T.C.
BALIKESİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

MOBBİNG İÇERİKLİ YARGI KARARLARININ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE SINIFLANDIRILMASI

ÖZLEM AYDIN

YÜKSEK LİSANS TEZİ

Jüri Üyeleri : **Dr. Öğr. Üyesi Kadriye ERGÜN (Tez Danışmanı)**
Dr. Öğr. Üyesi Kamil TOPAL
Dr. Öğr. Üyesi Tuğba TUNACAN

BALIKESİR, EYLÜL - 2020

ETİK BEYAN

Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak tarafımca hazırlanan "**Mobbing İçerikli Yargı Kararlarının Makine Öğrenmesi Algoritmaları ile Sınıflandırılması**" başlıklı tezde;

- Tüm bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Kullanılan veriler ve sonuçlarda herhangi bir değişiklik yapmadığımı,
- Tüm bilgi ve sonuçları bilimsel araştırma ve etik ilkelere uygun şekilde sunduğumu,
- Yararlandığım eserlere atıfta bulunarak kaynak gösterdiğimi,

beyan eder, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ederim.

Özlem AYDIN

Bu tez çalışması Balıkesir Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından 2019/096 nolu proje ile desteklenmiştir.

ÖZET

MOBBİNG İÇERİKLİ YARGI KARARLARININ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE SINIFLANDIRILMASI

YÜKSEK LİSANS TEZİ

ÖZLEM AYDIN

BALIKESİR ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

(TEZ DANIŞMANI: DR. ÖĞR. ÜYESİ KADRİYE ERGÜN)

BALIKESİR, EYLÜL - 2020

Bilgisayar teknolojileri ve internet kullanımındaki gelişmeler, üretilen veri boyutu ve çeşitliliğini artırmakla kalmamış bu verilere ulaşılmasını da kolaylaştırmıştır. Günümüzde birçok bilgiye internet üzerinden online olarak ulaşmak mümkün hale gelmiştir. Bu bilgilerin bir kısmı yapısal verilerden oluşurken bir kısmı yapısal olmayan (ses, görüntü, metin vs.) verilerden oluşur. Bu verilerin zamanında ve doğru analiz edilmesi kriz süreçlerinin yönetilmesi, karar aşamasında yön tayini, stratejik planların oluşturulması, ülkelerin yönetimi, ulusal ve uluslararası güvenlik gibi birçok konuda kritik öneme sahiptir.

Veri yığınlarını kullanılabilir özet bilgilere dönüştürebilmek için çeşitli madencilik yöntemleri uygulanmaktadır. Metin verileri üzerinde yapılan madencilik çalışmaları metin madenciliği uygulamalarıdır. Metin verileri yapısal olmayan verilerdir ve bilgisayarda işlenebilmeleri için bir dizi ön işlemden geçirilerek analize hazır hale getirilmelidirler.

Bu tez kapsamında, metin madenciliği yöntemleri kullanılarak mobbing içerikli yargı kararları incelenmiştir. Model python pogramlama dilinde yazılmıştır. Mobbing karaları güncel bir konu olması ve ispat edilmesi zor bir dava konusu olması sebebi ile tercih edilmiştir. Yüksek mahkemece verilen kararlar mobbing iddiasının kabul edilip edilmemesine göre iki sınıfa ayrılmıştır. Ardından makine öğrenmesi yöntemlerinden sınıflandırma analizi yapılarak gözetimli öğrenme gerçekleştirilmiştir. Oluşturulan model ile mobbing kararlarının sınıflandırılmasında %80 ve üzeri başarı sağlanmıştır.

NAHTAR KELİMELER: Metin madenciliği, makine öğrenmesi, gözetimli öğrenme, metin sınıflandırma, MLP sınıflandırıcı, gradient boost sınıflandırıcı, word2vec.

Bilim Kod / Kodları : 90618 Sayfa Sayısı : 79

ABSTRACT

CLASSIFICATION OF MOBBING TOPICAL JUDICIAL DECISIONS BY MACHINE LEARNING ALGORITHMS

MSC THESIS

ÖZLEM AYDIN

BALIKESIR UNIVERSITY INSTITUTE OF SCIENCE

INDUSTRIAL ENGINEERING

(SUPERVISOR: ASSIST. PROF. DR KADRİYE ERGÜN)

BALIKESİR, SEPTEMBER - 2020

Developments in computer technologies and internet usage have not only increased the size and variety of data produced, but also facilitated access to these data. Today, it has become possible to reach a lot of information online over the internet. While some of this information consists of structural data, some of it consists of unstructured data (sound, image, text, etc.). Timely and accurate analysis of these data has critical importance in many aspects such as managing crisis processes, determining the direction in decision-making, creating strategic plans, managing countries, national and international security.

Various mining methods are used to transform data stacks into usable summary information. Mining studies on text data are text mining applications. Text data are unstructured data and must be prepared for analysis by a series of preprocessing in order to be processed on the computer.

Within the scope of this thesis, judicial decisions involving mobbing were analyzed using text mining methods. Mobbing decisions have been preferred because it is a current issue and is a difficult subject to prove. The decisions given by the high court are divided into two classes according to whether the mobbing claim is accepted or not. Then, by performing classification analysis, one of the machine learning methods, supervised learning was realized. With the created model, a success of 80% or more has been achieved in classifying mobbing decisions.

KEYWORDS: Text mining, machine learning, supervised learning, text classification, MLP classifier, gradient boost classifier, word2vec.

Science Code / Codes : 90618 Page Number : 79

İÇİNDEKİLER

Sayfa

ŞEKİL LİSTESİ

Sayfa

TABLO LİSTESİ

Sayfa

ÖNSÖZ

Yüksek lisans eğitimim boyunca yardım ve desteklerini benden esirgemeyen, sabır ve özveri ile bana yol gösteren, saygı değer Dr. Öğretim Üyesi Kadriye ERGÜN' e sonsuz teşekkür ederim.

Ayrıca iş yerinde beni destekleyen değerli iş arkadaşlarıma yardım ve destekleri için teşekkür ederim.

Maddi ve manevi desteklerini hiçbir zaman benden esirgemeyen, aileme teşekkür ederim.

Balıkesir, 2020 **Özlem AYDIN**

1. GİRİŞ

Metin madenciliği bilgisayar teknolojilerinde yaşanan gelişmelere paralel olarak, yazınsal verilerin barındırdığı gizli anlamların araştırılmasına duyulan ihtiyaç sonucu ortaya çıkmış bir bilgi keşfi yolculuğu olarak tanımlanabilir. Uzun zaman süresince sayısal verilerin analizinde kullanılan bilgisayarlar teknolojileri, geliştirilen yeni yöntemler aracılığı ile metin verilerinin de bu teknoloji ile işlenebilmesine olanak tanımıştır. Çağımızda bilgi keşfine her alanda ihtiyaç duyulması nedeniyle metinsel verilerin işlenmesinin önemi artarak devam etmektedir. Metinsel verilerden faydalı bilgilerin çıkarılması işlemleri sırasında ortaya bazı zorluklar ortaya çıkmaktadır. Bu zorluklar temelinde veri yapılarından kaynaklanmaktadır.

Bilgisayarlar ile işlenen veri yapıları ikiye ayrılmaktadır. Bunlar; yapısal (structured) ve yapısal olmayan (unstructured) verilerdir. Yapısal veriler herhangi bir işleme tabi tutulmadan bilgisayar programları tarafından kullanılabilecek veriler iken yapısal olmayan veriler, bilgisayarların işleyebilmesi için bir dizi işlemden geçmesi gereken verilerdir. Metin madenciliği bu yapısal olmayan verileri analiz edilerek anlamlı sonuçlar çıkarmayı amaçlayan bir süreçtir ve aslında veri madenciliğinin bir alt uygulaması olarak ortaya çıkmıştır. Yapısal verilere finansal veriler, sensor datalar gibi bir tablonun satır ve sütunlarını oluşturabilecek nümerik veriler örnek verilebilir. Yapısal olmayan veriler ise yazınsal verilerdir.

Tarihteki ilk yazılı mahkeme kararı M.Ö. 1850 dolaylarında Sümerler tarafından kaleme alınmıştır [1]. Bu alanda metin verilerinin üretilmesi hem tarihsel olarak milattan öncesine dayanmaktadır hem de büyük miktarlarda veri üretilmektedir. 2019 yılında Yargıtay Hukuk Genel Kurulu ve Yargıtay Hukuk Dairelerine yeni intikal eden dosya sayısı 138.669 dur. Önceki yıllardan devreden dosyalar ile toplam dosya sayı 362.779 olarak gerçekleşmiştir. Bu kararlardan 232.416 tanesi onama, bozma, kısmi onama / bozma, gönderme, geri çevirme, ret, diğer başlıkları ile sistemden çıkarılmıştır. 130.363 adet dosya sonraki yıla devretmiştir [2]. Yargıtay Başkanlığının sitesinde yayınlanan bu istatistikler, üretilen yapılandırılmamış verinin boyutlarının büyüklüğü ile ilgili yaklaşık bir fikir

edinmemizi sağlayabilmektedir. Bu büyük miktardaki metinsel metin madenciliği yöntemleri kullanılarak işlenmesi, hukuk sisteminin verim ve etkinliğinin artırılmasını, iş yükünün azaltılmasını, hukuksal süreçlerin kısaltılmasını vs. sağlayabilecek sistemler üretilmesini sağlayabilir.

Bu çalışmada yargı kararları metin madenciliği uygulamaları ile incelenmiş ve makine öğrenmesi algoritmaları ile sınıflandırmaya tabi tutulmuştur. Yargı kararlarından mobbing içerikli olan Yargıtay Hukuk Genel Kurulu ve Yargıtay Hukuk Dairelerinin kararları ele alınmıştır. Bu kapsamda Kazancı İçtihat Bilgi Bankasından temin edilen 2013 ve 2019 yılları arasında Yargıtay Hukuk Genel Kurulu ve Yargıtay Hukuk Dairelerinin mobbing içerikli 461 kararı taranmış bunların 131 tanesi modele dâhil edilmiştir. Kararların mobbing olarak değerlendirilmesi ve değerlendirmemesi durumuna göre özellik çıkarımı yapılmıştır. Modelde Gözetimli Öğrenme (Supervised Learning) yöntemi benimsenmiş, 122 karar kendi içinde Yargıtayın nihai kararına göre mobbing varlığı kabul edilenler 1 (bir) etiket ile kabul edilmeyenler ise 0 (sıfır) etiketle sınıflara ayırılmıştır. Model oluşturulurken sınıflandırma algoritması olarak python sklearn linear kütüphanesinde bulunan "Lojistik Regresyon (Logistic Regression), Naive Bayes, Karar Ağaçları (Decision Tree), K En Yakın Komşu (K-NN), Destekçi Vektör Makinesi (SVM), Gradient Boosting Classifier, AdaBoost Classifier, Bagging Classifier, Random Forest Classifier, MLP Classifier" sınıflandırma algoritmalarından yararlanılmıştır. Yine sklearn linear kütüpanesindeki metindeki tokenleri sayarak çalışan "count vector" yöntemi, ters doküman frekansı "TFxIDF", doküman vektörleri yöntemleri ve "Doc2Vec" ile sayısallaştırılmıştır. Ön işleme aşamalarından geçirilen veri seti üçe ayrılmış birinci grup, öğrenme gerçekleştirilmesi için makine öğrenmesi algoritmaları kullanılarak oluşturulan sınıflandırma modeline gönderilmiştir. Modelden Yargıtay'ın kararlarında mobbing olarak kabul edilme ve edilmeme durumunu test ve doğrulama setlerini kullanarak tahminlemesi istenmiştir. Daha sonra her bir algoritma sonucuna göre bilgi çıkarımı yapılmıştır. Çıkan sonuçlar grafikler ile ifade edilmiştir. Kullanılan on adet sınıflandırma algoritmasından biri olan Random Forest Sınıflandırıcısı kesinlik (precision) ve duyarlılık (recall) oranları da göz önünde bulundurulduğunda %89 doğruluk (accuracy) oranı ile en iyi öğrenmeyi gerçekleştirmiştir. Doğrulama setinde ise MLP Sınıflandırıcısı % 91 doğruluk (accuracy) oranı ile en iyi sonucu verdiği gözlenmiştir.

Kararların içeriğini oluşturulması nedeniyle bu bölümde mobbing kavramına değinilmiştir. Mobbing kavramı ilk olarak akademik çalışmalarda 1960'lı yıllarda incelenmiş olup yapılan çalışmalar sonucunda biyoloji alanından, tıp alanına, oradan davranış bilimleri alanına geçmiş, yapılan akademik çalışmaların sonucu olarak hukuk dünyası bu kavramla ilgili düzenlemeye gitmek durumunda kalmıştır. Sonuç olarak mobbing kavramı ülkelerin kanunlarına girerek hukuk düzeni içinde yer edinmiştir.

Mobbing kavramının çalışmamıza dâhil olan kısmı işyerinde psikolojik tacizdir. Bunun dışında hayatın birçok alanında mobbingden söz edilebilir. Mobbing kavramı hukuk alanındaki diğer uyuşmazlıklara göre daha yeni olması ve tanımının kesin sınırlar içinde olmaması nedeni ile mevcut belirsizliğinden dolayı seçilmiştir. Bu çalışmada yargı kararları mobbingin varlığını kabul eden ve etmeyen kararlar olarak etiketlenmiş ve eğitim seti ile öğrenme gerçekleştirerek test setinde yer alan kararların bu etiketlere göre bir sınıfa atanması modellenmiştir. Bu sınıflandırma sonucunda modelin mobbingin varlığını kabul eden ve etmeyen yargı kararlarını ne oranda doğru atayacağını görmek amaçlanmıştır.

İşyerlerinde psikolojik taciz ifadesi, mevzuata doğrudan ilk olarak 6098 Türk Borçlar Kanunu "İşçinin kişiliğinin korunması" başlığı altında 417. madde ile girmiştir. Türk Anayasası, Türk Ceza Kanunu, 4857 sayılı İş Kanunu doğrudan mobbing ile ilgi bir düzenleme içermemektedir ancak içeriklerinde yer alan güvenceler ve düzenlemelerle birlikte mobbing ile ilişkilendirilmektedir. Ayrıca hukukun diğer kaynaklarından olan içtihatlar ve bilimsel görüşler karar aşamasında hakimlerce kullanılmaktadır. Özellikle Yargıtay 9. Hukuk Dairesinin 2007/9154 E. 2008/13307 K. sayılı 30.05.2008 tarihli kararı konuyu detaylı olarak el almıştır.

1.1 Literatür Taraması

Metin madenciliği veri madenciliği yöntemlerini kullanarak metinleri inceler. Metinler biçimsiz ve karmaşık veri yapılarına sahiptir. Bilginin üretildiği birçok ortamda yapısal verilerden ziyade yapısal olmayan veriler mevcuttur. Bu verilerin boyutu kadar yayıldığı yelpaze de çok geniştir. Geniş yelpaze çalışma alanlarının çeşitliliğini de artırmaktadır. Yukarıda belirtilen nedenlerle metin madenciliği potansiyeli yüksek bir alan olarak karşımıza çıkmaktadır. Son yıllarda yapılan metin madenciliği çalışmalarındaki artış bu potansiyelin keşfedildiğinin kanıtı niteliğindedir. Metin madenciliği ile ilgili son dönemlerde yapılan bazı çalışmalara aşağıda yer verilmiştir.

Kishor ve Kolhe (2017) tarafından yapılan araştırmaya göre, işletmelerin bilgilerinin %80'i metin dosyalarında kayıtlıdır [3].

Liu (2018) tarafından yapılan araştırmada Biyomedikal literatürler ve elektronik sağlık kayıtları incelenmiş Doğal Dil İşleme (NLP) ve metin madenciliği teknikleri kullanılarak yapılandırılmamış metinsel içeriklerin ilişkisi incelenmiştir [4].

Biyomedikal literatürde veri madenciliği çalışması yapan diğer araştırmacı ise Binkheder ve Samar (2019) dır. Binkheder; fenotipleme tanımlarının literatürden alınmasını, sınıflanmasını ve çıkarılmasını otomatikleştirmek için kural tabanlı ve makine öğrenme yöntemlerini birleştiren bir metin madenciliği önermiştir. Bu amaçla, fenotip ve laboratuvarlar gibi fenotipleme tanımlarının modalitelerinin kanıtı olan cümleleri açıklayan on boyutlu bir ek açıklama kılavuzu geliştirilmiştir. Çalışmada bir fenotiple ilişkili fenotip adaylarını tanımlamak ve sıralamak için ortak oluşum ve ilişkilendirme yöntemleri kullanılmıştır. Bu çalışma literatüre dayalı dernekler ve büyük ölçekli korpus ile yeni veri odaklı fenotipleme tanımlarının oluşturulmasına ve asgari uzman katılımı ile mevcut tanımların genişletilmesine katkıda bulunmuştur [5].

M'Bareck (2019) tarafından yazılan doktora tezinde üç Arkansas siyasetçisinin Twitter hesaplarından indirilen 354 adet silahla ilgili tweetler ve üç yerel gazetede bu politikacıların silah politikasıyla ilgili görüşlerini içeren 40 haber incelenmiştir. Çalışmanın sonucunda, politikacıların Twitter' daki söylemlerinin son derece kutuplaşmış sözcükler ve görüşlerden oluştuğu ve yerel gazetelerin haberlerinin gerçeklere dayalı ve tarafsız olduğu görülmüştür. Ayrıca silah politikaları ile ilgili Twitter' daki siyasi duyguların son derece olumsuz, korkulu ve tedirgin edici olduğunu gözlenirken, gazetelerin silah politikası konusunda son derece tarafsız bir haber anlayışını benimsediği tespit edilmiştir [6].

Toprak (2018) tarafından yapılan çalışmada, Türkiye'deki her il için yayımlanan haberlere bir konu modelleme yöntemi olan Gizli Dirichlet Tahsisi kullanılarak en yüksek frekansa sahip 10 konu belirlenmiştir. Veri seti olarak Hürriyet gazetesinin açık kaynak veri tabanında bulunan haberler kullanılmıştır. Türkiye genelindeki haberlerde üniversite ve spor ile ilgili haberin daha fazla olduğu görülmüştür. İl bazında ise simgeleşmiş bazı kişi, kurum veya değerleri içeren haberlerin sayısının oldukça fazla olduğu görülmektedir [7].

Hamde (2018) yılında yayımlanan yüksek lisans tezinde Türkiye'de faaliyet gösteren firmaların metin madenciliği teknolojisini kullanarak; şirketleri ve rakipleri hakkındaki ilgili bilgileri otomatik olarak okumak ve belgelere erişmek suretiyle yöneticilere, karar verme ve rekabet analizi konularında yardımcı olup olamayacağını konu almıştır [8].

Tekin (2018) talep önceliklerini belirlediği çalışmasında; talebin, belirlenen inisiyatife bağlı olması durumunda gerçeklikten uzaklaştığını ve kritik olmayan talebi yüksek öncelikli olarak girilebildiğini tespit etmiştir. Tespitin hatalı planlama ve müşteri memnuniyetsizliği ile sonuçlanabileceğini vurgulamıştır. Çalışmada metin madenciliğinde sıkça kullanılan algoritmaların karşılaştırması yapılmıştır. Veri seti üzerinde en iyi sonucu veren algoritma %74,5 F-Skoru değeri ile Rassal Orman algoritması olmuştur [9].

Tan (2018) tarafından yapılan yüksek lisans tez çalışmasında sosyal paylaşım platformu Foursquare'de bulunan yorumlara metin madenciliği ve duygu analizi teknikleri uygulayarak karar destek sistemi geliştirilmiştir. Veriler üzerinde doğal dil işleme ve metin madenciliği teknikleri kullanılmış ilgili mekân hakkında genel olarak belirtilen duygu ve düşüncenin bulunması hedeflenmiştir. Önerilen karar destek sistemi ile olumlu ve olumsuz görüşler gerçek zamanlı olarak belirlenmekte ve duygu analizleri otomatik bir şekilde yapılmıştır. Bu sayede sosyal medya kullanıcılarının binlerce yorumu okumadan gitmeyi planladıkları mekânlar hakkında karar vermelerine katkı sağlamıştır. Çalışmanın sonuçları çerçevesinde hizmet verenlerin de hizmet kalitesini iyileştirmesi desteklenmektedir [10].

Atan ve Çınar (2019)' a ait araştırmada BİST 30 firmaları ile ilgili farklı kaynaklarda yayınlanan 14.108 haber metninde duygu analizi yapılmıştır. Yayınlanan finansal piyasa haberlerinin duygu içerikleri ile finansal değerler arasında anlamlı ilişkilerin var olduğu ve Türk finansal piyasalarının değerlendirilmesinde önemli bir araç olarak Türkçe haber kaynaklarının da kullanılabileceği sonucuna ulaşmıştır [11].

Göker ve Tekedere (2017), FATİH projesine yönelik internet yorumlarını metin madenciliği yöntemleri ile otomatik tespitini amaçladıkları çalışmalarında makine öğrenmesi algoritmalarını kullanmışlardır. Çeşitli sınıflandırma algoritmalarının veri kümesi üzerinde başarı yüzdeleri değerlendirilmiş ve en iyi sonucu Minimal Optimizasyon Algoritmasının verdiği gözlemlenmiştir. Çalışmada FATİH projesine yönelik görüşlerin olumlu ya da olumsuz olma durumu %88,73 doğruluk oranı ile otomatik tespiti edilmiştir [12].

Yalçın ve Erduran (2018)' a ait çalışmalarında İpsala Meslek Yüksek Okulu son dönem 214 öğrencisine "Öğrenciliğinize göstermiş olduğunuz sorumluluk duygusu düzeyinizle iş hayatınızda başarılı olabileceğinizi düşünüyor musunuz?" sorusunu yöneltmişlerdir. Verilen cevap metinlerine kümeleme algoritmaları uygulayarak öğrenciler gruplara ayrılmış ve birliktelik analizi yapılarak öğrencilerin ortak veya farklı kelime kullanımları araştırılmıştır. Yazarlar çalışmalarının amacını 'eğiticilerin öğrencilerin kendi

davranışlarını yorumlamasını sağlayarak, bireysel sorumluluk alma konusu ile ilgili bakış açılarını öğrenmek' olarak tanımlamışlardır [13].

Cecchini (2005) çalışmasında iflas ve hileli işlemlerin metin veriler kullanılarak daha doğru tahmin edilebileceğini savunduğu tezinde metin veriler için 10-K raporlarını kullanmıştır. Kernel yöntemleri ile bilgi çıkarımının kullanıldığı çalışmada; iflas tahmini ve hileli işlemlerin tespiti yüksek doğrulukla gerçekleştirilmiştir. 2010 yılında yapılan çalışmada ise 10-K raporlarının yönetici tartışmaları ve analizleri kısmı için metin analizi yaparak hileli finansal raporlama yapan ve yapmayan işletmeler ile iflas riski taşıyan işletmeleri sınıflandırmışlardır. Destekçi Vektör Makineleri (Support Vector Machine-SVM) algoritması ile sınıflandırma yapılmış; model hileli finansal raporları %75, iflas riski taşıyan işletmeleri ise %80 başarı ile sınıflandırmıştır [14].

Yıldız (2016)' a ait makalenin ana temasını halen faaliyet gösteren bir üniversitenin bilişim sisteminde bulunan kuruma ilişkin görüş öneri ve şikâyetlerin bildirildiği ve kurumca da bu bildirimlere cevap verildiği 3961 mesaj oluşturmaktadır. Veriler Ki Kare, Bilgi Çıkarımı ve TF-IDF yöntemleri ile işlenmiştir. Kurumun genelinde ve bölümler düzeyinde var olan raporlama hizmetlerini iyileştirmesi ve yönetimin ilgili raporlama sisteminin yeniden düzenlemesine olanak sağlanabilecek sonuçlara varılmıştır [15].

Drury ve Roche (2019) makalesinde tarım alanında artan bilimsel yazılı materyallerin sayısındaki artışa dikkat çekerek bu metinlerin tarımsal sorunları çözme ya da bilgi çıkarımı için metin madenciliği yöntemleri kullanılarak analiz edilmesinin yüksek potansiyele sahip olduğunu vurguladıkları bir makale yayınlamışlardır [16].

Yıldız (2019)' a ait çalışma Endüstri 4.0 ile ilgili 2012' den 2018' e kadar dünya genelinde yapılan çalışmaları incelemeyi amaçlamıştır. Endüstri 4.0 ile ilgili metinler SCOPUS veri tabanından sağlanmıştır. Kullanılan program yazarlar arasındaki alıntıları da kullandığı için

Endüstri 4.0'a etkisi olabilecek farklı konular da incelenebilir. Çalışmada Bilimsel Haritalama, Metin Madenciliği Teknikleri ve Biyometrik analiz yöntemleri kullanılmıştır [17].

Hei (2019) tarafından kaleme alınan makalede epilepsinin görülen en yaygın nörolojik hastalık olduğu ve hasta odaklı bakım girişimlerini yönlendirmek için veriye ihtiyaç olduğu vurgulanmıştır. Çalışma epilepsi hastalarının durumları hakkında ne tartıştıklarını öğrenmeyi ve online hasta destek gruplarından tedavi ile ilgili temaları tanımlamayı amaçlamıştır. Üç çevrim içi destek grubundan 355.838 gönderinin toplanıp incelendiği analiz sonucunda epilepsi hastalarının endişelerinin metin madenciliği yoluyla öğrenilmesinin mümkün olduğu gözlemlenmiştir [18].

Chaix, Deléger vd. (2019) makalelerinde gıda mikrobiyal çeşitliliği hakkında yapılan bilimsel yayınları metin madenciliği tekniklerinin kullanarak incelemiştirler [19].

Kano, Fujita vd. (2019) çalışmalarında yerel yönetimlerin insan gücü ve bütçe kısıtlamaları nedenleri ile karmaşık bölgesel sorunlara doğru öncelik verilememesine vurgu yapmıştırlar. Ayrıca nesnel veri analizine ve kanıta dayalı politika oluşturma üzerinde durulmuşlardır. Yazarlar metin madenciliği yöntemlerini kullanarak bölgesel politikalar oluşturulurken öncelik verilecek sorunların doğru tespiti için kullanılabilecek bir sistem önermiştir [20].

Literatürde yargı kararlarını ele alan Türkçe çalışmalar bulunmamakla birlikte yabancı kaynaklarda bu konu özelinde çalışmalar mevcuttur. Castro, Calixto vd. (2019) yargı kararlarında metin madenciliği uygulamalarını genişletmeyi amaçladıkları çalışmalarında ontoloji temelli semantik analiz yaparak davalardaki cezaları aramak için akıllı ve otomatik bir sistem önermiştirler. Bu yöntem bir nevi yargı kararlarının simülasyonu gibi çalışmakta ve adalet hizmetlerinin daha hızlı verilebileceği bir sistem öngörmektedir [21].

Metsker, Trofimov, vd. (2019) idari kararlara ilişkin bir çalışma yapmış ve makine öğrenmesi algoritmaları ile yarı yapılandırılmış veri analizine dayanarak temyiz sonucunu tahmine yönelik bir model geliştirmişlerdir. Çalışmanın mevcut mevzuatın iyileştirilmesi, kamu idarelerinin üzerindeki yükün azaltılması ile sonuçlanacağı öngörülmüştür [22].

Aletras, Tsarapatsanis vd. (2016) doğal dil işleme yöntemlerini kullanarak Avrupa İnsan Hakları Mahkemesinin kararların incelemiş ve tahminleme yapmıştırlar. Yayınlanan kararları belirli bölümlerinde benzerlik olabileceği düşüncesinden yola çıktıkları çalışmada dosya içeriklerinin sınırlı bir kısmına erişebildiklerini vurgulamış ve bu nedenle yayınlanmış kararların metin bölümleri ile başvuru ve metin özetlerinin benzer olması şartıyla sınıflandırma yapılabildiğini vurgulamıştırlar [23].

Thammaboosadee, Silparcha (2008) Tayland Ceza Davası Yüksek Mahkeme karalarını metin madenciliği yöntemleri ile değerlendirdikleri çalışmalarında her olay için (suç unsurlarını, suçlamaları, cezaları (istisnaları, ağırlıklarını)) üç unsur tanımlamışlar. Karar Ağacı Algoritmaları ile bu unsurların ilişkileri araştırılmış ve adli karar destek siteminin yapılandırılması önerilmişler [24].

Sağun (2015)' a ait tezinde mobbing kavramının ortaya çıkışı, teknolojik gelişmeler sonucu geçirdiği değişim, bireyleri ruhsal ve fiziksel olarak nasıl etkilediği, hukukun bu alana nasıl ve neden yöneldiği neden hukuki düzenlemelere ihtiyaç duyulduğunu irdelemiştir [25].

Çopur (2017) tarafından hazırlanan yüksek lisans tezinde Türkiye'de iş yerlerinde mobbing davranışları ile çok fazla karşılaşıldığına değinmiştir. "Türk Hukuk Sisteminde yeterli düzenleme ve yaptırımlar yer alıyor mu, Türk mevzuatında doğrudan mobbingi konu edinen maddeler var mı yoksa hangi maddeler mobbing kapsamında değerlendirilebilir" sorularına yanıt aramış ve mobbingin önlenmesine yönelik öneriler sunmayı amaçlanmıştır [26].

2. METİN MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ

Bu bölümde metin madenciliğinin tanımı, metin madenciliği süreci, metin madenciliğinde kullanılan teknikler, metotlar, metinsel verileri sayısallaştırma teknikleri ve makine öğrenmesi başlıklarına değinilmiştir.

