Survey

Metin Madenciliği Tekniklerinin Spam Tespitinde Kullanımı: Literatür Taraması

Ayşe Irmak Erçevik

Bilgisayar Müh. Bölümü, TOBB ETÜ, Ankara, Türkiye E-mail: ayseirmakercevik@etu.edu.tr

Özet

Bu çalışmada spam tespitinde metin madenciliği tekniklerinin kullanımı incelenmektedir. Çalışmada sosyal ağ platformlarında karşılaşılan yorum (inceleme) spamları, e-posta spamları ve SMS spamlarının tespiti kapsamında yapılan çalışmaların amaçları, metin önişleme, öznitelik çıkarma ve sınıflandırma yöntemleri, kullanılan veri setleri, alınan sonuçlar ve elde edilen başarımları özetlenmiştir. Yapılan çalışma sonucunda, spam tespitinde davranışsal, dilsel ve anlamsal olmak üzere üç farklı öznitelik çıkarım yaklaşımının tercih edildiği görülmektedir. 2000'li yıllarda başlayan metin madenciliği uygulamalarında son yıllarda anlamsal özniteliklerin ön plana çıktığı ve çalışmaların doğal dil işleme yöntemine doğru kaydığı görülmektedir. Metin madenciliği teknikleri ise gittikçe ulaşılabilir olan kaliteli veri setleri ve hibrit kullanımlar ile belirli amaçlar dahilinde maliyet-etkin sonuçlar üretme, bilgi çıkarımı ve doğal dil işleme gibi teknolojileri destekleme potansiyelini sürdürmektedir.

Anahtar Sözcükler: metin madenciliği, spam tespiti, metin madenciliği teknikleri, metin önişleme, öznitelik çıkarma, sınıflandırıcı, doğal dil işleme.

1. Giriş

Internet üzerindeki hemen her etkileşimi olumsuz etkileme potansiyeline sahip olan spamlar, gereksiz kaynak tüketimine, verimsizliğe, güvenlik riskleri ve kayıplarına, dolandırıcılık, vb. gibi zararlara yol açmaktadır. Günümüzde bir spam ekonomisi oluşmuş olup, spam tespitinin ve spam filtrelemesinin gerekliliği güncelliğini korumaktadır.

Spam türleri

Spamlar; yorum spamları, WEB spamları, e-posta spamları ve SMS spamları olarak sınıflandırılabilmektedir. (Dixit, 2016). Facebook, Twitter, YouTube, vb. sosyal medya sitelerinde kullanıcıların dikkatini başka bir yöne çekmek, bir ürün veya konuyu pazarlamak (Kaddoura, 2022) abartılı bir şekilde öne çıkarmak, iftiraya varan şekilde eleştirmek veya karalamak amacıyla kullanılan, görüş bildirimleri İngilizce literatürde "review spam" olarak geçmektedir. Bu çalışmada "review spam", Türkçe literatürde "eleştiri" ve "görüş" anlamlarını daha iyi karşıladığı için "yorum spamı" olarak Türkçeleştirilmiştir. (Jindal, 2008), yorum spamlarını 3 tipte değerlendirmektedir. Tip 1 (hak edilmemiş görüşler): hak etmeyen olumlu eleştiriler veya itibarı zedelemek için kötü niyetli olumsuz yorumlar. Tip 2 (yalnızca markalarla ilgili yorumlar): ürünü değilde markayı, üreticiyi veya satıcıyı hedefleyen yorumlar. Tip 3 (yorum olmayanlar): yorum veya görüş içermeyen alakasız paylaşımlar (örn. Sorular, reklamlar). Web spam ise bazı web sayfalarını hak ettiklerinden daha yüksek sıralamak için arama motorlarını yanıltma amacıyla siteye alakasız kelimeler eklenmesi, içerik veya bağlantılar eklenen bir spam türüdür. E-posta ve

SMS spamları genellikle, kullanıcının bir talebi olmaksızın, toplu olarak reklam amacı ile gönderilen ve belli oranda da sahte reklamlara tıklanması suretiyle truva atı, casus yazılım, fidye yazılımları (https://www.statista.com/statistics/420391/spam-email-traffic-share/) gibi kötü amaçlı yazılımları yaymak için kullanılan güvenlik riski olan mesajlardır.

Spamların tespit edilerek filtrelenmesi gerek kişisel gerek kurumsal kullanıcılar için bir zorunluluk haline gelmiştir. Kimi iş süreçlerinde gelen spamlar gelen mesajlar arasından öncelikle filtrelenip, gerçek mesajlar işleme tabi tutulurken bazı uygulamalarda mesajlar önce sınıflandırılmakta ve sınıflandırılmış özeliklerine göre farklı spam tespiti yöntemlerine tabi tutulmaktadır. (Zhu, 2011) deki Ürün Yorum Madenciliği Sisteminden de görüldüğü gibi; sosyal ağ sitelerinde tüketici/müşteri yorumlarının etkin bir şekilde işleme alınabilmesi ve detaylı şekilde analiz edilmesi için yorum spamlarının tespit edilerek ve filtrelenmesi ilk aşamada yapılan bir faaliyet olarak göze çarpmaktadır.

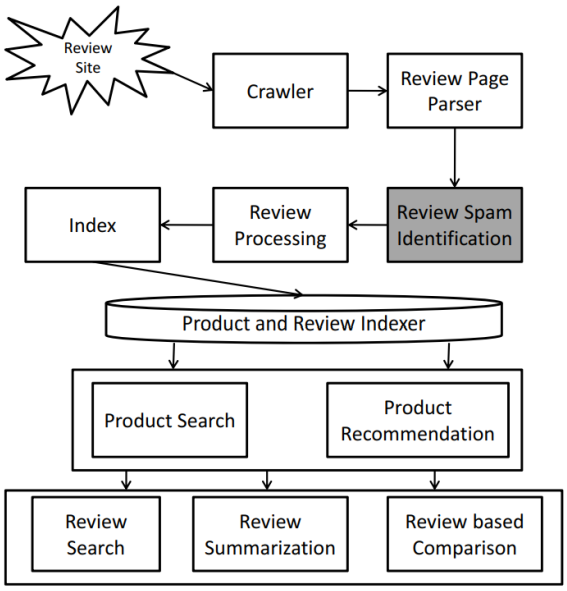

Figure 1. A Framework for Product Review Mining (Zhu, 2011)

Metin madenciliği ve spam tespiti

Metin madenciliği kuruluşun üretkenliğini artırma, istenmeyen e-postaları filtreleme, e-posta sınıflandırması, belge etiketleme, dolandırıcılık tespiti, garanti talebi işleme, rakip analizi, sosyal medyada tahmine dayalı modeller oluşturma vb. amaçlarla kullanılmaktadır. (Ranjan, 2021)

Yorum spamlarının tespitinde gönderilen metnin içeriğine, üzerine yorum yapılan nesneye/konuya ve göndericiye (spammer'a) göre farklı yöntemler uygulanabilmektedir. (Dixit, 2016), (Jindal, 2008)

Literatürde bazı çalışmalarda NLP gibi anlam analizine dayalı yöntemlerin metin madenciliği süreci içerisinde tanımlandığı görülmektedir. (Tandel, 2019), (Maheswari, 2015), (Gupta, 2020) Metin madenciliğinde karar ağacı kategorizasyonu, kümeleme, kategorizasyon gibi sıklıkla kullanılan tekniklerin (Lau, 2011) yanında doğal dil işleme gibi tekniklerin metin madenciliği sürecinde karşılaşılan sorunları azaltmada yardımcı olabileceği görülmektedir. (E.Zhong, 2012)

2. İncelenen Yayınların Seçimi

Yayın seçiminde aşağıdaki kriterler dikkate alınmıştır

a) Spam tespitinde metin madenciliği tekniklerinin kullanılmış olması,

b) Yayınlarda

i) kullanılan metin madenciliği model ve tekniklerinin,

ii) kullanılan veri seti ve öznitelik çıkarma tekniklerinin,

iii) karşılaşılan zorluklar ve sağlanan faydaların,

iv) spam tespitinde ulaşılan başarımın

yer almasına önem verilmiştir.

