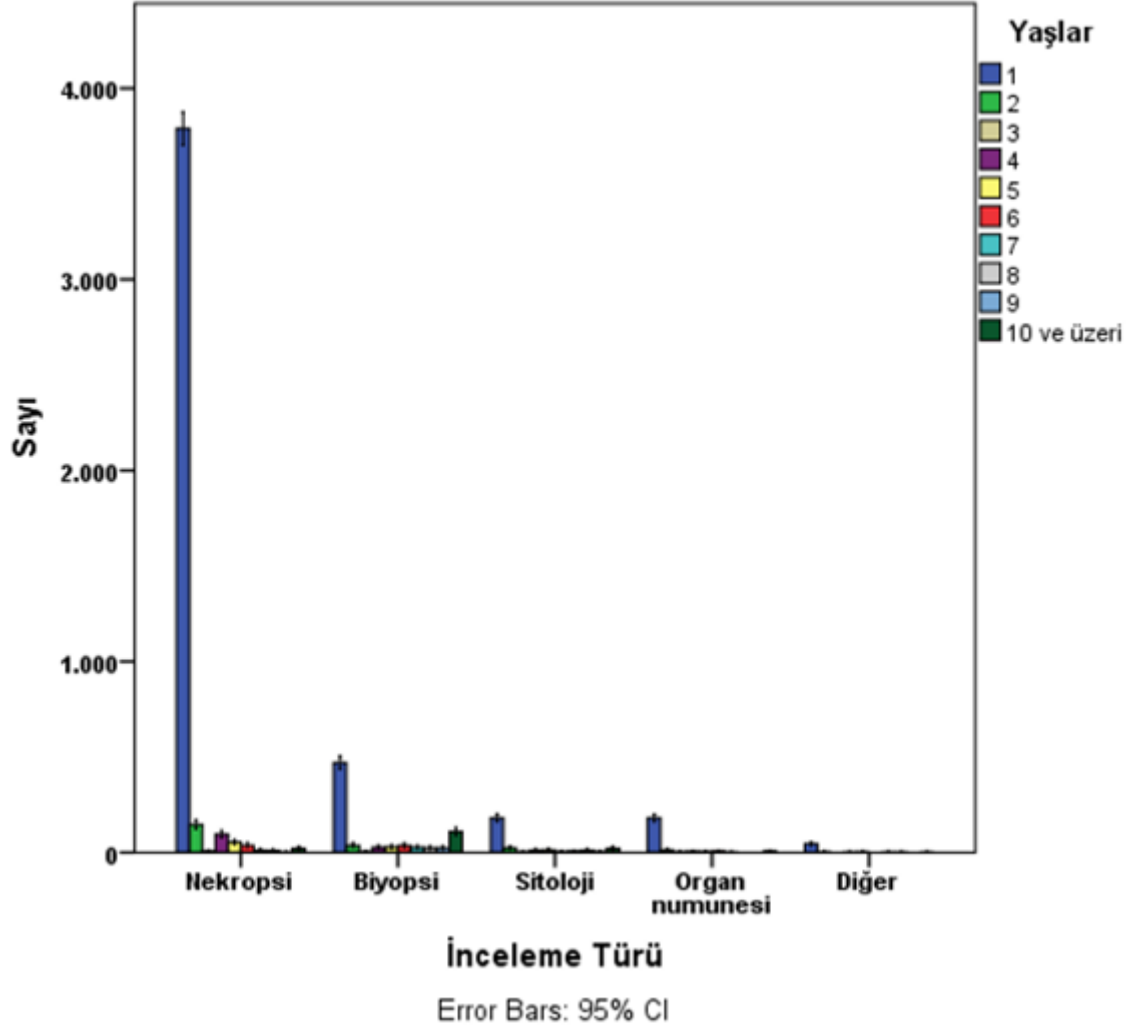
Şekil 4.14. 2000-2020 yılları arası hayvan yaş dağılımı grafiği

Bir diğer grafik olarak şehirlerden gelen hayvanların yaşlarının gözüktüğü grafik olan şehir-yaş grafiği Şekil 4.15'te verilmiştir. Bu grafiğe göre Burdur ilinden gelen hayvanların çoğunlukla 1 yaş ve altında olduğu görülmektedir.



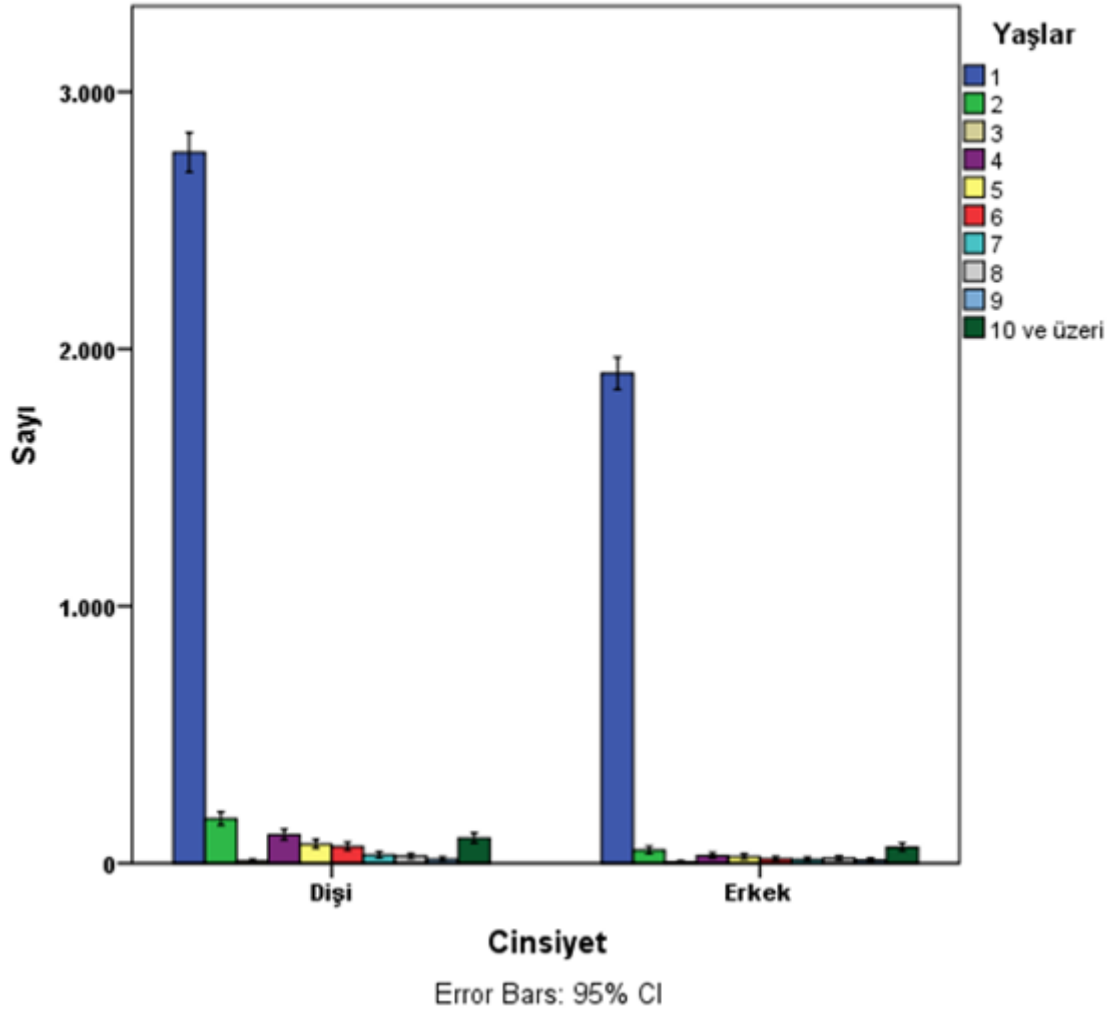
Şekil 4.15. Değişik illerden gelen hayvan yaş dağılımı grafiği