2.1 Metin Madenciliği

İnsanlık yazıyı icat ettiğinden bu yana metinsel verile üretmeye başlamıştır. İnsanoğlu metinleri bazen kil tabletlere, taşlara kazımış bazen de İnkalar gibi "khipus" adını verdikleri düğümlere kaydetmiştir. Değişmeyen şey ise metinlerden bilgi çıkarmanın sahip olduğu kritik önemdir. Yaşanan teknolojik gelişmeler ile metinleri depolama şekli de değişmiştir. Bilgisayar teknolojisi başlangıçta kamu kurumlarında kullanım alanı buldu ve veriler belleklerde saklamaya başlandı daha sonra bilgisayarlar işyerlerine ardından evlere girdi. Günümüzde ceplerimize girecek boyuta indirgenen bu teknolojinin yanı sıra ilerleyen internet teknolojisi ile verilerin hem miktarı hem de çeşitliliği artmıştır. Bu verilerin bilgi keşfi boyutunda işlenmesi ise ancak veri madenciliğinin gelişmesi ile mümkün olmuştur.

Metin madenciliği, veri madenciliğinin bir alt dalı olarak gelişen yazılı verileri analiz etmek için veri madenciliği yöntemleri kullanan bilimdir. Bu yöntem veri madenciliğinden özellikle veri ön işleme aşamasında ayrılır. Diğer aşamalar her iki yöntemde de büyük ölçüde benzerdir. Metin madenciliğinin anlaşılabilmesi için öncelikle uğraştığı veri yapısını anlaşılması gerekir. Veriler; yapısal, yarı yapısal ve yapısal olmayan veriler olarak üç gruba ayrılabilmektedir.

Yapısal veriler, veri tabanı ve veri ambarlarında tutulan ve SQL, OLAP gibi sorgulama yöntemleri ile sorgulanabilen veri türünü ifade eder.

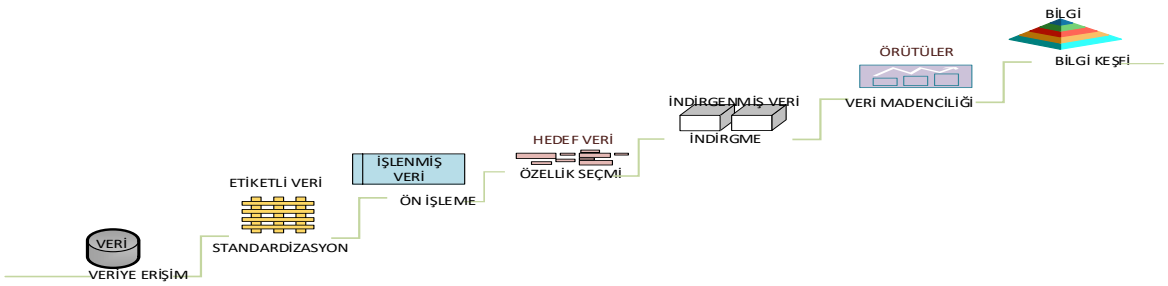
Yarı yapısal veriler ise metin, resim, grafik vs. olan belgelerdir. Belgelerin kim tarafından, hangi konuda ne zaman yazıldığı gibi bazı yapısal kısımları olmakla birlikte bir belgenin

içeriğinin tam olarak anlaşılması ancak bir insan tarafından okunması ile ortaya çıkarılabilmektedir.

Yapısal olmayan veri, önceden tanımlı bir veri modeli olmayan, SQL, OLAP gibi sorgulama yöntemleri ile sorgulanamayan ya da tanımlı bir modele uyarlanamayan ses, görüntü gibi akan verileri ifade etmek için kullanılır [29].

2.1.1 Metin Madenciliği Süreci

Metin madenciliği süreci temelde altı adıma ayrılabilir. Bunlar analiz edilecek metinlere erişim, metin ön işleme, metin dönüştürme, özellik seçimi, örüntü keşfi (veri madenciliği), sonuçların yorumlanması aşamalarıdır. Bu aşamalar Şekil 2.1 de gösterilmiş ve aşağıda açıklamalarına değinilmiştir [30].



Şekil 2.1: Metin madenciliği süreci.

1. Metinleri Bir Araya Getirme (Veriye Erişim)

Veri madenciliğinin ilk aşaması olarak analiz edilecek metinler toplanarak kullanılmak üzere kaydedilir. Dokümanlar önceden hazırlanmış olabilir, bir veri kaynağından alınabilir, problemin parçası olarak karşımıza çıkabilir ya da internet ortamından da toplanabilir [31].

2. Doküman Standardizasyonu

Toplanan dokümanların formatlarının farklılık gösterdiği veri kümelerinin olması durumunda metin formatlarının standartlaştırılması gerekir. Standartlaştırma işlemi metin madenciliği sürecinin verimli hale getirilebilmesi için önişleme aşamasında çeşitli kurallar

ve düzenlemeler tanımlayarak yapılır. Gerçek hayat veri kümelerinin elemanları Word, basit metin, resim vs. farklı formatlarda kaydedilmiş olabilir. Bu veri kümelerinde veri madenciliği yöntemlerini uygulayabilmek için metinler; CSV, XML, ARFF gibi formatlara dönüştürülmelidir [31].

3. Metin Ön İşleme

Farklı alan ve uygulamalardaki sonuçlara göre ön işleme toplam sürecin %80' ini kapsayabilmesi nedeniyle çok önemli bir safha olarak değerlendirilmektedir [32]. Metin verilerinin önceden işlenmesinde bazı özel hususların mevcudiyeti söz konusudur. Metinler kelimelerden, özel karakterlerden ve yapısal bilgilerden oluşmaktadır. Hangi ön işleme adımlarının gerektiği büyük ölçüde sonuçların ne amaçla kullanılacağına bağlı olarak değişmektedir. Genelde veriler, özel karakterler ve yapısal bilgiler (SGML etiketleri gibi) sembollerle değiştirilerek homojenleştirilmektedir. Noktalama işaretleri ve yapısal bilgilerin genellikle ayrı ayrı ele alınması gerekmektedir. Tez çalışmasında python dilinde yazılan bir kod yardımı ile noktalama işaretlerinin temizlenmesi, tüm harflerin küçük harfe dönüştürülmesi, boşluklar temel alınarak metinlerin tokenleştirilmesi aşamaları ön işleme adımları kapsamında gerçekleştirilmiştir.

Önişleme, doğal dil analizini de içerebilmektedir. Morfolojik analiz özellik vektörüne dahil edilebilecek veriler hakkında ayrıntılı bilgi verebilmektedir. Bu analiz, sözcükleri konuşma bölümleriyle değiştirerek verileri genelleştirmek için kullanılabileceği gibi belirli kelimelerin kombinasyonları yerine edat, isim gibi yapıları tanımlamakta da kullanılabilmektedir [31].

4. Verilerin filtrelenmesi

Sürecin bu noktasında keşif aşamasına odaklanılmakta, sonuç sayısını sınırlayarak gereksiz özellikler veri kümesinden ayıklamakta ve keşif aşamasında işleme için gerekli eforu azaltmakta kullanılmaktadır. Veri temizleme, keşif aşamasından önce veya sonra yapılabilmektedir. Genelde önerilen ne tür düzenler aradığımız konusunda net bir fikrin olmadığı durumlarda, temizleme işleminin ön işleme aşamasında daha sade tutulup post-proses aşamasında detaylandırılması yönündedir. Kullanılacak ön işleme adımları ve hangi

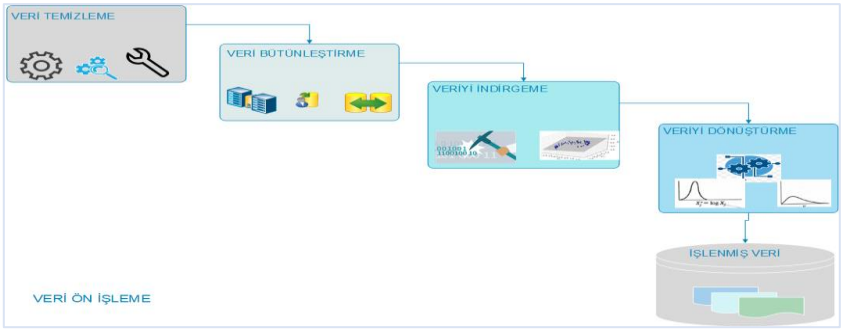
aşamada yoğunlaştırılacağı belirlenirken, verimlilik anlamında arama alanı ve keşif aşamasında ihtiyaç duyulan zamanın sınırlandırılmasında koleksiyonların boyutu da önemli bir konumundadır [33].

5. Metinlerde Özellik Seçimi

Bu aşama metin madenciliğinin en önemli aşamasıdır. Burada metinler için belirleyici olan önemli kelimeler ayrılır ve kümeye dahil edilir. Gürültülü veriler (önemsiz kelimeler) veri kümesinden ayıklanır. Bu basamakta yapısal olmayan veriler yapısal verilere dönüştürülmüş olur [34].

6. Veri Madenciliği

Yaşadığımız bilgi çağında gerek gerçek gerekse tüzel kişiler büyüt boyutlarda veri üretmekte, geleneksel istatistiksel yöntemler ise bu verileri analiz etmede yetersiz kalmaktadır. Geleneksel istatistiksel yaklaşım probleme karşı bir hipotez ortaya atar ve çeşitli analizler ile bu hipotezin doğruluğu veya yanlışlığını ispatlamayı amaçlar. Bu yaklaşım veri madenciliğinin temelini oluşturmak ile birlikte veri madenciliği, veriden gizli bilgilerin ortaya çıkarılmasını amaçlaması bakımından farklılaşır. Veri madenciliği ile kaynağından alınan ham veriler ön işleme aşamasından geçirilir. Eğer veri ses, görüntü, metin gibi yapılandırılmamış veriler ise ayrıca bilgisayarlarca işlenebilecek hale dönüştürülür yani sayısallaştırılır. Bu ön işleme aşamaları; eksik, hatalı, uygunsuz verilerin tespit edildiği ve çeşitli yöntemler ile bu eksiklik ve hataların giderildiği veri temizleme aşaması, varsa farklı veri kaynaklarından alınmış farklı yapılara sahip verilerin aynı türe dönüştürülmesi anlamına gelen veri bütünleştirme aşaması, veri setinde bazı verilerin amaç doğrultusunda analizin başarısını etkilemeyecek ya da artıracak şekilde çıkarılması anlamına gelen veri indirgeme, değişkenlerin ortalama ve varyanslarında büyük farklar olması durumunda normalize ya da standardize edilmesini içeren veri dönüştürme aşamaları olarak özetlenebilir [35]. Temel veri ön işleme aşamaları Şekil 2.2 de gösterildiği gibidir.



Şekil 2.2: Veri ön işleme adımları.

İşlenmiş verilere veri madenciliği yöntemleri uygulanarak bilgi çıkarımı gerçekleştirilir. Veri madenciliğinde uygulanan temel yöntemleri sınıflandırma, kümeleme, birliktelik kuralı çıkarma olarak üçe ayırabiliriz. Bu yöntemler çeşitli matematiksel ve istatistiksel algoritmalar vasıtası ile sonuçlar üretir.

Metin madenciliğinde, veri madenciliği yöntemleri kullanılmaktadır. Buradaki fark temelde veri kümesinin metinlerden oluşması ve ön işleme aşamasında dokümanların kelimelere ayrılması ve bu kelimelerin sayısallaştırma yöntemleri kullanılarak makinelerin işleyebileceği hale getirilmesi adımlarının eklenmesidir [30] [36].

7. Değerlendirme ve Yorum

Bu aşama bilgi keşfinin gerçekleştiği aşamadır. Veri madenciliği aşamasında ortaya çıkan sonuçlar değerlendirilerek yorumlanır, kullanıcıya anlaşılır ve uygun bir biçimde sunulur [34].

2.1.2 Metin Madenciliği Metotları

Metin madenciliği metotları bilgiye erişim (Information Retrieval) temelli metotlar ve bilgi çıkarımı (Information Extraction) temelli metotlar olarak ikiye ayrılabilirler. Bilgiye erişim temelli metotlar; terim temelli metot (term based method (tbm)), ifade temelli metot (phrase based method (pbm)), kavram temelli metot (concept based method (cbm)), örüntü sınıflandırma metodu (pattern taxonomy method (ptm)) dur. Bilgi çıkarımı temelli metotlar ise bilgi çıkarımı, sınıflandırma (kategorizasyon), kümeleme, metin özetleme, bilgi görselleştirme olarak sayılabilir. Bu metotlara ilişkin özet bilgilere aşağıda yer verilmiştir.

Terim temelli metot (Term Based Method (TBM)): Terim kavramı metinlerdeki anlamlı kelimelere tekabül eder. Bu yöntemde kelimeler ağırlıklandırarak matematiksel hesaplamalar yapılır. Uzun zamandır kullanılan bir yöntem olduğundan iyi bilinmesi analizlerde avantaj sağlamaktadır. Bu yöntem makine öğrenmesi ve bilgi çıkarımı çalışmaları ile ortaya çıkmıştır [37].

İfade temelli yöntem (Phrase Based Method (PBM)): İfadeler terimlere göre daha fazla anlam ve bilgi taşır, belirsizlik de daha azdır. İfadeler terimlere göre daha düşük istatistiksel oranlara sahiptirler, metinlerde görülme sıklığı daha azdır, içlerinde çok fazla gürültülü-gereksiz veri bulundurmazlar. Bu yöntemde terimler yerine ifadelerin metinleri temsil etmesi söz konusudur [37].

Kavram temelli yöntem (Concept Based Method (CBM)): Konsept temelli analiz cümle ve doküman düzeyinde yapılır. İstatistiksel yöntemler metin madenciliği uygulamalarında kelime veya kelime gruplarının (ifadelerin) metinlerde geçme sıklığını dikkate alırken dokümanları dikkate almaz. Bu yaklaşımın dezavantajı bir dokümanda aynı sıklıkta bulunan iki terimin anlamsal katkısı aynı olmayabileceği gerçeğidir. Bu nedenle yeni bir model geliştirilmiştir. Yeni modelin üç bileşeni bulunmaktadır. Birinci bileşen cümlenin semantik yapısını analiz eder. İkinci bileşen kavramsal bir bilgi grafiği (ontological graph (COG)) oluşturur. Bu yöntem ile anlamsal yapılar ve bileşenler tanımlanabilir, iki bileşene bağlı üst kavramlar ayrılabilir, standart vektör uzay modelini kullanarak kelime (özellik) vektörleri oluşturulabilir. Konsept temelli yöntemler cümlenin anlamı açısından önemli kelimelerin tespitinde çok etkilidir. Yöntem doğal dil işleme prensiplerine dayanarak çalışır [37].

Örüntü sınıflandırma yöntemi (Pattern Taxonomy Method (PTM)): Dokümanlar örüntüler baz alınarak analiz edilir. Örüntüye dayalı sınıflandırma veri madenciliğinde uzun zamandır kullanılan birliktelik analizi, sıralı örüntü madenciliği gibi teknikler kullanılarak yapılabilir [37].

Bilgi çıkarımı: Bilgi ayıklama, bilgisayarın metin içindeki anahtar ifadeleri ve ilişkileri belirleyerek yapılandırılmamış metni çözümlemesi için ilk adımdır. Bunu yapabilmek için metinde önceden tanımlanmış dizileri aramak için örüntü eşleştirme işlemi kullanılır. Bilgi çıkarım işlemi tokenleştirme, adlandırılmış varlıkların tanımlanması, cümle segmentasyonu, konuşma bölümü (part-of -speech) etiketlemesini içerir. Öncelikle ifadeler ve cümleler ayrıştırılır ve anlamsal olarak yorumlanır, daha sonra girilmesi gereken bilgi parçaları veritabanına aktarılır. Metin madenciliği uygulamalarındaki zorluk yapılandırılmamış veriler ile çalışılmasıdır. Bilgi çıkarımı bu sorunu çözen yöntemdir [38].

Kategorize etme (Sınıflandırma):

Metin kategorizasyonu (veya metin sınıflandırması), doğal dilde yazılmış belgelerinin içeriğine göre önceden tanımlanmış kategorilere atanmasıdır. Metinlerin otomatik olarak önceden tanımlanmış kategorilere ayrılması (veya sınıflandırılması), 2000'li yılların başından beri belgelerin dijital biçimde kullanılabilirliğinin artması ve bunları organize etme ihtiyacı nedeniyle artan bir ilgi görmüştür. Programlamanın sınıflandırma aşamasında metinlere kelime torbaları gibi bakılır ve bilgi çıkarımı işlemlerine girilmez. Bu aşamada metinlerde geçen kelimeler sayılır ve önceden belirlenmiş kelime haznesine (sözlük) dayanılarak dar anlam, geniş anlam, eş anlam, ilgili terimler vs. ye bakılıp ilişkiler belirlenmektedir. Amaç bir veya birden fazla sınıfa ait olabilecek metinleri sabit bir sınıfa atamaktır. Sınıflandırmaya dayalı öğrenme gözetimli öğrenme çeşididir. Amaç bilinen örneklerden (etiketli belgelerden) sınıflandırıcıları öğrenebilmek ve sınıflandırmayı bilinmeyen örneklerde (etiketsiz belgelerde) otomatik olarak yaptırabilmektir [39].

Sınıflandırma, serbest metin belgesine otomatik olarak bir veya daha fazla kategori atar. Kategorize etme, yeni belgeleri sınıflandırmak için girdi çıktı örneklerine dayandığı için denetimli öğrenme yöntemidir. Önceden tanımlanmış sınıflar, metin belgelerine içeriklerine göre atanır. Tipik metin sınıflandırma süreci ön işleme, indeksleme, boyutsal küçültme ve sınıflandırmadan oluşmaktadır. Sınıflandırmanın amacı sınıflandırıcıyı bilinen örneklere göre eğiterek daha sonra bilinmeyen örneklerin otomatik olarak kategorize edilmesidir. Naïve Bayesian sınıflandırıcı, En Yakın Komşu sınıflandırıcı, Karar Ağacı ve Destek Vektör Makineleri gibi istatistiksel sınıflandırma teknikleri, metni kategorilere

ayırmak için kullanılabilir. Otomatik sınıflandırma yaklaşımının temel bileşenleri, kategori çıkarma işlemi ve parametre seçim süreci olmak üzere iki işlemden oluşur [38].

Kümeleme: Benzer içeriğe sahip belge gruplarını bulmak için kümeleme yöntemi kullanılabilir. Kümelenmenin sonucu tipik olarak P kümeleri adı verilen bir bölümdür ve her küme bir dizi belgeden oluşur. Aynı kümedeki belgelerin içeriği daha benzerdir ve kümeler arasında kümelenmenin kalitesi daha farklıdır. Kümeleme tekniği benzer belgeleri gruplamak için kullanılmasına rağmen, kümeleme belgelerinde önceden tanımlanmış konuların kullanımı yerine anında kümelenmiş olduğu için sınıflandırmadan farklılaşmaktadır. Belgeler çoklu alt başlıklarda görünebileceğinden kümeleme, yararlı bir belgenin arama sonuçlarından çıkarılmasını sağlamaktadır.

Veri madenciliğinde K-araçları sık kullanılan kümeleme algoritmasıdır. Bu algoritmalar ile metin madenciliği alanında da iyi sonuçlar elde edilebilmektedir. Temel bir kümeleme algoritması, her belge için bir konu vektörü oluşturur ve belgenin her kümeye ne kadar iyi uyduğunun ağırlığını ölçer. Yönetim bilgi sistemlerinin organizasyonunda küme teknolojisi kullanılmaktadır [37].

Metin özetleme: Metnin temel anlamını ve önemli noktalarını koruyarak uzunluğunu ve detayını azaltmaktadır. Kullanıcı uzun bir belgenin ihtiyacını karşılayıp karşılamadığını anlamak için özeti okuma yolunu seçebilmektedir. Bazı durumlarda özet, belge kümesinin yerini alabilmektedir. Bilgisayarlar yerleri, insanları, zamanı tanımakta gayet başarılı iken anlamları kavramada zorlanmaktadır. Özetleme temelde üç aşamadan geçer. Bunlar:

1) Ön işleme aşaması ile orijinal metnin yapılandırılmış bir temsili elde edilir.

2) Özet yapıyı üretebilmek için algoritma uygulanır.

3) Özet yapıdan özetin son haline ulaşılır [38].

Bilgi görselleştirme: Metin madenciliği uygulamalarında görselleştirme ilişkili bilgilerin keşfini artırılabilir veya kolaylaştırılabilir. Tek tek belgeleri veya belge gruplarını temsil

etmek için belge kategorisini göstermek ve yoğunluk renklerini göstermek için metin bayrakları kullanılır. Görsel metin madenciliği büyük metin kaynaklarını görsel olarak hiyerarşik bir yapıya dönüştürür. Kullanıcı yakınlaştırma ve ölçeklendirme ile belgeyle birebir etkileşime geçebilmektedir. Bilgi görselleştirme, terörist ağları tanımlamak veya suçlar hakkında bilgi bulmak için hükümete uygulamalarına kullanılmaktadır. Bilgi görselleştirmenin amacı üç adıma ayrılmıştır:

1) Veri hazırlama adımı, görselleştirmenin orijinal verilerinin kararlaştırılmasından, elde edilmesinden ve orijinal veri alanını oluşturulmasından ibarettir.

2) Veri alanı oluşturulması adımı orijinal verilerden gerekli görselleştirme verilerinin analiz edilmesi, çıkarılması ve görselleştirmesini içermektedir. Veri analizi ve ekstraksiyonu olarak bilinir.

3) Görselleştirme eşleme, görselleştirme veri alanını görselleştirme hedefine eşlemek için belirli eşleme algoritmalarının kullanıldığı adımdır [37].

2.1.3 Metin Madenciliğinde Doküman Sayısallaştırma Yöntemleri

Metinsel verilerin algoritmalar tarafından işlenebilir hale gelebilmesi için öncelikle sayısallaştırılmaları gerekmektedir. Sayısallaşma için farklı sayısallaştırma ve ağırlıklandırma yöntemleri kullanılmak ile birlikte burada tez kapsamında kullanılan yöntemler olan kelime torbaları (bag of words), TF-IDF ve doc2vec' e değinilmektedir.

Eğitim setindeki bir belgenin T ile temsil edildiği ve C ise T ve C metinlerinin sınıflarını temsil ettiği durumda terim vektörü T:

$$T = (t_1, t_2, \dots, tp) \tag{2.1}$$

Burada 'p', koleksiyonun metindeki benzersiz terimlerin toplam sayısı ve 't_i', 'i' teriminin belgeyi karakterize etmek için göreceli önemini yansıtan ağırlıktır. Benzer şekilde C, belgeye atanan kategorileri temsil eden bir vektörü temsil etmektedir.

$$C = (c_1, c_2, \cdots cq)$$
(2.2)

‘q’ kategori sayısını ‘c_i’ ise 'i' kategorisinin önemini temsil etmektedir.

T ve C vektörleri için bir dizi çeşitli ağırlıklandırma yöntemleri kullanılabilir. Terim sıklığı (term frequency) ve ters doküman sıklığı (inverse document frequency) bunlara örnektir.

Terim sıklığı, bir terimin belgede geçme sayısı olarak tanımlanabilir. Ters belge sıklığı, belge koleksiyonundaki nadir terimlerin analizlere dahil edilmesine olanak sağlayan bir ağırlıklandırma yöntemidir.

‘S’ ile kategorize edilmesi gereken bir belgenin gösterilmesi durumunda ‘S’ i temsil eden vektör:

$$S = (s_1, s_2, \ldots, s_P) \quad (2.3)$$

burada ‘s_i’, ‘S ' deki i terimi için ağırlıktır.

‘S’ belgesi, bir benzerlik fonksiyonuna göre eğitim koleksiyonundaki her bir örneğe (yani belgeye) eşleştirilir. Bu işlev, her eğitim belgesi örneği için bir puan oluşturur. Puan ne kadar yüksek olursa, S ile belge örneği arasındaki benzerlik de o kadar yüksek olur.

$$\Delta(S, T) = \sum_{i=1}^{p} s_i t_i \quad (2.4)$$

Hesaplanan benzerlik değerlerine göre kategoriler tanımlanır [40].

Metin Analizi, sınıflandırma algoritmaları için önemli bir uygulama alanıdır. Bununla birlikte, ham veriler, bir dizi sembol sembolden oluşmaktadır ve algoritmalar doğrudan bu semboller ile beslenemezler, çünkü algoritmalar; çoğu değişken uzunluktaki ham metin belgelerinden ziyade sabit boyutlu sayısal özellik vektörleri işleyebilirler. Metin özelliklerinin sayısal özelliklere dönüştürülmesi için tokenleştirme, sayma ve normalizasyon adımları gerçekleştirilerek sayısallaştırma yapılır. Sayısallaştırma, bir metin belgesi koleksiyonunu sayısal özellik vektörlerine dönüştürme genel sürecidir. Bu özel stratejiye (tokenleştirme, sayma ve normalleştirme) Kelime Torbası veya "n-gram Torba" temsili denir. Bag of words doğal dil işleme ve bilgi çıkarımında kullanılan bir yoldur. Bu modelde belgeler, dilbilgisi kelime sırası gibi özellikler göz ardı edilerek sadece kelimelerin metinde geçme sıklığını korunmaktadır. Belgelerin çoğunluğunda korpusta kullanılan kelimelerin çok küçük bir alt kümesini kullanacağından, elde edilen matris birçok özellik değerine sahip ve sıfırlardan oluşan sparse matris olmaktadır. Aşağıda iki metinden oluşan bir veri kümesinin bag of words yöntemi ile sayısallaştırılma örneğine yer verilmiştir.

["Ali flim izlemeyi sever, çilek yemeyi sever"]

Tablo 2.1: Bag of words yöntemi sayısallaştırılma örneği.

Cümle	ali	çilek	film	izlemeyi	sever	yemeyi
0	1	1	1	1	2	1

Bow1= {"ali":1, "filim":1, "izlemeyi":1, "sever":2, "çilek":1, "yemeyi":1}

Birinci metinde sever kelimesi iki kez geçtiği için 2 değerini almış diğer kelimeler bir kez geçtiği için 1 değerini almıştırlar.

["Ayşe muz sever "]

Tablo 2.2: Bag of words ile kelimelerin sayılma yöntemi.

Cümle	ayşe	muz	sever
1	1	1	1

Bow2= {"ayşe":1, "muz":1, "sever":1}

Tablo 2.3: Bag of words yöntemi iki cümlenin kelime değerleri.

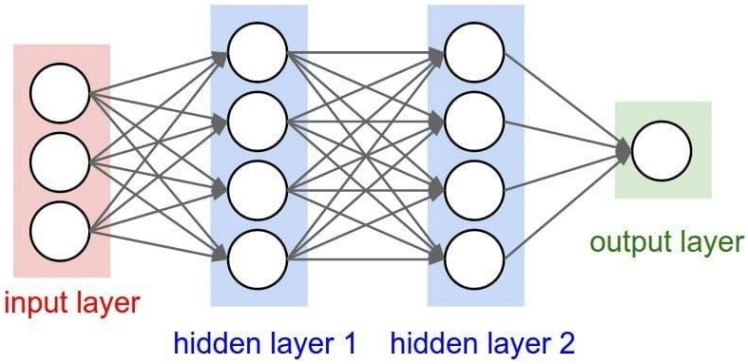
Cümle	ali	ayşe	çilek	film	izlemeyi	muz	Sever	yemeyi
0	1	0	1	1	2	0	2	1
1	0	1	0	0	0	1	1	0

$$Bow3 = Bow1 \cup Bow2$$

Bow3= {“ali”:1, “ayşe”:1, “çilek”:1, “filim”:1, “izlemeyi”:1, “muz:” 0, “sever”:3, “yemeyi”:1}

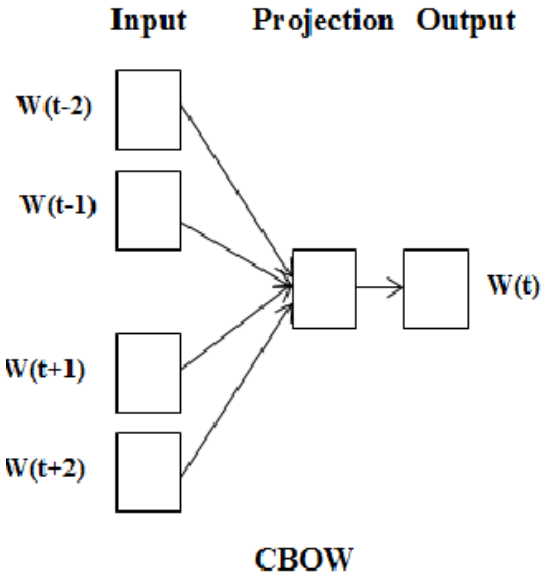
Bununla birlikte, terim frekansları metnin en iyi temsili olmayabilir. "ve", "ile", "gibi" vb. kelimeler metinlerde en sık kullanılan ve metnin analizinde faydalı olmayan durma terimleridir. Dolayısıyla, kelimenin metinde sık görülmesi, kelimenin daha önemli olduğu anlamına gelmez. Bu sorun için en yaygın yöntemlerinden biri normalleştirme yöntemi olarak kelimeleri ters doküman sıklığına (tf–idf) göre ağırlıklandırılmasıdır. Ayrıca, sınıflandırmanın özel amacı için, bir belgenin sınıf etiketini dikkate almak üzere denetimli alternatifler geliştirilmiştir. Hatta bazı problemler için frekanslar yerine n-gram ağırlıklandırma kullanılabilmektedir [41].