Çalışmada IEEE Xplore, Academia, ACM, SpringerLink, Google Scholar, Researchgate kaynaklarından, "spam detection" & "text mining" sorguları ile ulaşılan yayınlar arasından seçilen 22 yayın detaylı olarak incelenmiştir.

3. Metin Madenciliğine Genel Bakış

Metin Madenciliği Süreci

Literatürde metin madenciliği süreci tipik olarak aşağıdaki şekilde gösterilmektedir. (Maheswari, 2015)

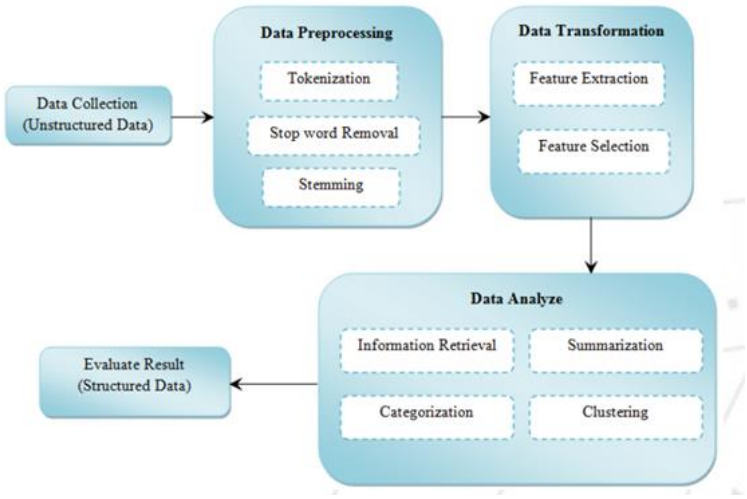

Figure 2. (Maheswari, 2015)'e göre Tipik Metin Madenciliği Süreci

Alanlar arası yapısı dolayısıyla metin madenciliğinin sınırlarını çok keskin bir şekilde çizmek çok kolay değildir. (Pasin, 2018)

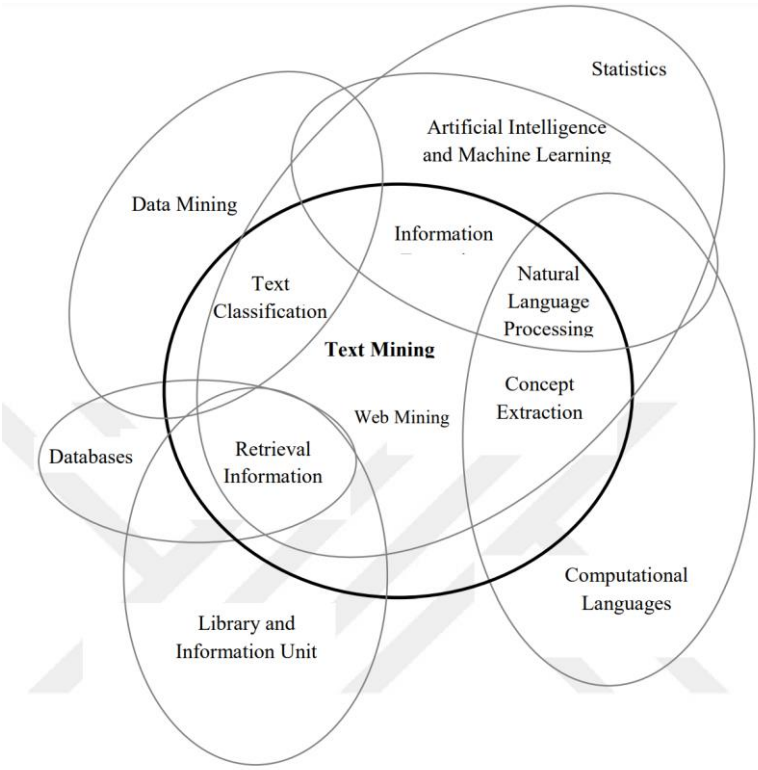

Figure 3. Metin Madenciliği ve İlgili Alanlar (Pasin, 2018)

(Hussain, 2019) metin madenciliği sürecini veri toplama, metin önişleme, belirteçleme (tokenizasyon), transformasyon, öznitelik seçimi şeklinde tanımlamaktadır. Buna göre yorum spamlarının tespitinde veri madenciliği sürecini ve bu aşamalarda yürütülen faaliyetler özetle aşağıdaki şekildedir.

1. Veri toplama

Sosyal ağ üzerinden (yapısal olmayan) analiz edilecek verilerin (yorumların) çekilmesi ve veritabanına atılması sağlanır.

2. Metin önişleme

Metin önişleme aşamasında köke indirgeme (stemming), etkisiz sözcük ayıklama (stop word removal) teknikleri veriyi temizleme ve analize hazırlama amacıyla kullanılmaktadır. Bu şekilde imla işaretleri, sayılar, linkler, kendi başına anlamı olmayan sözcükler ayıklanmakta ve sözcükler kök hallerine dönüştürülmektedir. Yine bu aşamada sözcük etiketleme (POS - part of speech tagging) kullanılarak belli bir bağlama veya metnin içinde belli görevlerde olan sözcükler etiketlenebilmektedir. (Hussain, 2019)

3. Belirteçleme (Tokenizasyon)

Bu yöntemde özellik olarak tek kelime veya kelime grubu kullanılır. Bu teknikte, bir kelime seçildiğinde uni-gram, iki kelime seçildiğinde bi-gram, üç kelime seçildiğinde tri-gram adlandırmaları yapılmakta ve tekniğe genel olarak n-gram denmektedir. Kısaltmalar, akronimler, eş anlamlılar ve belirsizlikler bu süreci zorlaştıran hususlardır. (Hussain, 2019)

4. Transformation

Doküman terim matrisi, seyrek (sparse) matris şeklinde n-gram modeli tarafından üretilen belirteçleri temsil etmek için kullanılır. Seyrek matris, terimlerin veya belirteçlerin sıklığını tanımlar. Literatür taraması ile araştırmacıların çoğunun dönüşüm için Basit Sayma ve TF/IDF tekniklerini birlikte kullandığı gözlemlenmiştir. (Hussain, 2019)

5. Öznitelik Seçimi

Öznitelik seçimi, veri setinden (review dataset) gereksiz kelimelerin çıkarılması, belirteçleme (tokenizasyon) ve dönüşüm tekniklerinden sonra en önemli adımdır. Ayrıca, bu teknik, metin özelliklerinin daha büyük boyutlu olması ve gürültülü özniteliklerin varlığı nedeniyle kullanışlıdır ve sınıflandırma performansını iyileştirmek için en uygun özellik kümesini seçmeye yardımcı olur. Bu aşamada metin sınıflandırmasında öznitelik seçimine yardımcı olan işlemler arasında Gini Index, Information Gain, Entropi, χ2-Statistic teknikleri sayılabilir. (Hussain, 2019)

4. Sınıflandırma (Metin Madenciliği) Yöntemleri

Kümeleme, Sınıflandırma, Bilgi alma (Information retrival), Konu bulma (topic discovery), Özetleme (Summarization), Konu çıkarma (Topic extraction), LDA yöntemleri ile sınıflandırma ve çıktıların üretilmesi sağlanmaktadır.