İnceleme türleri ile hayvan yaşları arasındaki ilişkiyi gösteren grafik Şekil 4.16'da gösterilmiştir. Nekropsinin en çok uygulanan inceleme türü olduğu kategori 1 yaş ve altı hayvanlardan oluşmaktadır. Nekropsi dışında kalan inceleme türlerinde de 1 yaş ve altı hayvanların çoğunlukta olduğu görülmektedir.



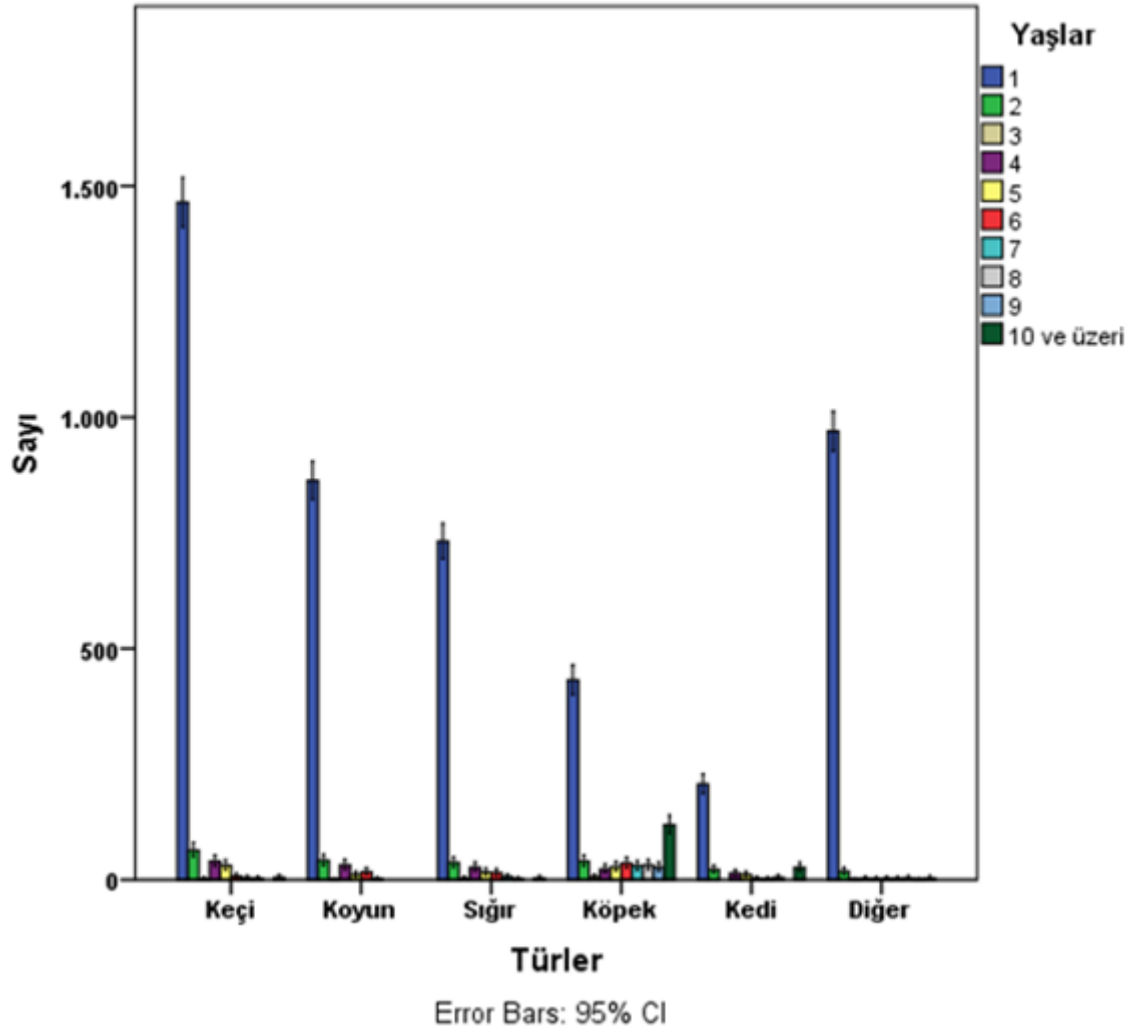
Şekil 4.16. Hayvan yaş gruplarında uygulanan inceleme türleri grafiği

Gelen hayvanların yaşlarının cinsiyete olan oranlarının gösterildiği grafik Şekil 4.17’de gösterilmiştir. Çoğunlukla gelen hayvanların 1 yaş ve altı olduğu gözükürken, dişi hayvanların erkek hayvanlara göre daha fazla geldiği görülmektedir.