Metni doğru analiz edebilmek için vektörlerin kelimeyi doğru temsil etmesi önemli bir etkendir. Bu amaçla 2013 yılında Google araştırmacısı Tomas Mikolov ve ekibi tarafından ‘word embedding’,’word2vec’ modeli geliştirilmiştir. Mikolov’ a göre Word2vec ile elde edilen vektörlerin, doğal dildeki kelimelerin söz dizimi ve anlambilim ilişkilerine benzer şekilde kümelenmektedirler. Bu model CBOW ve skip-gram olmak üzere iki teknik kullanmaktadır. Model kelimelerin girdi olarak alındığı ve gizli katmanda işlenerek bir çıktıya dönüştürüldüğü bir yapay sinir ağı gibi çalışmaktadır. Genel bir yapay sinir ağlarının işleme mantığı Şekil 2.3’ de verilmiştir. Toplam 100 kelimelik bir kez geçen kelime sayısının 70 olduğu bir veri setinde 100 kelimenin her biri için 70 er boyutlu one-hot vektörler oluşturulmaktadır. Daha sonra kullanıcı tarafından belirlenecek pencere boyutu parametresine göre kelime girdileri modele alınarak çıktı üretmektedir.



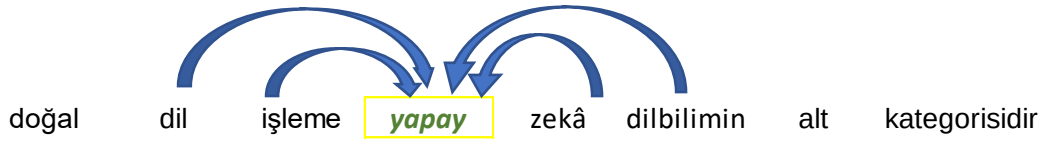
Şekil 2.3: Yapay sinir ağı modeli [42].

CBOW modeli, belirli bir pencerede kaynak kelimeleri verilen geçerli kelimeyi tahmin eder. Giriş katmanı kaynak kelimelerini ve çıkış katmanı geçerli kelimeyi içerir. Gizli katman, çıktı katmanında bulunan geçerli sözcüğü temsil etmek istediğimiz boyut sayısını içerir. CBOW yönteminin çalışma sürecine Şekil 2.4 de yer verilmiştir.

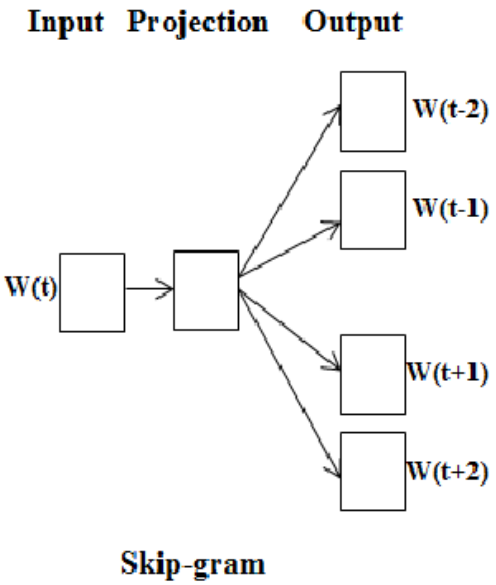


Şekil 2.4: CBOW modelinin çalışma yöntemi [43].

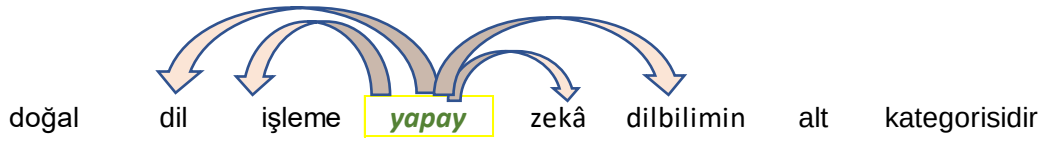
"Doğal dil işleme yapay zekâ ve dilbilimin alt kategorisidir." cümlesi CBOW yöntemiyle incelediğinde, pencere boyutunun iki olduğu varsayımı altında yapay kelimesinin tahmininde sağdaki ('dil', 'işleme') ve soldaki ('zeka', 'dilbilim') iki kelime girdi olarak sisteme dahil edilir ve çıktı olarak 'yapay' kelimesinin dönmesi beklenmektedir. Kelimenin sağında veya solunda pencere boyutu kadar kelime olmaması durumunda mevcut olmayan değerler sıfır alır.



Skip gram, geçerli sözcük verildiğinde belirli bir pencerede çevreleyen bağlam sözcüklerini tahmin eder. Giriş katmanı geçerli kelimeyi ve çıkış katmanı bağlam kelimelerini içerir. Gizli katman, giriş katmanında bulunan geçerli sözcüğü temsil etmek istediğimiz boyut sayısını içerir. Skip gram yönteminin çalışma sürecine Şekil 2.5 de yer verilmiştir.



Şekil 2.5: Skip-gram yönteminin çalışma mantığı [43].



Skip gram yöntemi ise sisteme girdi olarak ‘yapay’ kelimesi verilerek çıktıda sağındaki (‘işleme’, ‘dil’) ve solundaki (‘zeka’, ‘dilbiliminin’) kelimelerin elde edilmesi şeklinde çalışmaktadır [1], [2], [3].

2.2 Makine Öğrenmesi ve Metin Sınıflandırma

Makine öğrenmesi bilgisayarların bilgiden öğrenebilmesi fikrine dayanır. Geleneksel programlama yöntemlerinde bilgisayara bir komutun gerçekleştirilebilmesi için 'if-then-else' yapılarından oluşan çok sayıda emir girilmesi gerekmekteydi. Ama bu yöntem karmaşık problemlerin çözümünde efektif değildi ve yapay zekâ ile ilgilenen araştırmacılar makinelerin verilerden öğrenip öğrenemeyeceğini sorgulamaya başladı. Böylece makine öğrenmesi yöntemleri geliştirildi. Bu yöntemler güvenilir, tekrarlanabilir karalar ve sonuçlar üretebilmek için geçmiş verilerden öğrenme gerçekleştirerek yeni veriler oluşturmaktadır. Ayrıca farklı verilere de uyum sağlayabilmesi yönüyle de önem arz etmektedir. Makine öğrenmesi algoritmaları ile gerçekleştirilen güncel çalışmalara sürücüsüz araba projesi, Amazon ve Netflix tarafından kullanılan online tavsiye telifleri, müşteri yorumlarının değerlendirilmesi, yolsuzlukların tespiti vs. örnek verilebilir. Makine öğrenmesi yöntemleri finans sektörü, sağlık sektörü, kamu sektörü, taşımacılık, sosyal güvenlik sistemleri, yakıt, enerji sektörü gibi birçok sektör ve alanda bilgi çıkarımı, isabetli karar alma amaçları ile kullanılmaktadır.

Birçok makine öğrenmesi metodu bulunmak ile birlikte fazlaca kullanılanlar gözetimli öğrenme (supervised learning), gözetimsiz öğrenme (Unsupervised learning), yarı denetimli öğrenme (semisupervised learning), takviyeli öğrenme (reinforcement learning) yöntemleridir.

Gözetimli Öğrenme: Etiketli geçmiş verilere bakarak öğrenme gerçekleştirir ve yeni veriler için tahmin üretir. Gözetimli öğrenmede etiketlenmiş eğitim veri setlerinden oluşan bir girdi değişken ve beklenen bir çıktı değişkeni mevcuttur. Eğitim verilerini analiz edebilmek amacı ile girdi ve çıktı değişkenlerini eşleştirecek bir algoritma kullanılır. Bu algoritma aynı zamanda yeni verilerin ön görülmesi amacı ile de kullanılır. Gözetimli öğrenme ile kredi kartı yolsuzluklarının belirlenmesi, müşteri talep tahmini, potansiyel müşterilerin belirlenmesi gibi çalışmalar yapılabilmektedir. Bu yöntem ile sınıflandırma analizleri, regresyon analizleri, öngörü (tahmin) analizleri yapılabilmektedir. Metinlerin sınıflandırılması da gözetimli öğrenme ile gerçekleştirilebilmektedir. Tez kapsamında gözetimli öğrenme yöntemi kullanılarak doküman sınıflandırılması yapılmıştır.

Gözetimsiz Öğrenme: Geçmiş verilere dayalı bir etiket olmadan bütün datayı tarayarak verilerde gizli seyrek bir ağaç veya grafik gibi bazı yapıları bulup bilgi keşfetmeyi amaçlar. Market kampanyalarında benzer davranan müşterileri belirleme, müşteri segmentasyonu gibi çalışmalarda sıkça kullanılır. Kendi kendini düzenleyen haritalar, en yakın komşu haritalama, k- means kümeleme gibi yöntemleri kullanır. Kümelemenin metin madenciliği konularında kullanımına metinlerin bölümlerini konulara göre ayrıma, öge önerme, uç verileri tanımlama gibi örnekler verilebilir. Bu yöntem kapsamında: kümeleme, tüm veri setini gruplara bölerek diğer gruplara göre birbirine en fazla benzeyen elemanların bir araya toplanması ile boyut indirgeme, daha anlamlı sonuçlar alabilmek için veri setinde analizin başarısını hiç etkilemeyecek veya nispeten az etkileyecek verilerin ayıklanarak datanın işlemleri gerçekleştirilebilir.

Yarı Gözetimli Öğrenme: Gözetimli öğrenme ile aynı uygulamaları kullanmak ile birlikte eğitim seti etiketli ve etiketsiz (genellikle çoğunluğu etiketsiz) verilerden oluşan öğrenme yöntemidir. Bu yöntem ile sınıflandırma, regresyon, öngörü analizleri yapılabilmektedir. Bu öğrenme yöntemi yüz tanıma uygulamalarında kullanılmaktadır.

Takviyeli Öğrenme: Öğrenenini/kullanıcının davranışlarını analiz ederek deneme yanılma yöntemiyle öğrenenin kullanıcının en fazla kazancı (ödülü) elde edeceği seçeneği öğrenmektir. Algoritma hangi eylemin gerçekleştirileceğini söylemek yerine farklı senaryoları deneyerek en fazla kazancı sağlayacak seçeneği belirler. Amaç öğreneni/kullanıcıyı hedefe en kısa sürede ulaştırmaktır. Deneme yanılma ve ödül sistemini kullanması öğrenme yöntemini diğerlerinden ayırmaktadır. Bu yöntem robot, oyun, navigasyon teknolojilerinde kullanılmaktadır [4].

3. MOBBING (PSİKOLOJİK TERÖR) KAVRAMI VE YARGITAY

Bu bölümde çalışma kapsamındaki metinlerin konusunu oluşturan mobbing kavramı incelenmiştir. Kavram; tanımları, tarihsel gelişimi, ulusal ve uluslararası hukuksal düzenlemeler bağlamlarında ele alınmıştır. Ayrıca çalışmanın içeriğinde bulunan karar metinlerine ilişkin bilgilere ve bazı istatistiklere de bu bölümün ilerleyen başlıklarında yer verilmiştir. Son olarak karar verici konumunda olan Yargıtay'a ilişkin genel bilgiler, organizasyon yapısı, iş akış şemaları eklenerek bölüm sonlandırılmıştır.

3.1 Mobbing Kavramı Tanımı ve Tarihsel Gelişimi

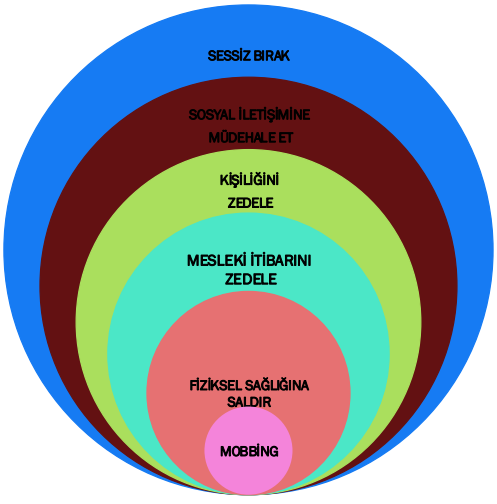
Mobbing kavramı Latince "kararsız kalabalık" anlamına gelen "mobile vurgus" sözcüklerinden türeyen "mob" sözcüğünün İngilizcede fiileştirmek için -ing ekini almasıyla oluşmuştur. Mobbing ise kelime anlamı itibarı ile "şiddet, kuşatma, taciz, rahatsız etme veya sıkıntı verme" anlamına gelmektedir [5]. Psikolojik taciz terimi ile literatürde 19. yüzyılda biyoloji ve etoloji alanında karşılaşılmaktadır. Bir etolojist olan hayvan gruplarının hareketlerini inceleyen Konrad Lorenz bir grup küçük hayvanın birleşerek tek bir büyük hayvana saldırmasını mobbing olarak tanımlamıştır [6]. Daha sonra İsveçli Doktor Peter Paul Heinemann çocukların ders saatlerinde birbirlerine ne yaptığını araştırırken terimi ödünç almıştır. Ardından ABD'li psikiyatr ve antropolog Carroll M Brodsky tarafından 1976 yılında "The Harassed Worker" isimli bir kitap kaleme almıştır. Kitap Kaliforniya İşçileri Tazminat İtirazları Kuruluna (California Workers' Compensation Appeals Board) ve Nevada Sanayi Komisyonuna; (Nevada Industrial Commission) başvuru sahipleri ile yapılan mülakatları içermektedir. Başvuru sahipleri işyerinde gördükleri çoğu vakada fiziksel olmamakla birlikte işverenlerin, iş arkadaşlarının ya da müşterilerin kötü muamele ya da aşırı iş üretilmesi yönündeki baskıları nedeni ile hasta ve çalışamaz hale geldiklerini bazı durumlarda sürekli veya total engellilik oluştuğunu iddia etmiştirler. Yazarın vardığı sonuca göre şiddet-taciz (harassment) bir kişi tarafından düşmanca ve saldırgan davranışlarla başka bir kişiye işkence etme, sinirini bozma, yıpratma değişik bir tepki almak amacıyla sürekli ve kalıcı girişimlerdir. Kişiyi günah keçisi çıkarma, kötüye kullanma, kişiye iş baskısı uygulama şeklinde psikolojik boyutta olabileceği gibi fiziksel boyutta da olabilir [7].

Yıldırının işyerinde vuku bulmasına ilişkin çalışmaların bir diğeri Heinemann'ın çalışmalarındakine benzer davranışların işyerlerinde görülme durumunu araştıran

Leyhmann'ın çalışmalarıdır. İş yaşamında mobbing ya da psikolojik terör; düşmanca ve etik olmayan, sistematik bir şekilde bir kişi veya bir gruba (genelde bir bireye) onu çaresiz ya da savunmasız bir duruma düşüren süreklilik arz eden iletişim ve davranışlar olarak tanımlanmıştır. Bu davranışlar çok sık tekrarlanır (istatistiklere göre: haftada en az bir kez) ve uzun bir zaman dilimi boyunca (istatistiklere göre en az altı ay boyunca) meydana gelmektedir. Bu aşırı sıklık ve uzun periyotlu düşman tavırlar nedeniyle psikolojik, psikosomatik ve sosyal acı ile sonuçlanmaktadır. Yazara göre tanım gereği yıldırı geçici çatışmalar üzerinde durmaz yani mobbing ne yapıldığı ya da nasıl yapıldığına odaklanmaz yapılan şeyin sürekliliği ve ne kadar süredir devam ettiğine odaklanır. Ayrıca çalışmada Leyhmann psikolojik terör aktivitelerini beş grupta toplamıştır. Bunlar;

1. Kurbanların yeterli iletişim kurma olasılıklarını engellemek (yönetici iletişim kurma imkânı vermez, sesiz bırakılır, işler ile ilgili sözlü saldırı vs.)
2. Sosyal temas sürdürülmesine müdahale (iş arkadaşlarının kurban ile iletişim kurmasına yönetici tarafından izin verilmez hatta yasaklanır, izole edilir, diğerlerinden uzak bir yere yerleştirilir vs.)
3. Kurbanın itibarının zedelenmesi (dedikodu, küçük düşürme, etnik kimlik, hareket tarzı, ya da konuşma şekli ile alay konusu etme)
4. Kurbanın mesleki durumunu olumsuz etkilemek (hiçbir görev vermemek, yada anlamsız görevler vermek vb.)
5. Kurbanın fiziksel sağlığına saldırmak (tehlikeli işler verilmesi, fiziksel tehdit veya saldırı, aktif şekilde cinsel saldırı vb.) [56]

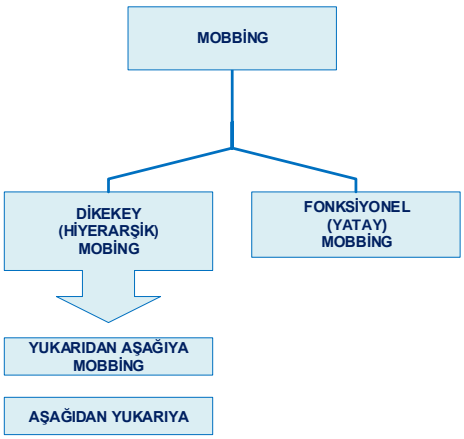
Leyhmann' ın yaptığı grup ayrımı Şekil 3.1' de gösterilmiştir.



Şekil 3.1: Leymann' a göre psikolojik terör aktivitelerine ait gruplar.

Yargıtay 22. Hukuk Dairesinin 22.05.2014 tarihli 2013/11788 esas sayılı 2014/14008 numaralı kararında mobbingi "Mobbing kavramının etimolojik anlamına ve tarihsel gelişimine bakıldığında; aynı ortamda bulunan veya aynı organizasyona bağlı olan bir veya birden fazla kimsenin bir kişiye belli bir amaçla, sistematik bir şekilde, yılgınlık, korku, tedirginlik, endişe, bunalım, bıkkınlık, sıkıntı veya kaygı oluşturacak söz, tutum veya davranışlarla psikolojik ve duygusal baskı kurarak onu belli şekilde davranmaya ya da davranmamaya, ortak alandan uzaklaştırmaya, güçsüzleştirmeye, değersizleştirmeye, aşağılamaya, küçük düşürmeye veya pasifize etmeye yönelik çabalarına mobbing denilir." şeklinde tanımlamıştır. Yine aynı karada mobbingle karıştırılabilecek bazı kavramlardan farkına da "Mobbingi; stres, tükenmişlik sendromu, işyeri kabalığı, iş tatminsizliği ya da doyumsuzluğu gibi olgulardan ayıran husus, belli kişinin belli bir amaca yönelik olarak hedef alınması, yapılan haksızlığın sürekli, sistematik ve sık oluşudur." ifadesi ile değinilmiştir [8].

Mobbingin farklı sınıflandırmaları yapılmaktadır. Şekil 3.2 de yer alan sınıflandırmada mobbing uygulayıcısının konumuna göre dikey (hiyerarşik) ve yatay (fonksiyonel) mobbing olarak ayrılması bu ayırımlardan biridir.



Şekil 3.2: Uygulayıcısına göre mobbing türleri.

İş ilişkisinde hiyerarşik olarak amir konumunda bulunanın/bulunanların ast konumunda bulunan/bulunanlara astı yıldırmak amacıyla uyguladığı sürekli sistematik her türlü kötü muamele yukarıdan aşağıya mobbingdir. Sert mizaçlı yöneticilerin katı ve kaba davranışları mobbing ile karıştırılmamalıdır. Zira bu durum literatürde işyeri kabalığı olarak tanımlanmaktadır. Mobbingin buradaki ayırt edici özelliği belirli bir veya birden fazla kişiyi hedef alarak uygulanması durumudur.

Aşağıdan yukarıya mobbing ise iş yerinde ast konumunda olanın/olanların üst konumunda olana/alanlara yıldırma amacıyla uyguladığı her türlü insan onuruna yakışmayan kötü muameledir. Fonksiyonel (yatay) mobbing ise aynı hiyerarşik aynı statüye sahip olanların uyguladığı psikolojik şiddettir [5].

3.2 Türk Hukuk Sisteminde Mobbing

Hukuk sistemimizde 2012 yılına kadar mobbing ile ilgili doğrudan bu düzenleme bulunmamaktaydı. 1 Temmuz 2012 tarihinde yürürlüğe giren 6098 sayılı Türk Borçlar Kanununu İşçinin Kişiliğinin korunması başlığında, madde 417' de kanun koyucu "İşveren, hizmet ilişkisinde işçinin kişiliğini korumak ve saygı göstermek ve işyerinde dürüstlük ilkelerine uygun bir düzeni sağlamakla, özellikle işçilerin *psikolojik ve cinsel tacize uğramamaları* ve bu tür tacizlere uğramış olanların daha fazla zarar görmemeleri için gerekli önlemleri almakla yükümlüdür." diyerek mobbinge ilişkin ilk doğrudan düzenlemeyi yapmıştır. Fakat doğrudan düzenleme olmaması mağdurların mahkemeye başvurması için bir engel değildir. Mobbing iddiası ile mahkemelerde kanunların genel hükümleri, kişilik haklarına saldırı hükümleri gibi hükümleri temel alınarak durdurma, tazminat, alacak davaları gibi davalar açarak hak talep edebilmiştir.

Anayasamızın eşitlik ilkesi kapsamında düzenlen 10. maddesi her Türk vatandaşının eşit haklara sahip olduğuna vurgu yapar. Bu madde mobbing davalarında kararlara kanuni temel teşkil etmiştir. Yine Anayasamızın kişinin korunmasına dair 17. maddesi, özel hayatın gizliliği, haberleşme hürriyeti, düşünce özgürlüğü, ekonomik ve sosyal haklar kapsamına giren genel düzenlemeleri mobbing ile ilişkilendirmiştir.

4857 sayılı iş kanununun içerdiği cinsel tacize ilişkin maddeleri, ahlak ve iyi niyet kurallarına uymayan haller başlıklı maddeleri, iş sağlığı ve güvenliği başlıklı maddeleri de genişletilerek mobbing davalarında karar aşamasında kullanılmıştır.

4721 sayılı Türk Medeni Kanunu kişiliğin üçüncü kişi veya kişiler tarafından saldırıya uğraması durumunda mobbing ile ilişkilendirilmektedir.

3.3 Mobbing Davalarının Genel İçeriği

Mağdurlar mobbing sebebi ile kişilik haklarına saldırı başlamamış fakat başlaması olası durumlarda önleme davası, sona ermiş saldırının zararı devam ediyor ise bu zararın tespit davası, kişilik haklarına saldırı devam ediyorsa, saldırının son erdirilmesi davası, zararın tazmini için maddi ve manevi tazminat davası açabilir.

Mobbingin davalara konu olan bir diğer yönü ise "iş sözleşmesini haklı nedenle derhal fesih hakkı" (4857 md.24 ve devamı) davalarına konu olmasıdır. Haklı nedenle fesih nedenleri 4857 sayılı kanunda açıkça tanımlanmamıştır ancak genel tanımlar yapılmıştır bu tanımlardan "Ahlak ve iyi niyete uymayan haller" nedeni mobbing ile ilişkilendirilebilmektedir. İşçi sözleşmesini haklı nedenle fesheder ise sözleşmeden kaynaklı kıdem tazminatı, fazla çalışma ücretleri, çalıştığı sürenin ücreti, ihbar tazminatı ödememe, var ise emeklilik hakkı gibi işçilik alacaklarını işverenden talep edebilir. Psikolojik taciz ahlak ve iyi niyete uymayan haller dolayısı ile haklı fesih nedenidir sözleşmeden doğan hakları talep etme imkânı sağlar.

Mobbing davalarında diğer sık karşılaşılan talep işe iade talepleridir. Mobbinge uğradığını iddia eden taraf iradesine fesat karıştırılarak istifaya zorlandığını ve istifa dilekçesini imzaladığını beyan etmiştir. Psikolojik taciz iddiasının ispatlanmasın durumunda haksız fesih de ispatlanmış demektir. İşe iade kararı ile sonuçlanır.

Tazminat davaları 4721 sayılı Medeni Kanun ve 6098 sayılı Türk Borçlar Kanunu kapsamında değerlendirilmektedir. Medeni Kanun, Birinci Kitap, Kişiler Hukuku, Kişilik başlığı altında düzenlenen 25. maddenin 3. fıkrası "Davacının, maddî ve manevî tazminat istemleri ile hukuka aykırı saldırı dolayısıyla elde edilmiş olan kazancın vekâletsiz iş görme hükümlerine göre kendisine verilmesine ilişkin istemde bulunma hakkı saklıdır." der. Türk Borçlar Kanunu İkinci Ayırım Haksız Fiillerden Doğan Borç İlişkileri başlığında düzenlenen 49. madde "Kusurlu ve hukuka aykırı bir fiille başkasına zarar veren, bu zararı gidermekle yükümlüdür. Zarar verici fiili yasaklayan bir hukuk kuralı bulunmasa bile, ahlaka aykırı bir fiille başkasına kasten zarar veren de bu zararı gidermekle yükümlüdür." şeklinde düzenlenmiştir. Aynı kanunun ikinci ayrım, Haksız

Fiillerden Doğan Borç İlişkileri, Tazminat, Kişilik Hakkının Zedelenmesi başlığı altında 58. maddede "Kişilik hakkının zedelenmesinden zarar gören, uğradığı manevi zarara karşılık manevi tazminat adı altında bir miktar para ödenmesini isteyebilir. Hâkim, bu tazminatın ödenmesi yerine, diğer bir giderim biçimi kararlaştırabilir veya bu tazminata ekleyebilir; özellikle saldırıyı kınayan bir karar verebilir ve bu kararın yayımlanmasına hükmedebilir." şeklinde bir düzenlemeye gidilmiştir.

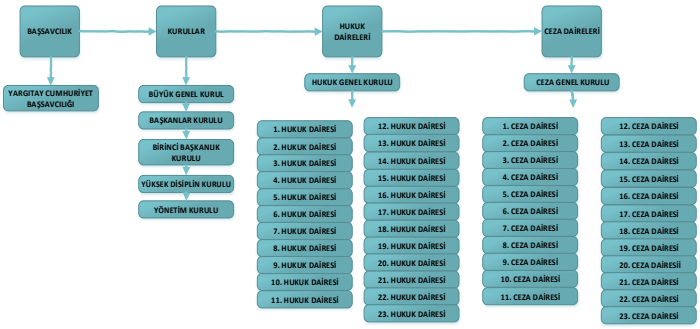
Yargı metinlerinde yaptığımız taramada sözleşmesini mobbing sebebiyle haklı nedenle feshettiğini iddia eden işçilik alacakların talep eden davacı işçi sayısı elimizdeki karaların %71' ini oluşturmaktadır. Davalar işçilik alacakları üzerine yoğunlaşmaktadır. Bu oranı %26 ile işe iade davaları takip etmektedir. Mobbing tazminatı talebi ile açılan dava sayısı oranı ise sadece %3 dür. Tazminat davalarının oranındaki düşüklüğün ispat güçlüğü ile mağdurlar ve avukatların mobbing konusunda yeterli deneyim ve bilgiye sahip olmaması nedeniyle olduğu düşünülmektedir.

3.4 Yargıtay'ın işleyişi

Tez kapsamında Yargıtay Hukuk Genel Kurulu ve Yargıtay Hukuk dairelerine intikal etmiş ve karara bağlanmış mobbing davaları incelendiğinden bu bölümde Yargıtay ile ilgili genel bilgilere yer verilmiştir. Bölüm kapsamında önce Yargıtay'ın organizasyon yapısı, organları, iş akışları ile ilgili genel ve çok özet bilgiler verilecektir. Bu bilgilerin çalışmanın kapsamı için gerekli olan kadarına yer verilmiş konuların tezin amacı açısından tüm yönleri ile verilmesi söz konusu değildir. Ayrıntılı bilgi için atıf ve kaynakçalara bakılabilir.

3.5 Çalışma Kapsamına Giren Yargıtay'a Ait Genel Bilgiler

10.01.1945 tarihinde 'Yargıtay' adını alan 6 Mart 1868' de "Divan-ı Ahkâm-ı Adliye" ismi ile kurulan bir yüksek mahkemedir. Anayasamızın Yüksek Mahkemeler başlığı altında 154. maddesinde "Yargıtay, adliye mahkemelerince verilen ve kanunun başka bir adli yargı merciine bırakmadığı karar ve hükümlerin son inceleme merciidir. Kanunla gösterilen belli davalara da ilk ve son derece mahkemesi olarak bakar." tanımı ile yer almaktadır. Yargıtay'ın karar organları Şekil 3.3'de görüldüğü gibidir.



Şekil 3.3: Yargıtayın karar organları.

Yargıtay'ın görevleri 2797 sayılı Yargıtay Kanununda şöyle sayılmıştır:

1. Adliye mahkemelerince verilen ve kanunun başka bir adli yargı merciine bırakmadığı karar ve hükümleri ilk ve son merci olarak inceleyip karara bağlamak,

2. Yargıtay Başkan ve üyeleri ile Yargıtay Cumhuriyet Başsavcısı, Yargıtay Cumhuriyet Başsavcı vekili ve özel kanunlarında belirtilen kimseler aleyhindeki görevden doğan tazminat davalarına ve bunların kişisel suçlarına ait ceza davalarına ve kanunlarda gösterilen diğer davalara ilk ve son derece mahkemesi olarak bakmak,

3. Kanunlarla verilen diğer işleri görmek.

Hukuk dairelerinin görev alanları Yargıtay Hukuk Daireleri, "Medeni Hukuk Daireleri", "Gayrimenkul Hukuku Daireleri", "Ticaret ve Borçlar Hukuku Daireleri", "İş ve Sosyal Güvenlik Hukuku Daireleri" olmak üzere dört "ihtisas alanı" altında toplanır. Ceza dairelerinin kapsamına ise Türk Ceza Kanunundan doğan davalara bakar.

Psikolojik terör davalarına Yargıtay iş bölümü kararlarına göre 9. Hukuk Dairesi ve 22. Hukuk Dairesi bakmaktadır. Çalışmamızda kullanılan kararlar bu dairelere aittir [9].

Hukuk Genel ve Ceza Genel Kurulları kanunda sayılan diğer görevlerinin yanı sıra Yargıtay dairelerinin bozma kararlarına karşı mahkemelerce verilen direnme kararlarını inceleyerek karar verme görevi vardır. Mobbing davalarında verilen yerel mahkemenin direnme kararını da Hukuk Genel Kurulu kesin karara bağlar.