5. Değerlendirme

Bu adımda modelin performansı ölçülmekte ve sonuç ROC, Precision & Recall, Accuracy, F1-F2 ölçümü vb. bazında değerlendirilmektedir.

Spam göndericinin davranışlarına bağlı tespit ve ilgili öznitelikler

(Jindal, 2008) yorumcu davranışlarına dayalı şekilde spam analizinde aşağıdaki 4 davranış şekillerini dikkate almayı önermektedir.

1. Yorumcunun ilk yazdığı yorum sayısının, toplam yazdığı ürün sayısına oranı

2. Yorumcunun yaptığı derecelendirmeyle ilgili özellikler: (yorumlayan tarafından verilen ortalama derecelendirme, derecelendirmede standart sapma ve yorumcunun her zaman yalnızca iyi, ortalama veya kötü derece verip vermediğini gösteren bir özellik

3. Yorumcunun birden fazla derecelendirme türü (yani iyi, ortalama ve kötü) verip vermediğini gösteren ikili özellikler. (Dört durum vardır: bir yorumcu hem iyi hem de kötü puan, iyi puan ve ortalama puan, kötü puan ve ortalama puan ve üç puanın hepsini vermiştir. Bu dört özellik, bir yorumcunun bir markanın ürünlerini övdüğü, ancak bir rakip markanın ürünlerini eleştirdiği durumlar içindir.)

4. Bir gözden geçirenin yukarıda tanımlı belli özelliklerine birlikte yapsamına dayalı bir yorumlama yüzdesi.

Spam göndericinin davranışlarına bağlı öznitelikler yorumun metin olarak içeriğinden çok yorumun metadatası ile ilgili olmaktadır. Yorumun ID si, Yorumcu ID si, verilen derece (rating) kelime sayısı, tarihi, zamanı gibi bilgiler spam tespiti yapmaya yardımcı olabilir. Literatür taramasından, araştırmacıların spam göndericiyi belirlemek için yorumcu özelliklerini kullandığı gözlemlendi. Ayrıca, istenmeyen e-posta gönderen kişi, etkinlik modeli, IP adresi, coğrafi konum ve profil özellikleri gibi ortak özellikleri

paylaşabilir. Bu özelliklerin yardımıyla veya bu özelliklerin bir kombinasyonu ile spam gönderen, spam ve spam olmayan yorumlar tanımlanabilmektedir. (Hussain, 2019)

(Hussain, 2019)de literatürden derlenmiş olan bazı yaygın bireysel spam gönderici karakteristikleri aşağıda yer almaktadır:

Maksimum inceleme sayısı: Mevcut çalışmalar, spam gönderenlerin çoğunlukla belirli bir günde birden fazla inceleme yazdığını göstermektedir.

Olumlu yorumların yüzdesi: Çoğu spam göndericinin olumlu ve olumlu yorumlar yazdığı gözlemlenmiştir, bu nedenle herhangi bir ürün veya hizmet hakkında yüksek oranda olumlu yorum, spam yorumlarının bir göstergesi olabilir .

İnceleme uzunluğu: Mevcut literatür, çoğu spam göndericinin bir ürün veya hizmet hakkında ayrıntılı incelemeler yazmadığını göstermektedir, bu nedenle bu yararlı incelemeci merkezli özellik, spam göndericilerin belirlenmesine yardımcı olabilir.

Yorumlayan sapması: Çoğunlukla spam gönderenlerin puanları, ortalama inceleme puanından sapar. Genellikle spam gönderenler ürün veya hizmetler için yüksek puan verir.

Maksimum içerik benzerliği: Benzer bir gözden geçiren tarafından farklı ürünler hakkında yapılan incelemelerin benzer metin içeriğinin, spam göndericinin güçlü bir göstergesi olduğu bulunmuştur.

Yorum Spamlarının Tespitinde Kullanılan Teknikler

Yorum Spamlarının Tespitinde Kullanılan Teknikler (Hussain, 2019) (Ott, 2011)da aşağıdaki şekilde verilmektedir.

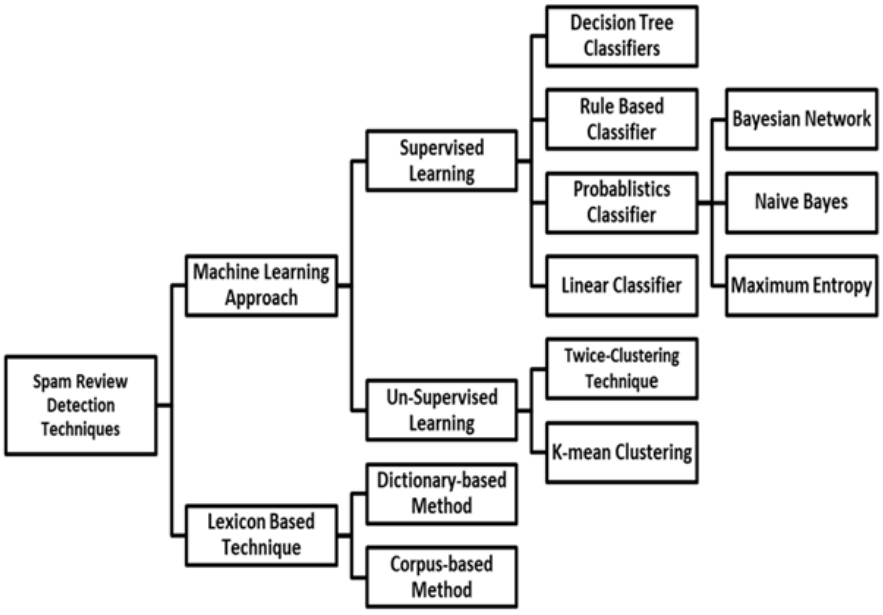

Figure 4. Yorum Spamları Tespit Tekniklerinin Sınıflandırılması [1]

4. İncelenen Çalışmalar

İncelenen çalışmaların amaçlarına göre veya kullanılan metin madenciliği tekniklerine göre sınıflandırılması mümkündür. Çalışmalarda önişleme ve öznitelik seçimi tekniklerinin birlikte farklı kombinasyonlarda kullanılması sebebiyle bu tekniklere göre bir gruplama yapılması tercih edilmemiştir. İncelenen çalışmalar bu bölümde sınıflandırıcılara göre gruplanarak özetlenmiştir. Herbir çalışmanın referansı, amacı, uygulanan önişleme teknikleri, dönüşüm/öznitelik seçimi teknikleri, sınıflandırıcı teknikleri, kullanılan veri setleri ve başarım metrikleri ve sonuçları ile birlikte özetlenmiştir. Karşılaşılan gerek amacı gerekse metodu ve sonuçları itibarıyla benzer olan çalışmaların referansları da belirtilmiştir.