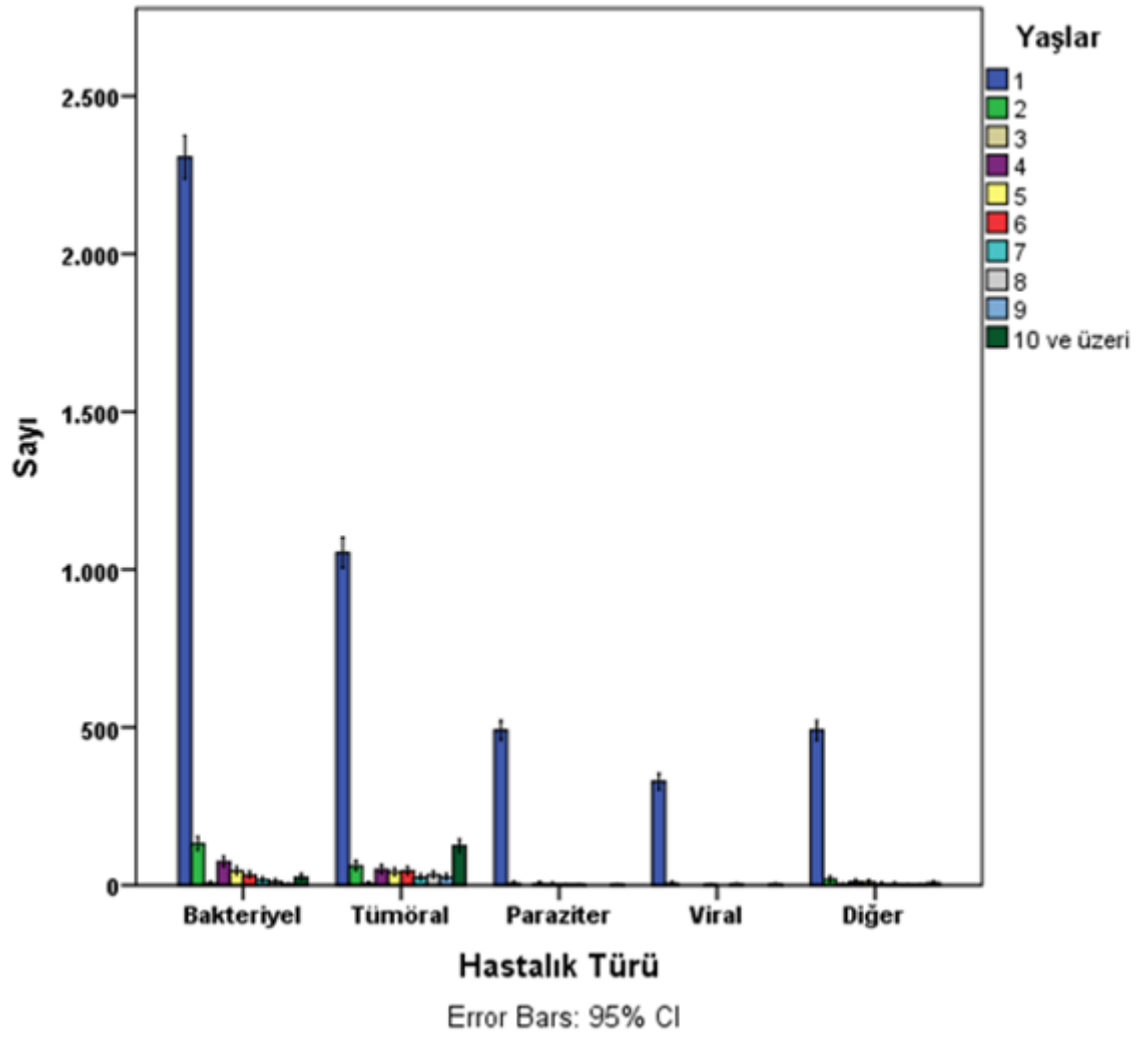
Şekil 4.17. Hayvan cinsiyetinin hayvan yaşına dağılımı grafiği

Şekil 4.18’de hayvan türlerinin yaşlara göre dağılımının gösterildiği grafik verilmiştir. Grafiğe göre çoğunlukla 1 yaş ve altı hayvan türlerinin geldiği öne çıkmaktadır.



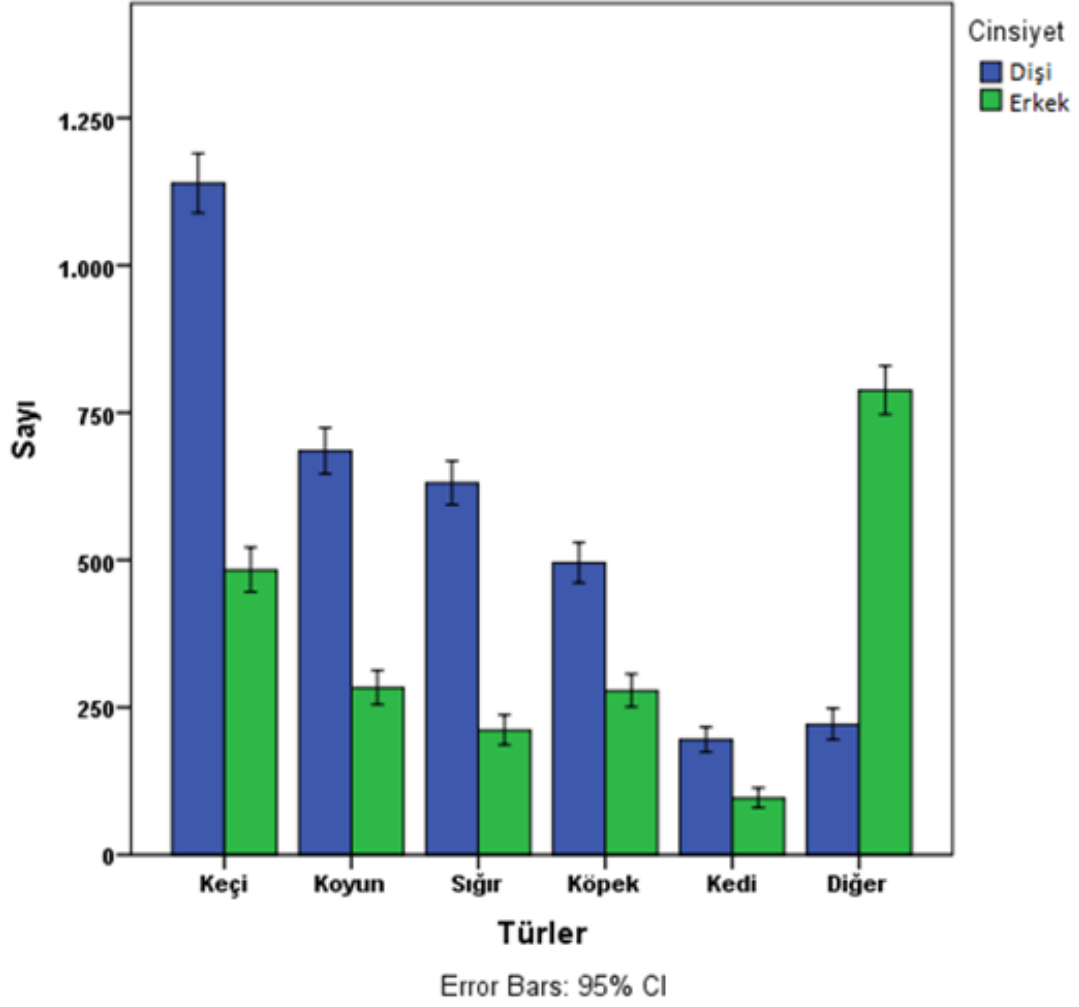
Şekil 4.18. Hayvan türlerinin hayvan yaşına dağılımı grafiği