3.6 Yargıtay Süreci ve İş Akışları

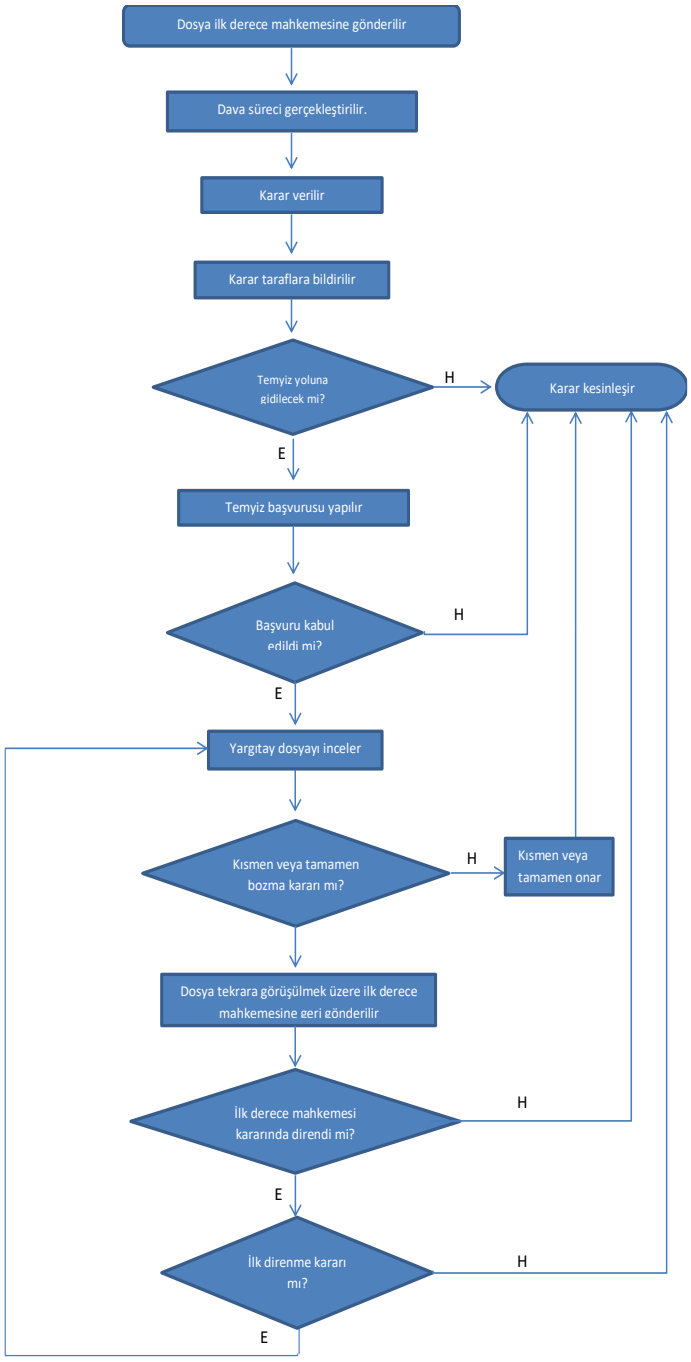
İlk Derece Mahkemeleri ve İstinaf Mahkemelerinin (Bölge Adliye Mahkemeleri (BAM)) bazı karar ve hükümleri doğrudan temyiz yolu ile bir üst mahkemeye gidebilir, bazıları için istinaf yoluna başvurulmadan temyiz yolu kapalıdır, bazılarının ise temyiz yolu tamamen kapalıdır. Temyize 6100 sayılı Hukuk Muhakemeleri kanununda temyiz kanun yolu, Temyiz başlığı altında 361. madde ve devamı maddelerde tanımlanmıştır. Temyize konu olabilecek ve olamayacak kararlar bu başlık altında tahdidi olarak sayılmıştır. İşyerinde psikolojik taciz davaları içeriğine ve kanuni tanımlara göre İstinaf Mahkemesine götürülebileceği gibi temyiz yolu ile Yargıtay'a götürülebilir.

Birinci Derece Mahkeme ve BAM kararlarından temyiz yolu açık olanların temyize götürülmesi ile ilgili süreç aşağıdaki şemada verilmiştir. Temyiz talebi için öncelikle kanunen kararı temyize götürmeye yetkili kişilerin bir dilekçe ile; kararı veren bölge adliye mahkemesi hukuk dairesine ya da Yargıtay'ın bozması üzerine hüküm veren ilk derece mahkemesine veya temyiz edenin bulunduğu yer bölge adliye mahkemesi hukuk dairesine ve de ilk derece mahkemesine başvuruda bulunması gerekir. Dilekçenin kabul edilmesi durumunda temyiz incelemesi başlar. Reddi durumunda ise ret kararına karşı temyiz yolu açıktır.

Sonraki süreçte dosyalar ilk olarak ön inceleme aşamasından geçer. Bu aşamada kanunda belirlenen dosyanın başka bir dairenin görev alanına girip girmediği, başvurunun süresinde yapılıp yapılmadığı, başvuru şartlarının yerine getirilip getirilmediği gibi davanın esasına girmeden anlaşılabilecek durumlar incelenir. Daha sonra dosyanın esastan incelenmesine geçilir. Gerek görülürse dava duruşmalı olarak yapılabilir. Aksi durumda dava dosya üzerinden incelenerek karara bağlanır.

Yargıtay kendisine intikal eden dosyaları kanunlara uygunluk bakımından inceler ve karar veya hükümleri yerinde bulursa mahkeme kararlarını onar. Karar veya hükümlerin kanuna aykırılık teşkil etmeyen basit maddi hatalarının olması durumunda ise kararın tamamen bozularak geldiği yer mahkemesine gönderilmesi yerine sözü geçen maddi hatayı düzelterek onama kararı verebilir.

Eğer kararlarda hukuka aykırılık görürse bozma yoluna gidebilir. Bozma tamamen olabileceği gibi kısmen de olabilir. Kısmen bozma durumunda Yargıtay'ın kanuna uygun bulmadığı kısmı tekrar görüşülmek üzere ilgili mahkemeye gönderilir. İlgili mahkeme Yargıtay kararına uyabileceği gibi kendi kararında direnebilir. Fakat bu direnme kararı bir kez verilebilir. Direnme kararları Yargıtay Hukuk Genel Kurulu tarafından karara bağlanır ve bu kararlar kesindir. Anlatılan hukuksal süreçlere ait akış şeması Şekil 3.5'deki gibidir [10].



Şekil 3.4: Mahkemelerin karar süreci.

4. UYGULAMA

Bu başlık altında başlangıçta tez kapsamında kullanılan makine öğrenmesi sınıflandırma algoritmaları Logistic Regresyon Algoritması, Gaussian Naive Bayes Algoritması, Karar Ağaçları (CART) Algoritması, K-En Yakın Komşu Algoritması, Destekçi Makineler Algoritması, Rassal Orman Algoritması, Bagging Classifier, Ada Boost Classifier, Gradient Boost Classifier ve MLP Classifier algoritmalarının hakkında genel bilgiler verilecektir.

Daha sonra tezde yer alan algoritmaların sınıflandırma başarısını ölçmek için seçilen doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skor (F1 score), ROC eğrisi yöntemlerinin hesaplanma şekilleri ve anlamlarına değinilecektir.

Sonra veri setine ilişkin genel bilgiler sunulacak, veri ön işleme ve görselleştirmelere değinilecek, uygulama kapsamında gerçekleştirilen adımlara ilişkin akış diyagramına yer verilecektir.

Ardından ölçme yöntemlerine göre test ve doğrulama setinde en yüksek başarıyı elde eden algoritma sonuçları sayısallaştırma metotları altında detaylandırılacaktır.

Son olarak tezde kullanılan on adet sınıflandırma algoritmasının başarılarına ilişkin ölçüm bilgileri tablolar ile özetlenecektir.

4.1 Tez Kapsamında Kullanılan Makine Öğrenmesi Algoritmaları

Makine Öğrenmesi yöntemi ile sınıflandırma yaparken pek çok algoritma kullanılabilmektedir. Fakat burada tez kapsamında kullanılan sınıflandırma algoritmalarına değinilecektir. Tez kapsamında on adet sınıflandırma algoritması kullanılmıştır. Bunlara ilişkin genel bilgilere aşağıda yer verilmiştir.

1. Logistic Regression Algoritması

Bağımlı değişkenini iki veya daha fazla düzeyli olduğu (0/1, Evet/Hayır, Doğru/Yanlış vs.) bağımsız değişken ile bağımlı değişken/değişkenler arasındaki ilişkiyi belirleyen bir sınıflandırma algoritmasıdır. Bu algoritmada bağımsız değişkeni bağımlı değişkene bağlayan logit, probit veya couchit bir link fonksiyonu bulunmaktadır. Verilere bu fonksiyonlar uygulanarak bağımlı değişkenin gerçekleşme olasılığı tahmin edilmektedir. Lojistik regresyonda gözlemlenen varyansın beklenen varyansdan büyük olması durum yaşanabilir. Buna aşırı yayılım (overdispersion) denmektedir [11] [12].

2. Gaussian Naive Bayes Algoritması

Tez uygulamasında Gaussian Naive Bayes algoritması, Python Sklearn kütüphanesinde bulunan ve Gauss ve Naive Bayes algoritmalarının birleştirilmesinden oluşmaktadır. Özelliklerin Gaussian olduğu varsayılır:

$$P(xi|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}}\exp(-\frac{(x_i\ \mu_y)^2}{2\sigma_y^2}) \quad (4.1)$$

$\boldsymbol{\sigma_y}$ ve μ_y değerleri maksimum olasılık kullanılarak tahmin edilmektedir [13].

3. Karar Ağaçları (CART) Algoritması

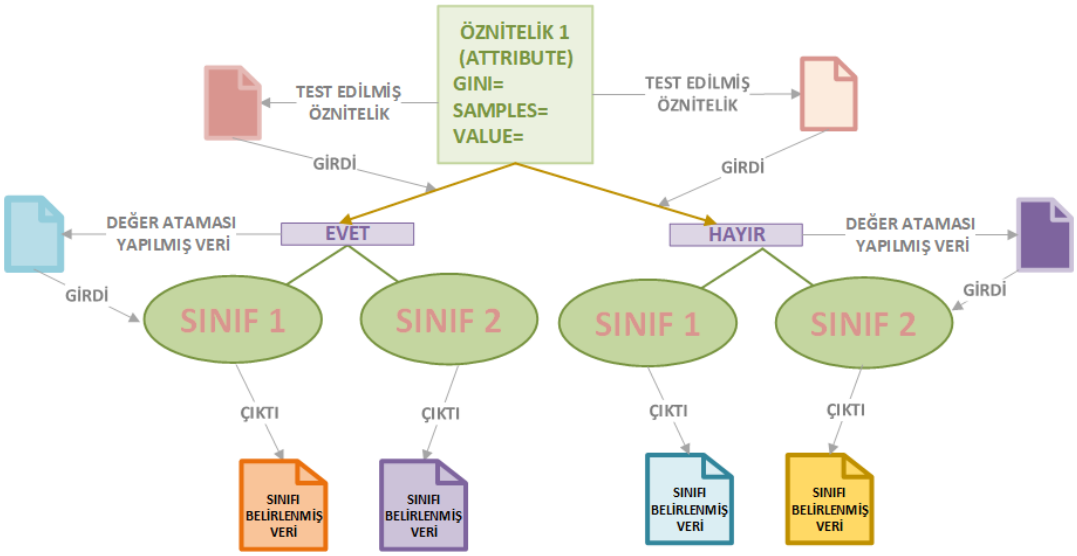
Karar Ağaçları (Decision Tree-DT' ler) sınıflandırma ve regresyon uygulamalarında kullanılan parametrik olmayan denetimli bir öğrenme yöntemidir. Amaç, veri özelliklerinden çıkarılan basit karar kurallarını öğrenerek bir hedef değişkenin değerini tahmin eden bir model oluşturmaktır. Çeşitli hareket şekilleri arasında seçim yapılması gereken problemler için kullanılmaktadırlar. Karar ağaçları kök düğüm, dal düğüm, yaprak düğüm olarak üç ana parçaya sahiplerdir.

Kök düğüm; örneğin öz niteliklerinin testini belirtir.

Dal düğüm; kök düğümden inen her bir dal düğüm kök nitelikteki özniteliğin olası değerine denk gelir.

Yaprak düğüm; örneğin sınıfını içerir.

Bir örnek, ağacın kök düğümünden başlayıp bu düğüm tarafından belirtilen özniteliği test ederek ve ardından verilen örnekteki öznitelik değerine karşılık gelen ağaç dalını aşağı kaydırarak sınıflandırılır. Karar ağaçlarının çalışma süreci Şekil 2.12' de gösterilmiştir.



Şekil 4.1: CART algoritmasının çalışma süreci.

Sınıflandırma ve regresyon ağaçları (CART, C&RT) 1984 yılında Breiman tarafından literatüre kazandırılmıştır. CART (Sınıflandırma ve Regresyon Ağaçları) C4.5'e çok benzer, ancak sayısal hedef değişkenleri (regresyon) desteklemesi ve kural kümelerini hesaplamaması bakımından farklılık gösterir. En iyi dallara ayrıma kriteri olarak entropiden yararlana algoritma ikili ağaçlar üreterek hem en uygun dallara ayırma değişkenini bulur hem de bu değişken ikiden fazla değer taşımakta ise hangi değerlere göre ikiden fazla değişkene ayrılacağını belirler. Herhangi bir t düğümündeki s dallara ayırma kriteri $\psi_{s/t}$ olarak gösterilir ise:

$$\psi_{\left(\frac{s}{t}\right)=} 2P_L P_R \sum_i^{\mu} |P(C_j|t_L) - P(C_j|t_R)| \tag{4.2}$$

t: dallanmanın yapılacağı düğüm

c: kriter

L: ağacın sol tarafı

R: ağacın sağ tarafı

$P_L P_R$: öğrenim kümesindeki bir kaydın sağda veya solda olma olasılığı

$P(C_j|t_L)$ ve $P(C_j|t_R)$: C_j sınıfındaki bir kaydın sağ veya solda olma olasılığı

Dallar ayırma işleminde veri kümesinde kayıp veri var ise algoritma bu kayıp verileri önemsemez. Hesaplanan $\psi_{s/t}$ değeri en yüksek nokta düğüm olarak seçilir ardından diğer işlemler karar ağaçları algoritmalarında olduğu gibi yapraklara ulaşana kadar tekrarlanır [14].

4. K-En Yakın Komşu Algoritması

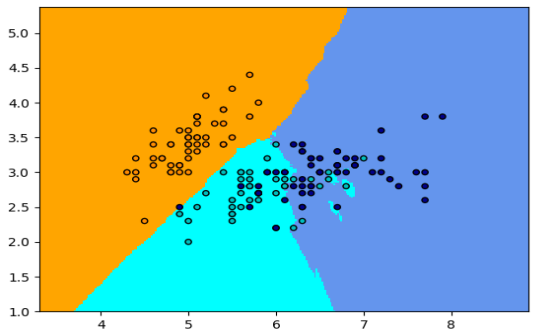
Veri kümesi içinde bulunan her bir öge için önceden kullanıcı tarafından ya da bazı hesaplama yöntemleri kullanılarak belirlenen bir k değeri kadar diğer ögelere olan mesafesinin ölçülmesi yoluyla hesaplanmaktadır. Gereğinden büyük bir k değeri belirlenmesi birbirinden farklı noktaların bir araya getirilmesi ile sonuçlanırken, küçük bir k değeri belirlenmesi benzer noktaların farklı alanlarda toplanması ile sonuçlanabilmektedir. Genel kabul gören k değeri 3.5 ve 7 dir. Bir örneğin en yakın komşuları standart Öklid mesafesi olarak tanımlanır. Herhangi bir x örneğinin

$$(a_1(x), a_2(x), \ldots . a_n(x)) \tag{4.3}$$

$a_{r(x)}$ özellik vektörü tarafından tanımlandığı varsayıldığında, burada rth, x özniteliğinin değerini ifade eder. d(x_i, x_j) arasındaki mesafe aşağıdaki şekilde hesaplanmaktadır.

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2} \tag{4.4}$$

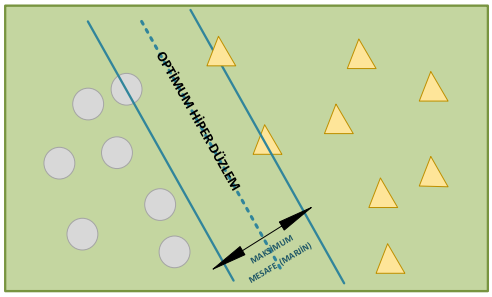
Aşağıdaki grafikte K-NN algoritmasının veriler üzerinde uygulandığında ortaya çıkan benzer verileri bir araya toplama şekli görselleştirilmiştir [15].



Şekil 4.2: K-NN algoritmasının verileri ayırma yöntemi [65].

5. Destekçi Makineler Algoritması

Bir destek vektör makinesi; sınıflandırma, regresyon veya diğer görevler için kullanılabilen, yüksek veya sonsuz boyutlu bir alanda bir hiper-düzlem veya bir dizi hiper-düzlem oluşturur. Sezgisel olarak, herhangi bir sınıfın en yakın eğitim veri noktalarına (fonksiyonel marj denir) en büyük mesafeye sahip olan hiper-düzlem ile iyi bir ayırma elde edilir, çünkü genel olarak marj ne kadar büyük olursa sınıflandırıcının genelleme hatası o kadar düşük olur. Destek vektör makinesi algoritmasının amacı, N-boyutlu bir uzayda (N-özellik sayısı) veri noktalarını ayrı ayrı sınıflandıran bir hiper-düzlem bulmaktır. İki veri noktasını ayırmak için birçok yöntem bulunmak ile birlikte her iki sınıfın veri noktaları arasındaki maksimum mesafeyi (marjı) sağlayabilecek düzlemi bulmaya odaklanılmaktadır. Veri noktalarını sınıflandırmaya yardımcı olan hiper-düzlemlerin sayısı özellik sayısına eşittir. SVM algoritmasının sonuca gidiş yöntemi Şekil 4.3'de verilmiştir [4].

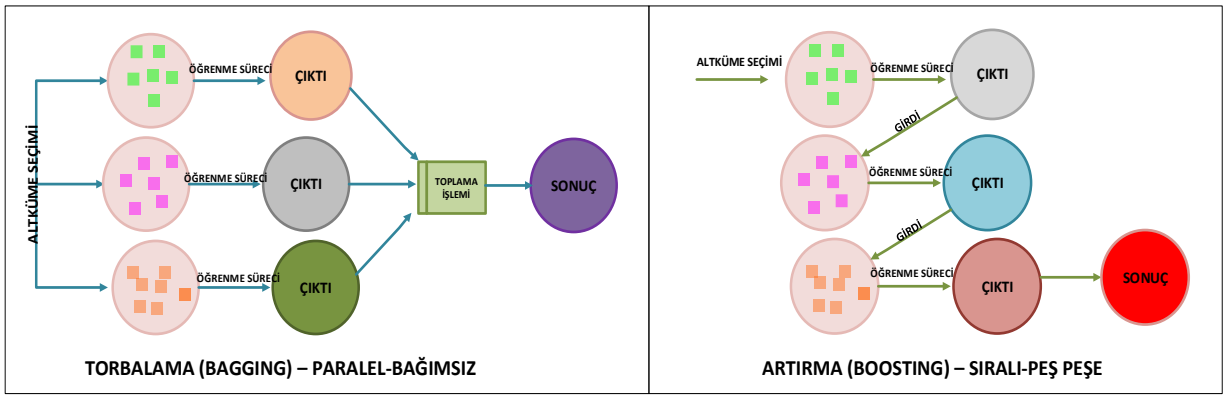


Şekil 4.3: SVM algoritması veri ayırma yöntemi.

6. Topluluk Öğrenmesi (Bagging, Rassal Orman, Ada-Boost, Gradient Boost)

Topluluklar, birkaç farklı sınıflandırıcıyı eğitip tek bir sınıf etiketi çıkarmak için kararlarını birleştirerek tek bir sınıflandırıcının doğruluğunu artırmak amacı ile tasarlanmışlardır. Bazı kaynaklarda 'uyum içinde birlikte öğrenme' [4] olarak da tanımlanan bu yöntem sınıflandırma ve regresyon problemlerinde daha iyi tahminlerde bulunabilmek amacıyla birden fazla makine öğrenmesi algoritmasının kullanılmasını ifade etmektedir. Topluluğa dayalı sınıflandırıcılar genelde benzer temel denklemlere sahip sınıflandırma algoritmalarından düşük varyanslı olanların bir araya getirilmesi ile oluşturulmaktadır. Algoritmaların rassal tahminlerinin ortalaması alınarak varyans azaltılır. Toplama yöntemler genelde iki sınıfa ayrılır. Bunlar torbalama yöntemleri (Bagging meta-estimator) ve artırma yöntemleri (Boosting methods) dir. Bagging, birçok bağımsız model inşa edip bazı model ortalama teknikleri kullanılarak modellerin birleştirildiği (ağırlıklı ortalama, çoğunluk oyu veya normal ortalama vs. yöntemler ile) basit bir toplama tekniğidir. Birçok

durumda torbalama yöntemi, altta yatan temel algoritmayı uyarlamaya gerek kalmadan, tek bir modele göre iyileştirmenin çok basit bir yolunu oluşturur. Artırma yöntemi ise modellerin tahminlerini bağımsız olarak değil hatalardan öğrenerek yaptığı varsayımına dayanılarak oluşturulan yöntemlerdir. Bu iki yöntemin çalışma farkı Şekilde 4.4' de özetlenmiştir.



Şekil 4.4: Topluluk öğrenmesi yöntemlerinin çalışma tekniği.

Torbalama yöntemleri birçok çeşide sahiptir, ancak genelde eğitim setinin rastgele alt kümelerini çizme yöntemleriyle ayrışırlar.

Topluluktaki her ağaç sisteminde rastgele bir alt küme seti her bir modelin girdisi olarak seçildiği ve topluluğun tahmini, bireysel sınıflandırıcıların ortalama tahmini olarak verildiği yöntemlere örnek olarak ***Rassal Orman Algoritması*** (Random Forest) verilebilir. Topluluktaki modeller birer karar ağacı oluşturur ve bunlar bir araya gelerek ormanları oluşturur.

Rassal olarak seçilen alt kümelerde özelliklerin önem kazanır. Bir ağaçta karar düğümü olarak kullanılan bir özelliğin göreli sırasının (yani derinliğinin), bu özelliğin hedef değişkenin öngörülebilirliğine göre nispi önemini değerlendirmek için kullanılabilir. Ağacın üstünde kullanılan özellikler, girdi örneklerinin daha büyük bir kısmının nihai tahmin kararına katkıda bulunduğu yöntemlerdir.

Verilerin denetlenmemiş dönüşümlerinin gerçekleştirildiği tamamen rassal ağaçların oluşturduğu bir orman kodlayarak, yüksek boyutlu ve seyrek bir ikili kodlamaya imkân veren bir k değerine göre öğrenmeye gidilen yöntemdir [16].

Torbalama sınıflandırıcısı (Bagging Classifier) Breiman tarafından toplulukları inşa etmek için toplama bootstrap kavramı ile tanıtılmıştır. Orijinal eğitim veri setinin bootstrap kopyaları ile farklı sınıflandırıcıların eğitiminden oluşturulmuştur. Yani, orijinal veri kümesinden (genellikle orijinal veri kümesi boyutunu koruyarak) örnekleri rassal seçerek (değiştirerek) her sınıflandırıcıyı eğitmek için yeni bir veri kümesi oluşturulmaktadır. Böylece, farklı veri alt kümeleri kullanılarak yeniden örnekleme prosedürü ile çeşitlilik elde edilmektedir. Son olarak, her bir sınıflandırıcıya bilinmeyen bir örnek sunulduğunda, sınıfı çıkarmak için çoğunluk veya ağırlıklı oy kullanılmaktadır.

Girdi:
S: Eğitim Seti, T:İterasyon sayısı, s: bootstrap boyutu, I: Zayıf öğrenci
Çıktı:
Uyarılmış sınıflar $h_t \in [-1,1]$ iken Bagged Classifier: $H_{(x)} = sign(\sum_{t=1}^{T} h_t(x))$
$\boldsymbol{for}\ t = 1\ to\ T\ do$
$S_t \leftarrow RandomSampleReplacement$
$h_t \leftarrow I(S_t)$
$\boldsymbol{end\ for}$

(4.5)

Artırma yöntemleri; 1990 yılında Schapire tarafından tanıtılan ve ARCing, uyarlanabilir yeniden örnekleme ve birleştirme olarak da bilinen yöntem zayıf bir öğrenciyi daha güçlü bir öğrenciye dönüştürülebiceği varsayımına dayanır. Bu yöntem hataların optimizasyonuna dayalı adaptif bir öğrenme yöntemi olarak da tanımlanabilmektedir. Zayıf tahmin ediciler kullanılarak tek bir sınıf değeri verecek şekilde modellenmektedir. Sırası gelen model kendinden bir önceki modelin hatalı sınıflandırdığı veriler üzerine uygulanarak kullanılırlar. Burada modeller birbirinden bağımsız değildir. Ada Boost Classifier ve Gradient Boosting Classifier bu yöntemi kullanan algoritmalardır.

Ada Boost, varyansın yanı sıra sapmayı da azaltan SVM de olduğu gibi marjları artıran bir algoritmadır. İlk başarılı boosting algoritması olarak prestijli Gödel ödülünü kazanmıştır. Ada Boost, her bir sınıflandırıcıyı seri olarak eğitmek için tüm veri kümesini kullanmakta ancak her turdan sonra mevcut yineleme sırasındaki bir sonraki tekrarda zor verilere daha

fazla odaklanmaktadır. Amaç hatalı sınıflandırılan verileri doğru sınıflandırmaktır. Ağırlıklar başlangıçta tüm örnekler için eşit olarak atanmaktadır. Her yinelemeden sonra, yanlış sınıflandırılan örneklerin ağırlıkları artırılır; doğru sınıflandırılmış örneklerin ağırlıkları azaltılır. Ayrıca daha sonra test aşamasında kullanılan toplam doğruluğuna bağlı olarak her bir sınıflandırıcıya başka bir ağırlık verilir; daha doğru sınıflandırıcılara daha fazla ağırlık verilir. Son olarak, yeni bir örnek gönderildiğinde, her sınıflandırıcı ağırlıklı bir oy verir ve sınıf etiketi çoğunluk tarafından seçilir. Algoritma şu şekildedir:

Girdi: Eğitim seti $S = \{x_i, y_i\},\ i = 1, \ldots\ldots, N$ ve $y_i \in \{-1,1\}; T =$ iterasyon sayısı; I= Zayıf öğrenci

Uyarılmış sınıflar $h_t \in [-1,1]$ ve sınıf oransal sınıf ağırlıkları h_t, α_t Boost algoritması $H_{(x)} = sing(\sum_{t=1}^{T} \alpha_t h_t(x))$

Şekil 4.5' de görüldüğü gibi bir daireler ve üçgenleri sınıflandırma problemimiz olduğunu varsayalım. İlk iterasyonda algoritma bütün noktalara eşit bir ağırlık değeri vermektedir. Bu değerin 1 olduğu varsayımı altında algoritma en iyi sınıflandırmayı yapmaya çalışmaktadır. Yapılan sınıflandırma sonucunun ikinci şekilde görüldüğü gibi gerçekleştiği varsayımı altında;

$D_1(i) \leftarrow \frac{1}{N}$ **for** $i = 1, \ldots., N$

for $t = 1\ to\ T\ do$

$h_t \leftarrow I\ (S, D_t)$

$\varepsilon_t \leftarrow \sum_{i, y_i \neq h_t, x_i} D_t\ (i)$

$if\ \varepsilon_t > 0.5\ then$

$T \leftarrow t - 1$

$\boldsymbol{return}$

edn if

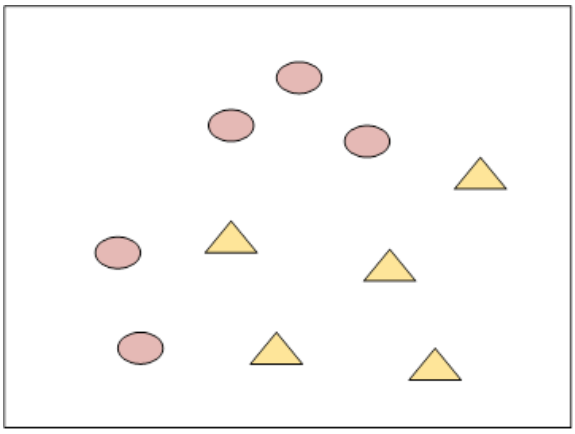
$\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

$D_{(t+1)}(i) = D_t(i).\beta^{1-[h_t(x_i) \neq y_i]}$ ***for*** $i = 1, \ldots., N$

Doğru bir dağılım için $D_{(t+1)}$ normalize ediliyor.