4.1. Sınıflandırmanın Geleneksel Makine Öğrenmesi Yöntemi ile Yapıldığı Çalışmalar

(Ott, 2011) çalışmasında yanıltıcı yorumların (opinion spams) tespiti amaçlanmıştır. Çalışmada problem 3 farklı şekilde ele alınmıştır: (1) n-gram tabanlı spam/gerçek yorum sınıflandırılması, (2) Psikolinguistik aldatma tespiti, (3) "yanıltıcı" yorum türünü "yaratıcı" metin türüne, "gerçek" yorum türünü "bilgilendirici" metin türüne eşleyerek yorumun türünün tespiti. Metin önişleme aşamasında etkisiz kelimeler yorumlardan silinmiştir (stop words removal) ve köke indirgeme işlemi (stamming) uygulanmış, tüm karakterler küçük harfe çevrilmiştir. Üç yaklaş (Mukherjee, 2013)ım için farklı öznitelik belirleme teknikleri kullanılmıştır. 1. yaklaşımda her bir yorum için UNIGRAMS, BIGRAMS+, TRIGRAMS+ lardan oluşan öznitelik kümeleri oluşturulmuştur. 2. yaklaşım için Linguistic Inquiry and Word Count (LIWC) aracı kullanılarak öznitelikler çıkarılmıştır. 3. yaklaşım olan tür tespiti için yaratıcı ve bilgilendirici yazılardaki her bir POS Tag'in sıklığının farklı olduğu düşünüldüğünden, Metinler içerdikleri POS-Taglerin frekansına göre tanımlanmıştır. Her üç yaklaşım içinde sınıflandırıcı olarak Naive Bayes ve Destek Vektör Makinesi kullanılmıştır. Veriseti olarak "Amazon Mechanical Turker ve TripAdvisor kullanılarak 400 "altın standart" spam, 400 gerçek (ham) otel yorumu toplanmıştır. (Ott Veritabanı). Performans değerlendirilirken her bir yaklaşımın Destek Vektör Makinesi üzerinden başarımı aşağıdaki şekilde ölçülmüştür. Psikolingusitik aldatma tespiti için çıkarılan özniteliklerle ve tür tespiti için çıkarılan özniteliklerle eğitilen sınıflandırıcının n-gram tabanlı spam/ham yorum sınıflandırıcısına kıyasla daha kötü çalıştığı gözlemlenmiştir. Çözüm yaklaşımlarına göre öznitelik çıkarımları ve Destek Vektör Makinesi (sınıflandırıcı) doğruluğu aşağıdaki gibi ölçülmüştür. (Ott, 2013) çalışmasında (Ott, 2011) ile aynı yöntemler ve veri seti kullanılmış ancak negatif gerçek ve negatif spam yorumlar çekilmiştir.

Approach	Features	Accuracy
GENRE IDENTIFICATION	POS_{SVM}	73.0%
PSYCHOLINGUISTIC DECEPTION DETECTION	$LIWC_{SVM}$	76.8%
TEXT CATEGORIZATION	$UNIGRAMS_{SVM}$	88.4%
	$BIGRAMS^{+}_{SVM}$	89.6%
	$LIWC{+}BIGRAMS^{+}_{SVM}$	**89.8%**
	$TRIGRAMS^{+}_{SVM}$	89.0%
	$UNIGRAMS_{NB}$	88.4%
	$BIGRAMS^{+}_{NB}$	88.9%
	$TRIGRAMS^{+}_{NB}$	87.6%

Table 1. (Ott, 2011) Başarım değerlendirme tablosu

(Mukherjee, 2013) çalışmasında Yelp Spam Yorum Filtresinin nasıl çalıştığının bulunması amaçlanmıştır. Çalışmada davranışsal öznitelik çıkarımı ve dilsel (linguistic) öznitelik çıkarımı olmak üzere iki farklı yöntemle öznitelik çıkarımının yapılması ve (gerçek veri olan) YELP veriseti üstünden bu iki yöntemin başarımlarının kıyaslanması planlanmıştır. "Linguistik Öznitelik çıkarımı yönteminde başlıca aşağıdaki teknikler kullanılmıştır. (1) her bir yorumun unigramları çıkarılmış ve her bir unigramın Information Gain'i hesaplanmıştır. (2) (Ott, 2011)'de kullanılan Bigarm +LIWC öznitelik çıkarımı uygulanmıştır. Davranışsal Öznitelik çıkarımı için; kullanıcıların, Maksimum Yorum Sayısı (MNR), Olumlu Yorumlarının Yüzdesi (PR), Yorumlarının Uzunluğu (RL), Yorumlardaki maksimum içerik benzerliği' değerleri hesaplanmıştır. Çalışmada Yelp veri setindeki Chicago çevresindeki 85 otel ve 130 restoran için yapılmış olan yorumlar kullanılmıştır (Yelp Veritabanı). (Ott, 2011) çalışmasında Destek Vektör Makinesi sınıflandırıcısında %90 başarım sağlayan LIWC+BIGRAMS öznitelik çıkarım yönetmi Yelp verisi üstünde uygulanmış ve yaklaşık %60'lık başarım elde edilmiştir. Bunun nedeninin AMT tarafından üretilen sahte yorumlar ile gerçek yorumlar arasındaki kelime frekansının farklı olması ve LIWC öznitelik seçiminin bu farktan etkilenmesi olduğu gözlemlenmiştir. Bu gözlem The Kyoto Language Modeling aracı kullanılarak yapılan bilgi teoremi analizi ile tespit edilmiştir. Anormal davranışların tespiti için Yelp datası üzerinde yapılan davranışsal öznitelik çıkarımı sonucunda ise %86'lık başarım (accuracy) elde edilmiştir.

(Kontsewaya, 2021) çalışmasında e-posta spam miktarını bir sınıflandırıcı kullanarak tespit ederek azaltmak amacıyla 7 farklı sınıflandırıcı ile e-posta spam tespiti yapılmıştır. Performansları değerlendirilip en yüksek doğruluk (accuracy) skoruna sahip sınıflandırıcılar belirlenmiştir. Önişleme aşamasında (1) metin içerisindeki tanımlanmayan değerler, boşluklar, tekrarlar, sayılar, noktalama işaretleri, etkisiz elemanlar silinmiş, (2) tüm karakterler küçük harfe çevrilmiş ve (3) belirteçleme yapılmıştır. Bunu takiben sözcük torbalama (bag of words) kullanılarak, belirlenen sözlüğün içindeki herbir kelime yorumun içinde bulunup bulunmamasına göre binary bir değer almıştır. Oluşan binary vektör yorumun öznitelik vektörü olmuştur. Çalışmada "kaggle.com sitesinden hazır bir veri seti kullanılmıştır. https://www.kaggle.com/datasets/karthickveerakumar/spam-filter. Veriseti 1:4 oranında train ve test olarak ayrılmıştır. Verisetinde 4360 gerçek, 1368 spam ileti bulunmaktadır. Sınıflandırıcı olarak "Naive Bayes, K-En Yakın Komşular, Destek Vektör Makinesi, SVM, Lojistik regresyon, Karar ağacı, Rastgele orman modelleri kullanılmıştır." Çalışmada kullanılan metrikler: precision, recall, accuracy, F-measure and ROC Area'dır. (Verma, 2020)'de incelenmiş olup, amacı, yöntemleri ve sonuöları itibarıyla bu çalışmanın bir benzeri olarak kabul edilebilir.)