Sınıflandırma sonucu olan hastalık türü kategorisinin gelen hayvan yaşlarına göre dağılımı Şekil 4.19’da verilmiştir. Diğer grafiklere benzer olarak 1 yaş ve altı hayvanların hastalık türleri kategorilerinde ilk sırada olduğu görülmektedir. Bunun yanında diğer yaş kategorilerinde genel olarak bakteriyel ve tümöral hastalık türlerinin gözüktüğü görülmektedir.



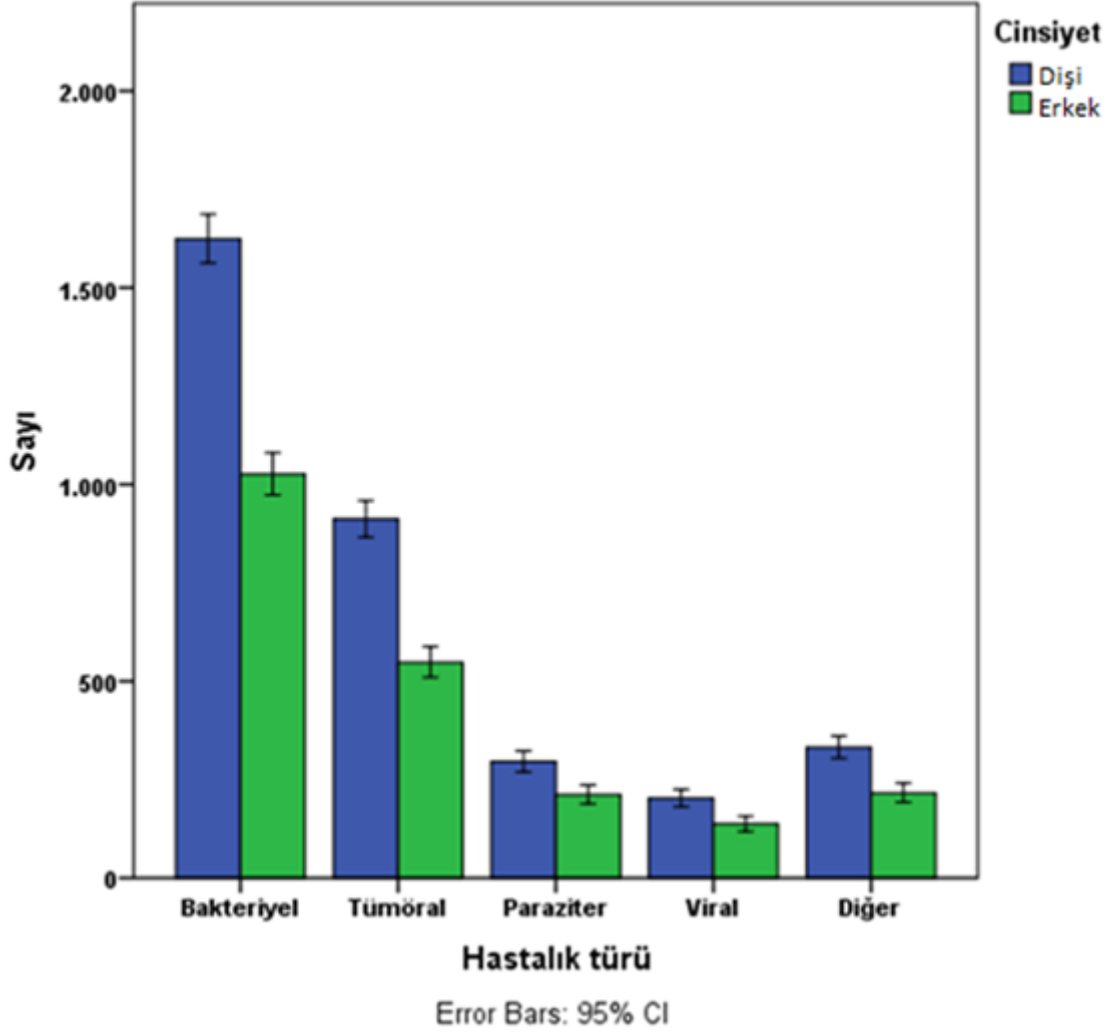
Şekil 4.19. Hastalık türlerinin hayvan yaşına dağılımı grafiği

Hayvanların cinsiyetlerinin hayvan türlerine olan oranı Şekil 4.20’de verilmiştir. Hayvan türlerinde dişi hayvanların daha çok geldiği gözükürken, en çok gelen hayvan türü olan keçi kategorisinde ise dişi sayısının erkek sayısının 2 katından bile fazla olduğu gözlemlenmektedir. Öte yandan diğer kategorisinde ise erkek oranının fazla olduğu görülmektedir. Diğer kategorisi içerisindeki muhabbet kuşu, tavuk ve balık türleri ise erkek hayvan oranının fazla olduğu türlerdir.