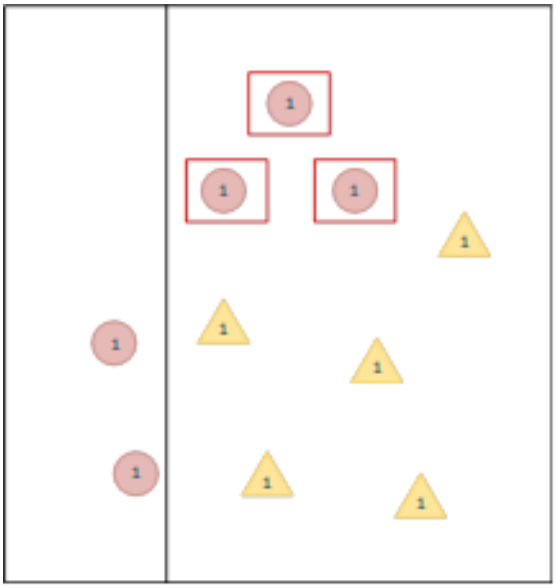
end for

(4.6)



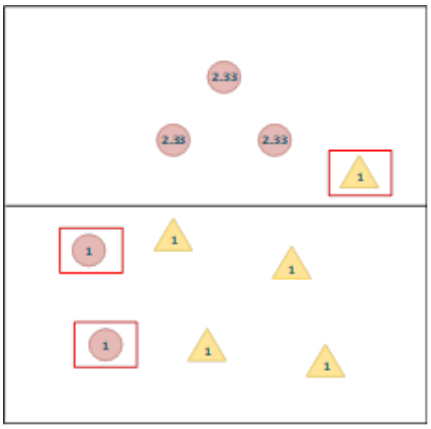
Şekil 4.5: Veri kümesi.

Birinci iterasyon sonucunda Şekil 4.6 da görüldüğü gibi kırmızı işaretli üç öge algoritma tarafından hatalı sınıflandırılmıştır. İkinci iterasyonda bu ögeler (toplam doğru sınıflandırılanlar / hatalı sınıflandırılanlar) oranı ile yeniden ağırlıklandırılmaktadır. Bu ağırlıklandırma %50- %50 dağılımlı bir sonuç getirmektedir. Yeni ağırlıklar 1*(7/3) = 2,33 olarak gerçekleşecektir.



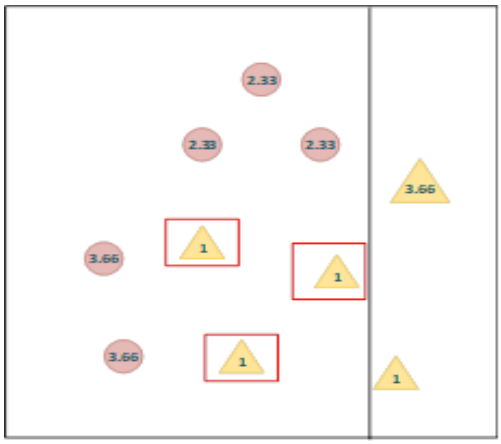
Şekil 4.6: Birinci iterasyon sınıflandırma sonucu.

İkinci iterasyonda en iyi çözüm yandaki gibi gerçekleşmektedir. Yine kırmızı ile belirtilmiş elemanlar hatalı sınıflandırılmıştır. Bu elemanlar bir sonraki iterasyon için yeniden ağırlıklandırılmalıdır. Yeni ağırlık (2,33+ 2,33+2,33+1+1+1+1) / 3 ≅ 3.66 olarak gerçekleşmektedir.



Şekil 4.7: İkinci iterasyon sınıflandırma sonucu.

Üçüncü iterasyonda en iyi çözüm Şekil 4.8'deki gibi gerçekleşmektedir. Doğru noktaların ağırlıkları 19 iken hatalı sınıflandırılan noktaların ağırlıkları 3 olarak gerçekleşmiştir. İterasyonlara istenilen başarı seviyesine ulaşılana kadar devam edilebilmektedir. Doğru sayısı arttıkça formülün katsayısı pozitif sonsuza doğru gidecektir, bütün sonuçları doğru veren bir çözümün ağırlığı sonsuz olarak gerçekleşmektedir- ln (∞) →∞.



Şekil 4.8: Üçüncü iterasyon sınıflandırma sonucu [67] [68] [69].

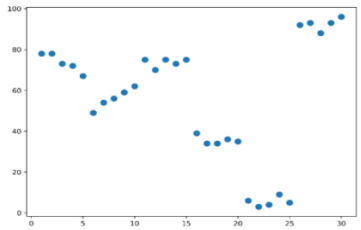
Gradient Boost, çeşitli pratik uygulamalarda önemli başarılar gösteren güçlü bir makine öğrenme tekniği ailesidir. Bunlar, farklı kayıp fonksiyonları ile öğrenilmesi gibi, uygulamanın özel ihtiyaçlarına göre son derece özelleştirilebilmektedirler. İstatistik bilimi ile bağlantı kurulabilmesi için, artırma yöntemlerinin gradyan-iniş temelli bir formülasyonu türetilmiştir. Boosting yöntemlerinin ve karşılık gelen modellerin bu formülasyonu, Gradient Boost (gradyan güçlendirme) makineleri olarak adlandırılmaktadır. Bu çerçeve aynı zamanda model hiperparametrelerinin temel gerekçelerini sağlamış ve model gelişimini daha fazla gradyanı artırmak için metodolojik

temel oluşturmuştur. Gradient Boost'da öğrenme prosedürü, bağımlı değişkeninin daha doğru tahminini sağlamak için art arda yeni modellere uymaktadır. Bu algoritma temelde öğrenenlerin tüm toplulukla ilişkili kayıp fonksiyonunun negatif gradyanı ile maksimum korelasyon gösterecek şekilde yapılandırılmasına dayanmaktadır. Uygulanan kayıp fonksiyonları rassal olarak seçilebilmektedir. Hangi kayıp fonksiyonunun uygulanacağı kullanıcıya bağlı olarak değişmektedir. Kullanıcı mevcut kayıp fonksiyonlarından birini kullanabileceği gibi kendisi de bir kayıp fonksiyonu belirleyebilmektedir. Gradient Boost algoritmasının Freidman tarafından yazılan şekli aşağıdaki gibidir. Fakat algoritmaya ait birçok parametrenin kullanıcı tarafından belirlenebilecek esnekliğe sahip olduğu hatırlanmalıdır [17] [18].

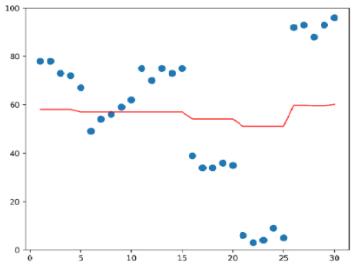
Gradient Boosting algoritması Şekil 4.10' da görüldüğü gibi ilk iterasyonda bir 'F' fonksiyonu üretmektedir. Bu fonksiyon tahmin üreteci olarak görev yapmaktadır. Tahminleri üreten bir "F" fonksiyonu oluşturur. 'h' fonksiyonu ise 'F' fonksiyonu tarafından tahmin değeri ile hedef değer arasındaki fark hesaplanarak oluşturulur. İkinci iterasyonda "F" ve "h" fonksiyonlarını birleştirir ve aynı hesaplamalar M değeri kadar tekrarlanır. Buradaki M değeri algoritmadaki nümerik optimizasyonunun çözümü için gerekli olan ve

$$P^* = \sum_{m=0}^{M} p_m \tag{4.7}$$

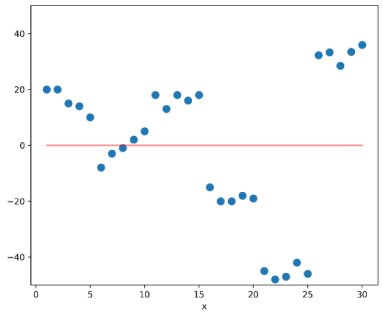
şeklinde ifade edilen parametrelerdir. Burada P_0 ilk tahmini $\{P_m\}_m^M$ her biri önceki adımların sırasına göre birbirini izleyen artımları ("adımlar" veya "takviyeler") ifade etmektedir. Adım sayısına ulaşana kadar 'F' fonksiyonu 'h' fonksiyonunu üreterek birleştirmektedir ve sürekli üstüne ekleyerek tahminler ile hedefler arasındaki farkı sıfıra eşitlemeye çalışmaktadır. 50 iterasyonda bir veri kümesi için tahmin ve tahmin ile hedef arasındaki ilişkiyi gösteren grafikler aşağıda verilmiştir.



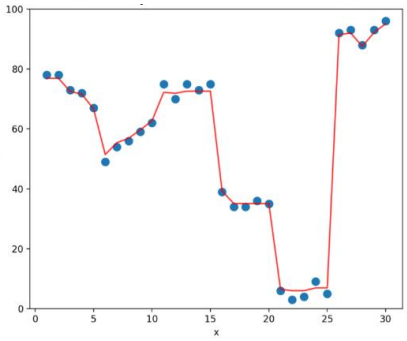
Şekil 4.9: Örnek veri kümesi.



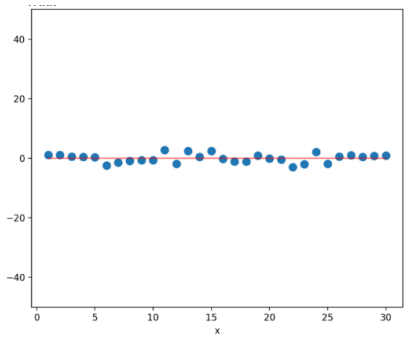
Şekil 4.10: Birinci iterasyon veri kümesi ve tahmin değeri.



Şekil 4.11: Birinci iterasyon hedef değer ile tahmin değeri farkı.



Şekil 4.12: Ellinci iterasyon veri kümesi ve tahmin değeri.



Şekil 4.13: Birinci iterasyon hedef değer ile tahmin değeri farkı [66].

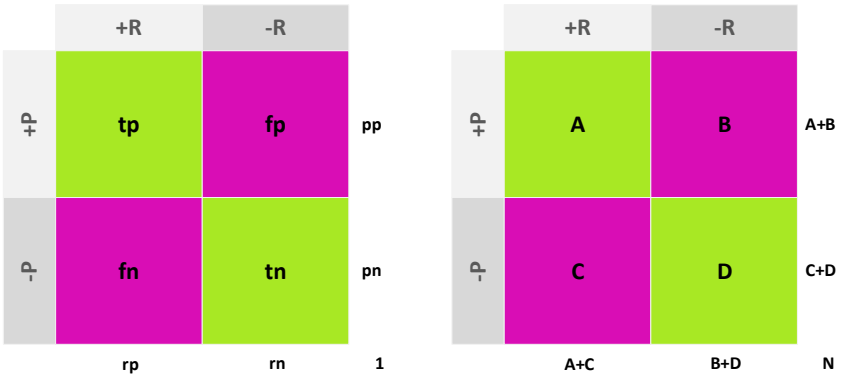
7. MLP Sınıflandırıcı

MLP sistemleri ileri beslemeli yapay sinir ağı sınıfındadırlar. Birden fazla katmandan oluşan ağ yapısına sahiptirler. MLP en az üç katmandan oluşur bunlar bir adet girdi katmanı bir adet gizli katman ve bir adet çıktı katanıdır. Giriş dışındaki katmanları doğrusal olmayan aktivasyon fonksiyonu kullanan birer nörondur. MLP Sınıflandırıcı algoritması ise bu temeller doğrultusunda python sklearn kütüphanesinde yer alan bir sınıflandırma algoritmasıdır [19].

4.2 Sınıflandırma Algoritmalarının Başarı Ölçme Yöntemleri

Makine öğrenmesi kapsamında yapılan sınıflandırma analizlerinde algoritmaların sınıflandırma başarılarını ölçmek amacıyla kullanılan farklı ölçütler olmak ile birlikte bu başlık altında sadece tezde kullanılan ölçütlere kısaca değinilmiştir.

Makine Öğrenimi deneylerinin sonuçlarını değerlendirirken etiketlerin konvansiyon + (pozitif) ve – (negatif) ile yapıldığı ikili sınıflandırmada ve bir sınıflandırıcının tahminlerinin dört hücreli bir olasılık tablosunda özetlendiği, iki boyutlu bir ikili sınıflandırma problemi bağlamında çeşitli ölçümlerin uygulanmaları yaygındır. Hata matrisi (confusion matrix) olarak bilinen bu tablonun genel görüntüsü aşağıdaki gibidir. İkili beklenmedik durum tablosunda sistematik ve geleneksel gösterimlere yer verilmiştir. Renk kodlaması ise yeşil olan hücreler doğru pembe olan hücreler hatalı durum oran veya sayılarını içerecek şekilde yapılmıştır.



Şekil 4.14: Hata matrisi.

Hücre ve kenar boşluğu etiketleri biçimsel olasılık ifadelerinden oluşabileceği gibi kenar boşluğu etiketlerinden hücre ifadeleri de türetilebilmektedir. Oransal veya tam sayı ifadeler ile kullanılabilirler. Ayrıca a, b, c, d veya A, B, C, D alfabetik sabit etiketleri veya 'Doğru ve Yanlış', 'Gerçek ve Öngörülen', 'Pozitif ve Negatif' terimlerinin baş harflerinden oluşan kısaltmalar kullanılabilmektedir. Matrislerde doğru pozitifler 'tp' ve yanlış pozitifler 'fp' (tp+fp yani A+B) tahmin edilenlerin sayısını ifade etmektedir. Doğru negatifler 'tn' ve yanlış negatifler 'fn' (tn+fn yani C+D) ile doğru pozitifler 'tp' ve yanlış pozitifler 'fp' (A+B) toplamı veri kümesindeki eleman sayısına (N) eşittir. Diğer yandan tp, fp, fn, tn ve rp, rn, pp, pn bileşik ve marjinal (bileşen) olasılıklar ifade edilmekte olup dört olasılık hücresi ve iki çift marjinal olasılık hücresi değeri 1'i vermektedir. Hata matrisinin tahminleri bir teorinin, bazı hesaplama kurallarının veya sistemlerinin (örneğin bir Uzman Sistem, Sinir Ağı vs.) tahminleri olabilmekte veya basitçe doğrudan bir ölçümlenmiş, hesaplanmış bir metrik, gizli bir durum, belirti veya işaretleyici de olabilmektedir. Hata matrisi verilerinden faydalanarak yapılan ve aşağıda tanımları verilen hesaplamalar bir modelin yaptığı tahminlerin başarısı hakkında fikir vermektedirler [20] [21] [22].

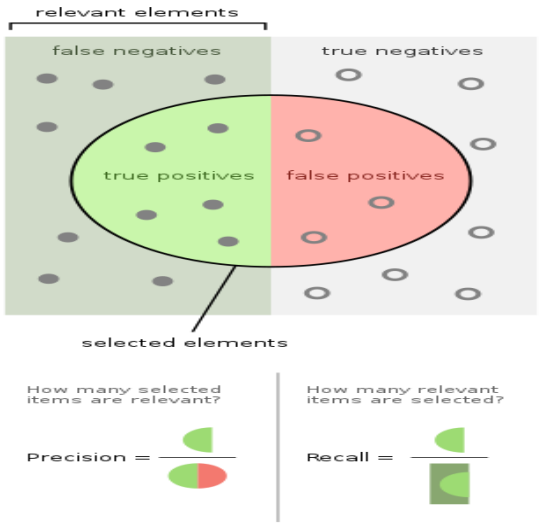
Hassasiyet veya Duyarlılık (Recall-Sensitivity); gerçekte pozitif olan etiketlerin pozitif tahmin edilen tüm etiketlere oranıdır. + P kuralı kaç tane ilgili vakayı aldığını yansıtması özelliği nedeni ile kullanılmaktadır. Hesaplamalı dilbilim/makine çevirisi bağlamında, duyarlılık kelime hizalamasının başarısını tahmin etmede büyük bir ağırlığı olduğu gösterilmiştir [23]. Duyarlılık ayrıca tüm Gerçek Pozitif (tp) vakaları tanımladığından temel olarak kabul edilir ve aynı zamanda ROC (Receiver Operating Characteristic)

analizinin üzerinde durduğu sac ayaklarından biridir. Bu bağlamda Gerçek Pozitif Oran (tpr) formülü:

$$tpr = tp/rp \quad \text{veya} \quad tpr = A/(A + C) \tag{4.8}$$

Kesinlik veya Güven (Precision-Confidence) doğru atanan pozitiflerin veri kümesindeki tüm pozitif atananlara oranı olarak ifade edilmektedir. Öngörülen Pozitif vakaların oranını göstermektedir. Makine Öğrenmesi, Veri Madenciliği ve Bilgi Erişiminde odak noktası olan bu oran ROC analizinde tamamen göz ardı edilmektedir. Gerçek Pozitif Doğruluk (True Positive Accuracy) (tpa) olarak da adlandırılabilen bu oranın formülü ve şekil olarak ifadesi aşağıdaki gibidir [24].

$$tpa = tp/pp \quad \text{veya} \quad tpa = A/(A + B) \tag{4.9}$$



Şekil 4.15: Kesinlik ve hassasiyet oranlarının hesaplanma yöntemi [51].

Bu iki ölçüt ve kombinasyonları sadece pozitif oranlara odaklanmakta negatif tahminlerin başarısı ile ilgili hiçbir veri içermemektedirler. Duyarlılık (precision) R kolonu ile hassasiyet ise +P satırı ile ilgili oranlamalar yapmaktadır. Bu durum tpr (recall) ve tpa (precision) nın harmonik ortalaması olan ***F- ölçütü*** için de geçerlidir. Harmonik ortalama, bir veri dizisinde bulunan ve diğerlerinden çok yüksek değere sahip elemanların ortalamaya etkisini azaltmak için kullanılmaktadır. Çünkü bu elemanlar çoğu zaman özel

bir durumla ortaya çıkmıştır ve bunların etkisini azaltmak, dizinin normal seyrini görmemize yardımcı olmaktadır. F ölçütü bu anomalileri düzenlediği için analizlerin doğru yorumlanmasında önem taşır.

$$F_1 = (\frac{2}{recall^{-1}+precision^{-1}}) \quad (4.10)$$

Gerçekte pozitif durum ile ilgili özellikli bir şart yoktur. Pozitif durumlar için hesaplanan istatistiksel oranlar tersine çevrilerek negatif durumların hesabında kullanılabilirler [25].

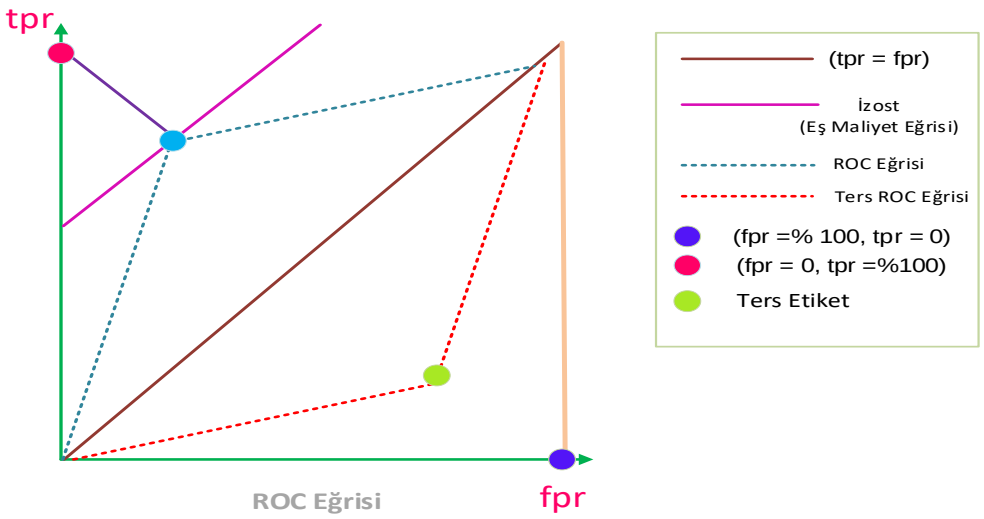
Ölçümlerin doğası ve eğim duyarlılıkları hakkında geometrik bilgiler vermek için ***ROC eğrisi*** kullanılmaktadır. Sinyal algılama teorisinde, alıcı işletim karakteristiği (Receiver Operating Characteristic - ROC) ROC eğrisi olarak tanımlanmaktadır. ROC eğrisi, ikili sınıflandırma sistemlerinde ayrım eşik değerinin farklılık gösterdiği durumlarda, duyarlılığın (recall), kesinliğe oranıyla ortaya çıkmaktadır. ROC daha basit anlamda doğru pozitiflerin, yanlış pozitiflere olan kesri olarak da ifade edilebilir. [65] ROC eğrisinin normal olmayan PN varyantına genişletilmesi analizleri özellikle kural öğrenme hedefine yöneltmiştir. ROC analizi, tpr oranını ve fpr oranını baz alırken PN, normalize edilmemiş TP ve FP'yi baz almaktadır. Normalizasyon tarafından yaratılan tek fark ölçekler ve degradelerde ortaya çıkmaktadır. Mükemmel bir sınıflandırıcının eksenlerin sol üst köşesinde sonuç vermesi belenmektedir çünkü bu noktada bütün sınıflar doğru tahmin edilmiştir (fpr = 0, tpr =%100). En kötü durum sınıflandırıcısı sağ alt köşede sonuçlanması olarak kabul edilmektedir. Bu durumda bütün etiketler hatalı atanmaktadır (fpr =%100, tpr = 0). Etiketlerin ters çevrilmesi gerekliliği mevcuttur. Rastgele bir sınıflandırıcının pozitif diyagonal boyunca (tpr = fpr) bir yerde puanlanması beklenir çünkü model aynı oranda pozitif ve negatif örnekler atar. Negatif köşegen için (tpr + c * fpr = 1), c eğriliği için sapmaya karşılık gelmektedir. ROC grafiği, sınıflandırıcıların veya parametrelerin karşılaştırılabilmesine ve bir anlamda tpr = fpr' ye en yakın ve en uzak seçeneklerin belirlenebilmesine imkân vermektedir. Optimal parametre veya model seçimi için gerekli koşullar aynı değildir. Fakat en yaygın koşul, eğrinin altında kalan eden alanı (AUC) (0,0) ile (1,1) segmentasyonunu optimize etmek amacı ile en aza indirmektir.

Parametreli bir model için, (0,0) ila (1,1) arasındaki segmentlerden oluşan monotonik bir fonksiyon olacaktır. Belirli bir maliyet modeli ve / veya doğruluk ölçümü, maliyete duyarlı olmayan bir model için c = 1 olacak olan bir izost gradyanını tanımlamaktadır ve bu nedenle başka bir yaygın yaklaşım, eğriye temas eden en yüksek izost çizgisi üzerinde teğet bir nokta seçmektir.

$AUC = (tpr - fpr + 1)/2$ veya

$AUC = (tpr + tnr)/2$ yahut

$AUC = 1 - (fpr+fnr)/2$ (4.11)

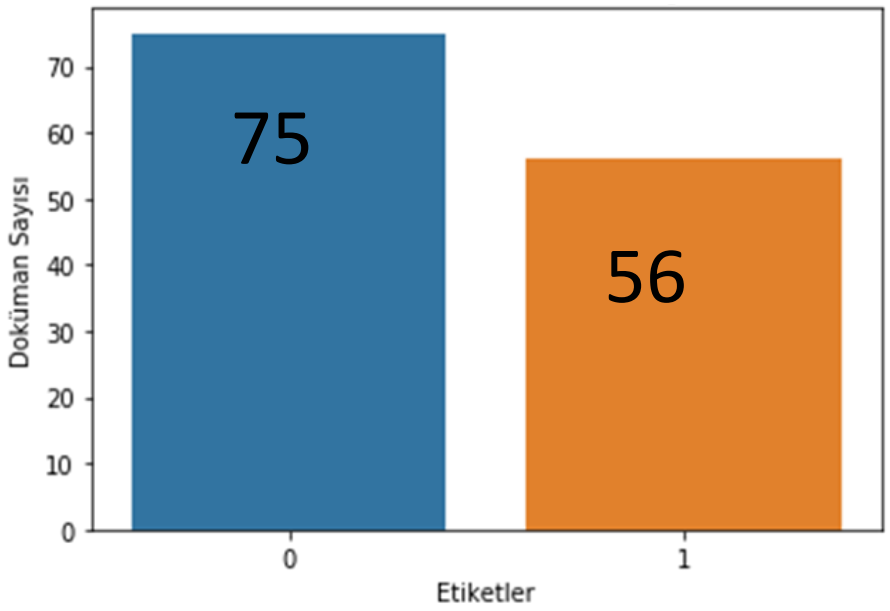


Şekil 4.16: ROC eğrisi ve hata matrisi ilişkisi [52].