Algorithm	Accuracy	Precision	Recall	F-measure	ROC area
KNN	0,90	0,91	0,63	0,74	0,95
Naive Bayes	0,99	0,97	0,99	0,98	1,00
Decision tree	0,94	0,82	0,96	0,88	0,95
SVM	0,98	0,98	0,95	0,96	1,00
Random forest	0,84	1	0,28	0,42	0,99
Logistic regression	0,99	0,98	0,96	0,97	0,95

Table 2. (Kontsewaya, 2021)Başarım Tablosu

(Ruskanda, 2019) çalışmasında ön işleme adımlarının denetimli spam sınıflandırıcı algoritmalarının performansı üzerindeki etkisi incelenmiştir. Doğru bir sınıflandırma sonucu için kullanılabilecek ön-işlem metodlarının kombinasyonu ve bu kombinasyonun hangi sınıflandırıcılarda iyi çalıştığının tespiti yapılmak istenmiştir. Çalışmada (1) Etkisiz kelimelerin silinmesi, (2) köke indirgeme ve lemmatization, (3) TF-IDF oranının hesabı yapılmış, sözcük çantası (bag of word) modeli ile her bir e-postanın temsil vektörü oluşturulmuştur. Çalışmada Ling-spam veri setleri kullanılmış olup, toplamda 962 gerçek/spam e-postadan oluşmaktadır. Çalışmanın sonuçlarına göre etkisiz sözcüklerin

çıkarılması ve stemming ön-işlemleri Naive Bayes, Destek vektör makinesi sınıflandırıcısının daha iyi sonuçlar vermesine yardımcı olurken, DCM sınıflandırıcısında ön-işlem yöntemi kullanılmadığı takdirde daha iyi sonuçlar elde edildiği gözlemlenmiştir.

(Ahmad, 2021) çalışmasında spam tweetlerin tespiti amaçlanmıştır. Önceki çalışmaların çoğunda, öznitelik çıkarımı için, belirli kelimelerin, mesaj biçimlerinin kullanılması veya aynı anda birden çok kullanıcıya mesaj gönderilmesi gibi davranışsal ve içerik tabanlı öznitelik çıkarım yaklaşımları kullanılmıştır. Ancak bu çalışmada bu özniteliklere ek olarak, sosyal ağların yapısı, kullanıcı bilgileri, mesajı alan ve gönderen arasındaki ilişki vb ek öznitelikler çıkarılmıştır. Problemin çözüm aşaması, problem spesifik önişleme adımlarından ve 4 farklı öznitelik çıkarma yönteminden oluşmaktadır. Tweetlerin önişlem aşamasında diğer çalışmalardaki önişlem aşamalarında olduğu gibi etkisiz kelimeler ve noktalama işaretleri çıkarılmıştır. Bunun yanında her bir tweet için, kelimelerin eş zamanlı dağılımı hesaplanmış, Type Token Ratio (TTR) and Mean Word Frequency oranlarına bakılarak kelime yoğunluğu ve zenginliği hesaplanmış ve tweet içerisinde yer alan tagging(etiketleme) dağılımına bakılmıştır. Çalışmada, (1 (Ahmad, 2021))Kullanıcı profil öznitelikleri, (2) kullanıcı etkinliğine dayalı öznitelikler, (3)içerik tabanlı öznitelikler, (4 (Watcharenwong, 2017)) iletişim tabanlı (graph tabanlı) öznitelikler çıkarılmıştır. Kullanıcı etkinliğine dayalı öznitelikler içerisinde, kullanıcının doğru cevap oranı ve kullanılan API çeşitlilik oranına bakılmıştır. İletişim tabanlı öznitelik çıkarımlarında ise, SRANK algoritması ile kullanıcının komşularıyla olan benzerliği hesaplanırken, "Ortak mahalle oranı" ile kullanıcının komşularıyla ortak arkadaş sayısına bakılmıştır. Daha sonrasında "özyinemeli öznitelik çıkarılması" algoritması kullanılarak, sınıflandırıcılar için en ayırt edici olan öznitelikler belirlenmiştir. Belirlenen ayırt edici öznitelikler kullanılarak, Çok Katmanlı Algılayıcı(MLP),Naive Bayes, Rastgele Orman, K-en yakın komşu ve Destek Vektör Makinesi üzerinden sınıflandırılma yapılmıştır. Çalışmada 2 milyon spam/gerçek tweetden oluşan Honeypot veriseti kullanılmıştır. Sınıflandırıcılar arasından Destek Vektör Makinesi diğer sınıflandırıcılara kıyasla daha iyi sonuç vermiştir. Accuracy skoru 0.96, precision skoru 0.98'dir.

(Watcharenwong, 2017) çalışmada kapalı Facebook grupları için spam tespiti yapılması amaçlanmıştır. Metin özniteliklerinden ve sosyal özniteliklerden oluşan 11 öznitelik çıkarılmış ve Rastgele Orman algoritması üzerinden kapalı facebook grupları için spam tespiti yapılmıştır. Önişlem aşamasında (1) Boşluklara göre belirteçleme, (2) Etkisiz kelime çıkarımı gerçekleştirilmiştir. Metin öznitelikleri kapsamında: her bir post'daki (1) toplam sözcük kelime sayısı ve (2) spam corpusundaki kelimelerle eşleşen "spam kelimelerinin" sayısı belirlenmiştir. Sosyal öznitelikler kapsamında: (1) bir mesajdaki URL sayısı, (2) gömülü videoların sayısı, (3) bir mesajın resim içerip içermediği, (4) bir mesajdaki beğeni sayısı, (5) bir mesajdaki hashtag sayısı, (6) yorum yapılıp yapılmadığı, (7) bir gönderinin paylaşılıp paylaşılmadığı, (8) bir mesajdaki etiketlenen kişi sayısı, (9) yayınlanma zamanı seçilmiştir. Çalışmada bir web tarayıcısı kullanılarak Facebook'tan taranan 1.200 etiketli gönderi üzerinden veriseti oluşturulmuştur. Rastgele Orman Spam sınıflandırılmasında en ayırt edici özniteliklerin post'da yer alan beğeni sayısı ve post'a yapılan bir yorumun olup olmaması olduğu tespit edilmiştir. Bunun yanında metin öznitelikleri; toplam kelime sayısı ve "spam kelime" sayısının yukarda belirtilen özniteliklere kıyasla daha az ayırt edici olduğu belirtilmiştir. Bunun yanında rastgele Orman sınıflandırıcısının precission F1 ve recall skorunun yaklaşık %98 olduğu gözlemlenmiştir.

