



**ORMAN SEKTÖRÜ ÇALIŞANLARINDA GÜRÜLTÜYE BAĞLI İŞİTME
KAYIPLARININ MAKİNE ÖĞRENMESİ ALGORİTMALARI
KULLANILARAK TAHMİNİ**

Işık DOĞRU

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2024

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Işık DOĞRU

28/06/2024

ORMAN SEKTÖRÜ ÇALIŞANLARINDA
GÜRÜLTÜYE BAĞLI İŞİTME KAYIPLARININ
MAKİNE ÖĞRENMESİ ALGORİTMALARI KULLANILARAK TAHMİNİ
(Yüksek Lisans Tezi)

Işık DOĞRU

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
Haziran 2024

ÖZET

Günlük 85 dB(A) veya daha fazla gürültüye maruz kalmak geri döndürülemez ve tedavi edilemez işitme kayıplarına yol açmaktadır. Aynı gürültü şiddeti ve aynı maruziyet süresi, farklı çalışanlardaki farklı seviyelerde işitme kayıplarına neden olabilmektedir. Bunun nedeni, çalışanların kişisel özelliklerindeki ve iş yerindeki davranışlarındaki farklılıklardır. Bu çalışmada cinsiyet, yaş, toplam çalışma süresi, sektörde çalışma süresi, gürültülü ortamda çalışma süresi, sigara kullanımı, gürültülü hobi ve yetersiz kulak koruyucu donanım kullanımının ağaç ürünleri imalatı sektöründe çalışanlardaki GBİK'e etkileri değerlendirilmiştir. İlk aşamada, bağımsız değişkenlerin çalışanlarda kaç dB'lik işitme kaybına neden olduğu araştırılmak amaçlanmıştır. Ancak varyasyon analizlerinde bağımlı değişken ve bağımsız değişkenler arasındaki korelasyon katsayısının düşük çıkması sebebiyle regresyon analizi gerçekleştirilmemiştir. İkinci aşamada, bağımsız değişkenlerin çalışanlarda GBİK'e neden olup olmadığı araştırılmıştır. Bağımlı değişkeni iki kategoriden oluşan veri kümesine uygun denetimli öğrenme sınıflandırma yöntemlerinden ikili lojistik regresyon, k-en yakın komşu algoritması, destek vektör makinesi, karar ağacı sınıflandırması, rastgele orman algoritması, Naive Bayes (Gauss) sınıflandırması, Naive Bayes (Bernoulli) sınıflandırması, stokastik gradyan inişi sınıflandırması ile çapraz doğrulama yapılarak modeller oluşturulmuştur. Modeller karşılaştırıldığında rastgele orman algoritması en iyi model seçilmiştir. Sonuçta sınıflandırma algoritmalarının çalışanlardaki GBİK'i sınıflandırmada başarılı; regresyon algoritmalarının GBİK'i önceden tahmin etmede başarısız olduğu görülmüştür.

Bilim Kodu : 90603
Anahtar Kelimeler : Makine öğrenmesi, GBİK, odyogram, frekans (hz), gürültü ölçümü, sınıflandırma
Sayfa Adedi : 67
Danışman : Prof. Dr. Erman ÇAKIT

PREDICTION OF NOISE-INDUCED HEARING LOSS IN THE FOREST SECTOR
WORKERS USING MACHINE LEARNING ALGORITHMS

(M. Sc. Thesis)

Işık DOĞRU

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2024

ABSTRACT

Exposure to noise levels exceeding 85 dB(A) on a daily basis can lead to permanent and untreatable hearing damage. Because of the diversity in personal traits among workers, various workers may experience different degrees of hearing loss even when exposed to identical noise intensity and duration. This research assessed the impact of gender, age, total years of work experience, years of experience in the sector, months of exposure to a noisy environment, smoking habits, engagement in noisy hobbies, and improper use of ear protection equipment on noise-induced hearing loss among employees in the forest sector. In the first phase, it was aimed to investigate how many dB of hearing loss the independent variables caused in employees. However, regression analysis was not performed due to the low correlation coefficient between the dependent variable and the independent variables in the variation analyses. The next phase of the research involves evaluating various classification algorithms to determine if eight independent variables are linked to NIHL in workers. Models were created by cross-validation with binary logistic regression, k-nearest neighbor algorithm, support vector machine, decision tree classification, random forest algorithm, Naive Bayes (Gauss) classification, Naive Bayes (Bernoulli) classification, and stochastic gradient descent algorithm from supervised learning classification methods suitable for a data set whose dependent variable consists of two categories. When the models were compared, the random forest algorithm was chosen as the best model. As a result, classification algorithms are successful in classifying NIHL in employees; regression algorithms have been shown to fail to predict NIHL.

Science Code : 90603
Key Words : Machine Learning, NIHL, audiogram, frequency (hz), noise measurement, classification
Page Number : 67
Supervisor : Prof. Dr. Erman ÇAKIT

TEŞEKKÜR

Çalışmamda değerli önerileri ile beni yönlendiren, bilgisini ve yardımını esirgemeyen danışman hocam Sn. Prof. Dr. Erman ÇAKIT'a teşekkürlerimi sunarım. Hayatımda oldukları süre boyunca sevgi, anlayış ve desteklerini her zaman hissettiğim babam Necip KARAHAN'a, annem Hatice KARAHAN'a, eşim Ömer DOĞRU'ya ve canım kızım Ömür Ülgen DOĞRU'ya çok teşekkür ederim.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
1. GİRİŞ.....	1
2. LİTERATÜR ÇALIŞMASI.....	9
3. MATERYAL VE METODOLOJİ.....	15
3.1. Veri Toplama	16
3.1.1. Araştırma grubu (örneklem).....	16
3.1.2. Onam formu	16
3.1.3. Odyometri testi değerlendirmesi ve gürültüye bağlı işitme kaybı	17
3.2. Veri Analizi	18
3.2.1. Değişkenlerin seçimi.....	19
3.2.2. Veri setinin sınırları.....	22
3.3. Makine Öğrenmesi	22
3.3.1. Karar ağacı algoritması	29
3.3.2. Rastgele orman algoritması.....	30
3.3.3. K-en yakın komşu algoritması	31
3.3.4. Destek vektör regresyonu/sınıflandırması.....	31
3.3.5. İkili lojistik regresyon	32
3.3.6. Naive Bayes sınıflandırması (Gauss)	34

	Sayfa
3.3.7. Naive Bayes sınıflandırması (Bernoulli).....	35
3.3.8. Stokastik gradyan inişi sınıflandırması	36
3.4. Makine Öğrenmesinde Çapraz Doğrulama	36
4. GÜRÜLTÜYE BAĞLI İŞİTME KAYBI TAHMİNİ UYGULAMASI .	39
4.1. Verilerin Dağılımı	39
4.2. Gürültüye Bağlı İşitme Kaybının Regresyon Algoritmaları ile Tahmini.....	44
4.3. Gürültüye Bağlı İşitme Kaybının Sınıflandırma Algoritmaları ile Tahmini.....	46
4.3.1. Test büyüklüğü % 20 için sınıflandırma modellerinin performansları	46
4.3.2. Test büyüklüğü % 25 için sınıflandırma modellerinin performansları	47
4.3.3. Test büyüklüğü % 30 için sınıflandırma modellerinin performansları	47
4.4. GBİK Sınıflandırmasında En İyi Modelin Seçimi	48
5. TARTIŞMA.....	57
6. SONUÇLAR VE ÖNERİLER.....	59
KAYNAKLAR	63
ÖZGEÇMİŞ	67

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 1.1. Ağaç ürünleri imalatı sektöründeki iş ekipmanlarının gürültü seviyeleri ..	8
Çizelge 2.1. GBİK alanında literatür araştırması tablosu	12
Çizelge 3.1. AUSE kaynaklarına göre işitme kaybı dereceleri.....	17
Çizelge 3.2. Gürültülü ortamda çalışma süresi kategorileri.....	20
Çizelge 3.3. Değişkenlerin adı, tanımı ve tipi.....	21
Çizelge 4.1. Bağımsız değişkenlerin istatistiksel anlamlılık seviyesi.....	45
Çizelge 4.2. Test büyüklüğü % 20 için sınıflandırma modellerinin sonuçları.....	47
Çizelge 4.3. Test büyüklüğü % 25 için sınıflandırma modellerinin sonuçları.....	47
Çizelge 4.4. Test büyüklüğü %30 için sınıflandırma modellerinin sonuçları.....	48
Çizelge 4.5. Hata matrisi.....	54

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Ağaç panel üretiminin süreç akış şeması.....	3
Şekil 1.2. Yongalama işleminin proses akışı	4
Şekil 1.3. OYL (MDF) üretiminde kullanılan liflerin proses akış.....	5
Şekil 3.1. Uygulama basamakları akışı.....	15
Şekil 3.2. Odyogram örneği	18
Şekil 3.3. Makine öğrenmesi türleri.....	23
Şekil 3.4. Hata matrisi örneği	24
Şekil 3.5. AİK eğrisi örneği	25
Şekil 3.6. Denetimsiz öğrenme için sarmalayıcı yaklaşım	27
Şekil 3.7. Karar ağacı yapı örneği.....	29
Şekil 3.8. Destek vektör makinesi sınıflandırması örneği	32
Şekil 3.9. Naive Bayes yapısı	34
Şekil 3.10. Çapraz doğrulama örneği.....	37
Şekil 4.1. Çalışanların cinsiyetlerine göre dağılımı	40
Şekil 4.2. Çalışanların yaş aralıklarına göre dağılımı	40
Şekil 4.3. Çalışanların toplam çalışma yılının çalışan sayısına göre dağılımı.....	41
Şekil 4.4. Çalışanların sektörde çalışma yılının çalışan sayısına göre dağılımı	41
Şekil 4.5. Çalışanların gürültülü ortamda çalışma süresi dağılımı	42
Şekil 4.6. Çalışanların sigara içme durumu dağılımı	42
Şekil 4.7. Çalışanların KKD kullanma durumu dağılımı.....	43
Şekil 4.8. Çalışanların gürültülü hobi sahibi olma durumu dağılımı	43
Şekil 4.9. Çalışanların GBİK olma durumu dağılımı	44
Şekil 4.10. Veri setine ait histogram.....	45

Şekil	Sayfa
Şekil. 4.11. Sıfırncı çapraz doğrulama değişkenlerin önem katsayısı	49
Şekil 4.12. Birinci çapraz doğrulama değişkenlerin önem katsayısı	49
Şekil 4.13. İkinci çapraz doğrulama değişkenlerin önem katsayısı	50
Şekil 4.14. Üçüncü çapraz doğrulama değişkenlerin önem katsayısı	50
Şekil 4.15. Dördüncü çapraz doğrulama değişkenlerin önem katsayısı	51
Şekil 4.16. Beşinci çapraz doğrulama değişkenlerin önem katsayısı	51
Şekil 4.17. Altıncı çapraz doğrulama değişkenlerin önem katsayısı	52
Şekil 4.18. Yedinci çapraz doğrulama değişkenlerin önem katsayısı.....	52
Şekil 4.19. Sekizinci çapraz doğrulama değişkenlerin önem katsayısı	53
Şekil 4.20. Dokuzuncu çapraz doğrulama değişkenlerin önem katsayısı	53
Şekil 4.21. AİK eğrisi grafiği.....	55

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

hz

Hertz

khz

Kilohertz

p

İstatistiksel anlamlılık seviyesi

r

Korelasyon katsayısı

R²

Determinasyon Katsayısı

Kısaltmalar

Açıklamalar

AİK

Alıcı İşletim Karakteristiği

AUSE

Amerikan Ulusal Standartlar Enstitüsü

FN

Yanlış negatif (False negative)

FP

Yanlış pozitif (False positive)

GBİK

Gürültüye bağlı işitme kaybı

GGM

Günlük gürültü maruziyeti

KKD

Kişisel koruyucu donanım

SGK

Sosyal Güvenlik Kurumu

TN

Gerçek negatif (True negative)

TP

Gerçek pozitif (True positive)

TS

Türk standartları

VWH

Von Willebrand hastalığı

YİT

Yönlü iplik tahtası

1. GİRİŞ

Ağaç ürünleri imalatı sektörü, birçok gercin ve ara mamulün üretiminde var olan, Türkiye'nin önemli gelir kaynaklarının başında gelen bir sektördür. Sektörün ana hammaddesi ağaçtır. Ülkemizde ormanlık alan, 23,1 milyon hektar olup ülke yüzölçümünün % 29,4'ünü oluşturmaktadır. SGK "2024-Şubat Sigortalı İstatistikleri"ne göre, NACE Rev. 2.16 koduyla ekonomik faaliyet gösteren sektör bünyesindeki 11 672 işyerinde; 56 833'ü erkek, 12 546'sı kadın olmak üzere toplam 69 379 çalışan istihdam edilmektedir. Çalışan sayısının onbinleri bulduğu bu sektör, iş kazası ve meslek hastalıkları açısından pek çok maruziyeti de bünyesinde barındırmaktadır. Sektördeki süreç ve faaliyetler incelendiğinde, sektör çalışanlarının maruz kaldığı en önemli meslek hastalığının GBİK olduğu görülmektedir.

Bir çalışanın kişisel gürültü maruziyet seviyesinin tespiti sadece TS EN ISO 9612-"Akustik-Çalışma ortamında maruz kalınan gürültünün belirlenmesi" standardı referans alınarak kişisel gürültü dozimetreleri yardımıyla yapılır. Bu ölçümde çalışana uygun ölçüm stratejisi (görev veya iş tabanlı) seçilmesi, sonucu doğrudan etkilemektedir. Ölçüm sonucunun, 8 saatlik en düşük maruziyet etkin değeri aşması halinde çalışanın gürültüye maruz kaldığı anlaşılmaktadır. Çalışanların gürültü maruziyetinden en belirgin olarak etkilendikleri durum ise işitme kayıplarıdır. İşitme kayıplarının tanısı sadece odyolog ve odyometristler tarafından gerçekleştirilen saf ton odyometrisi ile mümkündür. İşitme kaybı türlerinden geçici işitme kaybı, 16-48 saat gürültüden uzak kalınması ile gerileyip kaybolurken, tespit edilmiş bir kalıcı işitme kaybı geri döndürülemezdir. Kalıcı işitme kaybı, çok yüksek gürültü seviyesine maruziyetle aniden veya 8 saatlik en düşük maruziyet etkin değerini aşan sabit gürültü seviyesine 10-15 yıl süren maruziyetle geniş bir zaman diliminde, çok yavaş oluşabilir. Bu nedenle, gürültü maruziyetinin GBİK yaşanmadan önlenmesi oldukça önemlidir.

GBİK'in önlenmesi farklı bilim dallarının koordineli çalışmasıyla mümkündür. Gürültü maruziyetinin yüksek olduğu hallerde önce gürültüyü kaynağında yok etmeye yönelik önlemler alınması gerekmektedir. Gürültü, kaynağında elimine edilemiyorsa gürültülü ortamda önlemler alınmalıdır. Bu önlemler de gürültü maruziyetini azaltmıyorsa son olarak gürültünün etkisindeki çalışan üzerinde önlemler alınmalıdır. Tüm bu önlemler, sadece disiplinler arası çalışmayla hayata geçirilebilmektedir.

Bir işyerinde kullanılan iş ekipmanlarının ve ulaşım araçlarının türü, sayısı, yerleşimi ve kullanım süresi; mamül/yarımamül üretiminde kullanılan prosesler ve tesis yerleşimine göre değişmektedir. Bu nedenle GBİK araştırılırken sektör ve işyerindeki gürültü kaynakları iyi tanımlanmalıdır. Aşağıda ağaç ürünleri imalatı sektörü ve GBİK ile ilgili genel bilgilere yer verilmektedir.

Ağaç ürünleri

Ağaç ürünleri üretim sektöründe, temel hammadde olarak kullanılan odundan iki türlü üretim yapılmaktadır. Bunlardan ilki ham ağacın direk işlenmesi ile üretilen ürünlerdir. Diğeri ise ham ağaçtan elde edilen tahta talaşı, ahşap tozu ve odun tozunun sıkıştırılması ile üretilen ağaç panel ürünleridir.

Ham ağaç ürünleri: Ham ağaç ürünlerinin hammaddesi, tomruk adı verilen işlem görmemiş yuvarlak ağaç gövdeleridir. Bükme, yarma, kesme, soyma, biçme ve emprenyeleme gibi bir-biri ardına gerçekleştirilen işlemlerin sonunda tomruktan ham ağaç ürünleri elde edilir. Ham ağaç ürünlerinden en önemlileri kereste, masif parke, lamine parke, kaplama ve kontrplaktır.

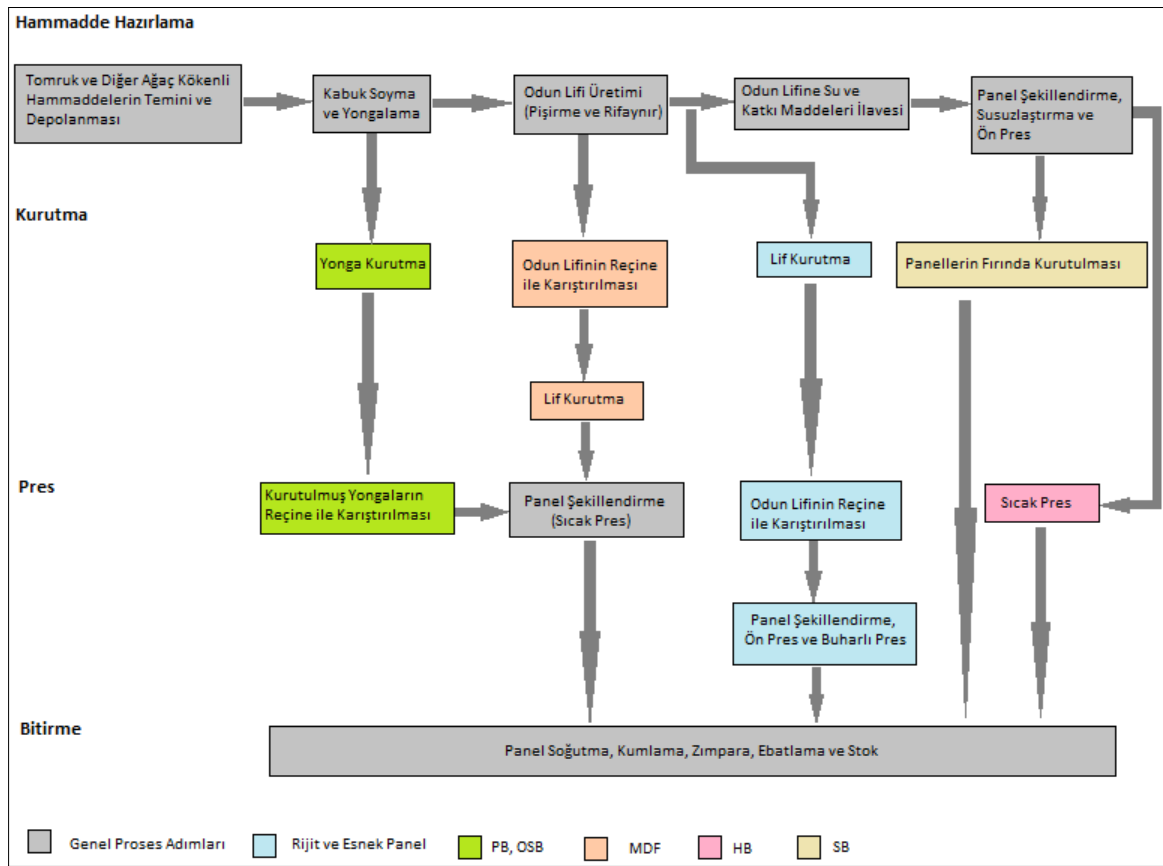
Kereste üretimi: Tomruk hammaddesine uygulanan ilk işlem kabuk soyma işlemidir. Bu işlem ile tomruk temizlenerek bir sonraki prosese hazır hale getirilir. Bu tomruklar daha sonra şerit ya da dairesel testereler yardımıyla çeşitli boyutlardaki keresteler haline getirilir.

Masif ve lamine parke üretimi: Masif parke, çok yıllık sert ağaçlardan ince şeritler kesilmesi ile üretilir. Lamine parke ise ağaç kaplama ve suya dayanıklı kontrplağın birleştirilmesiyle oluşur. Masif ve lamine parke üretiminde, kereste halindeki ara ürüne önce yan alma işlemi uygulanır. Yan alma, kereste üzerinde kalan ağaç yanlarının özel testerelerde temizlenmesidir. Son olarak, pürüzsüz yüzey elde etmek amacıyla ara ürüne planya makinesi ile planyalama işlemi uygulanır.

Kaplama ve kontrplak üretimi: Kaplama ve kontrplak üretiminde özel biçme, dilme ve soyma makineleri kullanılır. Bu makineler yatay veya dikey kesim yapar. Islatılarak yumuşatılmış tomrukların kabukları soyulur. Kabuğu soyulan tomruklardan, üretilmek istenen kaplama türüne göre biçme, dilme ve soyma yöntemleriyle çeşitli kalınlık ve büyüklükte

kaplama levhaları elde edilir. Hazırlanan bu kaplama levhalar bir araya getirilip preslenir ve zımparalanır.

Ağaç panel ürünleri: Ağaç paneller, özel tutkallar ile karışım haline getirilen odun veya odunlaşmış diğer bitki yongalarının, belli sıcaklık ve basınç altında yapılandırılması ve şekillendirilmesi ile üretilir. Yonga levha, sunta, suntalam, OYL (MDF), YİT (OSB) ağaç panel üretim sektöründe üretilen başlıca ürünlerdir. Şekil 1.1’de ağaç panel üretiminin proses akışı verilmiştir.