4.2 Veri Setine İlişkin Genel Bilgiler

Çalışma kapsamında Yargıtay Hukuk Genel Kurulu ve Yargıtay Hukuk dairelerince 2013-2019 yılları arasında verilmiş 461 adet mobbing içerikli karar taranmış fakat bunların yalnızca 131 tanesi çalışmaya dahil edilebilmiştir. Modele dahil edilmeyen karar içeriklerinde mobbing iddiasına ilişkin ayrıntıları içermediğinden elenmiştirler. Veri setine ilişkin genel bilgiler aşağıda özetlenmiştir.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 131 entries, 0 to 130
Data columns (total 2 columns):
Rating    131 non-null int64
Karar     131 non-null object
dtypes: int64(1), object(1)
memory usage: 2.2+ KB
None
```



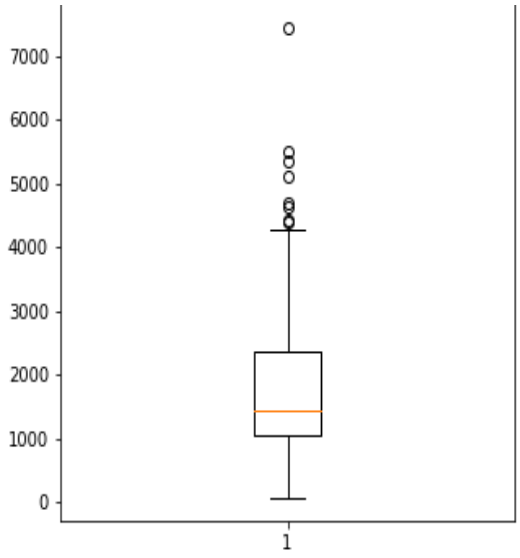
Şekil 4.17: Veri kümesindeki dokümanların etiketlere göre dağılımı.

131 adet karardan Yargıtay tarafından mobbing olarak kabul edilenler 1 (bir) ile kabul edil meyenler 0 (sıfır) ile etiketlenmiştir. Kararlardan 75 tanesi sıfır 56 tanesi bir etiketlidir.

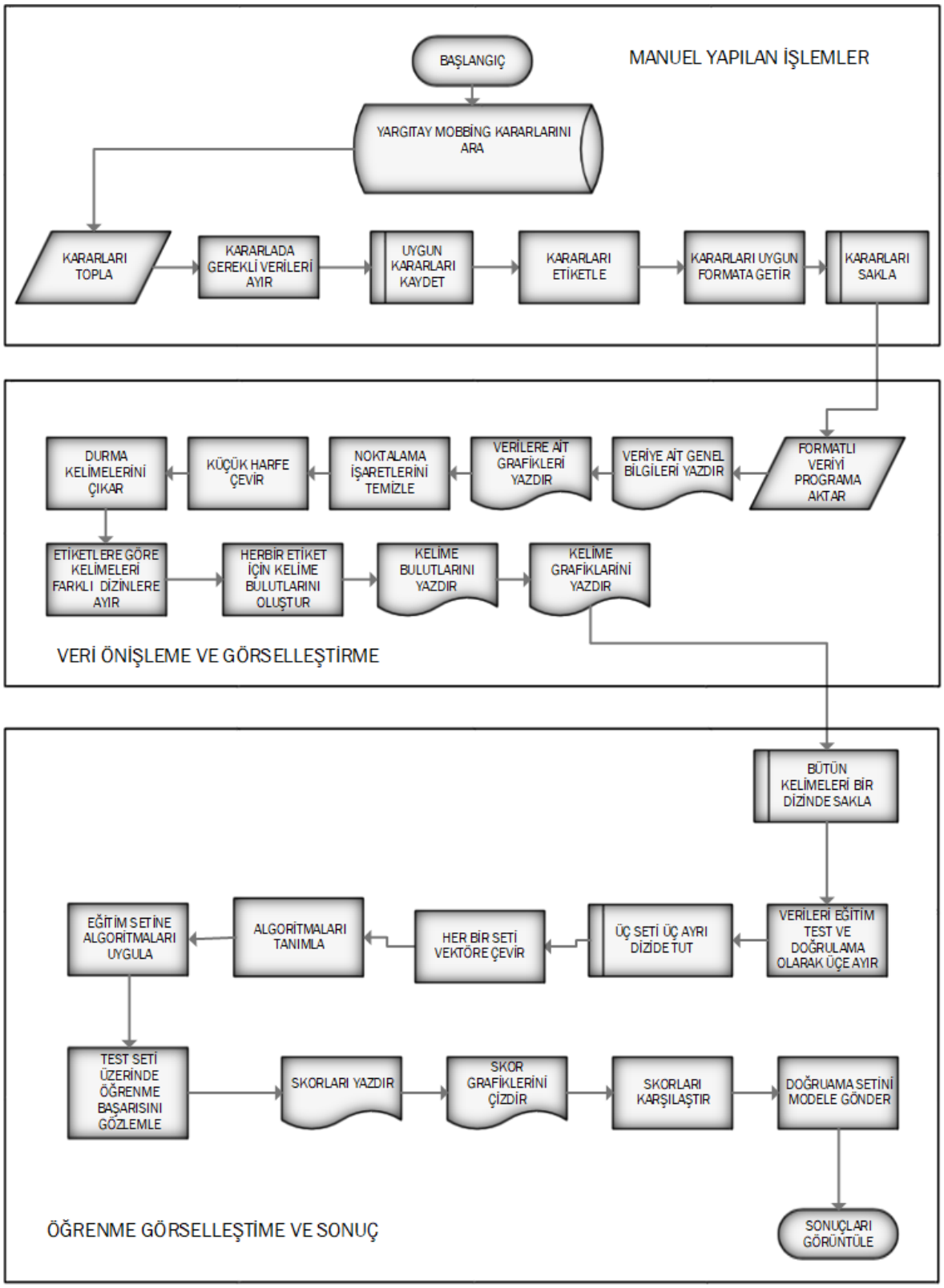
Veri setindeki dokümanların cümle uzunluklarına ilişkin bazı örnekler Şekil 4.18'de verilmiştir. Ayrıca veri setindeki cümle uzunluklarına göre sınıf duyarlılıklarını gösteren kutu-bıyık diyagramına da yer verilmiştir.

	Rating	Karar	phrase_len
0	0	vekilince mobbing uygulanıp uygulanmadığı husu...	1915
1	0	noterden ihtarname feshini işverenin psikoloji...	441
2	0	... yerinde beklenenin hedefleri konularak gör...	1022
3	0	tanığı sözleşmede imzası ... beyanında başkanı...	1040
4	0	mobbing istemine ilişkindir. edilmiştir. ... l...	916
5	0	metninden anlaşılacağı mobbingi davranışların ...	243
6	0	..-.. çalışma - asistanı - ipotek asistanı maa...	1112
7	0	tazminata kazanıp kazanmadığı hususu uyuşmazlı...	591
8	0	.. “şirketiniz ...genel müdürlügün'nde görev y...	476
9	0	aktini mesailerin karşılığının ödenmemesi göre...	580

Şekil 4.18: Veri setinden bazı örnekler.



Şekil 4.19: Cümle uzunluklarına göre sınıf duyarlılıklarını gösteren kutu-bıyık diyagramı.



Şekil 4.20: Analize ait akış şeması.

4.3 Veri Ön İşleme ve Görselleştirme

Ön işleme aşamasında önce noktalama işaretleri temizlenmiş, büyük harfler küçük harfe dönüştürülmüş, durma kelimeleri (stop words) ayıklanmıştır. Temizleme aşamaları tamamlanan verilerin mahkeme tarafından mobbingin varlığı kabul edilen kararları ile kabul edilmeyen kararlarına ait kelime bulutları çizdirilmiştir ve Şekil 4.2 ile 4.3 de gösterilmiştir. Bu işlem sonucunda yapılan değerlendirmede beklenildiği üzere kelime bulutlarının bir duygu analizi modelinde olduğu gibi birbirinin zıttı keskin kelimelerden oluşmadığı görülmüştür. Çünkü kararlar her iki etikette de benzer kelimeleri içermektedir. Bulutlara yakından bakıldığında banka ve mağaza kelimelerinin bulundurdukları görülmektedir. Bu durum incelemeye alınan yargı kararlarında dava tarafları arasında bu iki meslek grubunun da bulunduğunu göstermektedir. Ayrıca İstanbul kelimesinin de kelime bulutlarında yeri aldığı gözlenmektedir. Beyaz ve mavi yakalı çalışanların sayısının en fazla olduğu şehir olması nedeni ile bu ilimizin kelime bulutlarında yerini alması şaşırtıcı değildir.

Şekil 4.21: Mobbingin varlığını kabul etmeyen mahkeme kararlarına ait kelime bulutu.

Şekil 4.22: Mobbingin varlığını kabul eden mahkeme kararlarına ait kelime bulutu.

4.4 Kelime Torbaları (Bag of Words), Tf-Idf, Word2Vec Uygulamaları

Veri temizleme aşaması tamamlanan metinler Python programlama dili kütüphanelerindeki Bag of Words (count vectorizer), TF-IDF ve doc2vec yöntemleri kullanılarak vektörize edilmiştir.

Tablo 4.1'de bag of words yöntemi ile sayısallaştırılan metinlerde en sık geçen kelimelerden ilk 15 tanesi örnek olarak verilmiştir.

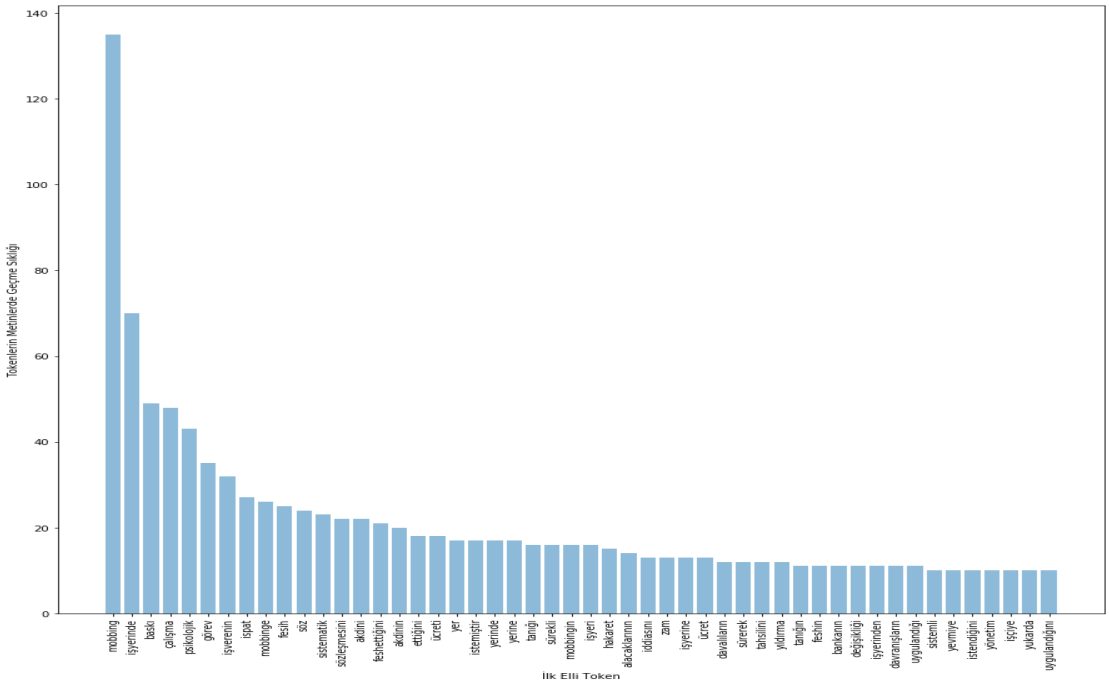
Tablo 4.1: Bag of words yöntemi en sık geçen 15 kelime

Mobbingin varlığı kabul edilmeyen kararlar (Yok)			Mobbingin varlığı kabul edilen kararlar (Var)			Bütün Kararların Toplamı (Bag of Words)		
İndeks	Terimler	Frekans	İndeks	Terimler	Frekans	İndeks	Terimler	Toplam Frekans
1	mobbing	135	1	mobbing	43	1	mobbing	178
2	baskı	49	2	çalışma	22	2	baskı	70
3	çalışma	48	3	baskı	21	3	çalışma	70
4	psikolojik	43	4	fesih	18	4	psikolojik	58
5	işverenin	32	5	psikolojik	15	5	fesih	43
6	ispat	27	6	akdinin	14	6	işverenin	40
7	mobbinge	26	7	ettiğini	13	7	ispat	36
8	fesih	25	8	tutanak	13	8	akdinin	34
9	sistematik	23	9	müdürünün	13	9	mobbinge	32
10	sözleşmesini	22	10	şube	13	10	ettiğini	31
11	akdini	22	11	istemiştir	12	11	akdini	30
12	feshettiğini	21	12	etmiştir	11	12	istemiştir	29
13	akdinin	20	13	feshin	11	13	feshettiğini	28
14	ettiğini	18	14	yerine	10	14	yerine	27
15	ücreti	18	15	verilmiştir	10	15	sözleşmesini	27

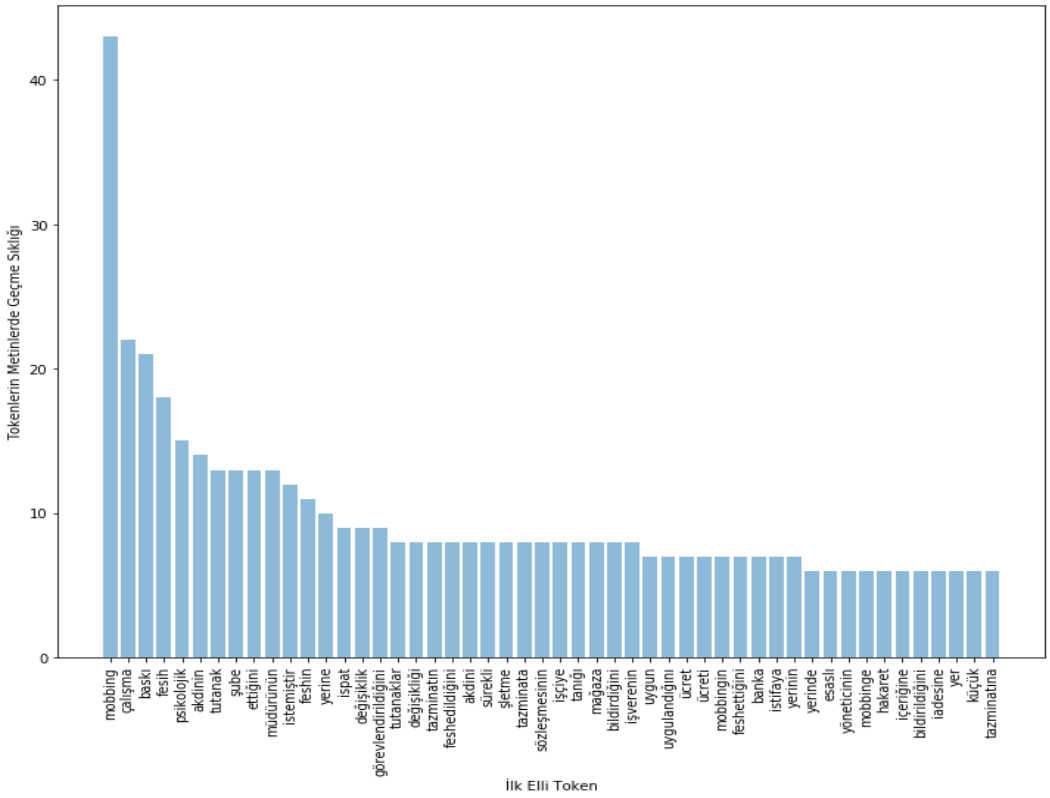
Şekil 4.23' de TF-IDF yöntemine göre mobbing varlığı kabul edilmeyen kararlarda geçen en yüksek ağırlık değerine sahip 50 kelimeye ait grafik paylaşılmıştır. Grafikte görüldüğü üzere ağırlığı en yüksek olan beş kelime 'mobbing, işyerinde, baskı, çalışma ve psikolojik' kelimeleridir.

Şekil 4.24' de aynı yöntemde mobbing varlığı kabul edilen kararlarda geçen en yüksek ağırlık değerine sahip 50 kelimeye ait grafik paylaşılmıştır. Bu grafikte en yüksek ağırlığa sahip beş kelime den üçü 'mobbing, işyerinde, baskı' şekil 4.6 ile aynı sıralamaya sahip iken dördüncü kelime olarak 'fesih' gelmiş beşinci kelime ise 'psikolojik' kelimesidir. İş

kanununa göre mobbingin varlığı durumunda iş akdinin haklı feshi söz konusudur. ‘fesih’ kelimesinin ilk beş arasına girmesi bu çerçevede değerlendirilebilir.



Şekil 4.23: Sıfır etiketli metinlerde en sık kullanılan elli kelime.



Şekil 4.24: : Bir etiketli metinlerde en sık kullanılan elli kelime.

TF-IDF modelinde n-gram 3 ile oluşturulan sözlük listesinden bazı örnekler aşağıda paylaşılmıştır. Listede sözlüğün içerdiği kelimeler (anahtarlar) ve bu kelimelerin sözlük değerleri (key value) bulunmaktadır. Kelimeler başlangıçta 1 gram daha sonra 2 gram ve son olarak 3 gram halinde sözlüğe dahil edilmiştirler. Bu analize ilişkin bazı örnekler aşağıda verilmiştir.

```
{'baskı': 1331,
 'şartlarında': 14894,
 'esaslı': 4163,
 'değişiklik': 3277,
 'akdini': 113,
 'feshettiğini': 4643,
 'şirkete': 15010,
 'memur': 8670,
 ….}
```

```
{'baskı şartlarında': 1433,
 'şartlarında esaslı': 14897,
 'esaslı değişiklik': 4164,
 'değişiklik akdini': 3278,
 'akdini feshettiğini': 130,
 'feshettiğini şirkete': 4677,
 'şirkete memur': 15013,
 …….}
```

```
{'baskı şartlarında esaslı': 1434,
 'şartlarında esaslı değişiklik': 14898,
 'esaslı değişiklik akdini': 4165,
 'değişiklik akdini feshettiğini': 3280,
 'akdini feshettiğini şirkete': 135,
 'feshettiğini şirkete memur': 4678,
 ……..}
```

TFIDF ve BOW modele ait eğitim ve test verilerine ait bilgileri içeren kod parçasının çıktısı aşağıdaki gibidir.