Şekil 4.20. Hayvan türlerinin hayvan cinsiyetine oranı grafiği

Son olarak ise hastalık türlerinin hangi cinsiyetlerde daha fazla görüldüğünü gösteren grafik Şekil 4.21’de verilmiştir. Grafiğe göre hastalık türlerinin genel olarak dişi hayvanlar üzerinde görüldüğü gözlemlenmektedir.



Şekil 4.21. Hastalık türlerinin hayvan cinsiyetine oranı grafiği

4.2. Karar Ağacı Metodu Sonuçları

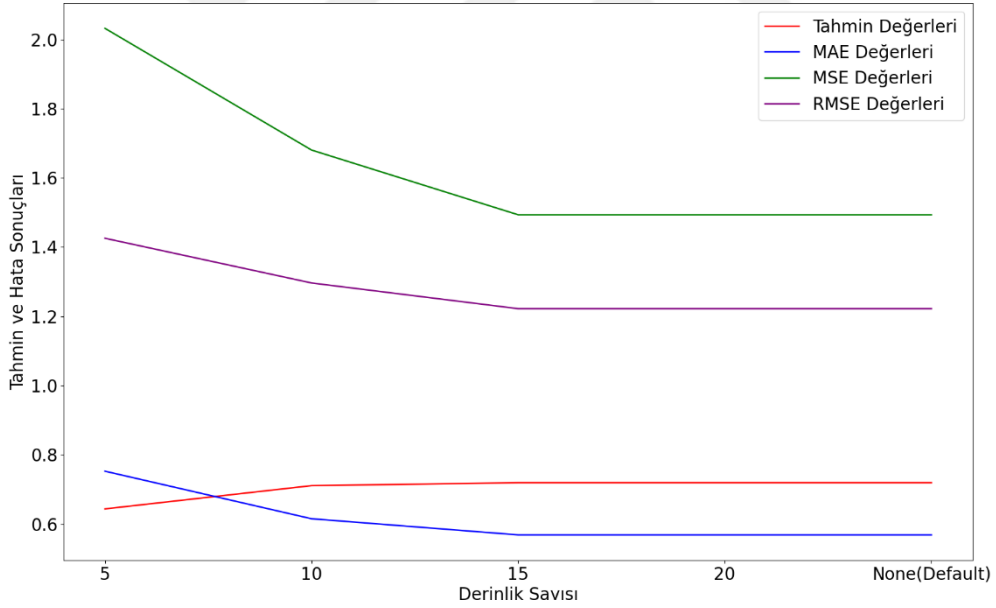
Karar ağacı sınıflandırma işlemi yapılırken parametre olarak girilen `test_size` değeri ile eğitim ve test veri setleri ayrımı yapılmaktadır. Bu çalışmada yüzde 30 oranında test veri seti oluşturulmuş olup (3853 eğitim seti, 1652 test seti, toplam 5505 veri) bu veri setleri ile eğitilen modellerin tahmin ve hata payı sonuçları Tablo 4.1’de verilmiştir. Tabloda farklı derinlikteki modellerden çıkan tahmini sonuç değerleri ile MAE, MSE ve RMSE hata payı değerleri verilmiştir. Tablodaki verilere göre derinlik 15 ve üzeri olarak tanımlandığında sonucun sabit olarak çıktığı gözlemlenmektedir. `max_depth` parametresinin varsayılan (default) değeri olan *yok* (None) değeri ağacın derinliğinin ulaşabileceği maksimum derinliğe ulaştığını gösterir. Bu değer sonuçları ise tabloda en son satıra değerleri eklenmiştir.

Tablo 4.1.Karar ağacı modeli sonucu çıkan değerler tablosu

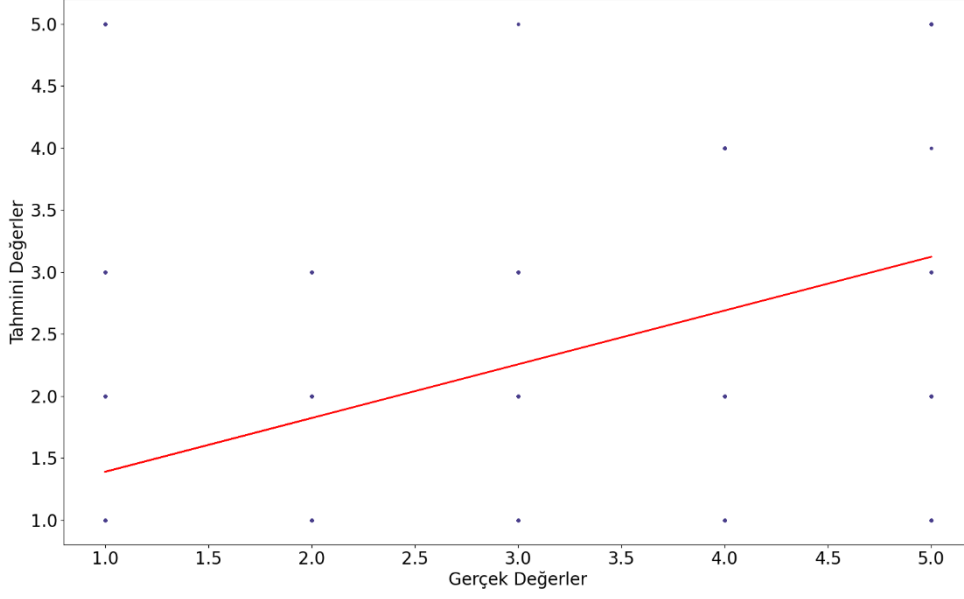
max_depth	Tahmin	MAE	MSE	RMSE
5	0,643	0,752	2,032	1,426
10	0,710	0,615	1,680	1,296
15	0,719	0,568	1,493	1,222
20	0,719	0,568	1,493	1,222
None	0,719	0,568	1,493	1,222

max_depth değeri karar ağacının derinlik sayısını belirten parametre olup, varsayılan (default) değeri yok(None) olarak girildiği zaman ağacın en derin noktasına kadar gitmektedir.

Tablodan çıkan sonuçlardan oluşan grafik Şekil 4.22’de verilmiştir. Grafikten de anlaşılacağı üzere derinlik sayısı 15 ve üzerinde tanımlandığında sonuçlarda ve hata paylarında sabit bir değer gözükmemektedir. Bu model için 15 derinlik değeri en iyi değer olarak görülüp, bu derinlik ile ilgili grafik Şekil 4.23’te verilmiştir.