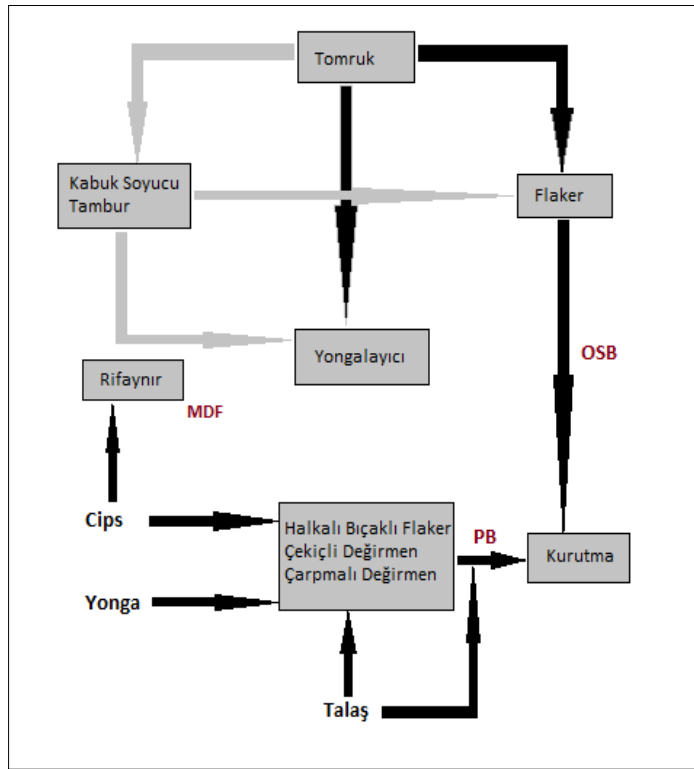


Şekil 1.1. Ağaç panel üretiminin süreç akış şeması (Çevre, Şehircilik ve İklim Değişikliği Bakanlığı [ÇŞİDB], 2020: 4)

Ağaç ürünleri üretim prosesleri

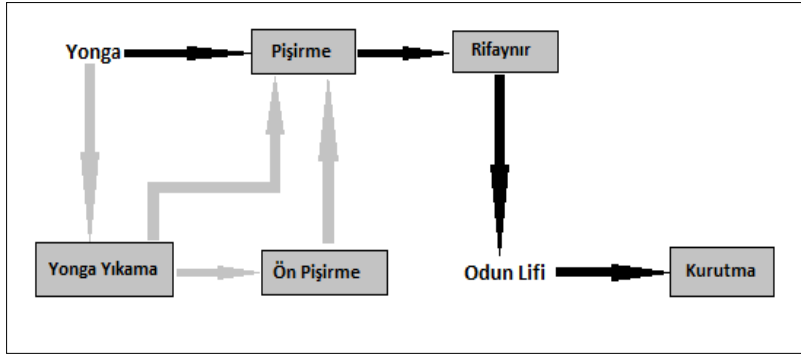
Ağaç ürünleri imalatı sektörü ev, bahçe, işyeri gibi günlük kullanım alanlarında kullanılan ara/nihai birçok ürünü üreten bir sanayi dalıdır. Sektörün ana hammaddesi olan odun, nihai ürünün türüne göre çeşitli yöntemlerle işlenmektedir. Bu yöntemlerin açıklamaları aşağıda sunulmaktadır.

Hammadde hazırlama ve yongalama: Tomrukların hammadde olduğu ürünlerin üretiminin ilk işlemi, döner tamburlarda gerçekleştirilen kabuk soyma işlemidir. Tomruklar, denge rutubetine ulaşabilsin diye en az 2 ay depoda bekletilir. 2 ayın sonunda bu tomruklar 2 aşamalı yongalama işlemine alınır. Birinci aşamada, tomruk, tamburlu (döner bıçaklı) yongalama makinelerinde 0,2 – 0,3 mm kalınlıkta ve 10 – 30 mm boyundaki yonga haline getirilir. İkinci aşamada, yongalar flakerlarda (halka şeklinde birleşik iki bıçak ve bir çekiç sisteminden oluşan değirmenler) kıyılarak daha da küçültülür. Şekil 1.2’de yongalama işleminin proses akışı verilmektedir.



Şekil 1.2. Yongalama işleminin proses akışı (ÇŞİDB, 2020: 5)

OYL (MDF) panel üretiminin hammaddesi odun lifidir. Odun lifi, yongalama makinesinden çıkan çok küçük boyutlu yongaların kazanlarda yüksek sıcaklıkta pişirilmesiyle elde edilir. Pişirilen yongalar rifaynır ünitesine alınır ve disklerin arasından geçerek liflendirme işlemi tamamlanır. Lifler son olarak kurutucuya yönlendirilir. OYL üretiminde kullanılan liflerin proses akışı Şekil 1.3’deki gibidir.



Şekil 1.3. OYL (MDF) üretiminde kullanılan liflerin proses akışı (ÇŞİDB, 2020: 6)

Panel üretimi: Tutkal veya reçine ile karışım haline getirilen kurutulmuş yongalar, istenilen panel ebadına uygun ölçülerdeki pres tablasına serilerek yüksek sıcaklık ve basınç altında preslerle sıkıştırılır. Bu presleme işlemi, plaka halindeki panelin 6 yüzeyinde hiç boşluk kalmayacak şekilde yapılmalıdır.

Bitirme işlemleri: Presten çıkan sıcak paneller, soğuduktan sonra son ebatlarına getirilmek üzere boyutlandırılır. Tüm bu işlemlerin sonunda üretilen paneller, hammadde olarak ya da daha fazla kaplama ve yüzey işleminin ardından son ürün olarak satışa sunulur.

Laminasyon işlemleri: Ağaç panellerin dayanıklılığını arttırmak ve kullanım ömrünü uzatmak amacıyla, bu panellere dekoratif kağıt, vinil, melamin gibi malzemeler kullanılarak laminasyon yapılır. Laminasyon malzemesi, ağaç panelin tek veya her iki yüzeyine silindirik pres kullanılarak yapıştırılır.

Gürültü ve ağaç ürünleri imalatı sektöründe gürültü kaynakları

Gürültü ve GBİK'in anlaşılabilmesi için bazı terimlerin anlaşılması gerekmektedir. Aşağıda bu terimlerden bazıları açıklanmaktadır.

Ses: Nesnelerin yaydığı titreşim ile oluşan dalgaların insan kulağı tarafından algılanmasıdır. Sesin şiddeti, desibel (dB) cinsinden ifade edilir.

Frekans (f): Bir saniyede tamamlanan ses titreşimidir. Frekans, hertz (Hz, 1/s) cinsinden ifade edilir.

Gürültü: İstenmeyen, rahatsız edici, niteliği ve niceliği bozulmuş ses veya seslerdir.

Ses, nesnel bir kavramken sesin verdiği rahatsızlık düzeyi kişiden kişiye değiştiği için gürültü öznel bir kavramdır (Sabancı ve Sümer, 2011). Çalışma hayatı boyunca gürültülü ortamlardaki mesleki gürültü, işitme hasarının yanı sıra birçok sağlık sorununa sebep olmaktadır.

Gürültünün sağlık üzerindeki etkileri

İnsan işitme organı, 20 - 20 000 Hz frekans aralığındaki sesleri algılayabilmektedir. Örneğin, günlük hayattaki standart bir konuşma sesi 500 - 20 000 Hz aralığında olup insan kulağının işitme aralığındadır. 20 Hz'in altındaki (infrasonik ses) ve 20 000 Hz'in üstündeki (ultrasonik ses) sesler, insan işitme organı tarafından algılanmadığı halde, insanlarda sağlık sorunlarına neden olabilmektedir (Öz ve Köse, 2020). Gürültü maruziyetinin insanlarda doğrudan işitme üzerinde veya diğer sistemler üzerinde etki bıraktığı geçmişte yapılan çalışmalarla kanıtlanmıştır.

Gürültü maruziyetinin doğrudan işitme üzerindeki etkileri:

- Şiddetli ve ani bir darbeli gürültünün yol açtığı kalıcı işitme kaybı,
- Kısa ya da uzun süreli ve şiddetli gürültünün yol açtığı geçici işitme kaybı,
- Kısa ya da uzun süreli ve şiddetli gürültünün yol açtığı kalıcı işitme kaybıdır.

Gürültü maruziyetinin diğer sistemler üzerindeki etkileri:

- Nörosensoriyal etkiler (Kulak çınlaması, baş dönmesi vb.),
- Uyku bozuklukları (Uykusuzluk, dinlenememe, baş ağrısı vb.),
- Kardiyovasküler sistem üzerine etkileri (Hipertansiyon, yüksek kolesterol vb.),
- Psikososyal etkileri (Sinirlilik, anksiyete, depresyona eğilimin artması vb.),
- Zihinsel etkilerdir (Dikkat dağınıklığı, motivasyon eksikliği, yorgunluk, iletişim problemleri vb.) (Kim, 2010).

Öznel bir kavram olması sebebiyle, gürültü insanların psikolojik ve otonom sinir sistemi üzerinde farklı etkiler gösterir. Bu nedenle yaş, cinsiyet, açık ten, boy, kilo, kalıtsal yatkınlık, sigara ve ilaç kullanımı gibi kişisel özellikler de işitme kaybının derecesini etkiler.

Gürültü hakkında yasal mevzuat

Gürültülü iş yerlerinde çalışanlarda, gürültüye bağlı etkilerin oluşmaması/azaltılması için ülkemizde Çalışma ve Sosyal Güvenlik Bakanlığı tarafından 28 Temmuz 2013 tarih ve 28721 sayılı Resmi Gazetede “Çalışanların Gürültü İle İlgili Risklerden Korunmalarına Dair Yönetmelik” çıkarılmıştır. Yönetmeliğin uygulanması bakımından dB cinsinden maruziyet eylem değerleri ve maruziyet sınır değerleri aşağıda verilmiştir:

- Günlük 8 saatlik çalışmada en düşük maruziyet eylem değeri = 80 dB(A)
- Günlük 8 saatlik çalışmada en yüksek maruziyet eylem değeri = 85 dB(A)
- Günlük 8 saatlik çalışmada maruziyet sınır değeri = 87 dB(A) (Çalışma ve Sosyal Güvenlik Bakanlığı [ÇSGB], 2013).

Maruziyet eylem değerleri esas alınarak yapılan kişisel gürültü maruziyeti hesaplamasında, çalışanın kişisel kulak koruyucu donanımı kullanıp kullanmadığı önemsenmezken; maruziyet sınır değerleri esas alınarak yapılan kişisel gürültü maruziyeti hesaplamasında çalışanın kişisel kulak koruyucu donanımı kullanıp kullanmadığı önemsenmektedir.

İşitme hasarı, günlük 8 saatlik çalışmada, 85 dB(A) veya daha yüksek seviyedeki gürültüye maruz kalınması ile oluşabilmektedir. Günlük 8 saatlik çalışmada, 85 ile 89 dB(A) seviyesindeki gürültü, uzun süreli maruziyetlerden sonra işitme hasarı oluşturabilmekte iken; 90 dB(A) ve daha yüksek seviyedeki gürültü, kısa süreli maruziyetlerden sonra işitme hasarı oluşturabilmektedir. Günlük 85 dB(A)'dan daha az gürültü maruziyet düzeylerinin gürültü-ilişkili işitme hasarı yaratması pek olanaklı değildir (ÇSGB, 2013).

Ağaç ürünleri sektöründe gürültü kaynakları

En gürültülü sektörler arasında yer alan ağaç işleri imalatı sektöründe, işyerlerinde kullanılan makine ve aletlerin, proseslerin ve çalışılan ürünün türü gürültü seviyesini etkiler. Sektörde kullanılan ekipmanların ortalama gürültü seviyelerinin, mevzuatta yer alan günlük maruziyet sınır değerinden yüksek olduğu Çizelge 1.1’de açıkça görülmektedir.

Çizelge 1.1. Ağaç ürünleri imalatı sektöründeki iş ekipmanlarının gürültü seviyeleri (ÇSGB, 2018: 31)

Makine	Gürültü seviyesi-dB(A)
Panel ebatlama makinesi, zımpara makinesi	97
Delik delme makinesi	98
Panel planya makineleri, dikey eksenli şekillendiriciler	100
Taşınabilir ağaç işleme aletleri	101
Testereler	102
Yüksek hızlı freze makineleri	103
Kenar bantlama ve çoklu kesim makineleri	105
Zıvana açma makinesi	107

Çalışmanın amacı

GBİK, geri dönüşü olmayan bir meslek hastalığıdır. Ancak disiplinler arası uygulamalarla önlenmesi oldukça kolaydır. Bu çalışmanın temel amacı çalışanlarda GBİK'e sebep olan faktörlerin önem sırasının belirlenip, bu faktörlerin ortadan kaldırılmasıyla GBİK'in oluşmasını engelleyecek bir mühendislik yaklaşımı geliştirmektir. Veri setinde, en gürültülü sektörler arasında yer aldığından ağaç ürünleri imalatı sektörü çalışanlarına yer verilmiştir. Çalışma yöntemi olarak en az hata ile tahmin yapmaya olanak sağladığından makine öğrenmesi uygulamaları tercih edilmiştir. Bu çalışma ile makine öğrenmesi algoritmaları kullanılarak en gürültülü sektörler arasında yer alan ağaç ürünleri imalatı sektörü çalışanlarında GBİK gerçekleşmeden işitme kayıpları önceden tahmin edilerek işitme kayıplarının önlenmesi/azaltılması hedeflenmiştir. Literatür araştırması ve istatistiksel analizlerin ardından GBİK'i etkileyen en önemli sekiz değişken tespit edilmiştir. Önem katsayısı en yüksek olan değişkenlerin etkisinin azaltılması veya bertaraf edilmesi ile gelecekte, çalışanlarda GBİK olasılığını önceden tahmin ederek işitme kayıplarını önleyen bir model geliştirmek amaçlanmıştır.

2. LİTERATÜR ÇALIŞMASI

Gelişen dünyada sanayileşmenin artmasıyla üretim artarken makineleşme de artmıştır. Artan makineleşme beraberinde iş kazalarının ve meslek hastalıklarının da artmasına sebep olmuştur. GBİK'in dünyada ve Türkiye'de en sık görülen meslek hastalıklarından biri olduğu bilinmektedir. Sanılanın aksine sadece yüksek şiddetteki gürültü GBİK'e sebep olmamaktadır. Yüksek şiddette olmayan gürültüye uzun süre maruz kalmak da GBİK'e sebep olabilmektedir. Gürültü, çalışanlarda GBİK gibi tedavisi olmayan bir meslek hastalığının yanı sıra psikolojik ve fiziksel pek çok bozukluğa yol açmaktadır. Bu nedenle, gürültü maruziyetinin ve GBİK'in önlenmesi konusunda bugüne kadar birçok araştırma yapılmıştır.

Bu bölümde hangi sektör çalışanlarında GBİK'in daha fazla yaşandığı ve GBİK'i en çok nelerin etkilediği ve bu değişkenlerin GBİK'i ne derece etkilediği ile ilgili dünya ve Türkiye'deki benzer literatür çalışmaları derlenmiştir.

Öztürk ve diğerlerinin (2007) ki kare analizi kullanarak döküm işkolundaki 392 çalışan üzerinde yaptıkları çalışmada, GBİK sıklığı % 23,5 olarak saptanmıştır. Çalışanların halen çalıştıkları bölümlerin gürültüsü ile çalışanlarda GBİK oluşması arasında istatistiksel olarak anlamlı bir fark bulunamamıştır. Gürültülü işte toplam çalışma süresi yaşa, sigara içme (paket-yıl) ve gürültü düzeyine göre düzeltilindiğinde GBİK riskinin 10 yıl üzerinde gürültülü işlerde çalışanlarda 5,9 kat; 15 paket - yıl üzerinde sigara kullanmak gürültülü işte toplam çalışma süresi, yaş ve gürültü düzeyine göre düzeltilindiğinde işitme kaybı 4,8 kat yüksek olduğu bulunmuştur.

Mutlu (2010)'nun 23 maden çalışanına ait kişisel gürültü maruziyet ölçüm ve odyometri testi sonuçlarından yararlanarak, aşamalı loglineer analizi kullanarak yaptığı çalışmada GBİK'i etkileyen en önemli unsurun, kişisel gürültü maruziyet ölçüm sonuçları olduğu tespit edilmiştir. Ayrıca kişisel maruziyet ölçüm sonucu 70-79 dB arasında olan grubun en riskli grup olduğu görülmüştür. Yaşı 37- 45 arasında değişen ve aynı işyerinde çalışma süresi 4-11 yıl arasında değişen çalışanların ise en riskli grubu oluşturduğu tespit edilmiştir.

Noma ve Ghani (2013)'ye ait çalışmanın ana katkısının, hibrit makine öğrenimi yaklaşımlarına dayalı olarak işitme kaybı semptomlarını doğru bir şekilde sınıflandırmak için bir semptom tespit modeli önermek olduğu görülmüştür. Sık desen büyümenin, özellik çıkarma amacıyla kullanılan denetimsiz bir yöntem olduğu; Naive Bayes modellerinin ise, sınıflandırma hedefine yardımcı denetimli modeller olduğu ortaya konulmuştur. Özellik çıkarmalı Naive Bayes Bernoulli modeli en düşük ortalama hata oranlarını verdiği için en iyi model seçilmiştir.

Erdoğan (2016) çalışmasında, genel GBİK sıklığı % 38,9 olarak tespit edilmiştir. GBİK sıklığı, 85dB ve üzeri gürültülü ortamda çalışanlarda %17,2; gürültü düzeyi daha yüksek olan ortamda çalışanlarda %19,5; erkekler çalışanlarda %19,9; 40 yaş ve üzerindeki erkek çalışanlarda %31,3; şuan çalıştıkları işte 5 yıldan fazla çalışanlarda %18,4; gürültülü işte toplam 15 yıl ve üzeri çalışanlarda %21,5; KKD hiç kullanmayanlarda %33,3; sigara içenlerde %20,4 ve toplam sigara tüketimi 15 paket/yıl ve üzeri olanlarda %37,2 olarak tespit edilmiştir. GBİK oluşumunu en çok etkileyen değişkenler gürültülü ortamda çalışma, gürültüye maruz kalınan süre ve KKD kullanımının yetersizliği olarak bulgulanmıştır.

Özdemir (2016) kaporta işletmelerinde, çalışanlardan alınan sekiz saatlik günlük gürültü maruziyeti ölçüm sonuçlarına göre yapılan çalışmada, çalışanların % 95'inin maruziyet sınır değerinden yüksek değerlerdeki gürültüye (ortalama 91,9 dB(A)) maruz kalarak çalıştığı ve çalışanların % 63,5'inde gürültüye bağlı işitme kaybının olduğu tespit edilmiştir. Bu oranın, ağaç ürünleri imalatı sektöründeki GBİK'li çalışan oranından daha yüksek olduğu tespit edilmiştir.

Önder (2018)'e ait, 22-53 yaş aralığındaki 109 maden çalışanı üzerinde, logistik regresyon analizi kullanılarak yapılan araştırmada; yaş, çalışma yılı, yapılan iş, kişisel gürültü maruziyet ölçüm sonuçları ile işitme kayıpları arasında anlamlı bir ilişki olduğu ispatlanmıştır. Kurulan model ile yaş değişkenindeki 1 birimlik artışın işitme kaybı olasılığını 1,112 kat artırdığı; ortamın eşdeğer gürültü seviyesindeki 1 dB'lik artışın işitme kaybı olasılığını 1,719 kat artırdığı; negatif katsayılı değişkenlerden darbe gürültü düzeyindeki 1 dB'lik artışın işitme kaybı yaşamama olasılığını 1,086 kat artırdığını; çalışanların maruz kaldıkları gürültü düzeyindeki 1 dB'lik artışın işitme kaybı yaşamama olasılığını 1,612 kat artırdığı tespit edilmiştir.

Zhao ve diğerkleri (2018) tarafından, 1 644 sanayi çalıřanı üzerinde yapılan çalıřmada, yař; cinsiyet; aynı iř yerinde çalıřma süresi (yıl) ve günlük kiřisel gürültü maruziyet seviyesinin GBİK üzerine etkisi arařtırılmak üzere, adaboost, destek vektör makineleri, rastgele orman ve çok katmanlı algılayıcı yapar sinir ağıları algoritmalarıyla sınıflandırma modelleri kurulmuřtır. Bu sınıflandırma modelleri arasında %80,14 doğruluk oranıyla destek vektör makineleri algoritması en iyi algoritma seçilmiřtir. Seçilen destek vektör makineleri algoritması modelinin 0,808'lik AİK puanı ise pozitif deęerler ile negatif deęerler arasında mükemmel bir ayırım olduęunu göstermiřtir.

Yin ve diğerkleri (2019) tarafından, gürültülü bir sektördeki 9 780 çalıřana ait iřitme testi ve kiřisel gürültü maruziyeti ölçüm sonuçlarının da içinde yer aldıęı 6 bağımsız deęiřkenin GBİK'e etkisinin arařtırıldıęı çalıřmada, rastgele orman algoritması kullanılarak modeller kurulmuřtır. İřitme testi sonucu iyi olan ancak kötü olduęu tahmin edilenler direnç grubu olarak ve iřitme testi sonucu kötü olan ancak iyi olduęu tahmin edilenler duyarlı grup olarak tanımlanmıřtır. 12 kHz'ın, 4 kHz'e göre daha duyarlı frekans olduęu ve kadınların GBİK'e daha dirençli olduęu tespit edilmiřtir.

Chen ve diğerkleri (2021) tarafından, GBİK'i tahmin etmek için makine öğrenimini kullanmanın avantajları ve sınırlamalarının arařtırıldıęı çalıřmada, ilgili geçmiř çalıřmalar analiz edilmiřtir. Sonuçta, GBİK'i tahmin etmede makine öğreniminin yüksek doğruluk ve düşük tahmin hatası ile modeller kurduęu tespit edilmiřtir. Ancak sınırlı örneklem boyutlarından, tek algoritma kullanımından, model oluřturmayla ilgili eksik raporlardan, pozitif ve negatif sınıfları ayırt etme performansının yetersizlięinden dolayı, AİK eęrilerinde kötü duyarlılık ve özgülük deęerleri ile karřılařılmıřtır. Bu nedenle model uyumunu artırmak için örneklem sayısının ve GBİK ile ilgili bağımsız deęiřken sayısının arttırılmasının önemi vurgulanmıřtır.

Fan ve diğerkleri (2022) tarafından yař, çalıřma süresi, yüksek frekans eřięi ve gen varyantlarının GBİK'e etkisinin arařtırıldıęı çalıřmada karar aęacı, lojistik regresyon ve rastgele orman algoritmaları kullanılarak sınıflandırma modelleri kurulmuřtır. Rastgele orman algoritması %91,1 doğruluk ve % 97,5 AİK puanı ile en iyi algoritma seçilmiřtir. VWH hastalıęının GBİK ile doğrudan iliřkili olduęu tespit edilmiřtir. Modelin iyi tahmin deęerinin, doğru tahmin edilen 73 duyarlılık gen varyasyonuna baęlı olduęu kanıtlanmıřtır.