```
BOW model:> Eğitim Seti Özeti:(83, 3046) Test Seti Özeti: (21, 3046)
TFIDF model:> Eğitim Seti Özeti: (83, 15270) Test Seti Özeti: (21, 15270)
```

Word embedding yöntemi ile vektörize edilirken girilen “vector_size=100, window=11, min_count=5, workers=4” parametreleri ile 352 özellik (kelime) için 100 boyutlu vektörler oluşturulmuştur. Veri setinin boyutu yeterince büyük olmadığından CBOW uygulanmıştır. Vektörize edilen kelimelerin listesi vektörlere ait çıktının bir bölümü aşağıda Tablo 4.2’ de -paylaşılmıştır.

Oluşturulan Doc2Vec modelinde ‘mobbing’ özelliği ile benzerlik gösteren yirmi kelime ve vektörleri aşağıda örneklenmiştir.

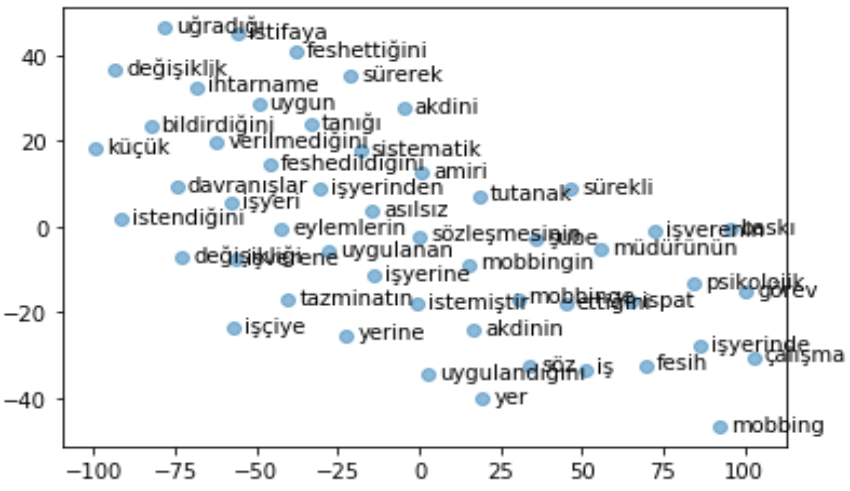
```
 'Mobbing' Özelliği İle En Fazla Benzerlik
Gösteren 20 Kelime ve Vektörleri
-----------------------------------------
[('görev', 0.9983108043670654),
 ('işyerinde', 0.998306393623352),
 ('işverenin', 0.9982203245162964),
 ('çalışma', 0.9981756806373596),
 ('baskı', 0.9980756044387817),
 ('fesih', 0.9977653622627258),
 ('psikolojik', 0.9977552890777588),
 ('ispat', 0.9973925352096558),
 ('mobbingin', 0.9973666667938232),
 ('iş', 0.9973581433296204),
 ('akdinin', 0.9973143339157104),
 ('işyerine', 0.9972798824310303),
 ('ettiğini', 0.9972113370895386),
 ('işverene', 0.9972089529037476),
 ('mobbinge', 0.9971983432769775),
 ('yerine', 0.9971903562545776),
 ('şube', 0.9970629215240479),
 ('sürekli', 0.9970242381095886),
 ('asılsız', 0.9969495534896851),
 ('yer', 0.9969112277030945),
 ('istemiştir', 0.9968994855880737),
 ('müdürünün', 0.9968841075897217),
 ('işyerinden', 0.996861457824707),
 ('uygulandığını', 0.9967484474182129),
 ('sözleşmesinin', 0.9966884255409241),
 ('söz', 0.9966813325881958),
 ('işyeri', 0.9966055750846863),
```

Tablo 4.2: Word2vec modelinde vektör oluşturulmuş özellikler.

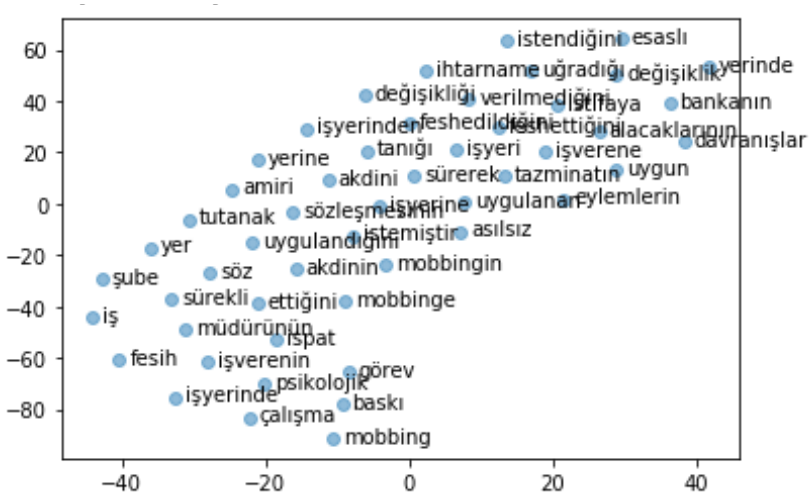
Sütun 1	Sütun 2	Sütun 3	Sütun 4	Sütun 5	Sütun 6	Sütun 7	Sütun 8	Sütun 9	Sütun 10
mobbing	çalışıldığını	iddialarını	tanığı	asistanı	uygulanması	faizi	uygulandığına	bankada	şube
sistematik	mobbinge	ispat	zam	maaşının	çalıştırıldığı	tahsilini	şahit	ikale	görevini
söz	uğradığını	etme	davacıyla	brüt	bulunmamaktadır	raporlu	ispatlayamadığı	bildirimi	değişikliğini
psikolojik	sürerek	fesih	ettiğini	ücreti	açıklandığı	yerinin	değiştirildiğini	baskılar	avukatlık
baskı	iddiaların	tanığın	uyguladığını	mesaiye	işçinin	banka	yaptırıldığını	müdürünün	planlı
küçük	beyanla	anlaşıldığından	maaş	verilerek	nedene	müdürlüğünde	sağlığının	görevlisi	çalışmasının
işyerinde	beyanlarına	iddiasını	uygulanan	bankanın	feshi	müdürlüğüne	bozulduğunu	zorlanması	kararlarında
yıldırma	dayanılarak	yukarda	bildiriminde	ücretlerinin	iradesinin	tutanak	değiştirilmesi	istifaya	öğretide
sindirme	uygulandığı	yanılgılı	göreve	uyguladığı	ayrılma	belirttiğini	emare	mahkemesinin	kurumdur
sistemli	kabulüyle	bozulmasına	işyerinden	davranışlarda	rencide	zararın	hakaret	akışına	beslemesi
söylenti	tazminata	iadesine	akdinin	baskının	sözler	iddiasına	görevi	yöneticinin	kasten
çalışma	sözleşmesinin	yerinde	tazminatına	yetkililerinin	feshettiğini	amiri	mobing	akdi	çıkarması
işçiyi	imzası	yeri	yerine	alacaklarının	ödenmediğini	bozulduğu	muameleye	ayrılmasını	amirleri
aşağılayıcı	ücret	değiştirilmek	ilişkindir	ücretli	alacaklarını	tutanağı	şirkette	davaya	tanımlanmıştır
davranış	feshedildiği	uygulandığını	davalıların	istemiştir	işyerine	rızası	söylendiğini	eylemlerin	emarenin
işyeri	işyerindeki	sözleşmesini	isteminde	belgelerden	baskıya	dosyada	hükmedilmesi	düşürücü	sağlığın
sürekli	niteliğindeki	işverene	onur	akdini	insan	anlaşılmakla	küfür	aşağılama	uğraması
süreklilik	davranışların	işçilik	kırıcı	feshettiği	çalışmakta	koşullarında	şikayette	davranışlara	tartışmasız
göstermeyen	sebebiyet	tanığının	kapsamına	zarara	birime	sebeplerle	belirtmiş	talebine	ayrımcılık
aralıklarla	tacize	ayrıldığını	beyanlarının	uyuşmazlıkta	atandığını	noterliğinden	raporuna	rahatsızlığını	tespitine
nitelendirilemez	uğradığı	duyuma	tanıkların	şartlarının	değişikliğinin	yevmiye	edilemeyeceği	teşhisi	ilkesine
iş	edildiğinin	dayandığı	iddialarının	esaslı	yönetici	bulunulduğunu	başlandığını	raporun	psikolojisinin
işyerinde	inandırıcı	dosyaya	ispatlanamadığı	koşullarının	birimin	ihtarnamesi	feshedildiğini	gönderildiğini	verilmemesi
yönetimi	delillerle	değişikliği	gerekçeyle	tazminatın	ayrıldığı	bildirdiğini	işverenden	istendiğini	tutulduğunu
yönetim	mobbingin	gerçekleştirilmesi	doğru	değişiklik	alınmadığını	ilişkisinin	talebin	görevlendirildiğini	noterliğinin
zorlandığını	unsurlarının	kanıtlar	bozulması	izne	personele	tehdidi	görevlendirme	uygun	bulunduğunun
görev	bozma	değiştirildiği	davranışlar	başkaca	çıkarıldığını	mağaza	noter	ücretinin	uygulanmaya
yer	nedenidir	kapsamından	edilmelidir	alacağının	bildirildiğini	şubede	işçiye	düşürüldüğünü	rahatsızlığı
verilmediğini	ihtarname	iddiasının	yıldırmak	ödenmemesi	bırakıldığını	feshin	feshinin	incelendiğinde	savunmasının
personelin	işverenin	edilemediği	kasıtlı	bahisle	ihtar	davalıkarşı	tazminatının	şubesinde	çıkışının

Grafik ve vektörler ayrıntılı olarak incelendiğinde mobbing kavramı ile en ilişkili kavramların özet olarak ‘mobbingin şartları’, ‘kim tarafından iddia edildiği’, ‘uygulanma şekli’, ‘uygulanma yeri’, ‘dava süreci’ başlıklarında toplandığı görülmektedir. Süreklilik, görev yerinde esaslı değişik, sistematiklik kelimelerinin mobbingin şartlarına tekabül etmektedir. ‘işveren’, ’müdür’, ‘işyeri’, ‘şube’ gibi kelimeler ise mobbingin uygulayıcıları oldukları iddia edilenler ve uygulanma yerine denk gelmektedir. Ayrıca iş akdinin haklı nedenle feshi, ücret, maaş, tazminat gibi kelimeler mahkeme sürecine tekabül etmektedirler ve ispat edilmiş mobbingin işçiye haklı feshin şartlarından yararlanma imkânı tanıması nedeniyle dava içeriğinde sıkça rastlanmasından kaynaklanmaktadır. Ayrıca model aynı köke sahip sadece çekim eki almış ‘mobbing’ özelliği ile ‘mobbinge’ ve ‘mobbingin’ kelimelerinin benzerliklerini sırası ile [0.995], [0.993] olarak başarılı bir şekilde tahmin etmiştir. Mobbing kavramı ile aynı anlama gelen ‘psikolojik şiddet’ kelimelerinin birbiri ile ilişkisini ise [0.994] olarak tahmin edilmiştir.

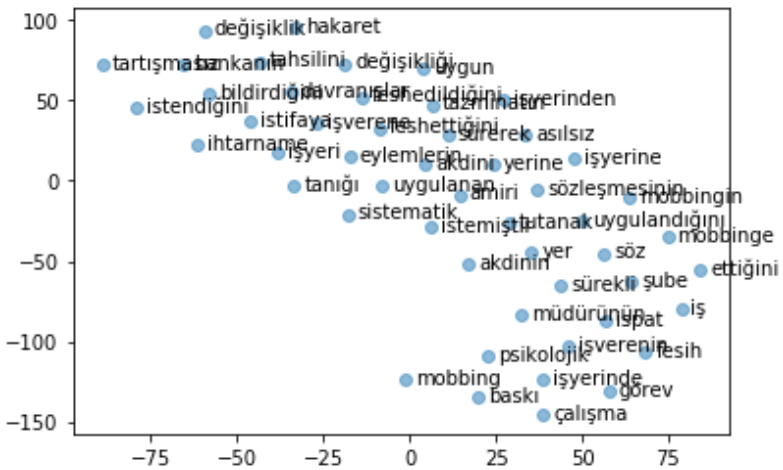
Türkçenin sondan eklemeli bir dil olması nedeniyle veri ön işleme aşamasında yapım ve çekim eki alan kelimelerinin ayıklanması işlemi (stemming) yapılabilmektedir. Fakat anlam kayıplarına neden olabileceği gerekçesi ile bu analiz kapsamında stemming işlemi yapılmamıştır. Bu nedenle kök ve türevleri kelimeler ayrı birer kelime olarak değerlendirilmiştir. Şekil 4.25’ de ‘mobbing’ özelliğine en yakın anlamlı elli özellik gösterilmiştir. Şekil 4.26 ve 4.27 de ise word2vec modelinin bulduğu kelimenin türevleri ‘mobbingin’ ve ‘mobinge’ kelimelerine ait en benzer elli özellikleri gösteren grafikler aşağıda verilmiştir. Grafikte benzer kelimeler bir araya kümelenmektedir. Grafiker incelendiğinde her üç grafikte de benzer kelimelerin yer aldığı görülmektedir.



Şekil 4.25: 'Mobbing' özelliğine en yakın elli özellik.



Şekil 4.26: 'Mobbingin' özelliğine en yakın elli özellik.



Şekil 4.27: 'Mobbinge' özelliğine en yakın elli kelime.

4.5 Makine Öğrenmesi Sonuçları

Bu bölümde Makine Öğrenmesine ait gözetimli öğrenme araçlarından sınıflandırma algoritmaları kullanılarak tahminleme ve sonuç çıkarımı yapılmıştır. Tüm uygulamalarda Python dilinde yazılan programlar kullanılmıştır. Makine öğrenmesi ile ilgili gelişmiş kütüphaneleri nedeni ile Python programlama dili tercih edilmiştir. Bu bölümde yapılan analiz sonuçlarından vektörleştirme yöntemlerine göre her bir algoritmada elde edilen sonuçlara ilişkin veriler paylaşılacaktır. Veri seti eğitim test ve doğrulama olmak üzere üç gruba ayırılmıştır. Eğitim seti ile öğrenme gerçekleştirilmiş test seti ile öğrenmenin gerçekleşme düzeyi test edilmiş ve doğrulama seti ile modelin daha önce görmediği veriler üzerindeki başarısı ölçülmüştür.

Başarılı sonuçlar belirlenirken test setinde ve doğrulama setinde algoritmaların doğruluk (accuracy) oranlarının yanı sıra hassasiyet oranları, doğruluk oranları ve F ölçütü sonuçları birlikte değerlendirilmiştir. Veri setinde etiketler dengesiz dağılmaktadır ve bu eşitsizlik durumu nedeniyle accuracy değeri modelin başarısını ölçmede tek başına yanıltıcı olabilmektedir. Yapılan analizler göstermiştir ki doğruluk oranı aynı iki algoritmanın ROC eğrisi analizi aynı çıkmayabilmektedir. Aynı şekilde doğruluk oranı diğer bir algoritmaya göre daha düşük gerçekleşen algoritmanın ROC eğrisi analizi daha iyi sonuç vermeyebilmektedir. Bu durumda algoritmaların analiz sonuçları değerlendirilirken doğruluk oranının yanında hassasiyet ve duyarlılık değerlerini temel alan ROCeğrisi analizi de dikkate alınmıştır.

Makine öğrenmesine ilişkin sonuç özetlerini veren tablolarda ‘c_0’ sıfır etiketli, ‘c_1’ bir etiketli verileri temsil etmektedirler. c_0 sınıf değeri Yargıtay’ın mobbingin varlığını kabul etmediği kararları, c_1 değeri mobbingin varlığını kabul ettiği kararları temsil etmektedir. Tablolar her bir sınıf değeri için Hatırlama (Recall), Hassasiyet (Precision), F1-skor, doğruluk oranı (accuracy) değerlerini ve ağırlıklı ortalama (weighted avg) makro ortalamayı (macro avg) içermektedir.

TF-IDF yönteminde test ve doğrulama setlerinde en iyi sonucu veren algoritma sonuç özetleri:

Tablo 4.3, 4.4 ve 4.5 incelendiğinde hem test setinde hem de doğrulama setinde uygulana on adet sınıflandırma algoritmasının yedi tanesi her bir set için en yüksek başarı oranlarının bu yöntemle vektörleştirilen verilerde elde edildiği görülmektedir. Yöntemin daha fazla algoritmada başarılı sonuçlar vermesinin sebebi veri kümesinin yapısından kaynaklanmaktadır. Data setindeki metinler anahtar kelimelerin bulunması ile daha iyi analiz edilebilecek türden oldukları için bu yöntemde algoritmalar daha iyi sonuç vermektedir. Yine aynı tablolar ışığında test seti üzerinde en başarılı sonuçları üreten algoritma %88,89 ile Random Forest Classifier olmuştur. Şekil 4.28'e göre analizde gerçekte sıfır etiketli veriler %89 duyarlılık oranı ile doğru atanmış iken bir etiketli olanların ise duyarlılık oranı %88 olarak gerçekleşmiştir. Şekil 4.28' e bakıldığında güven (precision) ve hassasiyet (recall) oranlarının harmonik ortalaması olan F1 skor değeri ise bu algoritmada sıfır etiket için %92 bir etiket için %82 olarak gerçekleştiği görülmektedir. Doğrulama setindeki başarısı ise %72,73 olarak gerçekleşmiştir. Tablo 4.3 incelendiğinde Doğrulama setinde en yüksek doğruluk oranına sahip algoritma ise %90,91 ile MLP Classifier algoritması olmak ile birlikte tablo 4.4 incelenerek güven ve duyarlılık oranları dikkate alındığında %81,82 ile SVM algoritmasının daha iyi sonuçlar verdiği gözlemlenmiştir. Bu algoritmanın test seti üzerindeki başarısı %85,19 olarak gerçekleşmiştir. Test ve doğrulama setlerindeki başarıları dikkate alınarak belirlenen yöntemin en başarılı algoritmalarına ilişkin özet bilgiler aşağıdaki gibidir.

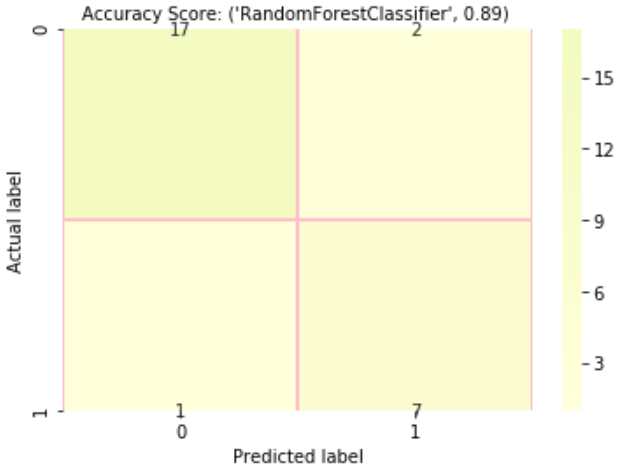
Test seti üzerinde gerçekleşen Random Forest Classifier sonuçlarının özetleri:

RandomForestClassifier -> ACC: %88.89

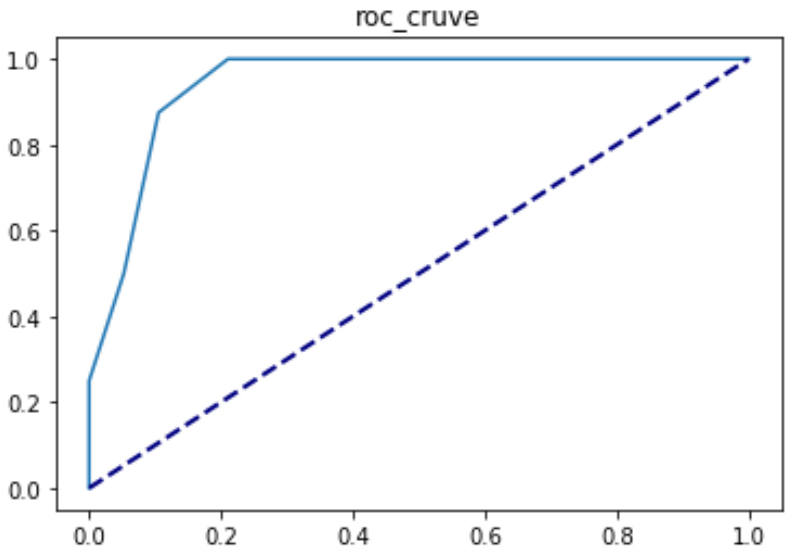
	precision	recall	f1-score	support
c_0	0.94	0.89	0.92	19
c_1	0.78	0.88	0.82	8
accuracy			0.89	27
macro avg	0.86	0.88	0.87	27
weighted avg	0.90	0.89	0.89	27

Şekil 4.28: TF-IDF yöntemi Random Forest algoritması test seti sonuç özeti

.



Şekil 4.29: TF-IDF yöntemi Random Forest algoritması test seti hata matrisi.



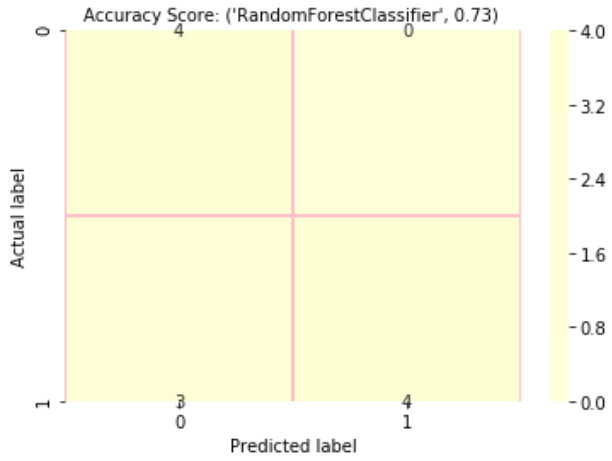
Şekil 4.30: TF-IDF yöntemi Random Forest algoritması test seti ROC eğrisi.

Doğrulma seti üzerinde gerçekleşen Random Forest Classifier sonuçlarının özetleri:

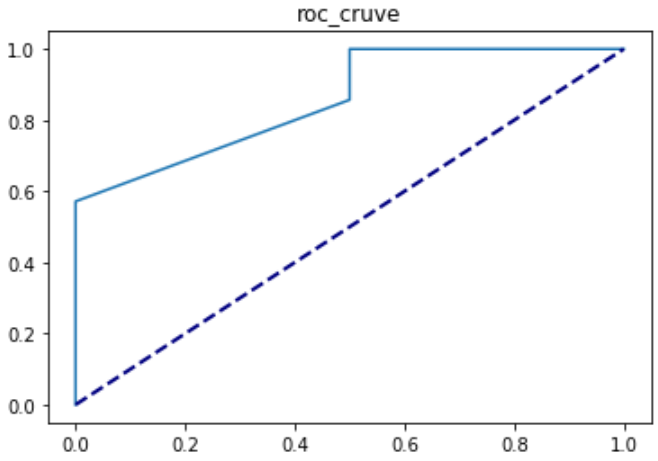
RandomForestClassifier -> ACC: %72.73

	precision	recall	f1-score	support
c_0	0.57	1.00	0.73	4
c_1	1.00	0.57	0.73	7
accuracy			0.73	11
macro avg	0.79	0.79	0.73	11
weighted avg	0.84	0.73	0.73	11

Şekil 4.31: TF-IDF yöntemi Random Forest algoritması doğrulama seti sonuç özeti.



Şekil 4.32: TF-IDF yöntemi Random Forest algoritması doğrulama seti hata matrisi.



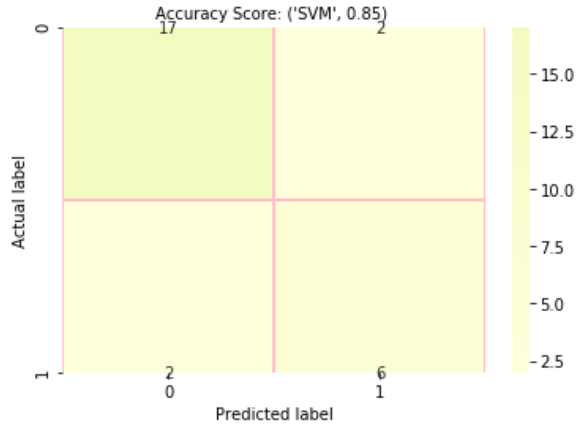
Şekil 4.33: TF-IDF yöntemi Random Forest algoritması doğrulam seti ROC eğrisi.

Test seti üzerinde gerçekleşen SVM sonuçlarının özetleri:

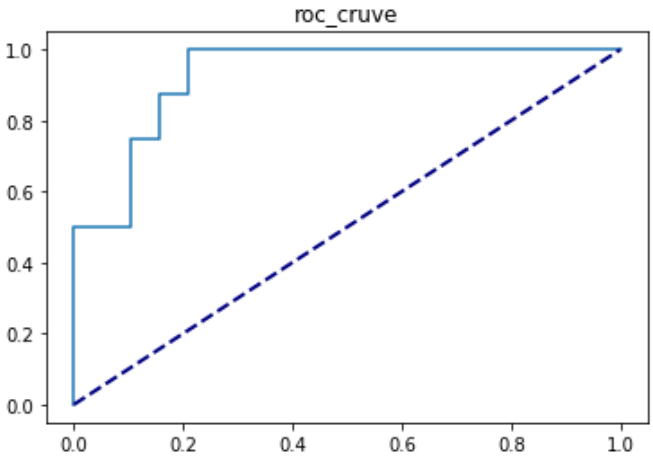
SVM -> ACC: %85.19

	precision	recall	f1-score	support
c_0	0.89	0.89	0.89	19
c_1	0.75	0.75	0.75	8
accuracy			0.85	27
macro avg	0.82	0.82	0.82	27
weighted avg	0.85	0.85	0.85	27

Şekil 4.34: TF-IDF yöntemi SVM algoritması test seti sonuç özeti.



Şekil 4.35: TF-IDF yöntemi SVM algoritması test seti hata matrisi.



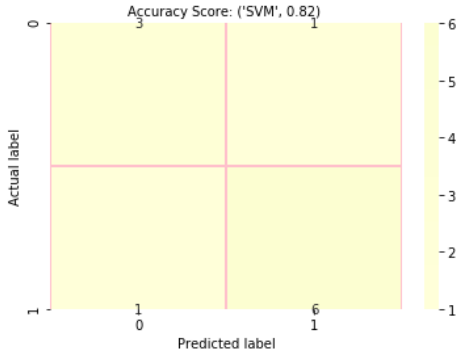
Şekil 4.36: TF-IDF yöntemi SVM algoritması test seti Roc eğrisi.

Doğrulama seti üzerinde gerçekleşen SVM sonuçlarının özetleri:

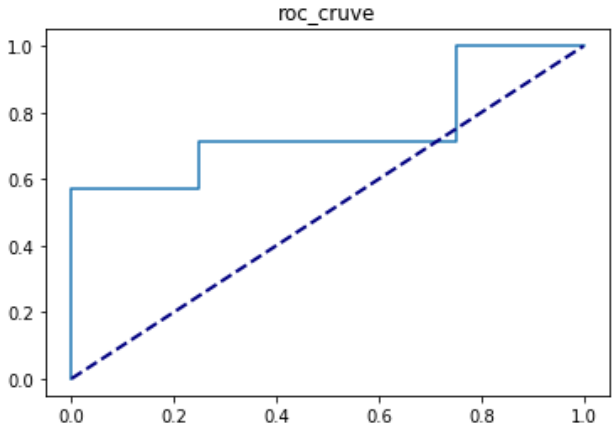
SVM -> ACC: %81.82

	precision	recall	f1-score	support
c_0	0.75	0.75	0.75	4
c_1	0.86	0.86	0.86	7
accuracy			0.82	11
macro avg	0.80	0.80	0.80	11
weighted avg	0.82	0.82	0.82	11

Şekil 4.37: TF-IDF yöntemi SVM algoritması doğrulama seti sonuç özeti.



Şekil 4.38: TF-IDF yöntemi SVM algoritması doğrulama seti hata matrisi.



Şekil 4.39: TF-IDF yöntemi SVM algoritması doğrulama seti ROC eğrisi.

Doc2vec yönteminde test ve doğrulama setlerinde en iyi sonucu veren algoritmaların sonuç özetleri:

Tablo 4.3 incelendiğinde uygulamada kullanılan on algoritmanın üçü en başarılı test seti sonuçları doc2vec yöntemimde gerçekleşmiştir. Bunlar %77,78 eşit doğruluk oranı ile KNN, MLP Classifier ve %74,07 ile Ada Boost Classifier' dır. Ayrıca Gaussian Bayes, KNN, Random Forest Classifier, MLP Classifier %77,78 oranı ile testinde setinde doc2vec yönteminin en yüksek doğruluk (accuracy) sonuçlarını elde eden algoritmalar olmuşturlar. Test setinde gerçekleştirilen bu başarı doğrulama setinde gerçekleştirilememiştir. Doğrulama setinde yöntemin en iyi doğruluk (accuracy) oranı sonuçlarını %63,64 ile Decision Tree (CART) ve Bagging Classifier gerçekleştirmiştir. Her ne kadar sayılan algoritmaların doğruluk oranı daha yüksek gerçekleşmiş olsa da talo 4.5 incelendiğinde ROC eğrisi analizinde Ada Boost Classifier test seti ve doğrulama üzerinde daha başarılı

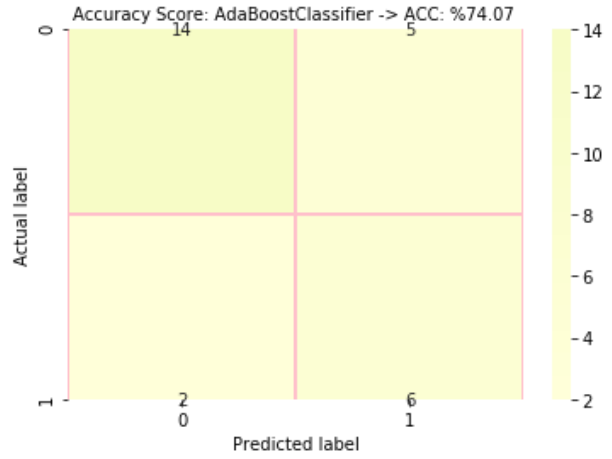
sonuçları veren algoritma olmuştur. Algoritmanın test ve doğrulma seti üzerindeki doğruluk oranları ise sırasıyla %74,07 ve %54,55 dir. Verilerin hassasiyet oranını ifade eden Recall değeri Şekil 4.40 incelendiğinde sıfır etiketli veriler için %74 bir etiketli veriler için %75 olarak gerçekleşmiş olup bu iki oran arasındaki makasın fazla açık olmaması nedeni ile daha anlamlı ROC eğrileri elde edilmiştir. Belki sayılan dört algoritmanın birlikte kullanılması ile daha iyi sonuçlar elde etmek mümkün olabilir. Ada Boost algoritmasına ilişkin özet bilgiler Şekil 4.40, 4.41 ve 4.42' de görüldüğü gibidir.

Test seti üzerinde gerçekleşen Ada Boost Classifier sonuçlarının özetleri:

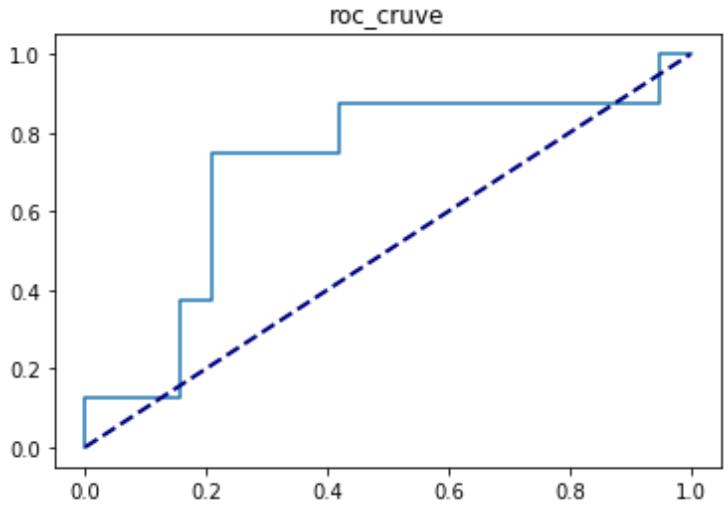
AdaBoostClassifier -> ACC: %74.07

	precision	recall	f1-score	support
c_0	0.88	0.74	0.80	19
c_1	0.55	0.75	0.63	8
accuracy			0.74	27
macro avg	0.71	0.74	0.72	27
weighted avg	0.78	0.74	0.75	27

Şekil 4.40: Doc2vec yöntemi Ada Boost algoritması test seti sonuç özeti.



Şekil 4.41: Doc2vec yöntemi SVM algoritması test seti hata matrisi.



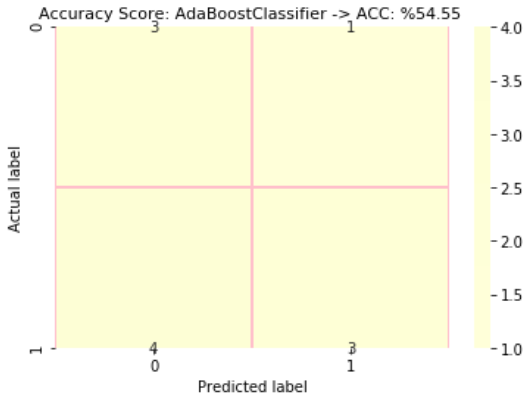
Şekil 4.42: Doc2vec yöntemi SVM algoritması test seti ROC eğrisi.

Doğrulama seti üzerinde gerçekleşen Ada Boost Classifier sonuçlarının özetleri:

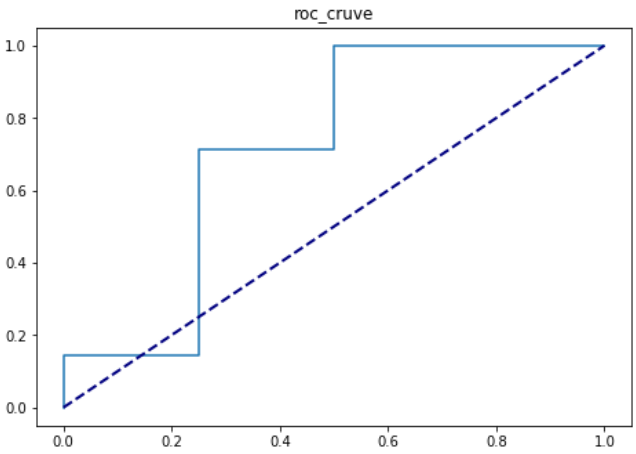
AdaBoostClassifier -> ACC: %54.55

	precision	recall	f1-score	support
c_0	0.43	0.75	0.55	4
c_1	0.75	0.43	0.55	7
accuracy			0.55	11
macro avg	0.59	0.59	0.55	11
weighted avg	0.63	0.55	0.55	11

Şekil 4.43: Doc2vec yöntemi Ada Boost algoritması doğrulama seti sonuç özeti.



Şekil 4.44: Doc2vec yöntemi Ada Boost algoritması doğrulama seti hata matrisi.



Şekil 4.45: Doc2vec yöntemi Ada Boost algoritması doğrulama seti ROC eğrisi.

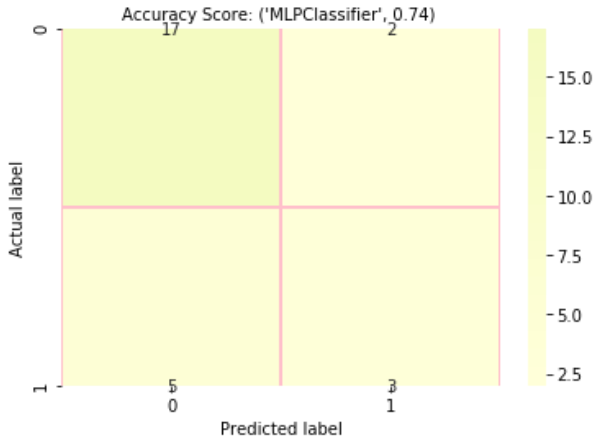
BOW yönteminde test ve doğrulama setlerinde en iyi sonucu veren algoritmaların sonuç özetleri:

Bu yöntemde test ve doğrulama setlerinde en iyi sonucu veren algoritma %74,07 ve %72,73 doğruluk oranları ile MLP Classifier olmuştur. Analize ilişkin özet bilgilere aşağıda yer verilmiştir. Uygulanan hiçbir algoritmanın en iyi doğruluk oranı bu yöntemde gerçekleşmemiştir.

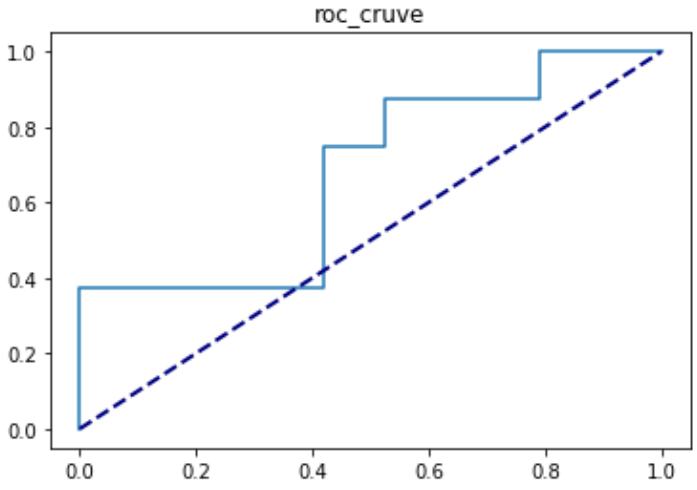
Test seti üzerinde gerçekleşen MLP Classifier sonuçlarının özetleri:

```
MLPClassifier -> ACC: %74.07
              precision    recall  f1-score   support

         c_0       0.77      0.89      0.83        19
         c_1       0.60      0.38      0.46         8

    accuracy                           0.74        27
   macro avg       0.69      0.63      0.65        27
weighted avg       0.72      0.74      0.72        27
```

Şekil 4.46: BOW yöntemi MLP Classifier test seti sonuç özeti.



Şekil 4.47: BOW yöntemi MLP Classifier test seti hata matrisi.

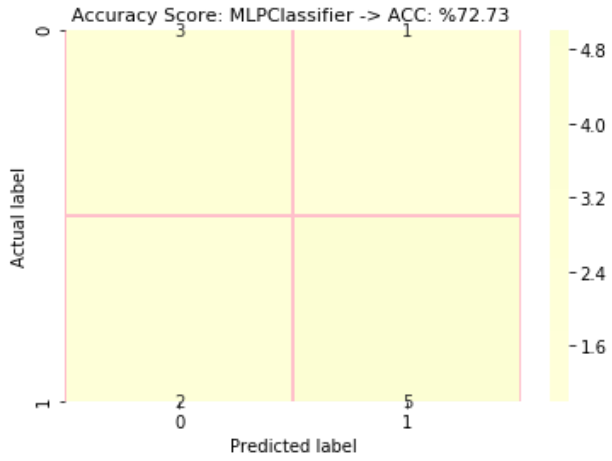


Şekil 4.48: BOW yöntemi MLP Classifier test seti ROC eğrisi.

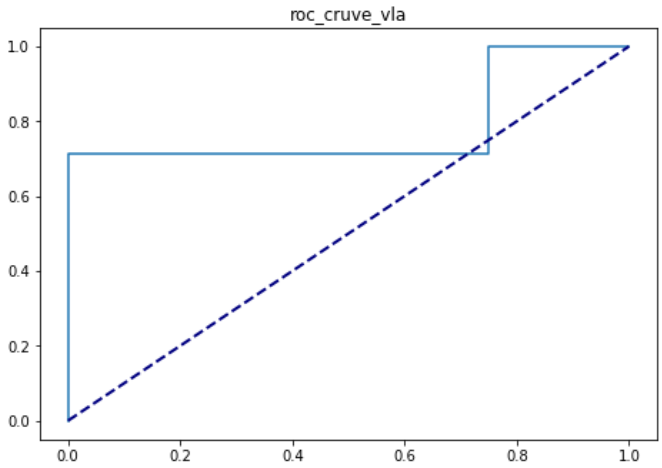
Doğrulama seti üzerinde gerçekleşen MLP Classifier sonuçlarının özetleri:

```
MLPClassifier -> ACC: %72.73
              precision    recall  f1-score   support

         c_0       0.60      0.75      0.67         4
         c_1       0.83      0.71      0.77         7