Şekil 4.22. Karar ağacı modeli tahmin ve hata payı grafiği



Şekil 4.23. Derinlik sayısı 15 olan model grafiği

4.3. KNN Metodu Sonuçları

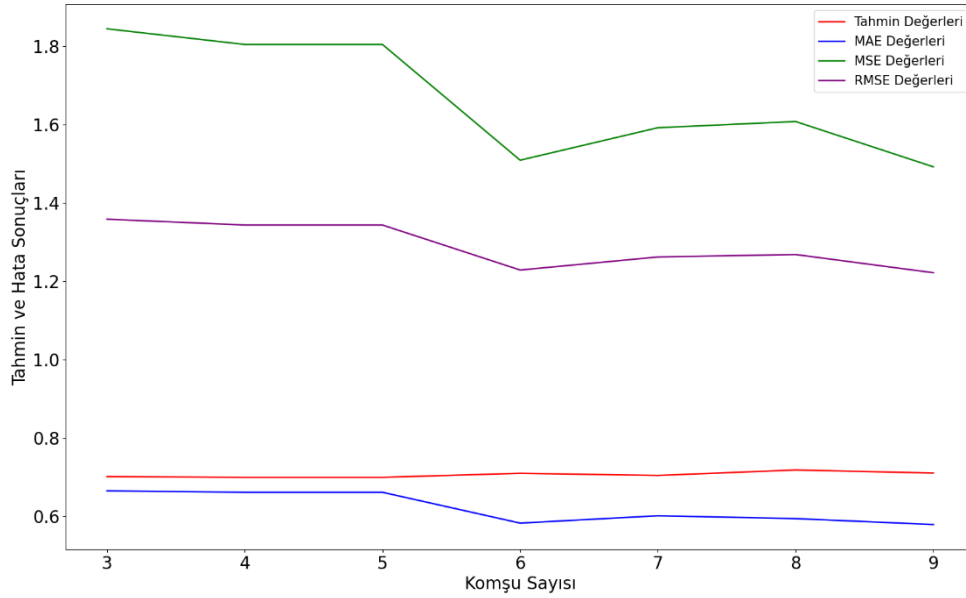
KNN sınıflandırma işlemi için daha önce karar ağacı modeli için hazırlanan eğitim ve test veri setleri aynı boyutta kullanılmaktadır. KNN modeli sonucu çıkan tahmin ve hata payı sonuçları Tablo 4.2’de verilmiştir. Tabloda farklı yakın komşu sayı değerlerdeki modellerden çıkan tahmini sonuç değerleri ile MAE, MSE ve RMSE hata payı yöntemleri verilmiştir. Tablodaki verilere göre komşu sayısı arttıkça genel olarak tahmin değeri daha iyi olarak çıkmakta, hata payı değerleri ise bir önceki değere göre düşük olarak gözükmemektedir. KNN modelinde varsayılan (default) komşu sayısı 5 olup varsayılan komşu sayısının üzerinde olan değerlerde daha iyi sonuç çıktığı görülmektedir.

Tablo 4.2.KNN modeli sonucu çıkan değerler tablosu

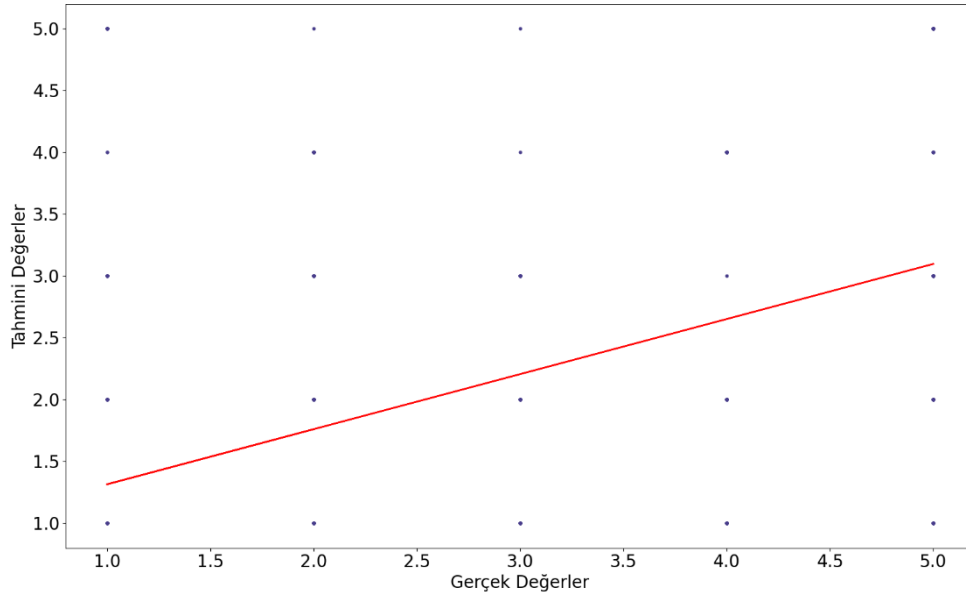
n_neighbors (k)	Tahmin	MAE	MSE	RMSE
3	0,700	0,665	1,844	1,358
4	0,699	0,661	1,804	1,343
5 (default)	0,699	0,661	1,804	1,343
6	0,709	0,582	1,508	1,228
7	0,704	0,601	1,591	1,262
8	0,718	0,594	1,607	1,268
9	0,710	0,579	1,492	1,221

n_neighbors (k) değeri yakın komşu sayısını göstermekte olup, varsayılan (default) değeri 5’tir.