(Fusilier, 2015) çalışmasında karakter n-gramları kullanarak spam yorum tespiti amaçlanmıştır. Bu çalışmada belli bir alan içerisinde kesin ve doğru yorumların yazım şekli (stili) ile spamların yazım şekli arasında farklılık olduğu düşünülmüştür. Bundan dolayı spam tespiti stilistik bir sınıflandırma görevi olarak ele alınmıştır ve karakter n-gram ile elde edilen sınıflandırma sonuçları ile word n-gram ile elde edilen sınıflandırma sonuç kıyaslanmıştır. Metinlerin önişleme aşamasında tüm noktalama işaretleri ve sayılar

kaldırılmıştır ancak daha önceki çalışmalardaki önişlem aşamalarının aksine etkisiz kelimeler bırakılmıştır. Bunun yanında, tüm harfler küçük harfe dönüştürülmüştür ve sadece alfabetik tokenlar oluşturulmuştur. Önişlem sonrası yorumları temsil etmek için bir Bag of Karakter n-gram (BOC) ve Bag of Word n-gram (BOW) kullanılmıştır; her iki durumda da ikili, ağırlıklandırma şeması uygulanmıştır. Oluşturulan temsil vektörleri Naive Bayes ve Destek Vektör Makinesine girdi olarak verilmiştir. Çalışmada, 1600 yorumdan oluşan Ott veri seti kullanılmıştır, vu versisetinin yanında spam yorumlar AMT ile üretilmiş, gerçekilen yorumlar TripAdvisor, Expedia, Hotels.com, Orbitz, Priceline, ve Yelp üzerinden alınmıştır. Çalışmanın sonucunda, Naive Bayes ve Destek vektör makinesi sınıflandırıcılarında karakter tabanlı temsillerin, kelime tabanlı temsillere göre olumlu ve olumsuz yanıltıcı yorumların(spam) tespitinde %2,3 ve %2,1'lük daha iyi performans sergilediği gözlemlenmiştir.

4.2. Sınıflandırmanın Kural Tabanlı Sınıflandırıcılar ile Yapıldığı Çalışmalar

(Saidani, 2020) çalışmada spam e-postalarının tespitindeki başarımı arttırmak için semantik analiz kullanılması amaçlanmıştır. İlk olarak e-postalar belirli alanlara göre (Sağlık, Eğitim,Finans…vb) kategorize edilmiştir. Her bir kategori altındaki e-postaların spam olup olmaması ayrımı ise manuel olarak belirtilen ve otomatik olarak çıkarılan kuralların sonucunda elde edilen öznitelikler üzerinden yapılmıştır. E-postaları belirli alanlara göre sınıflandırmak için uygulanan ön işleme aşamasında: Kural tanımlı olarak (1) Düzenli ifadeler (Regular expressions) kullanılarak anahtar kelime çıkarımı, (2) Arama ağacı uygulaması kullanılarak ayrı harfler içeren kelimelerin tespiti, (3) köke indirgeme (stemming), (4) etkisiz kelime ayıklama", (5) Alana göre e-posta sınıflandırması için her kategori için en yüksek IG'ye(Information Gain) sahip 500 terimi seçilmiştir. Yorumlarda bu kelimenin olup olmamasına göre binary öznitelik vektörü oluşturulmuştur. Alana göre semantik öznitelik çıkarmada ise; Uzmanlar tarafından oluşturulan düzenli ifadeler ile manuel kurallar belirlenmiştir. Otomatik oluşturulan öznitelikler ise CN2-sD yöntemi kullanılarak oluşturulmuştur. Alana göre semantik öznitelik çıkaraımı için oluşturulan kuralların her birinin binary çıkışı vardır. Her bir kural spam tespiti için tek başına zayıf bir sınıflandırıcı görevi görmektedir. Elde edilen anlamsal öznitelikler daha sonra alana özgü spam e-postaları tespit etmek için özel sınıflandırıcılar oluşturmak için kullanılmıştır." Çalışmada, Enron Corpus ve Ling-spam veri setleri birleştirilmiş ve model eğitilmiştir. Bunun yanında CSDMC2010 SPAM veriseti üzerinden de model eğitimi yapılmıştır. Alana göre e-posta sınıflandırılması için Naive Bayes, Karar ağacı, modelleri kullanılmıştır.

(Sandulescu, 2015)çalışmasında aynı kişi tarafından birden fazla isim kullanılarak yazılan sahte yorumları tespit etmek amaçlanmıştır. Birinci yaklaşım olarak, "bilgiye dayalı anlamsal benzerlik" ölçüsünü kullanan bir yöntem önerilmiştir. Bu çalışmada aynı kişi tarafından birden fazla adla yapılan yorumlarda, kişinin yeni ayrıntılarda hayal gücünün kısıtlı olduğu, yaptığı yorumların anlamsal olarak birbirine benzediği varsayılmıştır. Yorumlar arasındaki benzerliği tanımlarken WordNet'de yer alan eş anlamlı kelimelerden yararlanılarak iki yorum arasındaki benzerlik hesabı için bir formül geliştirilmiştir. Bu formülden elde edilen sonuç ile oluşturulan baseline modelden elde edilen sonuç kıyaslanmıştır. ikinci yaklaşım ise yorumları belirli bir Gizli Dirichlet Ayırımı (LDA) modeli kullanarak konu sınıflandırmasının yapılmasıdır. Yorumların altında yatan konu dağılımlarının benzerliği, yorumları spam veya gerçek olarak sınıflandırmak için kullanılmıştır. Önişlemede 1. yöntemin kıyaslandığı vektör tabanlı baseline'ını oluşturmak için Pos taging ve lemmatization yapılmış, etkisiz kelimeler silinmiş ve bir temsil vektörü oluşturulmuştur. İki vektör arasındaki kosinüs benzerliğinin yorumlar arasındaki benzerliği tanımladığı farzedilerek bu yaklaşım baseline olarak kabul edilmiştir. 2. yöntem için her bir yorumun kısıtlı POS etiketten oluşan bag of words ve bag of opinion temsilleri

oluşturulmuştur. Bag of opinion yaklaşımında herbir yorum için <özellik, duygu> (aspect-sentiment) ikilileri üretilmiştir. Bu ikililer LDA modelinde kullanılmıştır. Yorumların kısıtlı POS etiketten oluşan bag of words ve bag of opinion temsilleri konu tespiti için kullanılan LDA modeline girdi olarak verilmiştir. "Çalışmada "Trustpilot şirketi yorum veri seti (9.000 sahte ve gerçek yorum), Yelp veri seti (57K yorum), Ott veri seti (800 yorum)" kullanılmıştır." LDA modelleri anlamsal ve vektörel olanlardan daha düşük bir kesinlik skoru elde etmiştir, ancak dilden bağımsız olarak yorumların anlamlarını çıkarabildikleri görülmüştür. Bunun yanında, anlamsal benzerlik hesabında kullanılan WordNet'ler, İngilizce dışında başka diller için oluşturulmuş olsa da, hiçbiri İngilizce sürümünün derleme ölçeğiyle eşleşmez."