Selvaganesh ve diğerleri (2023) tarafından, yaşa bağlı olarak işitme kaybı tahmin edilerek erken teşhis ve tedavi imkanı sunan çalışmada, makine öğrenmesi sınıflandırma algoritmaları kullanılmıştır. Dinamik frekanstaki ses esas alınarak akustik simülasyon gerçekleştirilmiş ve tahmin işlemi için kullanıcı geri bildirimleri alınmıştır. Çalışmada önce mevcut frekansları belirlemek için iki taraflı bir simülasyon gerçekleştirilmiştir. Önyargılı olasılık sinir ağı modeli kullanılarak işitsel tepkiye uygun olarak sinyal uyarımı sonrasında yaş tahmini yapılmaktadır. Daha sonra akustik dalgalar ve beklenen yaş değerlendirilerek insandaki işitme bozukluğunun tespit edilmesi amacıyla düzenlenmiştir. Ekstrem öğrenme makinesi uygulanmıştır. Sonuçta işitme kaybı olan ve işitme kaybı olmayan kişileri sınıflandırmak için seçilen model % 99 oranında doğruluk göstermiştir.

Çizelge 2.1’de verilen tüm bu çalışmalar makine öğrenimi algoritması ve kurulacak model seçiminin, çalışmanın amacına, tasarımına, örneklemdaki değişkenlerin çeşidine göre yapılması gerektiğini savunmaktadır.

Çizelge 2.1. GBİK alanında literatür araştırması tablosu

Yazar, yıl	Çalışmanın amacı	Kullanılan yöntem
Öztürk ve diğerleri (2007)	Döküm işkolunda gürültüye bağlı işitme kayıpları sıklığı ve etkileyen etmenlerin değerlendirilmesi	Ki kare analizi
Mutlu (2010)	Madencilikte Gürültüye Bağlı İşitme Kayıplarının Tespiti: Taş Kırma Eleme Tesisi Örneği	Logaritmik doğrusal analiz
Noma ve Ghani (2013)	Makine Öğrenimi Algoritmalarını Kullanarak Odyometri Verilerinden İşitme Kaybı Belirtilerini Tahmin Etme	Sık desen büyüme alg., Naive Bayes (Bernoulli) sınıflandırması
Erdoğan (2016)	Denizli’de Üç Tekstil Fabrikasındaki Gürültü Düzeyinin Çalışanlar Üzerine Etkisi	Ki-kare, Mann-Whitney U testi
Özdemir (2016)	Küçük Ölçekli Kaporta İşletmelerinde Çalışanlarda Gürültüye Bağlı İşitme Kayıplarının Araştırılması	Tek yönlü varyans analizi, ki kare testi ve tek örneklem T testi metodu
Önder (2018)	Lojistik Regresyon Analizi ile Bir Yeraltı Metal Madeninde Gürültüye Bağlı İşitme Kayıplarının İncelenmesi	Makine öğrenmesi - lojistik regresyon

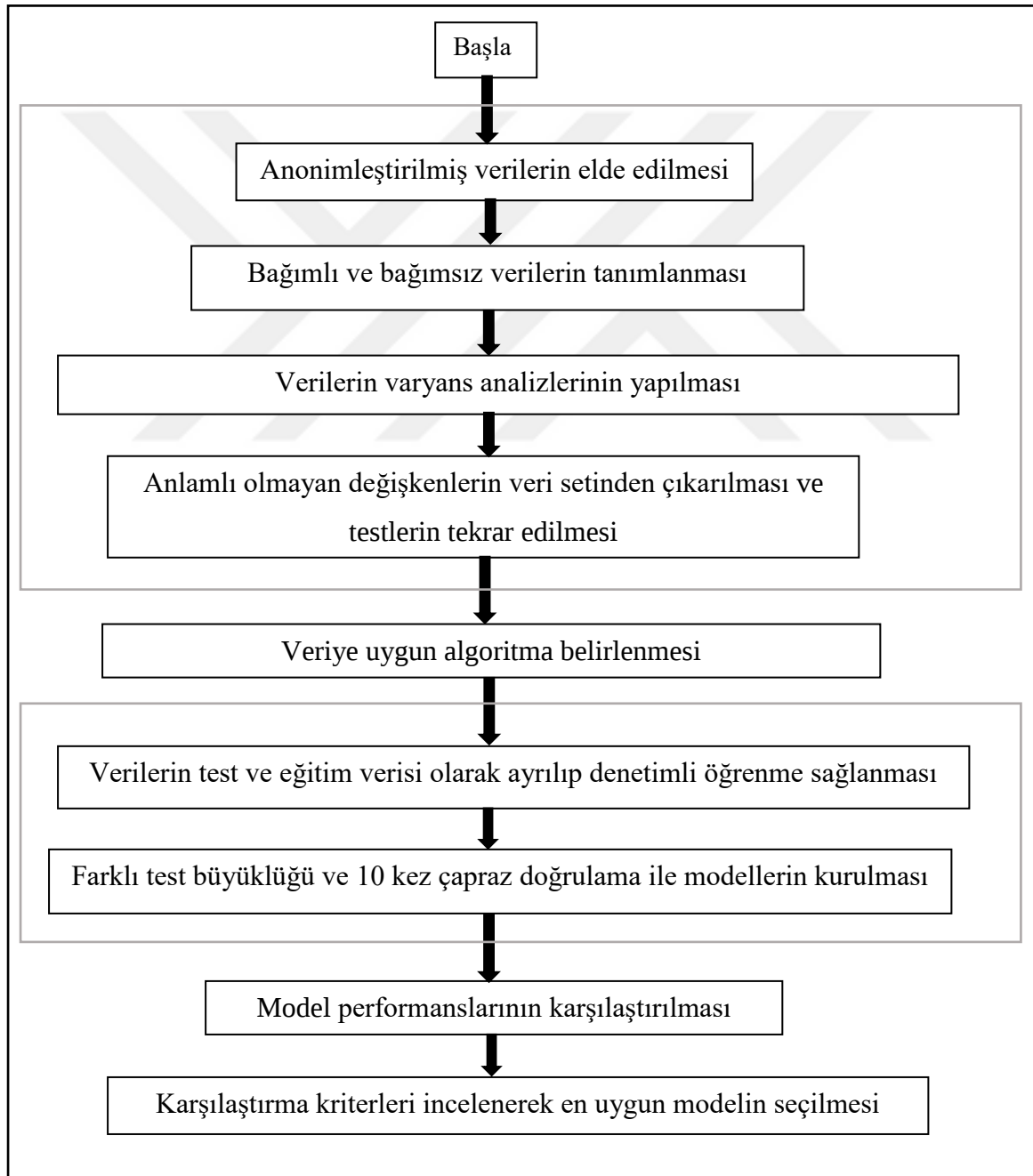
Çizelge 2.1. (devam) GBİK alanında literatür araştırması tablosu

Yazar, yıl	Çalışmanın amacı	Kullanılan yöntem
Zhao ve diğerleri (2018)	Karmaşık Endüstriyel Gürültüye Maruz Kalan Çalışanlarda İşitme Engelinin Tahminine Yönelik Makine Öğrenimi Modelleri: Pilot Çalışma	Uyarlanabilir artırma, çok katmanlı algılayıcı ağı, rastgele orman, destek vektör makinesi
Yin ve diğerleri (2019)	Gürültüye Bağlı İşitme Kaybı ve Bireysel Duyarlılık	Rastgele orman alg.
Chen ve diğerleri (2021)	Gürültüden Kaynaklanan İşitme Kaybını Tahmin Etmek İçin Makine Öğrenimini Kullanmanın Katkıları ve Sınırlamaları	Yapar sinir ağları, rastgele orman alg., destek vektör makinesi
Fan ve diğerleri (2022)	Makine Öğrenimi Modeli, Tüm Ekzom Sıralaması Yoluyla Gürültüye Bağlı İşitme Kaybındaki Genetik Çeşitliliğin Tanımlanması	Makine öğrenmesi - rastgele orman alg., lojistik regresyon, karar ağacı alg.
Selvaganesh ve diğerleri (2023)	Yeni Bir Önyargılı Olasılık Sinir Ağı ve Düzenleştirilmiş Ekstrem Öğrenme Makinesi tabanlı İşitme Kaybını Tahmin Sistemi	Önyargılı olasılık sinir ağı, düzenleştirilmiş ekstrem öğrenme makinesi



3. MATERYAL VE METODOLOJİ

Ağaç ürünleri imalatı sektöründe çalışanların cinsiyetinin, yaşının, toplam çalışma süresinin, sektörde çalışma süresinin, gürültülü ortamda çalışma süresinin, sigara kullanımının, gürültülü hobisinin ve KKD kullanımının, çalışanlarda GBİK'e etkisinin makine öğrenmesi algoritmalarıyla araştırılması için yapılan tez çalışmasında uygulanan yöntemin aşamaları Şekil 3.1'de sırasıyla sunulmuştur.



Şekil 3.1. Uygulama basamakları akışı

Veri setinin denetimli öğrenme sınıflandırma modelleri, Python programlama dili kullanılarak Jupyter Notebook'ta kurulmuştur. Python, kolay ve hızlı uygulama ve anlama olanağı sağlaması ve kapsamlı kütüphanesi nedeniyle özellikle makine öğrenimi ve yapay zeka alanlarında sıkça tercih edilen bir programlama dilidir.

3.1. Veri Toplama

Bu çalışmada, devlete ait bir iş hijyeni laboratuvarı tarafından, Türkiye'nin farklı illerinde, ağaç ürünleri imalatı sektörü çalışanlarından alınan anamnez formları, çalışanlar üzerinde yapılan işitme test ve analiz verileri kullanılmıştır.

3.1.1. Araştırma grubu (örneklem)

Çalışmanın araştırma grubunu, Türkiye'nin farklı illerinde ağaç ürünleri imalatı sektöründeki 1 477 çalışan oluşturmaktadır. Araştırma grubu, ön doktor muayenesinden geçen ve işitme testine uygun kişilerden oluşmaktadır.

Çalışanlara ait odyometri testi sonuçları, sağlık verisi olduğundan, Kişisel Verilerin Korunması Kanunu gereğince veri kullanıcısı tarafından anonimleştirilerek istatistiki veri statüsüne getirilmiştir. Bu istatistiki verilerin çalışmada kullanılmasında etik açıdan bir sakınca olmadığını gösteren 05.06.2023 tarihli “Etik Kurul İzni”nin, Gazi Üniversitesi Etik Komisyonu’ndan alınmasından sonra söz konusu veriler tez çalışmasında kullanılmıştır.

3.1.2. Onam formu

Devlete ait bir iş hijyeni laboratuvarı sağlık personeli tarafından hazırlanan “sosyodemografik özellikler”, “çalışma hayatına ilişkin özellikler”, “İş sağlığı ve güvenliği uygulamalarına yönelik özellikler”, “sağlık durumuna ilişkin özellikler”, “muayeneler” konu başlıklarındaki 63 soruluk anket formları katılımcıların rızası ile doldurulmuştur. 1 477 çalışana ait onam formlarındaki bu çalışmanın konusuna uygun soruların cevapları, gerekli izinler alındıktan sonra, ilgili laboratuvarın veri sorumlusu tarafından anonimleştirilerek istatistiksel analiz için uygun hale getirilmiştir.

3.1.3. Odyometri testi deęerlendirmesi ve gürültüye baęlı işitme kaybı

İşitme (odyometri) testi odyometrist ve odyologlar tarafından yetkili hekim nezaretinde sessiz kabinde gerçekleştirilmektedir. Çalışanın her iki kulağına farklı zaman dilimlerinde 500, 1 000, 3 000, 4 000, 6 000 ve 8 000 Hz frekanslarında sesler şiddeti artırılarak sırayla gönderilmektedir. Çalışan her sesi her işittiğinde elindeki düğmeye dokunmaktadır. Çalışanın hangi şiddetteki sesi duyduğu veya duymadığı odyogram deęerlendirmesine işlenmektedir. Genel işitme kaybı derecelendirmesi Çizelge 3.1'deki AUSE (ANSI) kaynakları dikkate alınarak yapılmaktadır.

Çizelge 3.1. AUSE kaynaklarına göre işitme kaybı dereceleri

İşitme eşik düzeyi	İşitme kaybı derecesi
-10 - 20 dB	Normal işitme
20 - 40 dB	Hafif işitme kaybı
40 - 70 dB	Orta derecede işitme kaybı
70 - 90 dB	Şiddetli işitme kaybı
90 + dB	Derin işitme kaybı

Ancak oluşan işitme kaybının gürültüye mi yoksa yaşlılığa mı baęlı olarak oluştuğunu anlamak oldukça güçtür. Bu nedenle bir işitme kaybı tanısının gürültüye baęlı olup olmadığını anlamak için odyogram üzerinden inceleme yapılmaktadır. Şekil 3.2'deki odyogramda görüldüğü gibi, 3 000 - 6 000 Hz arası frekanslardaki işitme kaybı, 500-2 000 Hz arası frekanslardakinden daha fazladır. Gürültüye maruziyet devam ettikçe 4 000 Hz altındaki sesler yani konuşma frekansları etkilenmeye başlamakta; 4 000 Hz civarındaki çentik daha derinleşip genişlemektedir. Örneğin, 140 dB şiddetindeki ani darbe gürültüsü ile meydana gelen işitme kaybında bu çentik derin ve dar; 90 dB şiddetindeki sabit gürültüye uzun süre maruziyet ile meydana gelen işitme kaybında bu çentik geniştir. 125-2 000 Hz arasındaki işitme kayıpları ise genellikle başka sebeplerden kaynaklanmaktadır.

Gürültüye bağlı işitme kayıpları, koklear sinir liflerine zarar vererek meydana gelmektedir. Bunun gibi iç kulak veya vestibülokoklear sinir hasarları şeklinde oluşan işitme kayıplarında, koklear sinir liflerinin yapısı bozulduğundan koklear sinir liflerinin eski haline dön­dürülmesi için herhangi bir tedavi bulunmamaktadır.

1 477 çalışana ait veri seti “cinsiyet”, “gürültülü ortamda çalışma süresi (ay)”, “sigara kullanımı”, “gürültülü hobi”, “kulak koruyucu donanım kullanımı” değişkenlerinde ve odyometri testi sonuçlarında yapılan yapay kodlamalar ile analize uygun hale getirilmiştir.

3.2.1. Değişkenlerin seçimi

Bu çalışmada ağaç ürünleri imalatı sektöründe çalışanların cinsiyetinin, yaşının, toplam çalışma süresinin (yıl), sektörde çalışma süresinin (yıl), gürültülü ortamda çalışma süresinin (ay), sigara kullanımının, gürültülü hobisi olup olmamasının ve yetersiz kulak koruyucu donanım kullanımının çalışanlarda gürültüye bağlı işitme kaybına etkileri değerlendirilmiştir.

Çalışmanın bağımsız değişkenleri;

Cinsiyet: Onam formlarındaki “sosyodemografik özellikler” konu başlığında yer alan “cinsiyetiniz nedir?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Kadın cinsiyeti “0”, erkek cinsiyeti “1” yapay değişken kodlaması ile analize dahil edilmiştir. Cinsiyet değişkeni kategoriktir.

Yaş: Onam formlarındaki “sosyodemografik özellikler” konu başlığında yer alan “doğum tarihiniz nedir?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Nominal bir değişken olan yaş değişkeni 18-60 aralığındadır.

Toplam çalışma süresi (yıl): Onam formlarındaki “çalışma hayatına ilişkin özellikler” konu başlığında yer alan “toplam kaç yıldır çalışıyorsunuz?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Toplam çalışma süresi değişkeni sürekli dir.

Sektörde çalışma süresi(yıl): Onam formlarındaki “çalışma hayatına ilişkin özellikler” konu başlığında yer alan “bu sektörde kaç yıldır çalışıyorsunuz?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Sektörde çalışma süresi değişkeni sürekli dir.

Gürültülü ortamda çalışma süresi (ay): Onam formlarındaki “sağlık durumuna ilişkin özellikler” konu başlığında yer alan “toplam kaç yıldır gürültülü ortamda çalışıyorsunuz?” sorusuna verilen cevapların aya çevrilerek toplanması ile Çizelge 3.2’deki gibi kategorize edilerek elde edilmiştir. Gürültülü ortamda çalışma süresinin ay cinsinden değerlendirilmesinin sebebi yasal mevzuat kaynaklıdır. Sosyal Sigortalar Kanunu Sağlık İşlemleri Tüzüğü’nün Meslek Hastalıkları Listesi’nde yer alan “Gürültü zararlarının meslek hastalığı sayılabilmesi için gürültülü işte en az iki yıl, gürültü şiddeti sürekli olarak 85 dB(A)’nın üstünde olan işlerde en az 30 gün çalışılmış olması gereklidir.” hükmü nedeniyle anemnez formlarında

gürültülü ortamda çalışma süresi ay cinsinden yazılmaktadır (Sosyal Sigortalar Kurumu [SGK], 1972). Gürültülü ortamda çalışma süresi değişkeni kategoriktir.

Çizelge 3.2. Gürültülü ortamda çalışma süresi kategorileri

Gürültülü ortamda çalışma süresi	Kategori
0-1 ay	1
2-24 ay	2
25-60 ay	3
61-120 ay	4
120 ve üzeri	5

Sigara kullanımı (evet/hayır): Onam formlarındaki “sağlık durumuna ilişkin özellikler” konu başlığında yer alan “sigara kullanıyor musunuz?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Bu değişken, ilgili soruya cevap “hayır” ise “0”, “evet” ise “1” yapay değişken kodlaması ile analize dahil edilmiştir. Sigara kullanımı değişkeni kategorik tir.

Gürültülü hobi (var/yok): Onam formlarındaki “sağlık durumuna ilişkin özellikler” konu başlığında yer alan “gürültülü hobiniz var mı?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Bu değişken, ilgili soruya cevap “yok” ise “0”, “var” ise “1” yapay değişken kodlaması ile analize dahil edilmiştir. Gürültülü hobi değişkeni kategoriktir.

Kulak koruyucu donanım kullanımı (var/yok): Onam formlarındaki “İş sağlığı ve güvenliği uygulamalarına yönelik özellikler” konu başlığında yer alan “İşyerinde koruyucu malzeme kullanıyor musunuz?” sorusuna verilen cevapların toplanması ile elde edilmiştir. Bu değişken, ilgili soruya cevap “yok” ise “0”, “var” ise “1” yapay değişken kodlaması ile analize dahil edilmiştir. Kulak koruyucu donanım kullanımı değişkeni kategoriktir.

Çalışmanın bağımlı değişkeni;

Odyometri testi sonucu: Çalışanlarda gerçekleştirilen işitme tarama testi (odyogram) sonuçlarının toplanması ile elde edilmiştir. İşitme testi yapılmadan önce her çalışana doktor tarafından genel bir fiziksel ve otolojik muayene yapılmıştır. Olası iletim tipi işitme kaybını bertaraf etmek için her çalışana otoskopik ve timpanometrik tarama yapılmıştır. Geçmişte anlık gürültüye maruz kalmış veya doğuştan gelen bir işitme kaybı olan çalışanlar gibi, bu

muayeneler sonucu uygun bulunmayan çalışanlar işitme tarama testine alınmamıştır. Gürültüye aynı sürede ve eşit mesafede maruz kalındığından GBİK çoğunlukla çift taraflıdır. Bu nedenle odyometri sonuçlarında herhangi bir tarafta bariz ve daha yüksek bir kalıcı işitme kaybı olan çalışanlar, iletimsel tipte işitme kaybına sahip olabileceğinden veri setine katılmamıştır. Çünkü iletim tipi işitme kayıpları genellikle yaş, hastalık veya genetik etkenler sonucu oluşan dış ve orta kulaktaki hasarlardan kaynaklanır.

Odyogram sonuçlarına göre sınıflandırma algoritma metodu ile analiz yapılmıştır. Kategorik bağımlı değişkenli sınıflandırma algoritmaları için bağımsız değişkenlerin işitme kaybına yol açıp açmaması incelendiğinden, bağımlı değişken odyometri test sonuçlarında 20 dB üzerindeki değerler için “işitme kaybı vardır (1)” ve 20 dB ve altındaki değerler için “işitme kaybı yoktur (0)” olarak iki kategorilidir.

Çalışmadaki değişkenlerin adı, tanımı ve değişken tipinin özeti Çizelge 3.3’te verilmiştir.

Çizelge 3.3. Değişkenlerin adı, tanımı ve tipi

Değişken adı	Değişken tanımı	Değişken tipi
Cinsiyet	Çalışanın cinsiyeti (Kadın/Erkek)	Kategorik
Yaş	Çalışanın yaşı	Nominal
Toplam çalışma süresi	Çalışanın işitme testi yapıldığı tarihe kadar yıl cinsinden toplam çalışma süresi	Nominal
Sektörde çalışma süresi	Çalışanın işitme testi yapıldığı tarihe kadar test yapılan iş yerinde yıl cinsinden toplam çalışma süresi	Nominal
Gürültülü ortamda çalışma süresi	Çalışanın işitme testi yapıldığı tarihe kadar test yapılan iş yerinde gürültülü ortamda ay cinsinden toplam çalışma süresi (0-120 ay 5 kategoriye ayrılmıştır.)	Kategorik
Sigara kullanımı	Çalışanın sigara kullanıp kullanmaması	Kategorik
Gürültülü hobi	Çalışanın iş hayatı dışında gürültülü olarak tanımlayabileceği hobisinin olup olmaması	Kategorik
KKD kullanımı	Çalışanın iş yaptığı süre boyunca kulak koruyucu donanım kullanıp kullanmaması	Kategorik
İşitme kaybı (var/yok)	Çalışanın kulaklarının birinde veya her ikisinde işitme kaybı olup olmaması (20 dB üzeri işitme kaybı vardır)	Kategorik

3.2.2. Veri setinin sınırları

Regresyon probleminde, başlangıçta, kişisel günlük gürültü seviyesinin de GBİK'i etkileme düzeyi de araştırılmak istenmiştir. 327 çalışana ait kişisel gürültü maruziyet ölçüm değerinin, dB cinsinden çalışanlara ait odyometri sonucuna etkisi incelenmek amacıyla istatistiksel varsayımlar araştırılırken kişisel gürültü maruziyet ölçüm değerinin GBİK'e etkisini inceleyen modelin varyans analizi sonuçlarında istatistiksel anlamlılık düzeyinin ($p=0.991$) 0.05'ten büyük çıkması sebebiyle model anlamlı çıkmamıştır. Çalışanların kişisel gürültü maruziyet seviyeleri ile GBİK arasında istatistiksel olarak anlamlı bir ilişki bulunamamıştır. Varsayımlar karşılanmadığı için ve çalışanların kişisel gürültü maruziyet seviyeleri ile GBİK arasında istatistiksel olarak anlamlı bir ilişki bulunmaması nedeniyle kişisel gürültü maruziyet ölçüm değerinin GBİK'e etkisini inceleyen modelden vazgeçilmiştir.