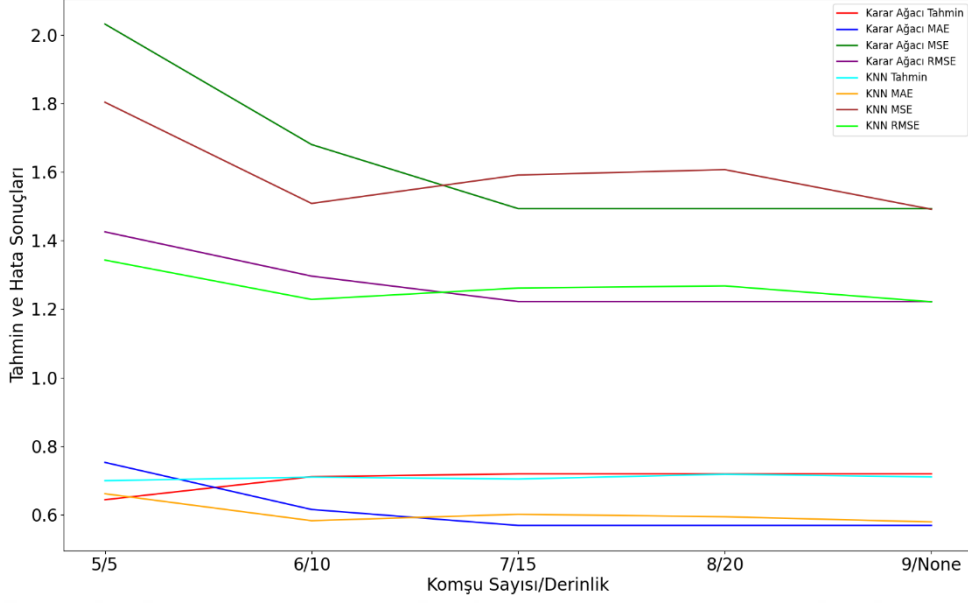
Tablodan çıkan sonuçlardan oluşan grafik Şekil 4.24'te verilmiştir. Grafikten de anlaşılacağı üzere komşu sayısı arttıkça çıkan sonuçlarda daha iyi değerler ortaya çıkmaktadır. KNN modeli için en yakın komşu sayısı 8 olan değer tahmin sonucu diğer değerlerin sonuçları arasında en yüksek olarak çıkmıştır. Komşu sayısı 8 olan değer sonuçları ile ilgili grafik Şekil 4.25'te verilmiştir.



Şekil 4.24. KNN modeli tahmin ve hata payı grafiği



Şekil 4.25. Komşu sayısı 8 olan model grafiği



Şekil 4.26. Model sonuç karşılaştırma Grafiği

İki modelin sonuçlarının karşılaştırma grafiği Şekil 4.26’da verilmiştir. Modellerin parametre değerleri birbirine eşit olmadığı için KNN modelinin komşu sayısı 5’ten küçük olan değerler grafiğe dahil edilmemiştir. Modellerin tahmin sonuçları neredeyse birbiriyle aynı olsa da karar ağacı modelinin çok küçük bir farkla daha iyi bir sonuç verdiği gözlemlenmektedir.

4.4. Tartışma

Gelişen teknolojilere bağlı olarak günümüzde patoloji alanında da sürekli gelişmeler olmaktadır. Patoloji alanında kullanılan cihaz ve aletlerin gelişip, değişmesi haricinde bilgisayar uygulamaları kullanımı da artıp gelişmektedir. Patolojik işlemler genellikle uzun süreli işlemler gerektirmektedir. Bu sürenin kısaltılması koyulacak teşhis ile uygulanacak tedavi yöntemleri için ekstra zaman kazandırabilir. Günümüzde dijital patoloji olarak isimlendirilen, patolojik numunelerin incelenmesinin bilgisayar ortamına taşınması, çeşitli uygulamalar kullanılarak sonuçların görselleştirilebilmesi ile bu sonuçların analiz edilebilmesini ve çıkan sonuçların saklanabilmesini sağlayan bir alan ortaya çıkmıştır (Barisoni vd., 2017; Bera vd., 2019; Özmen, 2021).

Patoloji alanındaki gelişmeler, hızlı ve doğru teşhis koyulmasını sağlamaktadır. Bu nedenle patoloji alanında kullanılan teknolojik alet ve cihazlar ile uygulanan teknikler hızlı ve doğru teşhis koyulmasına katkı sağlamaktadır. Kullanılabilecek gelişmiş bilgisayar uygulamaları ile patoloji alanında yapılacak çalışmalara olumlu katkı sağlayabileceği ve

daha kolay çalışmasına olanak sunmaktadır. Bunun yanında ağ ortamında bu çalışmaların paylaşımı mümkün olup teşhis ve tedavi işlemlerinde başka uzmanlardan bilgi alınabilmesi kolay bir şekilde mümkün olabilmektedir (Niazi vd., 2010; Abels vd., 2019; Chang vd, 2019; McCarty vd., 2006; Özmen, 2021).



5. SONUÇ

Bu çalışma sayesinde Burdur Mehmet Akif Ersoy Üniversitesi Veteriner Fakültesi Patoloji Anabilim Dalının 2000-2020 yılları arasında incelenen veriler kullanılarak analiz edilmiş olup, analizden sonra ise sınıflandırma işlemi yapılmıştır. Çıkan analiz sonuçları grafikler halinde verilip açıklanmıştır. Sınıflandırma işlemi için iki ayrı metot kullanılmış olup bu metotlardan çıkan sonuçlar karşılaştırılmıştır. Sınıflandırma işlemi yapılırken veri seti, %30 test seti ve %70 eğitim seti olmak üzere ikiye ayrılmıştır. Modellerin sonuçlarının elde edilmesi sırasında parametreler değiştirilerek en iyi sonucun çıkması açısından farklı değerler elde edilmiştir. Karar ağacı sınıflandırmasında derinlik parametresi değiştirilerek 5 ayrı sonuç elde edilirken, KNN sınıflandırmasında komşu sayısı parametresi değiştirilerek 9 ayrı sonuç elde edilmiştir. Karar ağacı metodunda derinlik arttıkça daha iyi sonuç elde edilirken 15 ve üzeri derinlik sayısında ise sabit bir oranda değerler elde edilmiştir. KNN metodunda ise komşu sayısı varsayılan değer olan 5'ten küçük alındığı zaman değerlerin daha düşük oranda tahmin edildiği görülürken, 5 ve üzeri komşu sayılarından eğitilen modelin sonuçları daha iyi değerlerden oluştuğu görülmektedir. Yapılan karşılaştırmada ise iki modelinde yaklaşık %70 oranında sonuçlar verdiği belirlenmiştir. Çok küçük bir farkla karar ağacı yönteminin daha iyi sonuç verdiği görülmüştür. Model parametrelerinin yüksek değerli tanımlanması sonucu hata payı değerlerinin giderek azaldığı ve bu azalma sonucunda tahmin sonuçlarında artış gözlenmiştir.

Bu çalışmada gerçekleştirilen analiz sonuçları ise grafiklerle gösterilmiştir. Grafik sonuçlarına göre gelen numunelerin tür, yaş, şehir, cinsiyet gibi verilerin birbirine olan oranları görülmektedir. Bu oranlar sayesinde daha etkili ve daha hızlı teşhis koyulma ihtimali artmaktadır.