4.3. Sınıflandırmanın Hibrit Yöntemle Yapıldığı Çalışmalar

(Sharmin, 2017), Bu çalışmada YouTube videolarının altında yer alan Yorum kısmındaki spam yorumların tespiti amaçlanmıştır. Spam yorumların tespiti için 4 farklı tekli sınıflandırıcı ve torbalama (bagging) yöntemini kullanan topluluk modeli (ensemble model) uygulanmış ve performansları karşılaştırılmıştır. Metin önişleme aşamasında etkisiz kelimeler yorumlardan silinmiştir (stop words removal) ve köke indirgeme işlemi (stamming) uygulanmıştır. Önişleme sonrası her bir kelimenin TF-IDF değeri hesaplanmıştır. Her bir yorumun kelime çantası (bag-of-words) modeline göre öznitelik vektörü oluşturulmuştur. Oluşturulan öznitelik vektörleri "Naive Bayes, 1-KNN, 3-KNN, Torbalama yöntemi kullanan topluluk modeli ve Destek Vektör Makinesi sınıflandırıcılarına girdi olarak verilmiştir. Çalışma için veriseti oluşturma aşamasında UCI Makine Öğrenmesi Sunucusu üzerinden YouTube'da en çok izlenen 10 video arasında yer alan beş video için yapılmış olan yorumlara ulaşılmış ve %51'i spam olan toplamda 1956 adet yorum toplanmıştır. Çalışma sonucunda modellerin MCC (Matthews Correlation Coefficient), Precission, Accuracy, Recall değerleri hesaplanmış, modellerin başarımlarının kıyaslanmasında MCC metriği kullanılmıştır. Alınan sonuçlara göre Torbalama yöntemi ile oluşturulmuş topluluk (ensemble) modelinin en iyi performansı gösterdiği belirlenmiştir. MCC metriğine göre modelin başarım değerleri şöyledir; 1-KNN 0.883, 3-KNN 0.857, NavenBayes 0.85, Bagging 0.902, SVM 0.851. ((Aiyar, 2018) Çalışması amacı, metodolojisi ve çıktıları itibarıyla (Sharmin, 2017) ile benzerlik göstermekte olup bu grupta değerlendirilmiştir.)

(Saeed, 2019) çalışmasında Arapça çevrimiçi spam yorumların tespitinin yapılması amaçlanmıştır. Kural tabanlı sınıflandırıcıyı makine öğrenme teknikleri ile entegre eden hibrid bir çözüm önerilmiştir. Önişleme aşamasında (1.) Metindeki boşluklara göre belirteçleme yapılmıştır. (2) ARapça olmayan belirteçler kaldırılmıştır. (3) etkisiz sözcükler kaldırılmıştır (4) hafif köke indirgeme (light stamming) uygulanmıştır. (5) Normalizasyon uygulanmıştır. Özniteliklerin seçilmesinde, N-gram tabanlı öznitelik çıkarımı yapılmıştır. Uni-gram, bi-gram ve Tri-gramlardan oluşan öznitelik kümesi kullanılmıştır. N-gramların polarity skoru hesaplanmıştır. Olumsuzlaşma işlemi (Negation handling) yapılarak olumsuz kelimeler içeren N_gramların polarity değerleri tersine çevrilmiştir. İçerik tabanlı öznitelik çıkarımı: her bir yorumdaki kelime sayısı, benzersiz kelime oranı, derecelendirme sapması hesaplanmıştır. Ver, seti olarak İngilizce 1.600 yorum incelemesine sahip Opinion spam corpus (DOSC & HARD) veri kümeleri kullanılmıştır. Veriler uzmanlar tarafından Arapçaya çevrilmiştir. SInıflandırıcı olarak; (1) Kural tabanlı sınıflandırıcı, (2) Majority Voting Topluluk sınıflandırıcısı, (3) Makine Öğrenmesi Sınıflandırıcıları ve (4) Stacking Topluluk sınıflandırıcısı kullanılmıştır. Her bir sınıflandırıcı için Olumsuzlaşma işlemi uygulamanın başarımı arttırdığı gözlmlenmiştir. Accuracy skoru bakımından en başarılı sonuç veren Stacking Topluluk sınıflandırıcısnın diğer başarım metrikleri için skoru : Accuracy- 95.25%, Recall- 91.75%, Precision- 98.66%, F1-Score- 95.08 şeklindedir.

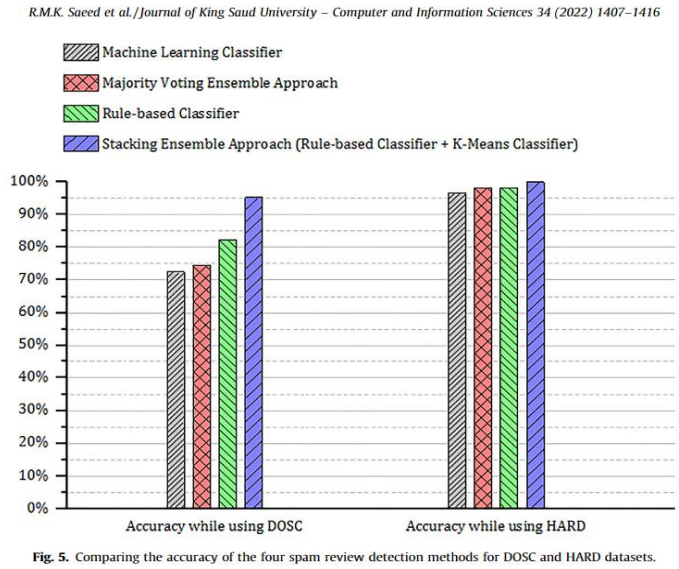

Fig 5. DOSC ve HARD veri setlerinde 4 farklı sınıflandırıcının başarımı

4.4. Sınıflandırmanın Derin öğrenme Teknikleri Yapıldığı Çalışmalar

(Shahariar, 2019) çalışmasında herhangi bir aldatıcı (spam)yorumun tespiti amaçlanmıştır. Çalışmada Hem etiketli hem etiketsiz yorumlar, Çok Katmanlı Algılayıcı (MLP), Evrişimli Sinir Ağı (CNN) ve Tekrarlayan Sinir Ağının (RNN) bir çeşidi olan LSTM modelleri üzerinde çalıştırılmıştır. Bunun yanında veriler geleneksek makine öğrenimi sınıflandırıcılarından Naive Bayes, K-En Yakın Komşu (KNN) ve Destek Vektör Makinesi üzerinden de çalıştırılarak modeller arası performans kıyaslanması yapılmıştır. Önişleme aşamasında, yorumlardaki; etkisiz elemanlar ve noktalama işararetleri çıkarılmıştır, tüm harfler küçük harflere dönüştürülmüş, tokenization ve stamming işlemleri uygulanmıştır. Bunun aynında destek vektör makinesi modelinde karar fonskiyonunun modifiye edilmiş hali olan Değiştirilmiş Aktif öğrenme" Algoritması kullanılarak etiketsiz veriler, etiketlenmiştir. Geleneksel makine öğrenimi algoritmaları ve MlP için; TF/IDF ve n-gram tabanlı öznitelik çıkarma işlemi uygulanırken, Cnn ve LSTM modelleri içinse Word2Vec kelime gömme yaklaşımı uygulanmıştır. Çlaışmada, Ott veriseti ve Yelp veriseti kullanılmıştır. ("https://www.yelp.com/dataset"=). Çalışmanın sonucunda Yelp veriseti üstünde Derin öğrenme algoritmalarının, geleneksel makine öğrenimi algoritmalarına kıyasla daha iyi performans gösterdiği gözlemlenmiştir.