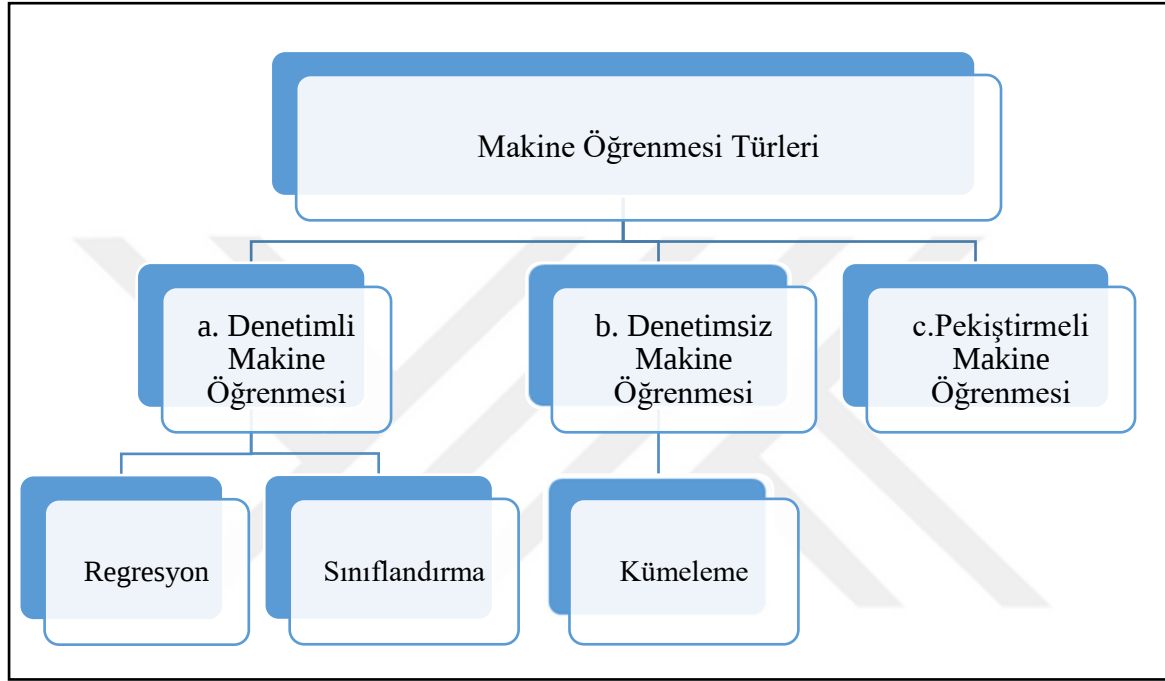
Sonuçta, dokuz değişken yerine hem örneklem sayısının az olması hem de istatistiksel anlamlılık düzeyinin düşük olması sebebiyle “kişisel gürültü maruziyet ölçüm değerleri” değişkeni veri setinden çıkarılarak sekiz değişken ile modeller kurulmuştur. Literatür incelemesinde ise bu çalışmaya konu sekiz değişkenden “cinsiyet”, “yaş”, “sektörde çalışma süresi”, “gürültülü ortamda çalışma süresi”, “sigara kullanımı”, “KKD kullanımı” değişkenlerinin daha önceki benzer çalışmaların analizlerine dahil edildiği; ancak “toplam çalışma süresi”, “gürültülü hobi” değişkenlerinin daha önceki benzer çalışmaların analizlerine dahil edilmediği görülmüştür.

3.3. Makine Öğrenmesi

Makine öğrenmesi makinenin verileri analiz ederek insan gibi öğrenmesidir. Makine önce seçilen algoritmayı kullanarak verinin özelliklerini belirlemektedir. Öğrendiği özellikler ile tahminde kullanacağı veri modelini oluşturmaktadır. Yani makine öğrenmesi, analiz ettiği eğitim verisi içerisindeki kural ve kalıpları içeren modelleri otomatik olarak türetmeyi hedeflemektedir.

Makine öğrenmesi tahmin, sınıflandırma, görüntü ve ses tanıma, otomatik dil çevirisi, tıbbi teşhis, online ürün tavsiyesi, e-posta spam ve kötü amaçlı yazılım filtreleme, çevrimiçi dolandırıcılık tespiti, borsa gibi pek çok alan ve amaçta geniş bir kullanım alanına sahiptir.

Makine öğrenimi algoritması türleri insan denetiminin dahil olma derecesine göre denetimli makine öğrenmesi denetimsiz makine öğrenmesi ve pekiştirmeli makine öğrenmesi olmak üzere Şekil 3.3'deki gibi üç türe ayrılmaktadır. Bu türlerin dışında, denetimli ve denetimsiz makine öğrenmesinin en iyi özelliklerini kullanarak hem etiketli hem de etiketsiz veriler üzerinde eğitilen yarı denetimli makine öğrenmesi diye de bir tür bulunmaktadır.



Şekil 3.3. Makine öğrenmesi türleri (Naqa ve Murphy, 2015)

Denetimli makine öğrenmesi, hem tahmin hem de sınıflandırma modellerinde kullanılan, sadece girdi-çıkıttı çiftinden oluşan veri setleri için uygun olan makine öğrenmesidir. Daha önce görülmemiş yeni veriler üzerinde en az tahmin hatasıyla tahminler yapabilen bir model oluşturmak için girdilerden bir eğitim veri seti oluşturmaktadır. Girdi verilerine eğitim verileri denir. Çıktılar ise tahmin etmek istediğimiz etiketler veya doğru yanıtlardır. Denetimli makine öğrenmesi, regresyon ve sınıflandırma olmak üzere iki ana amaç için kullanılmaktadır.

Regresyon, bağımlı (çıkıttı) ve bağımsız (girdi) değişkenler arasındaki ilişkiyi inceleyerek tahmin yapılmasını sağlayan istatistiksel bir yaklaşımdır. Bağımsız ve bağımlı değişkenler arasındaki ilişkiyi R^2 ve r değerlerini kullanarak incelemektedir. R^2 değeri bağımsız değişkenin regresyon modelinde bağımlı değişkendeki değişkenliğin ne kadarını açıkladığını temsil eden bir istatistiksel ölçüdür. R^2 değeri 0 ile 1 arasındadır ve 1'e ne kadar yakınsa kurulan

model o kadar başarılıdır. r değeri ise değişkenler arasındaki korelasyon katsayısıdır. r değeri (-1) ile $(+1)$ arasında bir değer almaktadır. Pozitif değerler bağımlı değişken ve bağımsız değişkenler arasında aynı yönlü doğrusal ilişkiyi; negatif değerler ise değişken ve bağımsız değişkenler arasında ters yönlü bir doğrusal ilişkiyi belirtmektedir. r değeri $+1$ 'e ne kadar yakınsa bağımlı değişken ve bağımsız değişkenler arasındaki ilişki o kadar yüksektir.

Regresyonla, bağımlı değişkenin sürekli olduğu problemlerde, değişkenler arasındaki bağlantıyı optimize eden en uygun fonksiyon belirlenir. En uygun model fonksiyonu problemde en iyi tahmini yapmaktadır.

Basit doğrusal regresyon çoklu doğrusal regresyon, kademeli doğrusal regresyon, polinom-sal regresyon, lojistik regresyon, ridge regresyon, lasso regresyon, kantil regresyon, elastik net regresyon, sıralı regresyon en bilinen regresyon analizi türleridir.

Sınıflandırma algoritmaları, bağımlı değişkenin kategorik olduğu problemlerde, verileri önceden ayarlanmış kategorilere atamak için tahmine dayalı hesaplamalar kullanılmaktadır.

Sınıflandırma algoritmaları, bağımlı değişkenin kategorik olduğu problemlerde, verileri önceden ayarlanmış kategorilere atamak için tahmine dayalı hesaplamalar kullanılmaktadır. Sonuçta hata matrisi (Şekil 3.4) ve AİK eğrisi elde edilir. Performans kriterleri ise doğruluk, kesinlik, duyarlılık, F-1 puanı ve AİK puanıdır.

	Tahmin değerleri	
Gerçek değerler	Gerçek pozitif (TP)	Yanlış negatif (FN)
	Yanlış pozitif (FP)	Gerçek negatif (TN)

Şekil 3.4. Hata matrisi örneği

Doğruluk değeri, gerçekten doğru tahmin ettiğimiz pozitif ve negatif değerlerin tüm değerlerin toplamına oranıdır. Doğruluk değeri modelin başarısını değerlendirme için tek başına yeterli değildir. Hesaplanması için gerekli formülasyon (3.1)'deki gibidir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.1)$$

Kesinlik, doğru tahmin edilen pozitif değerlerin, pozitif olarak tahmin edilen tüm değerlere oranıdır. Hesaplanması için gerekli formülasyon (3.2)'deki gibidir.

$$Kesinlik = \frac{TP}{TP+FP} \quad (3.2)$$

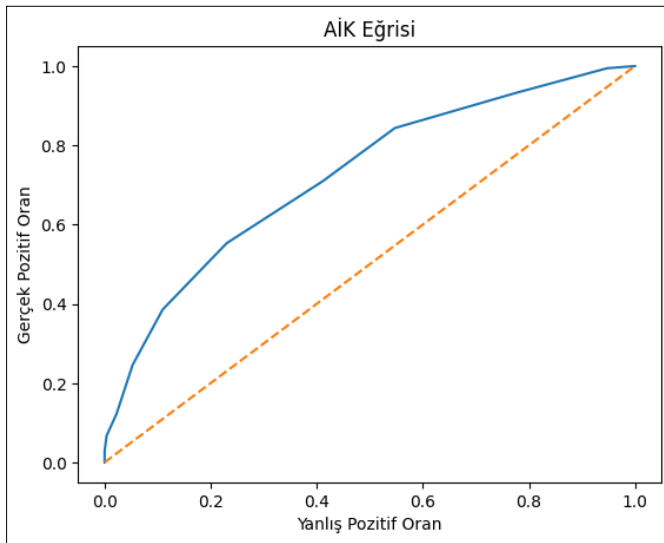
Duyarlılık, doğru tahmin edilen pozitif değerlerin, gerçekten pozitif olan tüm değerlere oranıdır. Hesaplanması için gerekli formülasyon (3.3)'deki gibidir

$$Duyarlılık = \frac{TP}{TP+FN} \quad (3.3)$$

F1 skoru kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır (3.4). Uç durumlar yok sayılmasın diye basit ortalama yerine harmonik ortalama ile hesaplanır.

$$F1 = 2 \times \frac{Kesinlik \times Duyarlilik}{Kesinlik + Duyarlilik} \quad (3.4)$$

AİK eğrisi aslında bir olasılık eğrisidir. AİK eğrisinin altında kalan alan ise pozitif ve negatif değerlerin birbirinden ayrılabilirliğinin derecesini vermektedir. Şekil 3.5'te görüldüğü gibi eğrinin yatay ekseninde “yanlış pozitif oran”, dikey ekseninde ise “gerçek pozitif oran” bulunmaktadır.



Şekil 3.5. AİK eğrisi örneği

AİK eğrinin altında kalan alan arttıkça 0 ve 1 sınıfları arasında ayırt etme performansı artmaktadır.

K-en yakın komşu algoritması, karar ağacı algoritması, Naive Bayes algoritması, destek vektör makinesi algoritması, rastgele orman algoritması, ikili lojistik regresyon, stokastik gradyan inişi sınıflandırması en bilinen sınıflandırma türleridir.

“Gerçek pozitif oran” (3.5) aslında duyarlılık değeridir. Yani hasta olan kişileri tahmin etme oranıdır. “Yanlış pozitif oran” (3.6) ise hasta olmayan kişiler için yanlış tahmin yapma oranıdır.

$$\text{Gerçek pozitif oran} = \text{Duyarlılık} = \frac{TP}{TP+FN} \quad (3.5)$$

$$\text{Yanlış pozitif oran} = \frac{FP}{FP+TN} \quad (3.6)$$

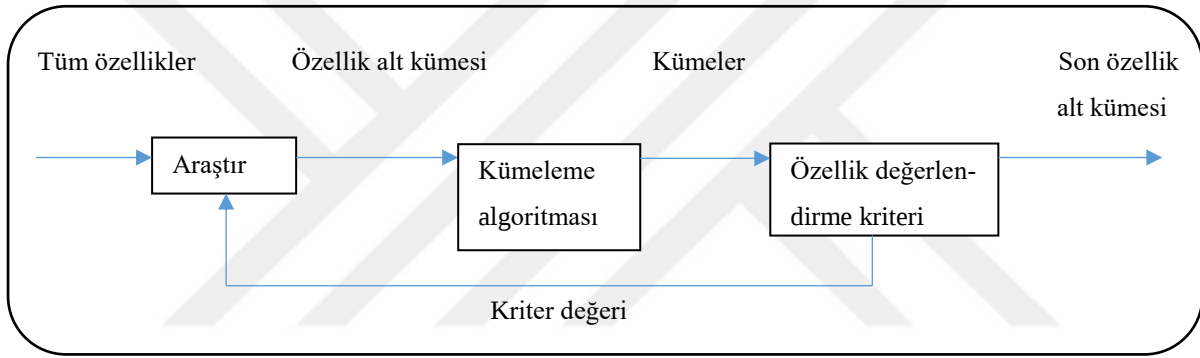
Burada, sınıflar arası ayırt ediciliği en iyi olan eşik değeri, en iyi F1 değerini vermektedir. En iyi F1 değeri de en iyi AİK eğrisini oluşturmaktadır. En iyi AİK eğrisinde, eğri altında kalan alan en fazladır. Bu alanın en fazla olduğu anda model sınıfları en başarılı şekilde ayırt etmiş demektir.

Sonuçta, AİK puanı bize modelin ne kadar verimli olduğunu söylemektedir. AİK eğrisi altında kalan alan, ne kadar fazla olursa modelin pozitif ve negatif sınıfları ayırt etme performansı o kadar iyi olur. Buna göre;

-AİK = 0,5	Ayrım yok.
-0,5 < AİK < 0,7	Zayıf ayırım.
-0,7 ≤ AİK < 0,8	Kabul edilebilir ayırım.
-0,8 ≤ AİK < 0,9	Mükemmel ayırım.
-AİK ≥ 0,9	Olağanüstü ayırım (Hosmer ve Lemeshow, 2013: 52-57).

Denetimsiz makine öğrenmesi, karşılık gelen etiket olmaksızın sadece girdiden oluşan veri setlerinde kullanılan makine öğrenmesidir. Model, yapı veya gruplamaları araştırmak için girdi verileri ile ilgilenir. Eğitim sonucunda, kümeleme ve boyut küçültme amacıyla bir model elde edilmektedir.

Kümeleme yönteminde, girdilerin çeşitli görüntüleri tarafından oluşturulan fotoreseptör faaliyetleri olduğunu düşündüğümüzde, olası tüm faaliyetler alanında, bu belirli girdiler, daha düşük boyuta sahip iki küme oluşturmaktadır. Denetimsiz öğrenmenin amacı bu ayrı, düşük boyutlu kümeleri bulmak ve karakterize etmektir. Denetimsiz öğrenmenin sarmalayıcı yaklaşım adımları Şekil 3.6'da gösterilmiştir.



Şekil 3.6. Denetimsiz öğrenme için sarmalayıcı yaklaşım (Dy ve Brodley, 2004)

K-ortalama kümeleme algoritması ve hiyerarşik kümeleme algoritması en sık kullanılan kümeleme algoritmalarıdır.

Pekiştirmeli makine öğrenmesi, çıktı örnekleri mevcut olmadığında ancak bazı performans kriterlerine göre çıktı örneklerini puanlamanın mümkün olduğu durumlarda pekiştirmeli öğrenme kullanılmaktadır. Bu öğrenme biçimi, hatalardan öğrenmektedir. Yani uzun süreli hafızası bu türün öğrenme şeklidir. Örneğin, kapsama alanının zayıf olduğu yeni bir bölgede, amaç cep telefonu için iyi bir sinyal alımı elde etmektir. Sinyal izlenirken telefonla birlikte hareket edilmektedir. Yeterli sinyale ulaşmak için, sinyalin en iyi olduğu yer bulunana kadar hareket edilmeye devam edilmektedir. Burada alınan bilgiler, doğrudan iyi bir karşılama elde etmek için nereye gitmemiz gerektiğini söylememektedir. Her gözlem, bundan sonra hangi

yöne gidilmesi gerektiğini de söylememektedir. Her gözlem, mevcut durumun iyiliğini değerlendirmeyi sağlamaktadır. Nereye gidilmesi gerektiğine karar verilmesi için etrafta dolaşılmalıdır. Doğru davranış örnekleri verilmemektedir.

Dünyadaki bir konum olan x ve x konumunda elde edilebilecek ödül (örneğin, telefon sinyal gücü) $R(x)$ için pekiştirmeli öğrenmenin amacı, maksimum $R(x^*)$ ödülünü veren x^* konumunu belirlemektir. Pekiştirmeli öğrenme sistemine R değeri ve herhangi bir eğitim örneği verilmemektedir. Bunun yerine harekete geçme (yani x değerlerini seçme) ve ortaya çıkan $R(x)$ ödülü gözlemlenmektedir. $R(x)$ değeri deterministik veya stokastik olabilir (Barto ve Dietterich, 2004).

Pekiştirmeli öğrenmede, denetimli öğrenmedeki gibi, x veri noktalarının çizildiği sabit bir $D(x)$ dağılımı yoktur. Bunun yerine araştırmacı x 'in değerlerini kendisi seçmektedir. Ayrıca, pekiştirmeli öğrenmede amaç, denetimli öğrenmedeki gibi x veri noktaları için çıktı (y) değerlerini tahmin etmek değildir. Bunun yerine amaç, maksimum ödülü veren tek bir x^* değerini bulmaktır. Bu nedenle, x değerlerinin tüm alanı boyunca ($D(x)$ 'e göre ağırlıklandırılmış) beklenen kaybı en aza indirmek yerine amaç, ödülü tek bir x^* noktasında maksimuma çıkarmaktır. Ödül $R(x)$ stokastik ise, amaç beklenen ödülü maksimuma çıkarmaktır, ancak beklenti, bir $D(x)$ olasılık dağılımına göre değil, tek x^* noktasındaki R 'deki rastgeleliğe göre alınmaktadır (Barto ve Dietterich, 2004).

Pekiştirmeli öğrenmenin bu basit biçimini kayıp fonksiyonunun en aza indirilmesi açısından formülü $L(x) = R(x^*) - R(x)$ şeklindedir. $L(x)$ değeri, x^* 'de alanabilecek ödül ile gerçekte alınan ödül arasındaki farktır (Barto ve Dietterich, 2004).

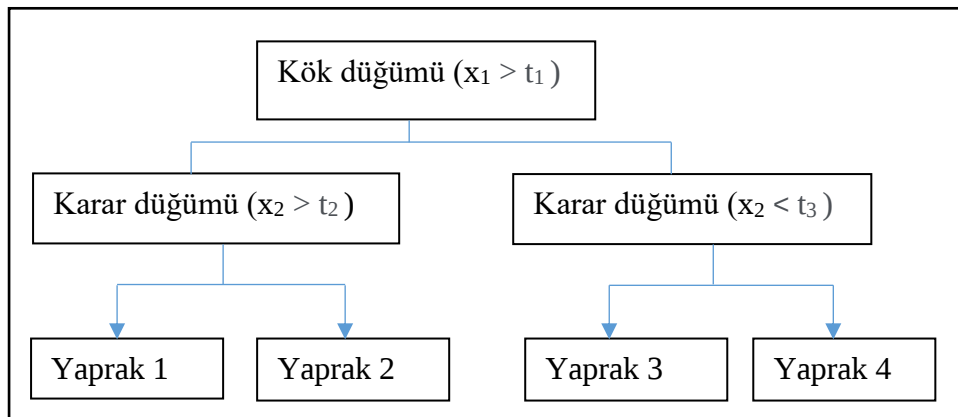
Çalışmada yalnızca denetimli makine öğreniminin sınıflandırma türünde algoritmalar kullanılmıştır. Denetimli makine öğrenimi veri etiketleme ile girdileri, kendisine karşılık gelen önceden tanımlanmış çıktı değerleriyle eşleştirmektedir. Bu yöntemle karmaşıklığı azaltarak basit bir kullanım sağlamaktadır. Denetimli makine öğrenimi ile olası bir sınırlı sonuç kümesi tahmin edilebilir, veriler tanımlanmış kategorilere ayrılabilir veya başka iki makine öğrenimi algoritmasından gelen sonuçlar birleştirilebilir.

Veri setinin sınıflandırma yöntemi için bu çalışmada kullanılan algoritmalar ise ikili lojistik regresyon, k-en yakın komşu algoritması, destek vektör sınıflandırması, karar ağacı sınıflandırması, rastgele orman algoritması, Naive Bayes (Gauss) sınıflandırması, Naive Bayes (Bernoulli) sınıflandırması, stokastik gradyan inişi sınıflandırmasıdır. Bu analiz yönteminde kullanılan nitelikler ise doğruluk, kesinlik, duyarlılık, AİK skoru, ortalama ağırlıklı F1 puanıdır. Ayrıca, modellerin sonuçları “AİK eğrisi” grafikleri ve “değişkenlerin önem katsayısı” tabloları ile görünür hale getirilmiştir.

Sekiz farklı sınıflandırma algoritması kullanılarak, % 20, % 25 ve % 30'luk test büyüklüklerinde, 10 kat çapraz doğrulama yapılarak toplam 24 sınıflandırma modeli kurulmuştur. Rastgele test verisi seçiminin önüne geçmek yani her analizde aynı verileri seçmek için rastgele durum sıfır olarak sabitlenmiştir.

3.3.1. Karar ağacı algoritması

Karar ağaçları, sınıflandırma ve regresyon için kullanılan parametrik olmayan denetimli bir öğrenme yöntemidir. Aynı zamanda hem ikili hem de çok sınıflı olduğu durumlarda sınıflandırma yeteneğine sahiptir. Karar ağacı, akış şemasına benzer Şekil 3.7.'teki gibi bir ağaç yapısıdır. Karar ağacı, örnekleri ağacın kökünden yapraklarına doğru sıralayarak sınıflandırmaktadır. Ağaçtaki her düğüm bazı değişkenleri test eder ve her dal bu değişkenin alabileceği değerlerden birine uygun düğümü gösterecek şekilde aşağıya doğru inmektedir. Karar ağacı algoritması, ağacın kökünde hangi değişken ile test edilmesi gerektiği sorusu ile başlayarak yukarıdan aşağıya doğru ağacı oluşturmaktadır (Bilekdemir, 2010).