Bu çalışmanın hayvan hastalık teşhislerinde daha etkili ve hızlı tanı koyulması amaçlanmaktadır. Çıkan analiz sonuçları ise gelen numune veya hayvanların bazı koşullarda ne gibi hastalıklarla karşılaşabileceğini göstermesi beklenmektedir.

KAYNAKLAR

- Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, et al., 2019. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. J Pathol.
- Barisoni L, Gimpel C, Kain R, Laurinavicius A, Bueno G, Zeng C, et al., 2017. Digital pathology imaging as a novel platform for standardization and globalization of quantitative nephropathology. Clin Kidney J.
- Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A., 2019. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol.
- Bounsaythip C., ve Esa R. R., 2001. Overview of Data Mining For Customer Behavior Modeling, VTT Information Technology Research Report, Version 1.
- Brown RW., 2015. Error Reduction and Prevention in Surgical Pathology, Springer.
- Carlton, W.W., McGavin, M.D., 1995. Thomson's Special Pathology, Mosby-Yearbook, Inc., Missouri.
- Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, et al., 2019. Artificial intelligence in pathology. J Pathol Transl Med.
- Cheville, N.F., 1999. Introduction to Veterinary Pathology, 2nd Ed. Iowa State University Pres.
- Emel G., Taşkın Ç., 2005. Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması, Eskişehir Osman Gazi Üniversitesi Sosyal Bilimler Dergisi, Cilt:6 Sayı:2.
- Jones, T.C., R.D. Hunt., 1993. Veterinary Pathology, Lea & Febiger, Philadelphia.
- Kahraman, M.M., 1996. Genel Patoloji Ders Notları, Uludağ Üniversitesi Veteriner Fakültesi.
- McCarty J, Minsky ML, Rochester N, Shannon CE., 2006. A proposal for the Dartmouth Summer Research Project on artificial intelligence, August 31, 1955. AI Magazine.
- Niazi MKK, Parwani AV, Gürcan MN., 2019. Digital pathology and artificial intelligence. Lancet Oncol.
- Özmen Ö., 2006. Veteriner Genel Patoloji Ders Notları, Mehmet Akif Ersoy Üniversitesi Veteriner Fakültesi.
- Özmen Ö., 2016. 2000-2015 Yılları Arasında Burdur'daki Rutin Patoloji Teşhisleri. VIII. Ulusal Veteriner Patoloji Kongresi, 1-3 Eylül 2016, Samsun.

- Özmen Ö., 2021. Veteriner Fakültesi Öğrencilerinin Uygulamalı Patoloji Laboratuvar Eğitimleri ile Bilgi Düzeylerinin Arttırılması ve Çağdaş Yaklaşımlar ile Mesleğe Hazırlanması, Bilimsel Araştırma Projeleri Komisyonu, Mehmet Akif Ersoy Üniversitesi Veteriner Fakültesi.
- Slauson, D.O., 1990. Cooper, B.J. Mechanisms of Disease A Textbook of Comparative General Pathology, 2nd Ed., Williams & Wilkins.
- Sütcü C., Aytekin Ç., 2018. Veri Bilimi, Paloma Yayınevi.
- URL-1, 2021. pandas – Python Data Analysis Library. <https://pandas.pydata.org>, (Erişim Tarihi: 21.12.2021).
- URL-2, 2021. pandas: powerful Python data analysis toolkit. <https://github.com/pandas-dev/pandas>, (Erişim Tarihi: 21.12.2021).
- URL-3, 2021. Pandas. <https://tr.wikipedia.org/wiki/Pandas>, (Erişim Tarihi: 21.12.2021).
- URL-4, 2021. sklearn.impute.SimpleImputer. <https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html>, (Erişim Tarihi: 22.12.2021).
- URL-5, 2021. sklearn.tree.DecisionTreeClassifier. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>, (Erişim Tarihi: 23.12.2021).
- URL-6, 2021. missingno. <https://github.com/ResidentMario/missingno>, (Erişim Tarihi: 22.12.2021).
- URL-7, 2021. NumPy. <https://numpy.org>, (Erişim Tarihi: 20.12.2021).
- URL-8, 2021. NumPy. <https://tr.wikipedia.org/wiki/NumPy>, (Erişim Tarihi: 20.12.2021).
- URL-9, 2021. Matplotlib: Visualization with Python. <https://matplotlib.org>, (Erişim Tarihi: 20.2021).
- URL-10, 2021. pandas.get_dummies. https://pandas.pydata.org/docs/reference/api/pandas.get_dummies.html, (Erişim Tarihi: 23.12.2021).
- URL-11, 2021. sklearn.model_selection.train_test_split. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html, (Erişim Tarihi: 24.12.2021).
- URL-12, 2021. sklearn.metrics.accuracy_score. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html, (Erişim Tarihi: 26.12.2021).
- URL-13, 2021. sklearn.neighbors.KNeighborsClassifier, <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier>, (Erişim Tarihi: 24.12.2021).

- URL-14, 2021. Öklid Uzaklığı, https://tr.wikipedia.org/wiki/Öklid_uzaklığı, (Erişim Tarihi: 23.12.2021).
- URL-15, 2021. Nearest Neighbors, <https://scikit-learn.org/stable/modules/neighbors.html>, (Erişim Tarihi: 23.12.2021).
- URL-16, 2021. sklearn.metrics.mean_squared_error, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html, (Erişim Tarihi: 24.12.2021).
- URL-17, 2021. sklearn.metrics.mean_absolute_error, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_absolute_error.html, (Erişim Tarihi: 24.12.2021).
- URL-18, 2021. Mean Absolute Error, https://en.wikipedia.org/wiki/Mean_absolute_error, (Erişim Tarihi: 25.12.2021).
- URL-19, 2021. Mean Squared Error, https://en.wikipedia.org/wiki/Mean_squared_error, (Erişim Tarihi: 25.12.2021).
- URL-20, 2021. Root Mean Squared Error, https://en.wikipedia.org/wiki/Root-mean-square_deviation, (Erişim Tarihi: 25.12.2021).
- Zorman M., Vili P., Kokol P., Peterson M., Sproggar M., ve Ojstersek M., 2001. Finding The Right Decision Tree's Induction Strategy For A Hard Real World Problem, International Journal of Medical Informatisc 1-2.