    accuracy                           0.73        11
   macro avg       0.72      0.73      0.72        11
weighted avg       0.75      0.73      0.73        11
```

Şekil 4.49: BOW yöntemi MLP Classifier doğrulama seti sonuç özeti.



Şekil 4.50: BOW yöntemi MLP Classifier doğrulama seti hata matrisi.

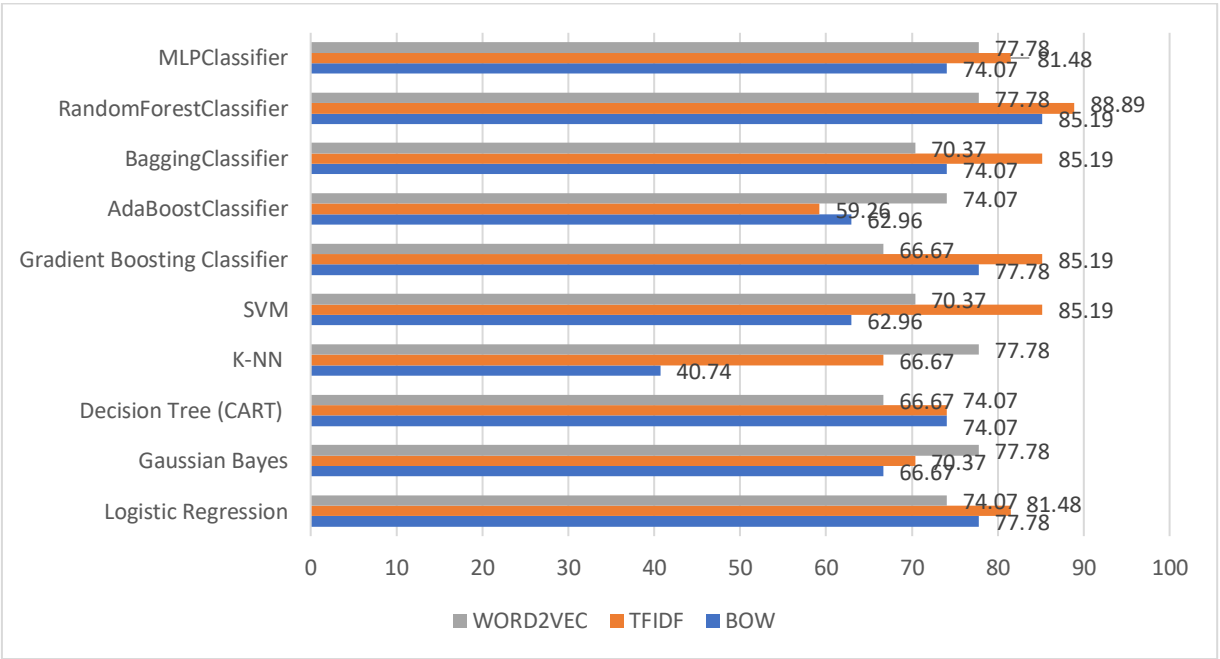


Şekil 4.51: BOW yöntemi MLP Classifier doğrulama ROC eğrisi.

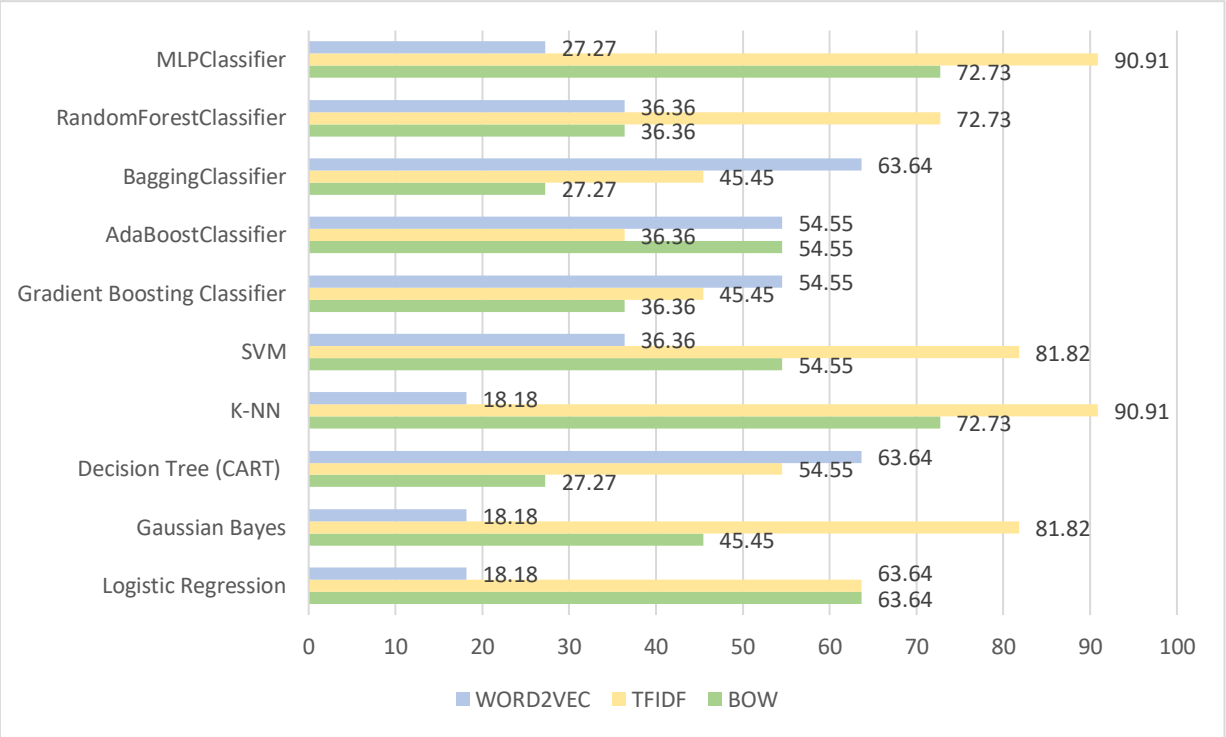
Yapılan analizler sonucunda üç (BOW, TFIDF, DOC2VEC) yöntemi ile sayısallaştırılmış dokümanlar test setinde %89 doğrulama setinde %90' a varan doğruluk (accuracy) oranları ile sınıflandırmıştır. Algoritmaların test ve doğrulama setlerine ait doğruluk oranlarının tablo ve grafiklerine aşağıda yer verilmiştir.

Tablo 4.3: Sayısallaştırma yöntemlerine göre test ve doğrulama seti doğruluk oranları.

ALGORİTMALAR	BOW		TFIDF		DOC2VEC	
	Test	Doğrulama	Test	Doğrulama	Test	Doğrulama
Logistic Regression	77,78	63,64	81,48	63,64	74,07	18,18
Gaussian Bayes	66,67	45,45	70,37	81,82	77,78	18,18
Decision Tree (CART)	74,07	27,27	74,07	54,55	66,67	63,64
K-NN	40,74	72,73	66,67	90,91	77,78	18,18
SVM	62,96	54,55	85,19	81,82	70,37	36,36
Gradient Boosting Classifier	77,78	36,36	85,19	45,45	66,67	54,55
Ada Boost Classifier	62,96	54,55	59,26	36,36	74,07	54,55
Bagging Classifier	74,07	27,27	85,19	45,45	70,37	63,64
Random Forest Classifier	85,19	36,36	88,89	72,73	77,78	36,36
MLP Classifier	74,07	72,73	81,48	90,91	77,78	27,27



Şekil 4.52: Test seti algoritmalara ait doğruluk oranları.



Şekil 4.53: Doğrulama setinde algoritmalara ait doğruluk oranları

Kullanılan BOW, TFIDF ve DOC2VEC yöntemlerine göre algoritmalar için ölçümlenen hassasiyet (Precision), hatırlama (Recall), fl-score değerleri test ve doğrulama seti üzerindeki sonuçlarına aşağıdaki tablolarda yer verilmiştir.

Tablo 4.4: Test seti güven (Precision), duyarlılık (Recall), f1-score değerleri.

ALGORİTMALAR	Sınıf Değeri	BOW			TFIDF			DOC2VEC		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Logistic Regression	c_0	0,78	0,95	0,86	0,85	0,89	0,87	0,73	1,00	0,84
	c_1	0,75	0,38	0,50	0,71	0,62	0,67	1,00	0,12	0,22
Gaussian Bayes	c_0	0,81	0,68	0,74	0,79	0,79	0,79	0,78	0,95	0,86
	c_1	0,45	0,62	0,53	0,50	0,50	0,50	0,75	0,38	0,50
Decision Tree (CAR	c_0	0,75	0,95	0,84	0,83	0,79	0,81	0,81	0,68	0,74
	c_1	0,67	0,25	0,36	0,56	0,62	0,59	0,45	0,62	0,53
K-NN	c_0	1,00	0,16	0,27	0,78	0,74	0,76	0,78	0,95	0,86
	c_1	0,33	1,00	0,50	0,44	0,50	0,47	0,75	0,38	0,50
SVM	c_0	0,74	0,74	0,74	0,89	0,89	0,89	0,70	1,00	0,83
	c_1	0,38	0,38	0,38	0,75	0,75	0,75	0,00	0,00	0,00
Gradient Boosting Cl	c_0	0,76	1,00	0,86	0,83	1,00	0,90	0,78	0,74	0,76
	c_1	1,00	0,25	0,40	1,00	0,50	0,67	0,44	0,50	0,47
Ada Boost Classifier	c_0	0,80	0,63	0,71	0,75	0,63	0,69	0,88	0,74	0,80
	c_1	0,42	0,62	0,50	0,36	0,50	0,42	0,55	0,75	0,63
Bagging Classifier	c_0	0,75	0,95	0,84	0,89	0,89	0,89	0,82	0,74	0,78
	c_1	0,67	0,25	0,36	0,75	0,75	0,75	0,50	0,62	0,56
Random Forest Clas	c_0	0,83	1,00	0,90	0,94	0,89	0,92	0,84	0,84	0,84
	c_1	1,00	0,50	0,67	0,78	0,88	0,82	0,62	0,62	0,62
MLP Classifier	c_0	0,77	0,89	0,83	0,89	0,84	0,86	0,76	1,00	0,86
	c_1	0,60	0,38	0,46	0,67	0,75	0,71	1,00	0,25	0,40

Tablo 4.5: Doğrulama seti hassasiyet (Precision), hatırlama (Recall), f1-score değerleri.

ALGORİTMALAR	Sınıf Değeri	BOW			TFIDF			DOC2VEC		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Logistic Regression	c_0	0,50	0,75	0,60	0,50	1,00	0,67	0,22	0,50	0,31
	c_1	0,80	0,57	0,67	1,00	0,43	0,60	0,00	0,00	0,00
Gaussian Bayes	c_0	0,33	0,50	0,40	0,67	1,00	0,80	0,14	0,25	0,18
	c_1	0,60	0,43	0,50	1,00	0,71	0,83	0,25	0,14	0,18
Decision Tree (CART)	c_0	0,17	0,25	0,20	0,40	0,50	0,44	0,50	1,00	0,67
	c_1	0,40	0,29	0,33	0,67	0,57	0,62	1,00	0,43	0,60
K-NN	c_0	1,00	0,25	0,40	1,00	0,75	0,86	0,14	0,25	0,18
	c_1	0,70	1,00	0,82	0,88	1,00	0,93	0,25	0,14	0,18
SVM	c_0	0,40	0,50	0,44	0,75	0,75	0,75	0,36	1,00	0,53
	c_1	0,67	0,57	0,62	0,86	0,86	0,86	0,00	0,00	0,00
Gradient Boosting Classifier	c_0	0,29	0,50	0,36	0,33	0,50	0,40	0,44	1,00	0,62
	c_1	0,50	0,29	0,36	0,60	0,43	0,50	1,00	0,29	0,44
Ada Boost Classifier	c_0	0,00	0,00	0,00	0,33	0,75	0,46	0,43	0,75	0,55
	c_1	0,60	0,86	0,71	0,50	0,14	0,26	0,75	0,43	0,55
Bagging Classifier	c_0	0,17	0,25	0,20	0,33	0,50	0,40	0,50	1,00	0,67
	c_1	0,40	0,29	0,33	0,60	0,43	0,50	1,00	0,43	0,60
Random Forest Classifier	c_0	0,36	1,00	0,53	0,57	1,00	0,73	0,20	0,25	0,22
	c_1	0,00	0,00	0,00	1,00	0,57	0,73	0,50	0,43	0,46
MLP Classifier	c_0	0,60	0,75	0,67	1,00	0,75	0,86	0,25	0,50	0,33
	c_1	0,83	0,71	0,77	0,88	1,00	0,93	0,33	0,14	0,20

5. SONUÇ VE GELECEK ÇALIŞMALAR

Metin madenciliği yeni bir veri madenciliği alanıdır. Veri madenciliğinin ilişkili olduğu diğer bilgisayar bilimleri (makine öğrenmesi, doğal dil işleme, derin öğrenme vs.) ile uyumlu çalışabilmesi kullanım alanlarını ve yapılan analizlerin etkinliğini artırmaktadır. Büyük miktarlardaki yapısal olmayan dataların sahip olduğu gizli bilgileri keşfetmek amacıyla kullanılmaktadır. Keşfedilen ilginin kullanıcıları iş insanları, bilim insanları, tüketiciler olabileceği gibi ülkeler, politika yapıcılar, ulusalar arası kurumsal organizasyonlar, kamu kurumları olabilmektedir.

Bu çalışmada öncelikle bir hukuk veri tabanından mobbing içerikli Yargıtay kararları elde edilmiştir. Ardından karalar mobbingin varlığının kabulü ve kabul edilmemesi durumuna göre etiketlenmiştir. Etiketlenen kararlar ön işleme adımlarından geçirilmiştir. Elde edilen etiketli temizlenmiş veri seti makine öğrenmesi yöntemleri ile gözetimli öğrenmeye tabi tutulmuştur. Metinler sayısallaştırılırken kelime torbaları, TF-IDF ve Doc2Vec yöntemleri kullanılmıştır. Çalışmada mobbing içerikli mahkeme kararlarının otomatik sınıflandırılması için bir model oluşturulması amaçlanmıştır. Modelin sınıflandırma başarısı Tablo 4.3, 4.4, 4.5'de gösterilmiştir. Model üç metin sayısallaştırma yönteminde ve test ile doğrulama setlerinde en yüksek başarıyı Random Forest Classifier (%89), SVM (%82), MLP Classifier (%74), Ada Boost (%74) algoritmaları ile elde etmiştir.. Böyle bir sınıflandırmanın yapılabilir olması dava dosyalarının yargı süreci sonucunda nasıl sonuçlanabileceğine ilişkin bir sınıflandırmanın modellenebileceğini göstermiştir.

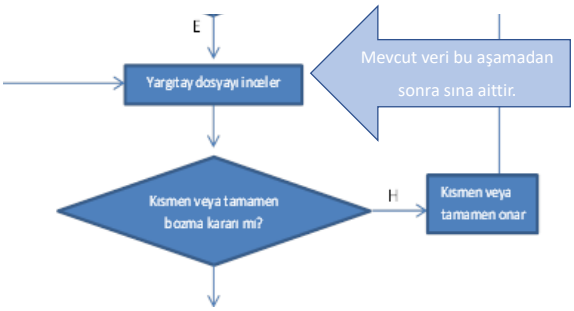
Bu modelin gerçek hayat problemlerinde uygulanma faydaları yargıda iş yükünün hafifletilmesi ve efektif karar verme mekanizmalarının geliştirme yönünde olacağı düşünülmektedir.

Dava dosyalarını hazırlarken gerek savunmalarını hazırlayan müdafi avukatları gerekse iddianamelerini hazırlayan müşteki avukatları emsal kararları tarayarak davanın gidişatını etkiletilmektedirler. Bu çalışma dava sonucunda verilebilecek kararın tam olarak kestirilebileceğini iddia etmemek ile birlikte emsal kararların makine öğrenmesi yöntemleri ile işlenmesi durumunda karar verici ve icracılara olası gidişata ilişkin yararlı

bilgileri sağlanabileceğini göstermektedir. Kaldı ki bu emsal kararlara avukatlar bireysel çabaları ile ulaşmakta ve dava ile ilgili karaların tümüne ulaşıp yararlı olanları seçebilmeleri her durumda mümkün olmamaktadır. Anahtar kelimeler ile taranan bu kararların arka planda bir makine öğrenmesi algoritması ile sınıflandırmaya tabi tutulduğu bir ara yüz programı ile kullanıcısının faydalı bilgilere erişiminin sağlanabilecek bir model oluşturularak hukuk sisteminde verimin artırılması mümkün görülmektedir.

Ayrıca ilgili kanununda alternatif çözüm yolları olarak tanımlanan arabuluculuk, uzlaştırmacılık gibi hukuksal çözüm yolları mevcuttur. Arabuluculuk belirli lisanslara sahip hukuk fakültesi lisans mezunları ve kanunda belirtilen diğer şartlara uygun kişiler tarafından yapılır. Uzlaştırmacılık hukuk fakültesi lisans mezunları ya da kanunda tahdidi olarak sayılan diğer fakültelerinin en az lisans düzeyinde mezunları tarafından icra edilir. Her iki çözüm yolunun uygulayıcıları da tarafları dava yoluna gitmeden sulh olmaya ikna ederken kendilerine intikal etmiş dava konuları ile ilgili mahkemelerin daha önce verdikleri karalardan faydalanmaktadır. Önerilen sistem bu hali ile uygulayıcılara fayda sağlayabilecek durumdadır. Sistemin geliştirilebilmesi durumunda bu çözüm yollarına ek olarak sistem üzerinden otomatik kararlar yazılabileceği makine öğrenmesi temelli yeni bir alternatif çözüm yolunun oluşturulabileceği öngörülmektedir.

Mevcut kullanılan veriler Şekil 3.4’ de verilen iş akış şemasının Şekil 5.1’ de gösterilen basamağından sonrası yazılan kararlardan oluşmaktadır. Dosya içeriğinde bulunan talep dilekçeleri, deliller, şahit ifadelerini içeren duruşma tutanakları, yerel mahkeme karalarının ayrıntıları gibi verilere erişim sağlanamamaktadır. Bu verilere ulaşılması halinde çalışmanın daha sağlıklı olacağı ve yeni boyutlar kazanacağı öngörülmektedir.



Şekil 5.1: Dosyanın Yargıtay'a ulaşması aşaması.

Gelecek çalışmalarda Şekil 5.2' de verile dava dosyalarının, ilk derece mahkemesine verilen ve yargı sürecini başlatan dava talep dilekçesi aşamasından başlayarak kararın kesinleştiği aşamaya kadar takip edilebileceği veriler elde edilerek mahkeme kararlarının otomatik tahmini üzerine çalışılacaktır.

Şekil 5.2: Dosyanın, ilk derece mahkemesine gönderilmesi aşaması.

Sonuçlar ayrıca çok disiplinli bir çalışma grubu oluşturarak davacı ve davalı tarafların eğitim, yaş, cinsiyet, aile durumu, çalışma düzeyi, medeni durumu sağlık bilgileri gibi verilere ulaşılabilmesi durumunda taraf profilleri çıkarılarak dosyada iddia edilen eylemlere yatkınlık düzeyleri, yapılan fiilleri tekrarlama eğilimleri gibi bilgilerin önceden kestirilebileceği kanısını oluşturmaktadır.

6. KAYNAKLAR

[1] N. S. Kramer, *Tarih sümerde başlar*. İstanbul: Kabalcı, 2019.

[2] K. P. Kolhe and R. K. Hingole, *Elements of casting technology*. Lap Lambert Academic Publishing, 2017.

[3] L. Sijia, "Algorithms for Relation Extraction from Biomedical Texts", Doctoral dissertation, The State University of New York, Buffalo, 2018. [Online].Available: https://search.proquest.com/docview/2194375787/200B6E0C5CCE4E86PQ/1?accountid=15410

[4] S. H. Binkheder, "Biomedical Literature Mining and Knowledge Discovery of Phenotyping Definitions", Doctoral dissertation, Indiana University, Indiana, 2019. [Online].Available: https://scholarworks.iupui.edu/handle/1805/20201

[5] M. M'Bareck,Lemine, "Political speech on twitter: a sentiment analysis of tweets and news coverage of local gun policy", Doctoral dissertation, University of Arkansas, Arkansas , 2019. [Online].Available: https://search.proquest.com/docview/2217120730/CFAA5F84C61644FEPQ/1?accountid=15410

[6] K.Toprak, "Metin madenciliği yöntemleri kullanarak illere göre haber analizi", Yüksek lisans tezi, Karadeniz Teknik Üniversitesi, Trabzon, 2018.

[7] M. A. Hamde, "Kurumsal belgelere (Metin verilerine) metin madenciliği tekniği ile erişimin değerlendirilmesi",Yüksel lisans tezi, İstanbul Üniversitesi, İstanbul, 2018.

[8] M. C. Tekin, "Yazılım geliştirme taleplerinin metin madenciliği ile sınıflandırılması ve önceliklendirilmesi", Yüksek lisans tez, İstanbul Maltepe Üniversitesi, İstanbul, 2018.

[9] F. G. Tan, "Metin madenciliği teknikleri ile sosyal ağlarda bilgi keşfi", Yüksek lisans tezi, Süleyman Demirel Üniversitesi, Ispartta, 2018.

[10] S. Atan and Y. Çınar, "Borsa istanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (sentiment) analizi ile incelenmesi", *Ankara Üniv. SBF Derg.*, cilt 74, no 1, 1–34, 2019, [Online]. Available: https://dergipark.org.tr/en/download/article-file/642997.

[11] H. Göker and H. Tekdere, "FATİH projesine yönelik internet yorumlarını metin madenciliği yöntemleri ile otomatik tespiti", *BTD*, cilt 10, no 3, 2017, [Online] Erişim adresi: 10.17671/gazibtd.331041.

[12] E. B. Yalçın and G. Y. Erduran, "Öğrencilerin bireysel sorumluluklarına bakış açılarının metin madenciliği ile analizi", *Akad. Bakış*, cil 66, 113–121, 2018, [Online] Erişim adresi: https://scholar.google.com.tr/citations?user=huwrt50AAAAJ&hl=tr#d=gs_md_cita-d&u=%2Fcitations%3Fview_op%3Dview_citation%26hl%3Dtr%26user%3Dhuwrt50AAAAJ%26citation_for_view%3Dhuwrt50AAAAJ%3Au-x6o8ySG0sC%26tzom%3D-180.

[13] M. Cecchini, "Quantifying the risk of financial events using kernel methods and information retrieval", *ProQuest Diss. Theses*, 2005. [Online].Available: https://www.semanticscholar.org/paper/Quantifying-the-risk-of-financial-events-using-and-Cecchini-Koehler/aa6be1e0456f17c36491d9c5e23fa4af84a09322

[14] O. Yıldız, "Metin madenciliğinde anahtar kelime seçimi bir üniversite örneği", *Yönetim Bilişim Sist. Derg.*, cilt 2, no 1, 29–50, 2016, [Online] Erişim adresi: https://dergipark.org.tr/en/download/article-file/347705.

[15] B. Drury and M. Roche, "A survey of the applications of text mining for agriculture", *Computer Electron Agr.*, vol.163, August 2019, doi: 10.1016/j.compag.2019.104864.

[16] T.Yıldız, "Examining the concept of industry 4.0 studies using text mining and scientific mapping method", *Procedia Comput SCI*, Vol. 158, pp. 498-507, 2019, doi.org/10.1016/j.procs.2019.09.081

[17] K. He *et al.*, "Understanding the patient perspective of epilepsy treatment through text mining of online patient support groups", *Epilepsy Behav*, vol. 94, pp. 65–71, 2019, doi: 10.1016/j.yebeh.2019.02.002.

[18] E. Chaix, L. Deléger, R. Bossy, and C. Nédellec, "Text mining tools for extracting information about microbial biodiversity in food", *Food Microbiol*, vol. 81, pp. 63–75, 2019, doi: 10.1016/j.fm.2018.04.011.

[19] E. Kano, Y. Fujita, and K. Tsuda, "A method of extracting and classifying local community problems from citizen-report data using text mining", *Procedia Comput SCI*,vol. 159, pp.1347-1356, 2019, doi: 10.1016/j.procs.2019.09.305.

[20] A. P. C. Júnior *et al.*, "Ontology applied in the judicial sentences", *Conf. 2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, October 2017, doi: 10.1109/CHILECON.2017.8229731.

[21] O. Metsker, E. Trofimov, S. Sikorsky, and S. Kovalchuk, "Text and data mining

techniques in judgment open data analysis for administrative practice control", in *5th International Conf. EGOSE,* St. Petersburg, Russia, November 2018. doi: 10.1007/978-3-030-13283-5_13.

[22] N. Aletras, D. Tsarapatsanis, D. Preoţiuc-Pietro, and V. Lampos, "Predicting judicial decisions of the European court of human rights: A natural language processing perspective", *PeerJ Comput. Sci.*, 2016, doi: 10.7717/peerj-cs.93.

[23] S. Thammaboosadee and U. Silparcha, "A framework for criminal judicial reasoning system using data mining techniques", in *2nd IEEE International Conference on Digital Ecosystems and Technologies,* 2008, doi: 10.1109/DEST.2008.4635219.

[24] A. Sagun, "İşyerinde psikolojik tacizin (mobbing) hukuksal temelleri ve sonuçları", Yüksek lisans tezi, Gazi Üniversitesi, Ankara, 2015.

[25] F. Çopur, "Mobbingin çalışanlar üzerindeki etkileri: Türk Hukuk Sistemi'nde mobbing", Yüksek lisans tezi, Süleyman Demirel Üniversitesi, Isparta, 2017.

[26] G. Akman, "Türkiye'de mobbingle ilgili düzenlemeler ve bir kamu üniversite hastanesinde taşeron çalışanlara yönelik mobbing," Yüsek lisans tezi, Gazi Üniversitesi, Ankara, 2014.

[27] N. Bilge, "Mobbingin mağdur, aile, örgüt ve toplum üzerindeki etkileri", Yüksek lisans tezi, İstanbul Atılım Üniversitesi, İstanbul, 2014.

[28] K. Ergün, "Metin madenciliği yöntemleri ile ürün yorumlarının otomatik değerlendirilmesi," Dktora tezi, Sakarya Üniversitesi, Sakarya, 2012.

[29] A. Hotho, A. Nürnberger, and G. Paaß, "A Brief Survey of Text Mining," *LDV Forum,* vol. 20, no. 1, pp. 19–62, 2005, doi: 10.1111/j.1365-2621.1978.tb09773.x.

[30] L. Francis and M. Flynn, "Text Mining Handbook", *Casualty Actuarial Society E-Forum,* Spring, 2010. [Online]. Available: https://www.casact.org/pubs/forum/10spforum/Francis_Flynn.pdf

[31] H. Mannila, "Data mining: Machine learning, statistics, and databases", in *8th International Conference on Scientific and Statistical Data Base Management,* Stockholm, Sweden, June 1996, doi: 10.1109/SSDM.1996.505910.

[32] H. Ahonen, O. Heinonen, M. Klemettinen, and A. I. Verkamo, "Applying data mining techniques for descriptive phrase extraction in digital document collections", in *Research and Technology Advances in Digital Libraries,* Santa Barbara, CA, USA , April 1998, doi: 10.1109/adl.1998.670374.

[33] E. Pasin, "Investigation of text mining methods on turkish text", Yüjsek lisans tezi,

Dokuz Eylül Üniversitesi, İzmir, 2018.

[34] F. Popowich, "Using text mining and natural language processing for health care claims processing", *SIGKDD Explor Newsl*, vol. 7, no. 1, pp. 59–66, 2005, doi: 10.1145/1089815.1089824.

[35] C. Clifton, "Data mining", *Encyclopedia Britannica*. [Online]. Available: https://www.britannica.com/technology/data-mining.

[36] S. V. Gaikwad, A. Chaugule, and P. Patil, "Text Mining Methods and Techniques", *IJCA*, vol. 85, no. 17, pp. 42–45, 2014, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.428.8805&rep=rep1&type=pdf.

[37] K. Ergün, "Metin Madenciliği, Bilgiye Erişim ve Bilgi Çıkarımı", *Rapid Miner ile Uygulamalı Veri Madenciliği*, Pusula Yayıncılık, 2017, pp. 250–265. [Online] Erişim adresi: http://kergun.baun.edu.tr/veri_madenciligi_hafta11.pdf

[38] S. Niharika, V. S. Latha, and D. R. Lavanya, "A survey on text categorization", *IJCTT*, vol. 1, no. 1, pp. 39-45, 2012. [Online]. Available: http://www.ijcttjournal.org/Volume3/issue-1/IJCTT-V3I1P108.pdf

[39] W. Lam, M. Ruiz, and P. Srinivasan, "Automatic text categorization and its application to text retrieval," *IEEE T Knowl Data En*, vol. 11, no. 6, pp. 865–789, 1999, doi: 10.1109/69.824599.

[40] J. Sivic and A. Zisserman, "Efficient visual search of videos cast as text retrieval," *IEEE Trans. Pattern Anal*, vol. 31, no. 4, pp. 591–606, 2009, doi: 10.1109/TPAMI.2008.111.

[41] Ö. Doğan, "Derin öğrenme nedir? Yapay sinir ağları ne işe yarar?" [Online] Erişim adresi: https://teknoloji.org/derin-ogrenme-nedir-yapay-sinir-aglari-ne-ise-yarar/.

[42] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space", *arXiv*, vol. 3, pp.1-12, January 2013. [Online]. Available: https://arxiv.org/pdf/1301.3781.pdf

[43] M. T. Sübay, "Türkçe kelime vektörlerinde görülen anlamsal ve biçimsel yakınlaşmalar", Yüksek lisans tez, İstanbul Maltepe üniversitesi, İstanbul, 2019.

[44] "Python Word Embedding using Word2Vec." https://www.geeksforgeeks.org/python-word-embedding-using-word2vec/.

[45] T. Mitchell, *Machine learning*. Portland: McGraw-Hill, 1997.

[46] B. Marr, "How quantum computers will revolutionize artificial ıntelligence,

machine learning and big data", *Forbes*, 2017, [Online]. Available: https://www.forbes.com/sites/bernardmarr/2017/09/05/how-quantum-computers-will-revolutionize-artificial-intelligence-machine-learning-and-big-data/#2063dd85609b.

[47] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms", in *3rd International Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, March 2016, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7724478.

[48] J. Fiaidhi and K.-C. Yow, "A Review of Machine Learning Algorithms for Text-Documents Classification," *JAIT*, vol. 1, no. 1, 2010, [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/30773019/jait0101.pdf?1362339584=&response-content-disposition=inline%3B+filename%3DJournal_of_Advances_in_Information_Techn.pdf&Expires=1600148703&Signature=AjlppCoysGqUZWcO83I6zqwTuuY5r7W3ykjKl6FJioYFqHtcZik13KEvjJ5Du82Yv5ZqBXXxUXLv2F6TKuxt7K4aIOKxTEXaPnVoDvy89otjxcBhMUCJ92ZVQhPPFRG8LrI-X6LjsJJ~PX4bkLJnbrHkEmGxU0iifV6xdLqWqMiKCoUFBU4JjIG3aB07TloYKcoy8z0c-LDwvQs~Xgytjs5mTM3EMcF6sKiI-iVwvIxY0tV9jltO0DmMrMihPSXREQ4mFEkK1GV1QXNC9kZZIZAboechNUl5zM~5lWwOavCJyOJf1O6hzBUQ-n~40pRcnb98aMyQ8~xTwG0cItK~dA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA#page=6.

[49] X. Zhu, "Semi-Supervised Learning Literature Survey," 200AD. [Online]. Available: https://minds.wisconsin.edu/handle/1793/60444.

[50] N. Gürsakal, *Makine öğrenmesi: makine öğrenmesi ve derin öğrenme*. Bursa: Dora, 2017.

[51] P. Tınaz, H. Ergin, and F. Bayram, *Çalışma psikolojisi ve hukuki boyutlarıyla işyerinde psikolojik taciz*. istanbul: Beta, 2008.

[52] K. Lorenz, *The natural science of the human species: an introduction to comparative behavioural research*. Lodon: MIT Pres, 1997.

[53] P. C. Sexton and C. M. Brodsky, "The Harassed Worker.," *Ind. Labor Relations Rev.*, 1977, doi: 10.2307/2522527.

[54] H. Leymann, "The content and development of mobbing at work," *Eur. J. Work Organ. Psychol.*, 1996, doi: 10.1080/13594329608414853.

[55] *Yargıtay 22. Hukuk dairesinin 22.05.2014 tarihli 2013/11788 esas sayılı*

2014/14008 nolu karar. 2014.

[56] "Yargıtay," *Yargıtay*. https://www.yargitay.gov.tr/isbolumu.

[57] "Hukuk usulü muhakemeleri kanunu", [Online] Erişim adresi: https://www.alomaliye.com/1927/06/18/1086-sayili-kanun-hukuk-usulu-muhakemeleri-kanunu/.

[58] E. R. Ziegel, M. Stokes, C. Davis, and G. Koch, "*Categorical Data Analysis Using the SAS Syste,,*" SAS Institute, USA, 1996, doi: 10.2307/1271334.

[59] D. R. Cox, "Some Remarks on Overdispersion," *Biometrika*, vol.70, no. 1, pp. 269-274, 1983, doi: 10.2307/2335966.

[60] "Naive Bayes," *Skylearn libbrary*. https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes.

[61] G. Silahtar, *Sınıflandırma teknikleri ve algoritmaları: Veri madenciliği kavram ve algoritmaları*. İstanbul: Papatya, 2016.

[62] "K Neighbors Classifier," *Sklearn library*. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html.

[63] "Ensemble methods," *Sklearn library*. https://scikit-learn.org/stable/modules/ensemble.html.

[64] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches", *IEEE T Syst Man and Cy Part C: Applications and Review*s. Vol. 42 , no. 4 , July 2012, doi: 10.1109/TSMCC.2011.2161285.

[65] Y. Freund, "A short introduction to boosting", *JSAI*, vol. 15, no. 5, pp. 771–780, 1999, [Online]. Available: https://cseweb.ucsd.edu/~yfreund/papers/IntroToBoosting.pdf.

[66] S. Ozker, "Boosting algoritmaları nasıl çalışır?" https://medium.com/@sertacozker/boosting-algoritmaları-nasıl-çalışır-edac1174e971.

[67] T. Hastie, R. Tibshirani, and J. Friedman, *Elements of Statistical Learning 2nd ed.* 2009.

[68] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial", *Front. Neurorobot.*, vol. 7, no. 21, december 2013, doi: 10.3389/fnbot.2013.00021.

[69] "MLP Classifier," *Sklearn library*. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html.

[70] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient

(MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, vol.21, no. 6, pp.2-13, 2020, doi: 10.1186/s12864-019-6413-7.

[71] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sens Env.*, vol. 62, no. 1, pp. 77–89, 1997, [Online]. Available: https://doi.org/10.1016/S0034-4257(97)00083-7.

[72] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, Vol. 27, no. 8, pp. 861-874, June 2006, 2006, doi: 10.1016/j.patrec.2005.10.010.

[73] A. Fraser and D. Marcu, "Measuring word alignment quality for statistical machine translation," *Comput. Linguist*, vol. 33, no. 3, pp. 293–303, 2007, doi: 10.1162/coli.2007.33.3.293.

[74] "F1 score," *wikipedia*. https://en.wikipedia.org/wiki/F1_score.

[75] D. M. W. Powers, "Evaluation: from precision, recall and f-factor," *Tech. Rep. SEI-07-001*, pp. 1–24, January 2007.

ÖZGEÇMİŞ

Kişisel Bilgiler

Ad Soyad	:Özlem AYDIN
Doğum Tarihi ve Yeri	: 01.01.1984 - Elazığ
E-posta	:ozlem.aydin@balikesir.edu.tr

Öğrenim Bilgileri

Derece	**Okul/Program**	**Yıl**
Yüksek Lisans	Balıkesir Üniversitesi/ Endüstri Mühendisliği	2020
Lisans	Kafkas Üniversitesi/İktisadi ve İdari Bilimler Fakültesi	2006