Şekil 3.7. Karar ağacı yapı örneği

Karar ağacı, sınıflandırma yaparken eğitim veri setini kullanarak, sınıf etiketleri bilinmeyen yeni verileri sınıflandırmak için kullanılabilecek şekilde sınıf etiketinin bir modelini oluşturmaktadır. Modelden yola çıkarak önce ağaç oluşturarak sınıflandırma yapar, her karar düğümünde bir karar verilmektedir. Özelliğin bağımlı değişken kümesi için değer aralıkları ($x_2 > t_2$) karar düğümü noktasında kullanılmaktadır (Ali ve diğerleri, 2012). Verilen karar modeli iyileştirmediginde o değere sahip dal budanmaktadır.

Uygulanmasının kolaylığı ve daha kolay anlaşılır olması nedeniyle diğer sınıflandırma algoritmalarıyla karşılaştırıldığında en sık kullanılan sınıflandırma algoritmasıdır.

3.3.2. Rastgele orman algoritması

Rastgele orman algoritması, hem sınıflandırma ve hem de regresyon için kullanılan parametrik olmayan denetimli bir öğrenme yöntemidir.

Rastgele Orman yönteminde, veri setinden yer değiştirmeli şekilde yeni bir eğitim veri seti oluşturulmaktadır. Rastgele özellik seçimi kullanılarak yeni eğitim setinden birçok sınıflandırılmış ağaç oluşturulmaktadır. Oluşturulan bu ağaçlar budanmaz. Girdi verisinden yeni bir objeyi sınıflandırmak için girdi verisini ormandaki her bir ağaca yerleştirmektedir. Her bir ağaç bir sınıflandırma vermektedir. Ormandaki en yüksek oya sahip olan sınıflandırmadaki ağaç seçilmektedir (Breiman ve Cutler, 2009).

Rastgele Orman, karar ağaçlarından oluşan bir topluluk oluşturmaktadır. Temel karar ağaçları arasında çeşitliliği sağlamak için Breiman, torbalama veya rastgele altuzay yöntemleriyle iyi çalışan rastgeleleştirme yaklaşımını seçmektedir. Rastgele ormanda her bir ağacı oluşturmak için Breiman önce eğitim setindeki kayıtlar N ise, N sayıda kayıt rastgele örneklenmektedir. Bu örnek ağacı büyütmek için gerekli eğitim setini oluşturmaktadır. M giriş değişkeni varsa, her düğümde M değişkeni arasından rastgele m değişken seçilecek ve düğümü bölmek için bu m özniteliklerindeki en iyi bölünme kullanılacak şekilde bir $m \ll M$ sayısı seçilmektedir. Ormanın büyümesi sırasında m değeri sabit tutulmaktadır. Her ağaç mümkün olan en geniş ölçüde büyütülmektedir. (Breiman ve Cutler, 2009).

Orman oluşturulduktan sonra, yeni bir örneği sınıflandırmak için ormanda yetişen tüm ağaçlar arasında çalıştırılmaktadır. Her ağaç, oy olarak kaydedilen yeni örnek için sınıflandırma

sağlamaktadır. Tüm ağaçlardan alınan oylar birleştirilmekte ve maksimum oyların sayıldığı sınıf yeni örneğin sınıflandırması olarak ilan edilmektedir. (Kulkarni ve Sinha, 2013)

3.3.3. K-en yakın komşu algoritması

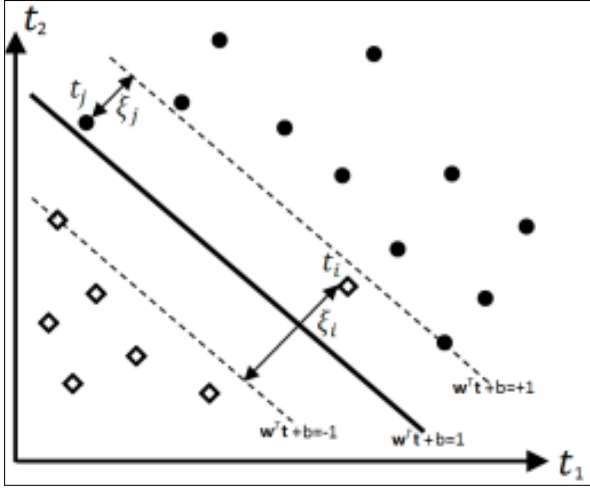
K-En Yakın Komşu Algoritması hem sınıflandırma hem de regresyon için kullanılan denetimli makine öğrenmesi algoritmasıdır. Amacı, yeni bir örnek geldiğinde varolan öğrenme verisi üzerinde sınıflandırma yapmaktır. Algoritma, yeni bir örnek geldiğinde, onun en yakın K komşusuna bakarak örneğin sınıfına karar vererek etiketi tahmin etmektedir (Bilekdemir, 2010). Kısaca, sınıfı bilinmeyen veri için eğitim setindeki verilerle kıyaslama yapılarak bir uzaklık ölçümü yapılmaktadır. Bulanık uzaklığa göre sınıfı bilinmeyen veri en uygun sınıfa atanmaktadır.

İki nokta arasında, $x_1 = (x_{11}, x_{12} \dots x_{1n})$ ve $x_2 = (x_{21}, x_{22} \dots x_{2n})$ olan iki nokta için uzaklık ölçümü çoğunlukla aşağıdaki formüle (3.7) göre yapılır (Hu ve diğerleri, 2016):

$$|x_1 - x_2| = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (3.7)$$

3.3.4. Destek vektör regresyonu/sınıflandırması

Destek vektör regresyonu/sınıflandırması, sınıflandırma, regresyon ve aykırı değerlerin tespiti için kullanılan bir dizi denetimli öğrenme yöntemidir. Bu algoritmanın sınıflandırma yöntemi, iki kategoriden oluşan bağımlı değişken verilerinin, birbirlerini ayırması için paralel en uzak çizgiyi çizmektedir. Yani sınıf etiketleri genellikle pozitif ve negatif olarak adlandırılan veri seti içerisinde en uygun hiper düzlemi bulmaktadır (Qi ve diğerleri, 2013). Destek vektör makine algoritması ile Şekil 3.8'deki gibi ortadan geçen paralel bir çizgi elde ederiz ve bu çizilen paralel aralığın kestiği noktalar destek noktalarıdır.



Şekil 3.8. Destek vektör makinesi sınıflandırması örneği (Karagül, 2014)

Destek vektör makinesinin primal model formülasyonu (İplikçi, 2013);

t_i = değişkenleri girdi vektörleri,

y_i = değişken çıktıları,

α = Lagrange parametreleri

C = ceza parametresi olmak üzere aşağıdaki gibidir (3.8):

$$\min w, b \quad \frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i$$

$$y_i (w^T t_i + b) \geq 1 - \xi_i, i=1, \dots, N \quad (3.8)$$

$$\xi_i \geq 0, i = 1, \dots, N$$

3.3.5. İkili lojistik regresyon

Lojistik regresyon, regresyon analizlerinden daha çok sınıflandırma için kullanılan denetimli öğrenme yöntemidir. İki seçenekli sonuçtan oluşan bağımlı değişkenler bağımsız değişkenler etkisinde atamalar yaparak sınıflandırmaktadır. Bağımsız değişken ile bağımlı değişken arasındaki denklem olarak logit fonksiyonu kullanılmaktadır. Logit fonksiyonu, x 'den değerlerinden bağımsız olarak y için yalnızca 0 ile 1 arasındaki değerleri döndürmektedir (Bircan, 2004).

Lojistik regresyonda, bağımsız değişkenlerin kategorik olarak düzenlenmiş bağımlı değişkeni ve aralarındaki ilişki gücünü açıklamak amacıyla model kurulmaktadır. Bu analiz sonucunda, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin gerçekleşme olasılığı ortaya konulmaktadır (Umar ve Bashir, 2019, s. 403). Ve bu olasılık İkili lojistik regresyon analizlerinde 0 veya 1 değerini almaktadır.

Lojistik regresyon modelleri, lojistik dağılım gösteren ve Y bağımlı değişken ve X bağımsız değişken olduğu bir fonksiyonla (3.9) ifade edilmektedir. Burada Y, 0 ile 1 arasında değerler almaktadır.

$$P_i = E(Y = 1|X_i) = 1 / [1 + e^{-(\beta_1 + \beta_2 X_i)}] \quad (3.9)$$

Burada, $\beta_1 + \beta_2$ ifadesi X_i yerine Z_i yazılarak aşağıdaki fonksiyon (3.10) elde edilmektedir;

$$P_i = 1 / [1 + e^{-Z_i}] \quad (3.10)$$

Fonksiyonda (8)'deki Z_i değeri $-\infty$ ile $+\infty$ arasında yer almaktadır. Fonksiyon bu durumda, P_i ile Z_i arasında doğrusal bir ilişki olmadığından fonksiyonu doğrusal hale getirmek için oranlama işlemi yapılmaktadır (Gujarati, 2010).

İkili lojistik regresyon analizlerinde, bağımlı değişken sadece 0 veya 1 değerini almaktadır. 0 veya 1 değeri, mevcut olayın gerçekleşip gerçekleşmeme olasılıklarını açıklamaktadır. Buradan bahis oranı tanımlanmaktadır. Bahis oranı, olayın gerçekleşme olasılığının gerçekleşmeme olasılığına olan oranıdır (3.11) (Agresti, 1996)

$$P_i / (1 - P_i) = (1 + e^{Z_i}) / (1 + e^{-Z_i}) = e^{Z_i} \quad (3.11)$$

Bahis oranının logaritması alındığında doğrusal hale getirilmiş logit model fonksiyonu (3.12) elde edilmektedir:

$$L_i = \ln [P_i / (1 - P_i)] = Z_i \quad (3.12)$$

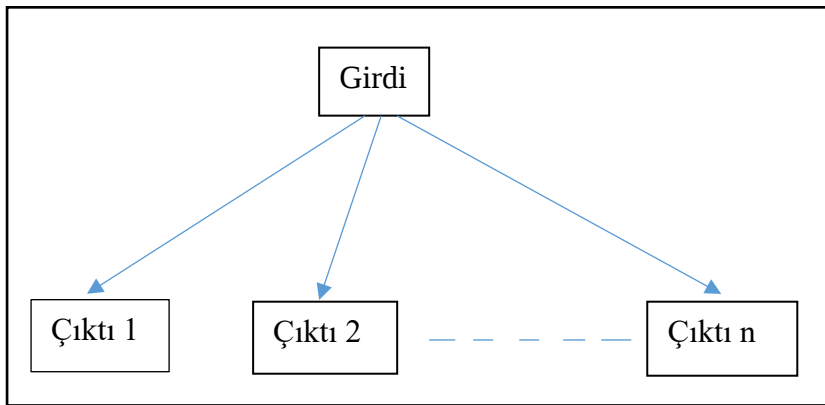
Fonksiyonda $Z_i = \beta_1 + \beta_2 X_i$ yerine yazılırsa; $L_i = \beta_1 + \beta_2 X_i$ fonksiyonu elde edilir. Fonksiyondaki β_1 değeri sabit terimdir. β_2 ise eğim katsayısıdır ve X 'deki 1 birimlik değişme karşısında L 'deki değişimi açıklamaktadır (Gujarati, 2010).

Lojistik regresyon analizinde bağımlı değişken 0 ile 1 olarak iki kategoriden oluştuğundan doğrusallık, süreklilik, eşvaryanslık koşullarını ve normallik varsayımını karşılama zorunluluğu yoktur. Bu nedenle bu yöntem en çok tercih edilen sınıflandırma algoritmalarından biridir.

3.3.6. Naive Bayes sınıflandırması (Gauss)

Naive Bayes yöntemleri, Bayes teoremini, sınıf değişkeninin değeri verilen her bir özellik çifti arasında "Naive" koşullu bağımsızlık varsayımıyla uygulamaya dayalı bir dizi denetimli öğrenme sınıflandırma algoritmasıdır. Naive Bayes-Gauss, Naive Bayes-Bernoulli ve Naive Bayes-Multinomial olmak üzere üç çeşitten oluşan bu sınıflandırma için kullanılan Naive Bayes-Gauss algoritmasında özelliklerin olabirliğin Gauss olduğu varsayılmaktadır (Cover, 1965).

Naive Bayes (NB) Sınıflandırıcısı, Bayes teorisini temel alarak olasılık tabanlı sınıflandırma yapar. Bayes ağı bağımlı değişkenler ile bağımsız değişkenler arasındaki olası ilişkileri, koşullu olasılıklar kullanarak modellemektedir. Şekil 3.9'deki gibi "girdi"ler düğüm; ilişkiler ise kesikli çizgilerle ifade edilmektedir. Eğer düğümler kesikli çizgilerle bağlı değil ise değişkenler koşullu bağımsız varsayılmaktadır.



Şekil 3.9. Naive Bayes yapısı

Bağımsız ve rasgele iki olayda, birinci olayın gerçekleştiği durumda ikinci olayın gerçekleşme olasılığı olan $P(A \cap B)$ 'nin formülü (3.13) aşağıdaki gibidir.

$$P(A \cap B) = P(A/B) P(B) = P(B/A) P(A) \quad (3.13)$$

Naive bayes ise;

$P(A)$ =bağımsız değişken

$P(B)$ =bağımlı değişken

$P(A|B)$ =bağımsız değişken gerçekleştiği durumda bağımlı değişkenin gerçekleşme olasılığı

$P(B|A)$ =bağımlı değişken gerçekleştiği durumda bağımsız değişkenin gerçekleşme olasılığı
iken A ve B olayları için koşullu olasılığı aşağıdaki gibi formüle (3.14) eder:

$$P(B|A) = \frac{P(B|A) P(A)}{P(B)} \quad (3.14)$$

Burada bağımsız değişken (X) birden fazla olduğunda ($X = \{ X_1, X_2, \dots, X_n \}$) ortak koşullu olasılığı aşağıdaki gibi formüle (3.15) edilmektedir. Bu formülde çıkan sonuçlardan en büyük sonucu veren seçilerek sınıflandırma yapılmaktadır.

$$P(X_i|Y_j) = \prod_{i=1}^n P(X_i|Y_j) \quad (3.15)$$

Sonuçta Naive Bayes sınıflandırması (3.16) formülü ile yapılır (Luis ve diğerleri, 2008).

$$Y' = \operatorname{argmax} (P(Y=y_j)) \prod_{i=1}^n P(X = x_i|Y = y_j) \quad (3.16)$$

3.3.7. Naive Bayes sınıflandırması (Bernoulli)

Çok değişkenli Bernoulli dağılımlarına göre dağıtılan veriler için Naive Bayes-Bernoulli sınıflandırması uygulanmaktadır. İkili sınıflandırmalar için, Bernoulli olasılığını kriter alarak, 0-1 değer ataması yapar. Naive Bayes-Bernoulli sınıflandırmasında sifıra çok yaklaşarak; multinominal naive bayes sınıflandırmasında ise sifıra çok yaklaşımadan cezalandırma yapılmaktadır.

3.3.8. Stokastik gradyan inişi sınıflandırması

Eğitim veri kümesindeki her örnek için hatayı hesaplar, ancak model yalnızca tüm eğitim örnekleri değerlendirildikten sonra parametreleri birer birer güncellemektedir. Sık güncelleme yaptığı için iyileştirme oranı yüksektir (Yavuz, 2021). Varsayılan öğrenme oranı çizelgesini kullanarak en iyi sonuçları elde etmek için, verilerin sıfır ortalamaya ve birim varyansa sahip olması gerekmektedir.

Algoritmanın kullandığı formül (3.17) ise aşağıdaki gibidir (Mei ve diğerleri 2018):

$$Q(w) = (1 \div n) \sum_{i=1}^n Q_i(w) \quad (3.17)$$

3.4. Makine Öğrenmesinde Çapraz Doğrulama

Denetimli makine öğrenmesinde algoritmaların model performanslarını artırmak için çeşitli yöntemler uygulanmaktadır. Bu yöntemlerden biri olan çapraz doğrulama, verilerin birden çok kez rastgele bölünüp ve her defasında değişik bir veri kümesinin test verisi olarak kullanılmasıdır.

Çapraz doğrulama sayesinde, veri setinin tamamı test verisi olarak kullanılmaktadır. Böylece, modelin tahmin veya sınıflandırma başarısı artmaktadır.

Örneğin, 5 katlı çapraz doğrulama yöntemi kullanıldığında eğitim seti daha fazla sayıda alt kümeye bölünmekte ve modelin performans değerlendirmesi bu alt kümelere göre yapılmaktadır.

Çapraz doğrulama, veri setini önce k parçaya ayırmakta, sonra k tane parçayı birer kez test verisi olarak kullanmakta ve k-1 tane parçayla da modeli eğitmektedir. Şekil 3.10'da gösterildiği gibi bu işlem k kez tekrarlandığında tüm veri seti, test verisi olarak kullanılmış olur.

1		n-1	n
Eğitim seti			Test seti
1		n-1	n
Eğitim seti		Test seti	Eğitim seti
.			
.			
.			
1		n-1	n
Test seti	Eğitim seti		

Şekil 3.10. Çapraz doğrulama örneği (Temel ve diğerleri, 2012)



4. GÜRÜLTÜYE BAĞLI İŞİTME KAYBI TAHMİNİ UYGULAMASI

Bu bölüm ağaç ürünleri imalatı sektöründe çalışanların cinsiyetinin, yaşının, toplam çalışma süresinin, sektörde çalışma süresinin, gürültülü ortamda çalışma süresinin, sigara kullanımının, gürültülü hobinin ve KKD kullanmamanın çalışanlarda GBİK'e etkisinin makine öğrenmesi algoritmalarıyla tahmini için yapılan tez çalışmasının uygulama bölümüdür.

Makine öğrenmesinde modelleme yaparken izlenen temel adımlar aşağıdaki gibidir:

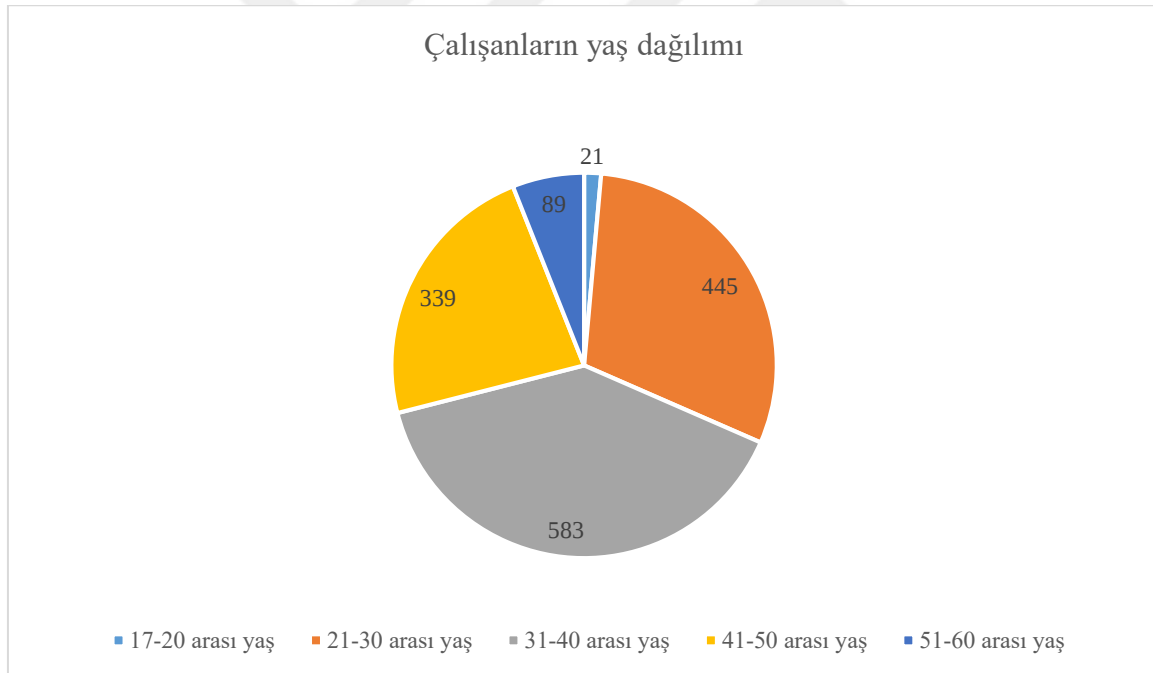
- Veri toplama: Bağımsız ve bağımlı değişkenlerin ve örneklem sayısının belirlenmesi.
- Veri ön-işleme: Veri temizleme ile regresyon tanımlayıcılarının (artık değer, uç değer vb.) gözden geçirilmesi. Uygun olmayan veya tekrarlı verilerin veri setinden çıkarılması. Veri dönüştürme ile kullanılan algoritma programının doğru çalışma şartlarını sağlayan dönüşümün yapılması (logaritmik, üstel dönüşüm vb.). Veriyi eğitim ve test kümelere ayırma ile deneyim için kullanılacak veri ile modelin kurulacağı verinin dağıtılması. Dağıtım yapılırken her defasında aynı veri kümelerinin kullanılması için sabitleme değeri seçilmesi.
- Eğitim verisi üzerinde modelin kurulması: Sabitlenen eğitim verisi ile algoritma parametrelerinin seçilerek modelin kurallarının belirlenmesi.
- Test verisi üzerinde modelin değerlendirilmesi: Sabitlenen test verisi ile modelin tahmin oranlarının değerlendirilmesi.
- Değerlendirme sonucuna göre uygulamaya geçilmesi: En ideal tahmin oranlarını veren modelin seçilmesi ve ileriye dönük tahmin sürecinde kullanılması.

4.1. Verilerin Dağılımı

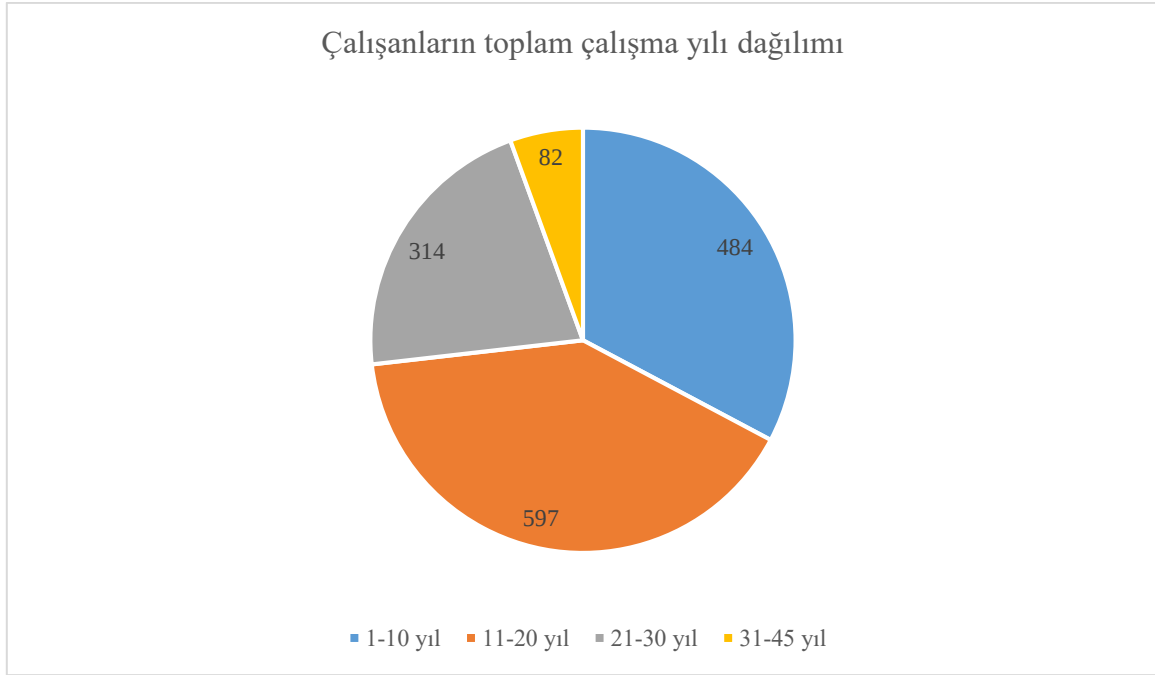
Çalışmanın veri setindeki 1 477 çalışanın, cinsiyet dağılımı Şekil 4.1'de, yaş dağılımı Şekil 4.2'de, toplam çalışma süresi (yıl) dağılımı Şekil 4.3'te, sektörde çalışma süresi (yıl) dağılımı Şekil 4.4'te, gürültülü ortamda çalışma süresi (ay) dağılımı Şekil 4.5'de, sigara kullanım dağılımı Şekil 4.6'da, gürültülü hobi dağılımı Şekil 4.7'de ve kulak koruyucu donanım kullanımı dağılımı Şekil 4.8'de, GİBK dağılımı Şekil 4.9'da gösterilmiştir.



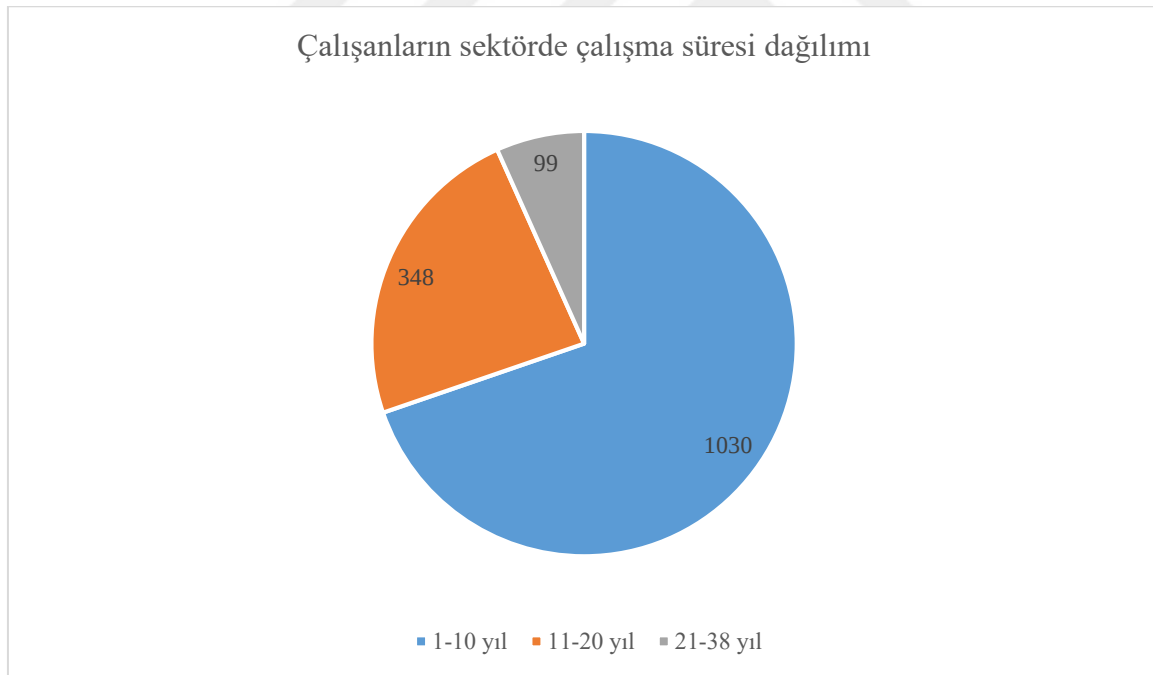
Şekil 4.1. Çalışanların cinsiyetlerine göre dağılımı



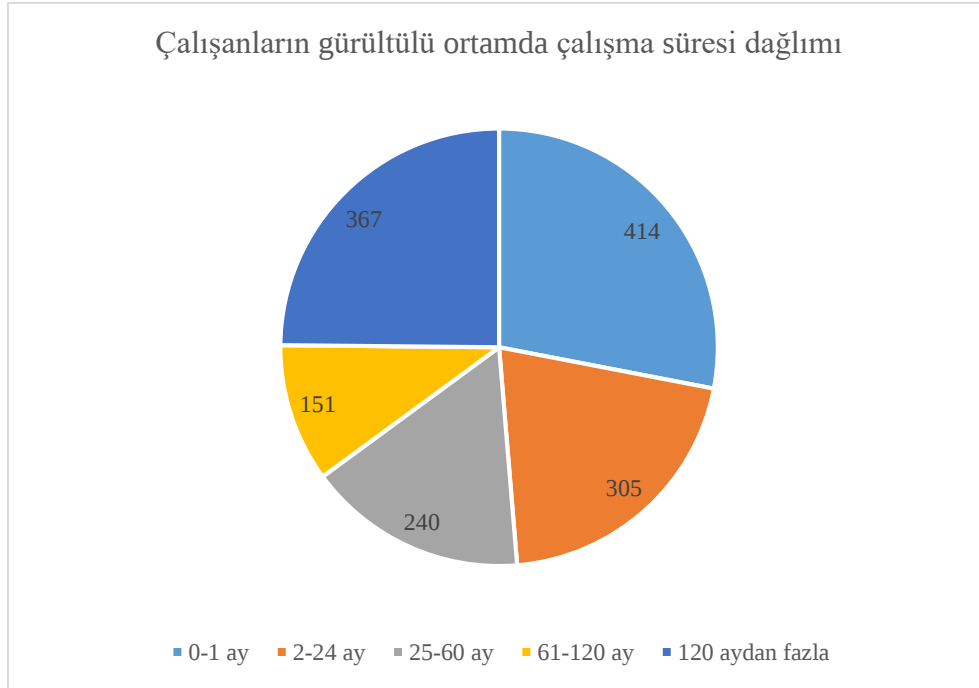
Şekil 4.2. Çalışanların yaş aralıklarına göre dağılımı



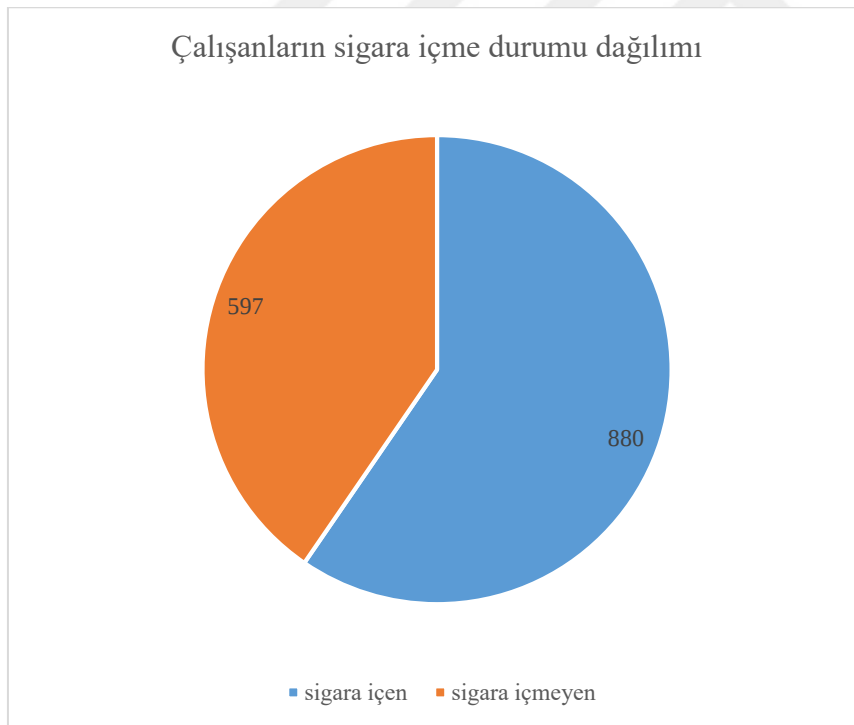
Şekil 4.3. Çalışanların toplam çalışma yılının çalışan sayısına göre dağılımı



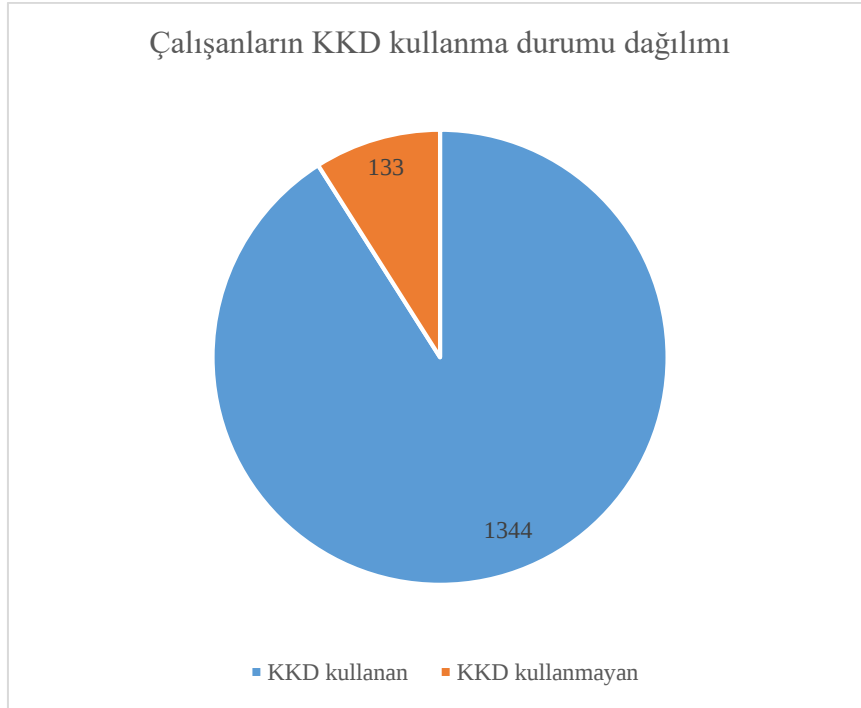
Şekil 4.4. Çalışanların sektörde çalışma yılının çalışan sayısına göre dağılımı



Şekil 4.5. Çalışanların gürültülü ortamda çalışma süresi dağılımı



Şekil 4.6. Çalışanların sigara içme durumu dağılımı



Şekil 4.7. Çalışanların KKD kullanma durumu dağılımı



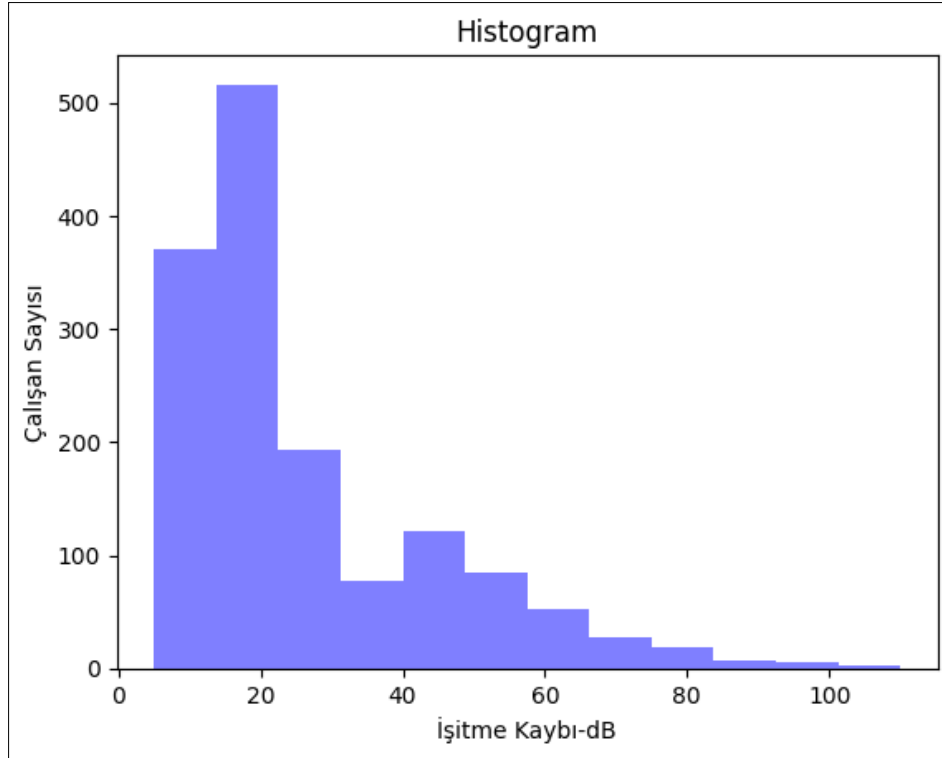
Şekil 4.8. Çalışanların gürültülü hobi sahibi olma durumu dağılımı



Şekil 4.9. Çalışanların GBİK olma durumu dağılımı

4.2. Gürültüye Bağlı İşitme Kaybının Regresyon Algoritmaları ile Tahmini

Çalışmanın ilk aşamasında 8 bağımsız değişkenin çalışanlarda kaç dB'lik işitme kaybına neden olduğunun araştırılırken veri kümesi ile 8 farklı algoritma kullanılarak 3 farklı test büyüklüğünde ve 10 defa çapraz doğrulama yapılarak regresyon modelleri oluşturmak hedeflenmiştir. Öncelikle veri setinin, regresyon analizinin varsayımlarını karşılayıp karşılamadığı araştırılmıştır. Bu varsayımlardan ilki olan normallik varsayımını araştırmak için histogram analizi gerçekleştirilmiştir. Şekil 4.10'daki histogramda veri setinin normal dağılmadığı, sağa çarpık olduğu görülmektedir.



Şekil 4.10. Veri setine ait histogram

Veri setinin, regresyon analizinin diğer bir varsayımı olan doğrusallık varsayımını karşılayıp karşılamadığını araştırmak için varyans analizi gerçekleştirilmiştir. Ancak varyans analizinde, Çizelge 4.1’de görüldüğü gibi “toplam çalışma yılı”, “sektörde çalışma yılı”, “KKD kullanımı” ve “gürültülü hobi” değişkenlerinin istatistiksel anlamlılık seviyesinin (p) 0,05’ten büyük çıkması sebebiyle bu değişkenlerle kurulacak model anlamlı çıkmamıştır. Bu yüzden, bu 4 değişken veri setinden çıkarılarak varyans analizi tekrar edilmiştir.

Çizelge 4.1. Bağımsız değişkenlerin istatistiksel anlamlılık seviyesi

Değişkenler	p değeri
Cinsiyet	0,000
Yaş	0,000
Toplam çalışma yılı	0,000
Sektörde çalışma yılı	0,060
Gürültülü ortamda çalışma süresi	0,151
Sigara kullanımı	0,000
KKD kullanımı	0,853
Gürültülü hobi	0,620

Tekrar edilen varyans analizinde kalan bağımsız değişkenlerin p değerlerinin 0,05'ten küçük çıkması sebebiyle kalan değişkenlerle kurulacak model anlamlı çıkmıştır. Ancak R^2 değerinin % 23,8 çıkması, dB cinsinden GBİK'teki değişimin sadece % 23,8'inin cinsiyet, yaş, toplam çalışma süresi (yıl), sektörde çalışma süresi (yıl), gürültülü ortamda çalışma süresi (ay), sigara kullanımı, gürültülü hobisi olup olmaması ve yetersiz kulak koruyucu donanım kullanımı ile açıklanabildiğini göstermektedir. Buradan korelasyon katsayısı (r) % 49 elde edilir. % 49'lük r değeri, bağımlı değişken ile bağımsız değişkenler arasında orta düzeyde ilişkiyi tanımlamaktadır. Bu nedenle GBİK'in, regresyon algoritmaları kullanılarak tahminlemesi yapılamamıştır.

4.3. Gürültüye Bağlı İşitme Kaybının Sınıflandırma Algoritmaları ile Tahmini

Çalışmanın ikinci aşamasında sekiz bağımsız değişkenin çalışanlarda işitme kaybına neden olup olmadığı araştırılırken veri kümesi ile sekiz farklı algoritma kullanılarak üç farklı test büyüklüğünde ve 10 defa çapraz doğrulama yapılarak ikili lojistik regresyon, k-en yakın komşu algoritması, destek vektör sınıflandırması, karar ağacı sınıflandırması, rastgele orman algoritması, Naive(Gauss) Bayes sınıflandırması, Naive Bayes (Bernoulli) sınıflandırması, stokastik gradyan inişi sınıflandırması algoritmaları ile sınıflandırma modelleri oluşturulmuştur.

Farklı algoritmalar ve farklı test büyüklüğündeki tüm sınıflandırma modellerinin, doğruluk, duyarlılık, kesinlik, f-1 puanı ve AİK puanları karşılaştırılarak en iyi sonucu veren model GBİK'i sınıflandırmak için seçilmiştir.

4.3.1. Test büyüklüğü % 20 için sınıflandırma modellerinin performansları

Veri setiyle % 20 test büyüklüğünde kurulan sınıflandırma modellerinin karşılaştırması Çizelge 4.2'de görüldüğü gibidir. Tüm test büyüklüklerinde, performans kriterleri karşılaştırıldığında en iyi sınıflandırma modelini % 25 test büyüklüğünde, rastgele orman algoritmasının verdiği görülmektedir.

Çizelge 4.2. Test büyüklüğü % 20 için sınıflandırma modellerinin sonuçları

Sınıflandırma algoritmaları	Doğruluk	Duyarlılık	Kesinlik	F1 puanı	AİK puanı
İkili Lojistik Regresyon	0.7331	0.56	0.71	0.63	0.7052
K-En Yakın Komşu Algoritması	0.7128	0.55	0.68	0.60	0.6855
Destek Vektör Makinesi	0.7027	0.39	0.75	0.52	0.6523
Karar Ağaçları Sınıflandırması	0.6655	0.47	0.61	0.53	0.6336
Rastgele Orman Algoritması	0.7399	0.54	0.74	0.62	0.7211
Naive Bayes Sınıflan.(Gauss)	0.6757	0.65	0.59	0.62	0.6710
Naive Bayes Sınıf.(Bernoulli)	0.6959	0.67	0.61	0.64	0.6921
Stokastik Gradyan İnişi Sınıf.	0.6757	0.50	0.62	0.56	0.6476

4.3.2. Test büyüklüğü % 25 için sınıflandırma modellerinin performansları

Veri setiyle % 25 test büyüklüğünde kurulan sınıflandırma modellerinin karşılaştırması Çizelge 4.3'te görüldüğü gibidir.

Çizelge 4.3. Test büyüklüğü % 25 için sınıflandırma modellerinin sonuçları

Sınıflandırma algoritmaları	Doğruluk	Duyarlılık	Kesinlik	F1 puanı	AİK puanı
İkili Lojistik Regresyon	0.7108	0.53	0.68	0.60	0.6826
K-En Yakın Komşu Algoritması	0.6811	0.48	0.64	0.55	0.6491
Destek Vektör Makinesi	0.6811	0.37	0.71	0.48	0.6311
Karar Ağaçları Sınıflandırması	0.6514	0.56	0.57	0.57	0.6368
Rastgele Orman Algoritması	0.7054	0.53	0.68	0.59	0.6770
Naive Bayes Sınıflan.(Gauss)	0.6757	0.63	0.59	0.61	0.6689
Naive Bayes Sınıf.(Bernoulli)	0.6838	0.63	0.61	0.62	0.6747
Stokastik Gradyan İnişi Sınıf.	0.6757	0.45	0.64	0.53	0.6392

4.3.3. Test büyüklüğü % 30 için sınıflandırma modellerinin performansları

Veri setiyle % 30 test büyüklüğünde kurulan sınıflandırma modellerinin karşılaştırması Çizelge 4.4'te görüldüğü gibidir.

Çizelge 4.4. Test büyüklüğü %30 için sınıflandırma modellerinin sonuçları

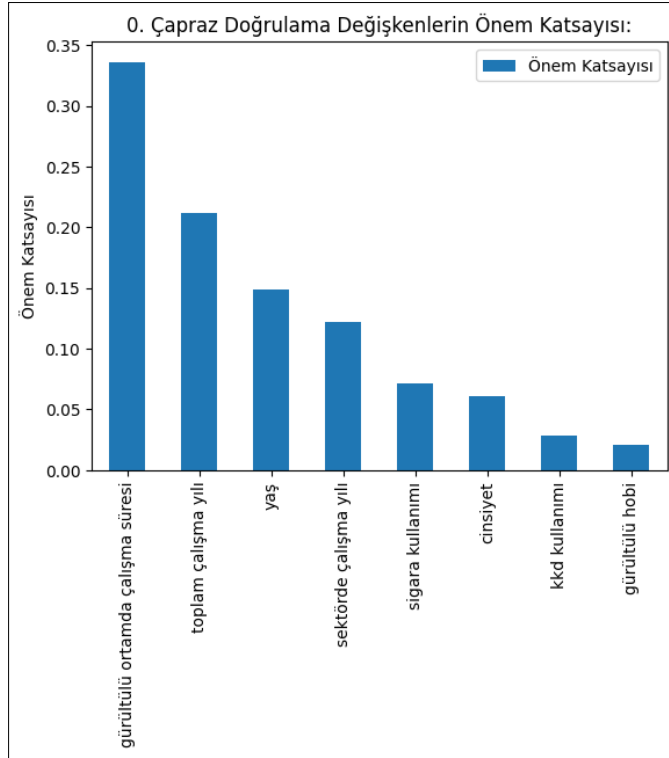
Sınıflandırma algoritmaları	Doğruluk	Duyarlılık	Kesinlik	F1 puanı	AİK puanı
İkili Lojistik Regresyon	0.6937	0.54	0.64	0.59	0.6682
K-En Yakın Komşu Algoritması	0.6757	0.41	0.66	0.50	0.6322
Destek Vektör Makinesi	0.6712	0.35	0.68	0.46	0.6194
Karar Ağaçları Sınıflandırması	0.6892	0.50	0.65	0.57	0.6590
Rastgele Orman Algoritması	0.6914	0.53	0.64	0.58	0.6654
Naive Bayes Sınıflan.(Gauss)	0.6712	0.61	0.59	0.60	0.6611
Naive Bayes Sınıf.(Bernoulli)	0.6667	0.66	0.58	0.61	0.6655
Stokastik Gradyan İnişi Sınıflan.	0.6577	0.55	0.58	0.57	0.6407

4.4. GBİK Sınıflandırmasında En İyi Modelin Seçimi

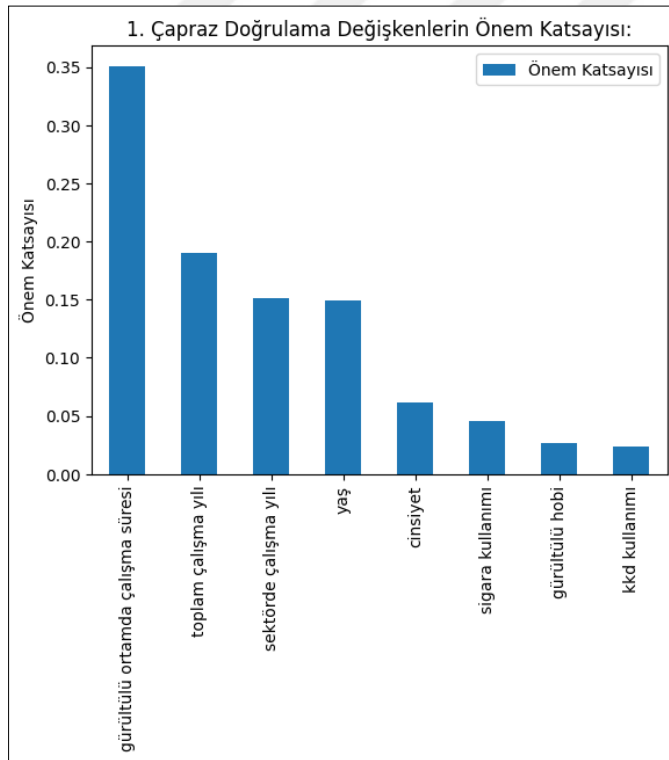
Sınıflandırma algoritmalarında en iyi model seçimi doğruluk, duyarlılık, kesinlik, f-1 puanı değerleri karşılaştırılarak yapılır. Bu performans kriterlerinde en yüksek değeri veren olan algoritma en iyi algoritmadır. Buna göre;

- Test büyüklüğü % 20 iken en iyi doğruluk oranı olan algoritma rastgele orman algoritmasıdır.
- Test büyüklüğü % 25 iken en iyi doğruluk oranı olan algoritma ikili lojistik regresyondur.
- Test büyüklüğü % 30 iken en iyi doğruluk oranı olan algoritma ikili lojistik regresyondur.
- En yüksek duyarlılık değeri ise test büyüklüğü % 20 iken % 67 ile Naive Bayes (Bernoulli) sınıflandırmasıdır.
- Bu üç algoritma ve üç farklı test size karşılaştırması içinde en yüksek doğruluk oranı (%73,99) ile % 20 test büyüklüğündeki rastgele orman algoritmasıdır.

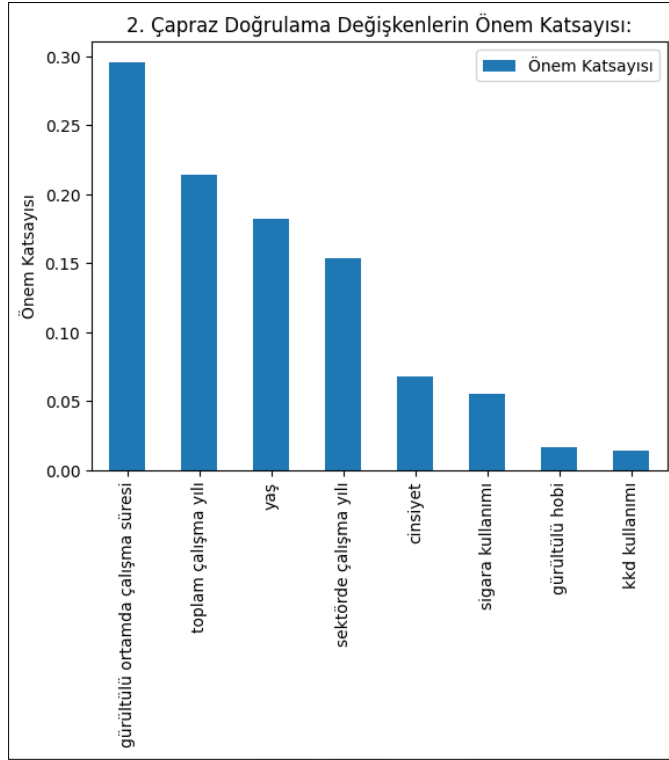
GBİK sınıflandırmasında test büyüklüğü % 20 olan rastgele orman algoritması en iyi model seçilmiştir. Bu modelde değişkenlerin 10 defa çapraz doğrulamasına ait önem katsayıları Şekil 4.11, Şekil 4.12, Şekil 4.13, Şekil 4.14, Şekil 4.15, Şekil 4.16, Şekil 4.17, Şekil 4.18, Şekil 4.19, Şekil 4.20'deki gibidir.



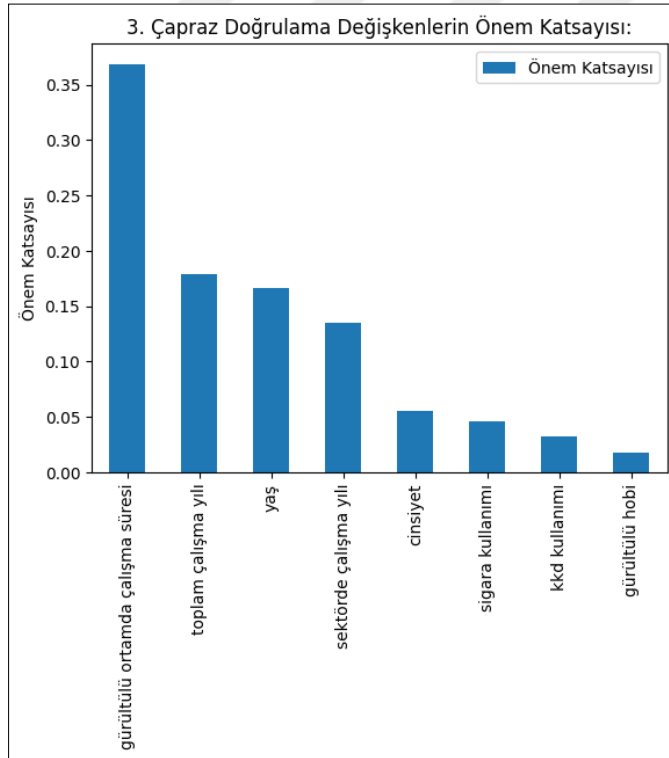
Şekil. 4.11. Sıfırıncı çapraz doğrulama değişkenlerin önem katsayısı



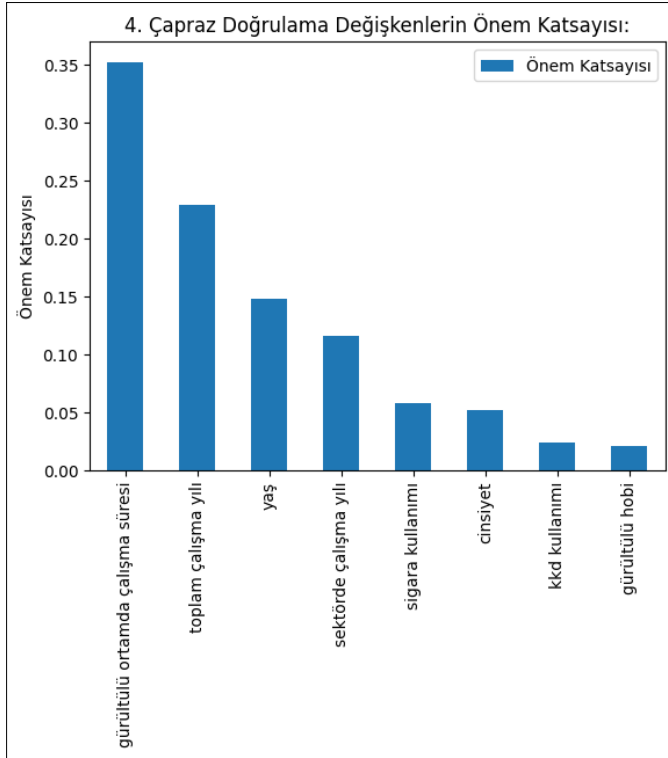
Şekil 4.12. Birinci çapraz doğrulama değişkenlerin önem katsayısı



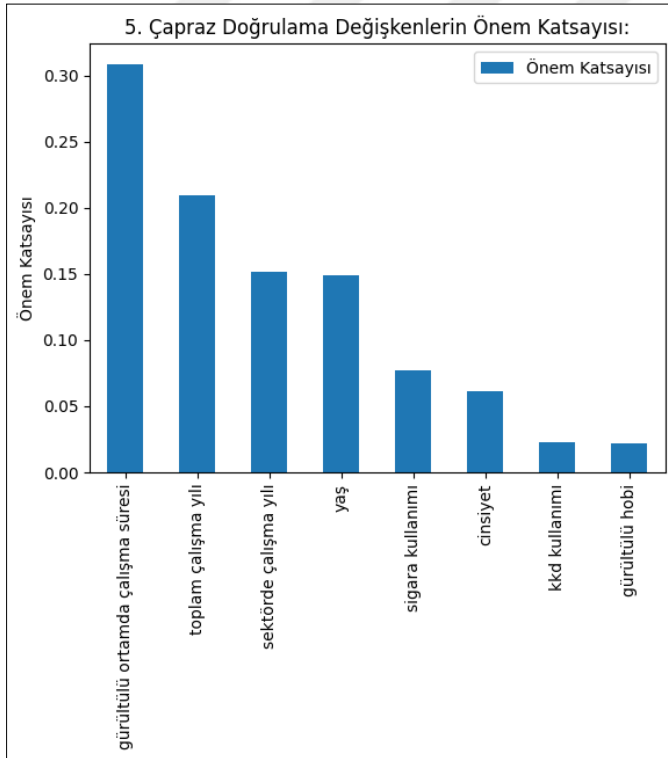
Şekil 4.13. İkinci çapraz doğrulama değişkenlerin önem katsayısı



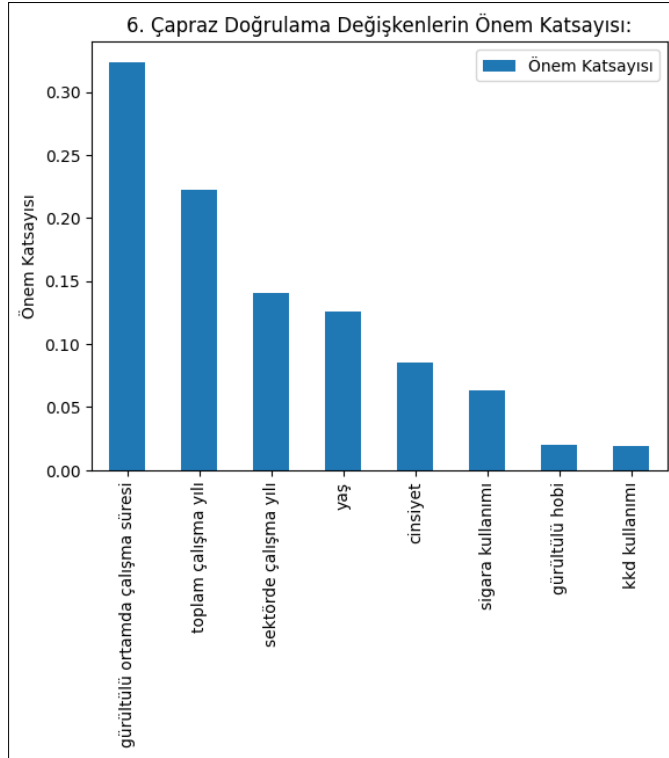
Şekil 4.14. Üçüncü çapraz doğrulama değişkenlerin önem katsayısı



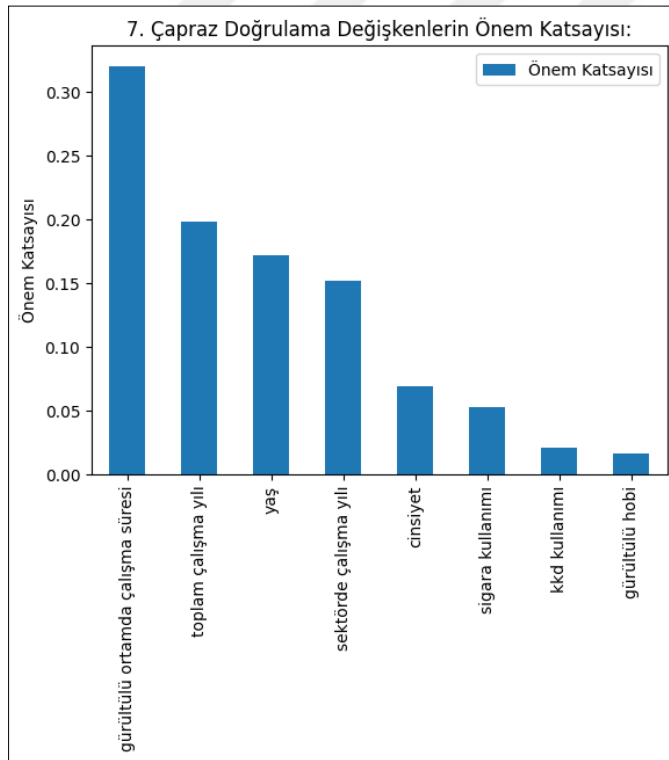
Şekil 4.15. Dördüncü çapraz doğrulama değişkenlerin önem katsayısı



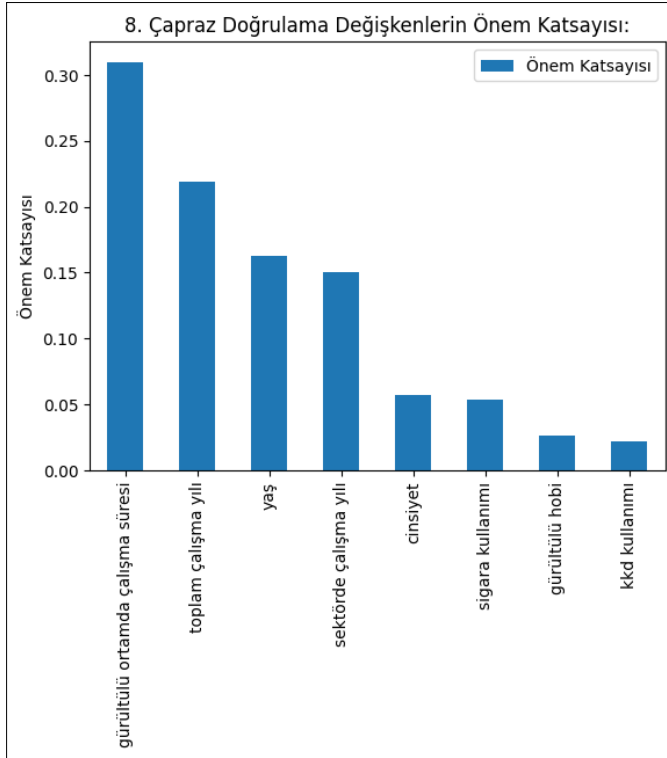
Şekil 4.16. Beşinci çapraz doğrulama değişkenlerin önem katsayısı



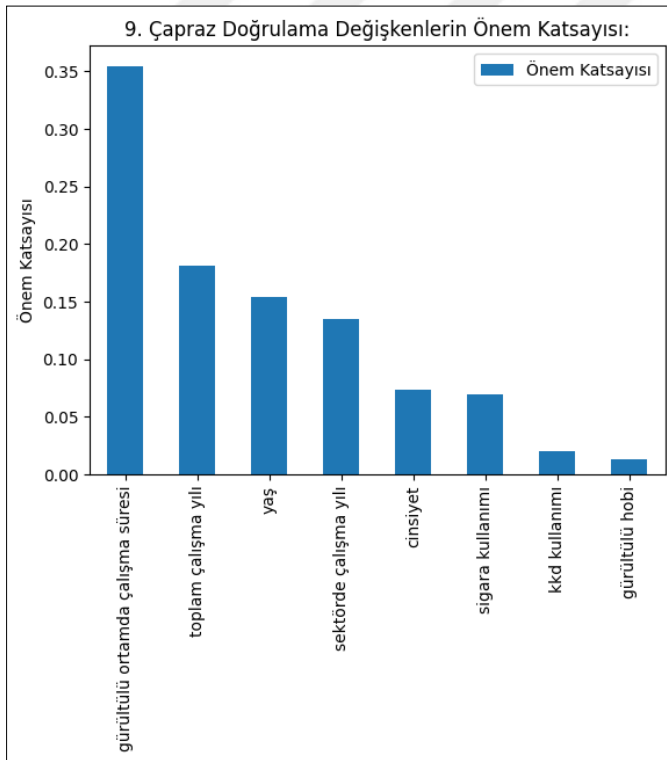
Şekil 4.17. Altıncı çapraz doğrulama değişkenlerin önem katsayısı



Şekil 4.18. Yedinci çapraz doğrulama değişkenlerin önem katsayısı



Şekil 4.19. Sekizinci çapraz doğrulama değişkenlerin önem katsayısı



Şekil 4.20. Dokuzuncu çapraz doğrulama değişkenlerin önem katsayısı

Rastgele orman algoritmasına göre GBİK'i etkileyen en önemli bağımsız değişken yaklaşık %34'lük oranla "gürültülü ortamda çalışma süresi"; GBİK'i etkileme derecesi en az olan bağımsız değişkenler de % 1'lik oranla "gürültülü hobi" ve "KKD kullanımı"dır.

Çizelge 4.5 incelendiğinde rastgele orman algoritmasına göre % 20'lik test verisi olan 296 kişiden;

- 219 kişiyi işitme kaybı olup olmadığı konusunda doğru tahmin etmiştir.
- Gerçekten işitme kaybı olmayan 177 kişiden 155 kişiyi doğru tahmin etmiştir.
- Gerçekten işitme kaybı olan 119 kişiden ise 64 kişiyi doğru tahmin etmiştir.

Çizelge 4.5. Hata matrisi

	Tahmin Edilen Negatif	Tahmin Edilen Pozitif
Gerçek Negatif	155	22
Gerçek Pozitif	55	64

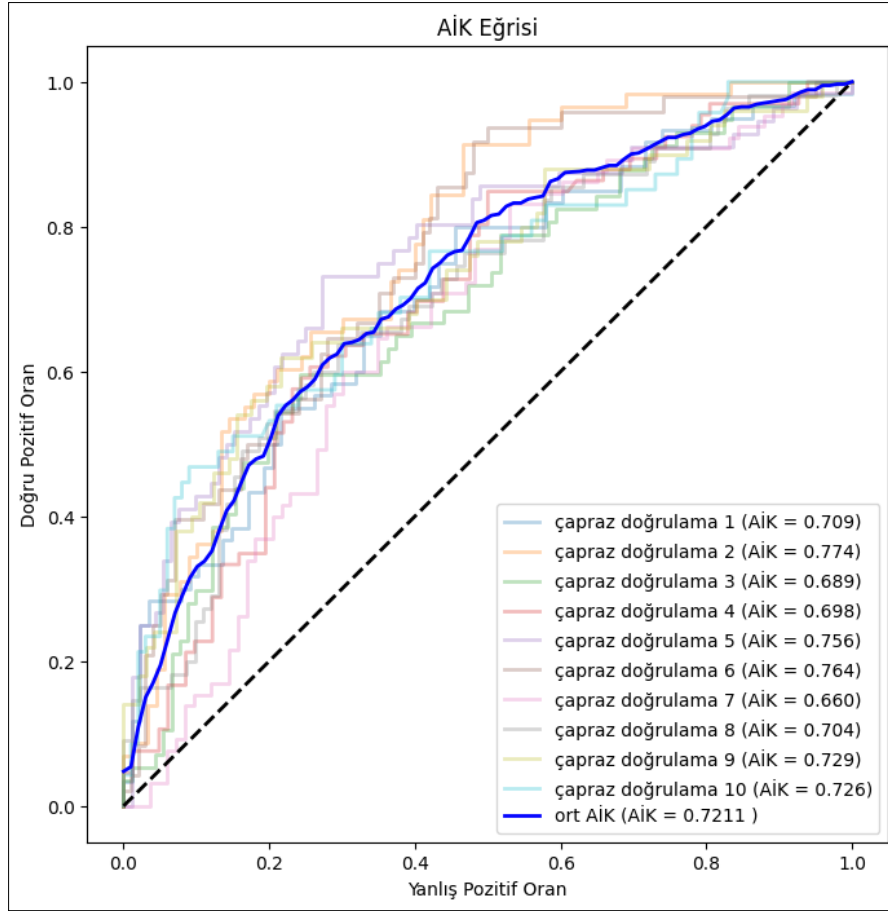
Doğruluk = $219/296=0,7399 \approx 0,74$ (tüm doğru tahmin edilenlerin tüm örnekleme oranı)

Kesinlik = $64/(64+22) = 0,7442 \approx 0,74$ (doğru pozitif tahmin edilenlerin, toplam pozitif tahmin edilenlere oranı)

Duyarlılık = $64/(64+55)=0,5378 \approx 0,54$ (doğru pozitif tahmin edilenlerin, toplam gerçek pozitiflere oranı)

F1 Puanı = $2 * Kesinlik * Duyarlılık / (Kesinlik + Duyarlılık) = 2 * 0,74 * 0,54 / (0,74 + 0,54) = 0,62$

Şekil 4.21 incelendiğinde test büyüklüğü % 20 iken ve 10 kere çapraz doğrulama yapılan rastgele orman algoritması % 72,11'lik değerle en yüksek AİK puanını vermiştir. Bu değer, pozitif ve negatif değerler arasında "kabul edilebilir ayrım" olduğunu gösterir.



Şekil 4.21. AİK eğrisi grafiği



5. TARTIŞMA

Bu çalışmada, ağaç ürünleri imalatı sektöründe çalışan 1 477 çalışanda cinsiyet, yaş, toplam çalışma yılı, sektörde çalışma yılı, gürültülü ortamda çalışma süresi, sigara kullanımı, KKD kullanımı, gürültülü hobinin GBİK üzerine etkisini araştırmak için hem regresyon analizi hem de sınıflandırma analizi yapılmıştır. Kullanılan regresyon algoritmaları analizlerinde R^2 değeri 0,2851 olarak bulunmuştur. Bu değer, dB cinsinden GBİK'teki değişimin sadece % 28,51'i cinsiyet, yaş, toplam çalışma süresi (yıl), sektörde çalışma süresi (yıl), gürültülü ortamda çalışma süresi (ay), sigara kullanımı, gürültülü hobisi olup olmaması ve yetersiz kulak koruyucu donanım kullanımı ile açıklanabilmektedir. Regresyon analizinde bulunan korelasyon katsayısı ($r = \%53,395$) ise GBİK ile bağımsız değişkenler arasında orta düzeyde bir ilişki olduğunu göstermektedir. Bağımlı değişken ve bağımsız değişkenler arasında zayıf yakın ilişki olması veri setinin doğrusal regresyon modellemesine uygun olmadığını göstermektedir. Bunun sebebi sağlık verilerinde bağımlı değişkenin, hiç fark edilmeyen veya hasta tarafından beyan edilmemiş/eksik beyan edilmiş birçok değişkenden etkilenmiş olması ihtimalidir. Bu ihtimal verilerde doğrusallığı bozabilmektedir.

Kullanılan sınıflandırma algoritmaları arasında % 73,99 (yaklaşık % 74)'luk doğruluk ile 10 defa çapraz doğrulama yapılan test büyüklüğü % 20 olan rastgele orman algoritması en iyi algoritma seçilmiştir. Rastgele orman algoritmasının %72,11'lik AİK puanı GBİK yaşayanlar ile GBİK yaşamayanlar arasında kabul edilebilir bir ayırım olduğunu göstermiştir. Bu tez çalışmasında, sınıflandırma algoritmaları ile 8 farklı bağımsız değişkenin GBİK üzerine etkisi araştırılırken en benzer çalışmanın Zhao ve diğerleri (2018)'ne ait olduğu görülmüştür. Zhao ve diğerleri (2018) çalışmalarında bu çalışmaya oranla daha yüksek doğruluk ve AİK skoru elde etmişlerdir. Daha yüksek doğruluk elde etmelerinin nedenleri arasında bağımsız değişken sayılarının az ve örneklem sayılarının fazla olması olabileceği gibi çalışanlardan alınan günlük kişisel gürültü maruziyet ölçümleri değişkeni de doğruluk puanının artmasına katkıda bulunmuştur. Bu tez çalışmasında da çalışanlardan alınan günlük kişisel gürültü maruziyet ölçümlerinin GBİK üzerine etkisi araştırılmak istenmiştir. Ancak, 327 çalışana ait günlük kişisel gürültü maruziyet ölçüm sonuçları ile yapılan analizlerde istatistiksel anlamlılık seviyesi (p) 0,05'ten büyük çıktığı için bu bağımsız değişkenle kurulacak model anlamlı çıkmamıştır.

En yüksek doğruluk oranına sahip olan rastgele orman algoritmasının duyarlılık değeri (% 53) doğruluk oranı (% 73,99) kadar yüksek değildir. Bu demek oluyor ki toplamda en iyi tahmini yapsa da, pozitifler içinde en çok doğru pozitif tahmini yapan algoritma rastgele orman algoritması değildir. Test büyüklüğü % 20 ile en yüksek duyarlılık değerine (% 67) sahip olan Naive Bayes (Bernoulli) sınıflandırması, pozitifler içinde en çok doğru pozitif tahmini yapmıştır. Sağlık verilerinde gerçek hasta olan kişi sayısını doğru tahmin etmek, gerçek hasta olmayan kişi sayısını tahmin etmekten daha önemlidir. Bu nedenle, benzer performans kriterlerinde olan modeller arasında, duyarlılık değeri en yüksek olan model en iyisidir. Çünkü gerçek toplam pozitifler içinde en çok doğru pozitif tahmini yapan algoritma sayesinde en çok işitme kaybı yaşayan çalışan sayısına ulaşılmaktadır.

Bilindiği üzere gürültülü ortamda bulunma sonrasında işitme kaybının, gürültü kesildikten sonra belirli zaman içerisinde işitme eşiklerinin gürültüye maruziyetten önceki seviyeye geri dönmesine geçici eşik değişikliği denmektedir ve birçok geçici eşik değişikliği ilk iki günde (en az 14 saatte) düzelmektedir. Bu çalışmada olduğu gibi Zhao ve diğerleri (2018) de çalışmalarında odyogramları, çalışanın mesleki gürültüye maruz kalmasından en az 16 saat sonra ölçerek geçici eşik düşmesi ihtimalini bertaraf etmişlerdir. En az 14 saatlik şartın sağlanmadığı durumlarda 80 dB (A) 'in üzerindeki gürültü düzeylerinde standart kulak koruyucu (tüm mesai boyunca) kullanılması şartıyla test gün boyu yapılabilir. Böylece çalışmalarıyla makine öğrenimi algoritmalarının, çeşitli karmaşık endüstriyel gürültülere maruz kalan işçilerde gürültü kaynaklı işitme bozukluğunu tahmin etmek için potansiyel araçlar olduğunu göstermişlerdir.

6. SONUÇLAR VE ÖNERİLER

Bu çalışmayla GBİK'i etkileyen bağımsız değişkenlerin etki düzeyi araştırılarak GBİK'in önceden tahmin edilerek önlenmesi hedeflenmiştir. Bunun için regresyon ve sınıflandırma modelleri kurulmak istenmiştir. Ancak regresyon analizi için yapılan varyans analizinde, bağımsız değişkenler ile dB cinsinden bağımlı değişken arasında orta düzeyde ilişki bulunmuştur. Bu nedenle regresyon modellerinin, bu çalışmanın veri seti kullanıldığında, GBİK'i tahmin etmede yetersiz kaldığı sonucuna varılmıştır.

Sınıflandırma modelleri incelendiğinde performans kriterleri en yüksek çıkan model rastgele orman algoritması olmuştur. Modelin, pozitif ve negatif sınıfları ayırt etme performansı "kabul edilebilir" derecede olduğundan; sınıflandırma modellerinin bu çalışmanın veri seti kullanıldığında GBİK'i sınıflandırmada yeterli olduğu sonucuna varılmıştır. Ayrıca seçilen modelde GBİK'i etkileyen en önemli değişkenin "gürültülü ortamda çalışma süresi" değişkeni olması, modelin bu çalışmanın hedefini karşılamada da yeterli olduğunu göstermiştir. En iyi sınıflandırma modeli sonuçlarına dayanarak, ağaç ürünleri imalatında çalışanlarda yaşanan GBİK'i en çok "gürültülü ortamda çalışma süresi"nin etkilediği kanıtlanmıştır. Bu nedenle, GBİK'i azaltmak/önlemek için gürültü maruziyetini azaltacak önlemler alınmalıdır. Bu önlemler aşağıdaki gibidir:

- Gürültüyü direk kaynağında yok etmeye yönelik alınması gereken önemler: Gürültü yayan makinelerin ve yapı parçalarının periyodik kontrolleri, bakım ve onarımları düzenli ve eksiksiz yapılmalıdır. Titreşim sonucu gürültü yayan yapılara titreşim sönümleyici malzemeler yerleştirilmelidir. Kesme işlerinde çelik yerine karbon bileşimli uç (karbit uç) kullanmasıdır. Çeliğe göre daha dayanıklı bir malzeme olan bu bileşimler, sürtünmeyi azalttığından gürültüyü de azaltacaktır. Ayrıca spiral kesiciler düz kesicilere göre gürültü seviyesini en az 5 dB düşürmektedir.
- Gürültü ortamında alınması gereken önlemler: Gürültü maruziyetini arttıran iş ekipmanları, diğer iş ekipmanlarından ayrı, yalıtılmış bölümlerde kullanılmalıdır. Gürültüyü artırıcı etkiye sahip olması nedeniyle yankılanmayı ortamda azaltmak için çalışma sahası akustik soğurucu malzemeler ile kaplanmalıdır. Bu malzemeler doğru yerleştirildiği takdirde gürültü seviyesinde 20 dB'e kadar azalma sağlamaktadır.

- Gürültüye maruz kalan çalışanda alınması gereken önlemler: İş istasyonlarında rotasyon uygulamalıdır. Gürültü seviyesine uygun ve ergonomik kulak koruyucu donanımlar kullanılmalıdır. Dinlenme saatleri gürültü maruziyet süresini azaltacak şekilde düzenlenmelidir.

Seçilen sınıflandırma modelinin, değişkenlerin önem katsayıları incelendiğinde, erkek çalışanlarda GBİK yaşama olasılığının kadınlardakinden; sigara kullananların GBİK yaşama olasılığının sigara kullanmayanlardan daha yüksek belirlenmiştir.

Gelecekteki çalışmalarda:

- İnsanlarda yaşlılığa bağlı olarak gelişen işitme sinirindeki veya iç kulaktaki hasarlar, işitme kaybına neden olmaktadır. GBİK değil de genel işitme kaybı ile ilgili tahmin modelleri geliştirmek istendiğinde, analiz sonucunda genel işitme kaybını en çok etkileyen değişkenin “yaş” değişkeni olması beklenirdi. Ancak GBİK araştırılırken doğrudan yaşa bağlı işitme kaybını bertaraf edebilmek için 4 000 Hz’de çentik şeklinde oluşan odyometri sonuçları, bağımlı değişken olarak belirlenmiştir. Sonuçta en iyi sınıflandırma algoritması modelindeki en önemli değişken “gürültülü ortamda çalışma süresi” olarak tespit edilmiştir. “Yaş” değişkeni ise GBİK’i dördüncü sırada etkilemektedir. Ancak bu halde dahi yaşa bağlı iletim tipi işitme kaybı ile GBİK birbirinden tamamen ayıramamaktadır. Bu ayrımı en yüksek seviyede yapabilmek için ileride yapılacak benzer çalışmalarda TS ISO 1999 “Akustik - Gürültüye Bağlı İşitme Kaybının Tahmini Standardı”ndan yararlanılmalıdır. Bu standartta hayatı boyunca mesleki gürültüye maruz kalmamış insanların odyometri sonuçlarından elde edilen “işitme eşiği seviyelerinin istatistiksel dağılımının desibel cinsinden seçilmiş değerleri tablosu” yardımıyla kalıcı işitme eşiği seviyesi hesaplandığından, yaş ile birlikte oluşabilen iletim tipi işitme kaybı bertaraf edilmektedir. Sonuçta çıkan kalıcı işitme eşiği seviyesi, yalnızca gürültü kaynaklı işitme kaybını vermektedir.
- Bu çalışmada sağlık verisi kullanıldığından, GBİK’i etkileyen değişkenler ile çalışanlardaki dB cinsinden işitme kaybı arasındaki korelasyon katsayısı ($r = 0,53395$) orta seviyededir. Yani, bağımlı değişken ile bağımsız değişkenler arasındaki doğrusallık orta düzeydedir. Bağımsız değişkenlerin, bağımlı değişkeni düzensiz şekilde etkilediği bu tarz sağlık verilerinde regresyon analizi yaparken, doğrusallığın düşük olması nedeniyle, analiz için bağımsız değişkenler ve bağımlı değişken arasında pozitif yönde orta seviyeli bir ilişkinin olduğu doğrusal regresyon modelleri yerine doğrusal olmayan regresyon model-

leri tercih etmek daha uygun olacaktır. Örneğin, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin n'inci derece polinom olarak modellendiği polinom regresyonu ile verilere bir polinom fonksiyonu uyarlanarak doğrusal olmayan ilişkiler daha doğru açıklanabilecektir.

- Bu çalışmada üçüncü aşama olarak, 327 kişisel gürültü maruziyet ölçüm değeri, dokuzuncu bağımsız değişken olarak eklenerek GBİK'in kaç dB etkilendiği araştırılmak istenmiştir. Ancak yapılan varyans analizi testlerinde p değerinin 0,05'ten büyük olması sebebiyle kurulması planlanan modellerin istatistiksel olarak anlamlı çıkmadığı görülmüştür. Örneklem sayısının az olması p değerinin düşük çıkmasına sebep olmuş olabileceğinden, benzer çalışmalarda kişisel gürültü maruziyet ölçüm değeri örneklem sayısının artırılması, p değerini 0,05'in altına düşürebilir. Sonuçta kurulacak modeller istatistiksel olarak daha anlamlı çıkacaktır
- En iyi sınıflandırma modelinin AİK puanı (%72,11) GBİK olanlar ve GBİK olmayanlar arasında kabul edilebilir bir ayrım olduğunu göstermiştir. Bu ayrım sonucu söz konusu veri setinin analizi için sınıflandırma algoritmalarının regresyon algoritmalarından daha uygun olduğunu kanıtlamaktadır. Gelecek çalışmalarda sınıflandırma modelleri algoritmaları karşılaştırırken doğruluk, kesinlik, duyarlılık, AİK skoru, ortalama ağırlıklı f1 puanı değerlerinde benzerlik olduğu durumda; duyarlılık değeri yüksek olan algoritmayı sınıflandırma algoritması olarak seçmek gerekmektedir. Çünkü duyarlılık değeri doğru tahmin edilen hasta sayısının toplam gerçek hasta sayısına oranı vermektedir. Sağlık verilerinde gerçekten hastayı bulmak gerçekten hasta olmayanı bulmaktan daha önemli olduğundan, duyarlılık değeri daha yüksek olan model tercih edilmelidir.

Bu çalışmayla, çalışmaya konu olan veri seti için GBİK analizinde sınıflandırma algoritmalarının başarılı; regresyon algoritmalarının yetersiz olduğu savunulmaktadır. Daha başarılı analiz sonuçları için uygun doğrusallaştırma yöntemleri kullanıldıktan ve yaşa bağlı işitme kaybını ortadan kaldıracak hesaplamalar yapıldıktan sonra regresyon analizinin yapılması önerilmektedir. Çalışmanın modeli, ağaç ürünleri imalatı sektörü dışındaki diğer sektörlerde uygulanabilir. Bu çalışmanın modeline, uygulanacak sektörde kullanılan hammadde ve proseslere uygun başka değişkenler eklenebilir. Veri toplama aşamasında, uygulanacak sektörde kimyasalla çalışma var ise kimyasalla ne kadar süre çalışıldığına ve çalışanlarda iç kulağı etkileyen ototoksik ilaç kullanımı var ise bu tür ilaçların kullanımına ait değişkenlerin eklenmesiyle daha kapsamlı modeller kurulabilir.



KAYNAKLAR

- Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley & Sons Incorporated, 377-381.
- Ali, J., Khan, R., Ahmad and N. and Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues*, 9(5), 272–278.
- Barto, A. G. and Dietterich, T. G. (2004). *Reinforcement Learning and its Relationship to Supervised Learning*. Massachusetts University.
- Bilekdemir, G. (2010). *Veri madenciliği tekniklerini kullanarak üretim süresi tahmini ve bir uygulama*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir, 24-34.
- Bircan, H. B. (2004). Lojistik regresyon analizi: Tıp verileri üzerine bir uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (2), 185-208.
- Chen, F., Zuwei, C., Grais, E. M. and Zhao, F. (2021). Contributions and limitations of using machine learning to predict noise-induced hearing. *International Archives of Occupational and Environmental Health*, (94), 1097–1111.
- Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 14(3), 326-334.
- Çalışma ve Sosyal Güvenlik Bakanlığı. (2013). *Çalışanların Gürültü ile İlgili Risklerden Korunmalarına Dair Yönetmelik*. Ankara: Çalışma ve Sosyal Güvenlik Bakanlığı.
- Çalışma ve Sosyal Güvenlik Bakanlığı. (2018). *Ağaç ürünleri imalatı sektöründe iş sağlığı ve güvenliği rehberi*. Ankara: Çalışma ve Sosyal Güvenlik Bakanlığı.
- Dy, J. G. and Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research*, (5), 845-889.
- Erdoğan, A. (2016). *Denizli’de üç tekstil fabrikasındaki gürültü düzeyinin çalışanlar üzerine etkisi*. Uzmanlık Tezi, Pamukkale Üniversitesi Sağlık Bilimleri Enstitüsü, Denizli, 57-64.
- Gujarati, D. N. (2010). *Temel ekonometri*. (Çev. Ü. Şenesen ve G.G. Şenesen), İstanbul: Literatür Yayıncılık, 442-444.
- Hosmer, D. W. and Lemeshow, S. (2013). *Applied logistic regression*. Canada: John Wiley & Sons Incorporated, 52-57.
- Hu, L. Y., Huang, M. W., Ke, S. W. and Tsai, C. F. (2016). The distance function effect on k- nearest neighbor classification for medical dataset. *Springer Plus*. 5(1), 1-9.

- İnternet: Ağaç Ürünleri Üretimi Sektörel Uygulama Kılavuzu. Web: <https://cevresehiriklim-kutuphanesi.csb.gov.tr/ShowPDF/fa915272-6c66-49f1-b99a-c2f76c77c645>, Son Erişim Tarihi: 05.01.2024.
- İnternet: ASA/ANSI S3.6-2018 (R2023): Specification For Audiometers, Web: <https://blog.ansi.org/asa-ansi-s3-6-2018-r2023-audiometers-specification/#gref>, Son Erişim Tarihi: 05.01.2024.
- İnternet: Breiman, L. and Cutler, A. (2009). Random Forest. Web: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm, Son Erişim Tarihi: 08.01.2024.
- İnternet: Fan, B., Wang, G., Han, H., Li, B., Wang, L., Qu, C., Zhang, X., Liu, H., Sun, Z., Zhou, Y., Zhang, Y. and Wu, W. (2022, August). Machine learning model identifies genomic variation in noise-induced hearing loss via wholeexome sequencing. *Research square*, 2. Web: <https://www.researchsquare.com/article/rs-1829497/v2>, Son Erişim Tarihi: 05.01.2024.
- İnternet: Mei, S., Montanari and A., Nguyen, P. (2018). A mean field view of the landscape of two-layer neural networks. Web: <https://www.pnas.org/doi/10.1073/pnas.1806579115> Son Erişim Tarihi: 20.06.2024.
- İnternet: Selvaganesh, N. Shanthi, D. and Rajesh Pandian, N. (November, 2023). A Novel Biased Probability Neural Network (BPNN) and Regularized Extreme Learning Machine (RELM) Based Hearing Loss Prediction System. *Iraqi Journal for Computer Science and Mathematics*. Web: <https://www.iasj.net/iasj/download/2469b219d3bd1612>, Erişim Tarihi: 01.07.2024.
- İplikçi, S. (2013). *Makine Öğrenmesi Yüksek Lisans Ders Notları*, Denizli: Yayınlanmamış.
- Karagül, K. (2014). The classification of the firms traded in İstanbul stock exchange by using support vector machines. *Pamukkale University Journal of Engineering Sciences*, 20(5), 174–178.
- Kim, K. S. (2010). Occupational hearing loss in Korea. *Journal of Korean medical science*, (25), 62-69.
- Kulkarni, V. Y. and Sinha, P. (2013). Random forest classifiers: a survey and future research directions. *International Journal of Advanced Computing*, 36(1), 1444-1153.
- Luis, E., Gonzales, J. P. N., Castanon, M. A. G. and Menendez, R. M. (2008). Fault diagnosis of industrial systems with bayesian networks and neural networks, advances in artificial intelligence. *Springer-Verlag*, 5317, 998-1008.
- Mutlu, A. (2010). *Madencilikte gürültüye bağlı işitme kayıplarının tespiti: taş kırma eleme tesisi örneği*. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 62-65.
- Naqa, I. E. and Murphy, M. J. (Editors). (2015). *Machine learning in radiation oncology, Switzerland: Springer*, 3-11.

- Noma, N. G. and Ghani, M. K. A. (2013). Predicting hearing loss symptoms from audiometry data using machine learning algorithms. *Australian Journal of Basic and Applied Sciences*, 7(9), 35-43.
- Önder, S. (2018). Lojistik regresyon analizi ile bir yeraltı metal madeninde gürültüye bağlı işitme kayıplarının incelenmesi. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 33(3), 11-22.
- Öz, H. M. ve Köse, E. (2020). Gürültü önleyici akustik malzemelerin performans düzeylerinin incelenmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (18), 1-10.
- Özdemir, M. (2016). *Küçük ölçekli kaporta işletmelerinde çalışanlarda gürültüye bağlı işitme kayıplarının araştırılması*. Uzmanlık Tezi, Çalışma ve Sosyal Güvenlik Bakanlığı, Ankara, 31-34.
- Öztürk, A., Ergör, G., Demiral, Y., Ergör, A. ve Tapçı, N. (2007). Döküm işkolunda gürültüye bağlı işitme kayıpları sıklığı ve etkileyen etmenlerin değerlendirilmesi. *Türk Tabipleri Birliği Mesleki Sağlık ve Güvenlik Dergisi*, 9(31), 40-46.
- Qi, Z., Tian, Y. and Shi, Y. (2013). Robust twin support vector machine for pattern classification. *Pattern Recognit. Pattern Recognition*, 46(1), 305-316.
- Sabancı, A. ve Sümer, S. K. (2011). *Ergonomi*. (3. Baskı). Ankara: Nobel Akademik Yayıncılık, 233-286.
- Sosyal Sigortalar Kurumu. (1972). *Sosyal Sigorta Sağlık İşlemleri Tüzüğü*. Ankara: Çalışma ve Sosyal Güvenlik Bakanlığı.
- Temel, G., Erdoğan, S. ve Ankaralı, H. (2012). Sınıflama modelinin performansını değerlendirmede yeniden örnekleme yöntemlerinin kullanımı. *Bilişim Teknolojileri Dergisi*, 5(3), 1-8.
- Umar, İ. K. and Bashir, S. (2019). Investigation of the factors contributing to truck driver's involvement in an injury accident. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 26(3), 402-408.
- Yavuz, A. (2021). *Derin öğrenme algoritmaları ile trafik işaret ve levhalarının tanımlanması*. Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli, 39-40.
- Yin, S., Wang, H., Jiang, Z. and Huang, S. (2019). Noise-induced hearing loss and its individual susceptibility. *Laryngorhinootologie*, 98(2), 159.
- Zhao, Y., Li, J., Zhang, M., Lu, Y., Xie, H., Tian, Y. and Qui, W. (2018). Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise. *Ear and Hearing*, 40(3), 690-699.





Gazili olmak ayrıcalıktır