Method	Train Test Ratio	Technique	Accuracy
CNN	90:10	word2vec	95.56%
LSTM	80:20	word2vec	96.75%
MLP	5-fold	Unigram+Bigrams+Trigrams	93.19%

Tablo 3. YELP veri setinde derin öğrenme başarımı (Shahariar, 2019)

Cross Validation	Classifier	Technique	Accuracy
10-fold	SVM	Unigram	91.73%
5-fold	KNN	Bigrams	90.75%
10-fold	NB	Unigram	91.75%

Tablo 4. YELP veri setinde geleneksel sınıflandırıcıların başarımı (Shahariar, 2019)

(AbdulNabi, 2021) çalışmasında, Bert Transformation modeli kullanılarak spam e-posta tespiti amaçlanmıştır. Önceden eğitilmiş BERT Modeli spam ve spam olmayan (ham) e-postaları ayırt etmek üzere "fine-tune" edilmiştir ve metinlerin etmsil vektörleri oluşturulurken kelime gömme yaklaşımı kullanılmıştır. BERT Modelinin başarımı, BiLSTM , DNN ,K-En Yakın komşu (K-NN) ve Naive Bayes modellerinin başarımıyla kıyaslanmıştır. Çalışmada kullanılan önişlem aşamasında,etkisiz kelimeler ve moktalma işaretleri çıkarılarak tokenization yapılmıştır. Her bir token için TF/IDF oranı hesaplanmş ve her bir e-posta için TF/IDF vektörü oluşturulmuştur. Daha sonra bu temsil vektörleri K-NN ve Naive Bayes sınıflandırıcılarına girdi olarak verilmiştir. BERT,BiLSTM ve DNN modelleri içinse kelime gömme yaklaşımı kullanılmıştır. Çalışmada veri seti olarak Açık kaynak UCI makine öğrenmesi sunucusu üzerinden ulaşılan (Spambase) veri seti ve Kaggle 'da bulunan (Spam Filter) veri seti kullanılmıştır. Çalışmanın sonucunda spam e-posta tespitinde BERT Transformer modelinin diğer modellere kıyasla daha yüksek doğruluğa ve F1 Puanına sahip olduğu gözlemlenmiştir.

Model	Accuracy	F1 Score
KNN	0.9310	0.9081
NB	0.9540	0.9408
BiLSTM	0.9650	0.9556
Bert Base Cased	**0.9730**	**0.9696**

Table 5. Eğitim verisi ile test sonuçları (AbdulNabi, 2021)

5. Değerlendirme ve Sonuç

Yapılan incelemelerde spam tespitinde davranışsal, dilsel ve anlamsal olmak üzere üç farklı öznitelik çıkarım yaklaşımının tercih edildiği görülmüştür. Çalışmalarda 2008-2015 arasında dilsel öznitelik çıkarımının, 2011-2018 arasında dilsel öznitelik çıkarımının yanında davranışsal ve anlamsal öznitelik çıkarımının (regular expression) (kural tabanlı modeller ile) tercih edildiği, 2018 sonrası çalışmalarda ise anlamsal öznitelik çıkarımında NLP tekniklerinin daha yoğun kullanılmaya başlandığı görülmektedir. Kullanılan sınıflandırma algoritmaları bakımından incelendiğimizde ise öznitelik çıkarımı yaklaşımlarındaki gelişmelere paralel olarak, 2008-2011 yılları arasında makine öğrenimi ve kural tabanlı daha sonra ise hibrit ve derin öğrenme algoritmalarının daha yoğun kullanıldığı görülmektedir. Bu çalışmada metin madenciliği tekniklerinin kullanımına odaklanılmış olup, derin öğrenme modeli ve doğal dil işleme tekniklerinin yer aldığı bazı çalışmalar konuyu bütünsel olarak değerlendirebilmek için incelemeye dahil edilmiştir. Çalışmalarda kullanılan veri setleri incelendiğinde ise, 2012 öncesinde gerçek spam verisine erişimin zor olduğu, Amazon Mekanical Turker platformu kullanılarak üretilen spam verilerinin de gerçeğe göre daha kolay ayırt edilebilir olduğu görülmüştür. 2000'li yıllarda başlayan metin madenciliği uygulamalarında son yıllarda anlamsal özniteliklerin ön plana çıktığı ve çalışmaların doğal dil işleme yöntemine doğru kaydığı görülmektedir.

References

[Online]. - **Spam Filter.** - https://www.kaggle.com/karthickveerakumar/spam-filter) .

AbdulNabi Spam email detection using deep learning techniques [Journal]. - [s.l.] : Procedia Computer Science 184(2):853–858, 2021. - Vol. DOI 10.1016/j.procs.2021.03.107..

Ahmad Spam detection on Twitter using a support vector machine and users' features by identifying their interactions [Journal]. - [s.l.] : Multimedia Tools and Applications (Springer) 80(8):11583–11605, 2021. - Vols. DOI 10.1007/s11042-020-10405-7..

Aiyar Shreyas N-Gram Assisted Youtube Spam Comment Detection [Journal]. - [s.l.] : Procedia Computer Science, 2018.

Dixit Snehal SURVEY ON REVIEW SPAM DETECTION [Journal] // International Journal of Computer and Communication Technology. - 2016.

E.Zhong Discovering Spammers in Social Networks [Journal]. - [s.l.] : Proceedings of the AAAI Conference on Artificial Intelligence, 2012.

Fusilier Detection of opinion spam with character n-grams [Journal]. - [s.l.] : In: Gelbukh A, ed. Computational Linguistics and Intelligent Text Processing., 2015.

Gupta Aaryan Comprehensive review of text-mining applications in finance [Journal]. - [s.l.] : Springer Open, 2020.

http://archive.ics.uci.edu/ml [Online]. - Spambase. - http://archive.ics.uci.edu/ml.

Hussain Naveed Spam Review Detection Techniques: A Systematic Literature Review [Journal]. - [s.l.] : Appl. Sci. 2019, 9, 987; doi:10.3390/app9050987, 2019.

Jindal Nitin Opinion Spam and Analysis [Journal]. - 2008.

Kaddoura Sanaa A systematic literature review on spam content detection and classification [Journal]. - [s.l.] : PeerJ Comput. Sci. 8:e830 DOI 10.7717/peerj-cs.830, 2022.

Kontsewaya Evaluating the effectiveness of machine learning methods for spam detection [Journal]. - [s.l.] : Procedia Computer Science 190(3):479–486, 2021. - Vol. DOI 10.1016/j.procs.2021.06.056..

Lau Raymound Text mining and probabilistic language modeling for online review spam detection [Journal]. - [s.l.] : ACM Transactions on Management Information Systems, 2011.

Liao S.H. Data mining techniques and applications – A decade review from 2000 to 2011 [Journal]. - [s.l.] : Expert Systems with Applications, 2012.

Maheswari M. Uma Text Mining: Survey on Techniques and Applications [Journal]. - [s.l.] : International Journal of Science and Research (IJSR) , 2015.

Mukherjee What Yelp fake review filter might be doing? [Journal]. - [s.l.] : In Proceedings of the International Conference on Web and Social Media, Cambridge, MA, USA,, 2013.

Ott [Online]. - Ott Veritabanı. - https://myleott.com/op-spam.html.

Ott Myle Finding Deceptive Opinion Spam by Any Stretch of the Imagination [Journal]. - [s.l.] : Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 309–319, 2011.

Ott Myle Negative Deceptive Opinion Spam [Journal]. - 2013.

Pasin Ezgi INVESTIGATION OF TEXT MINING METHODS ON Turkish Text [Book Section] // Graduate School of Natural and Applied Sciences of Dokuz Eylül University. - 2018.

Ranjan Nihar M. Text Analytics: An Application of Text Mining [Journal]. - [s.l.] : Journal of Journal of Data Mining and Management, 2021.

Ruskanda Study on the effect of preprocessing methods for spam email detection [Journal]. - [s.l.] : Indonesia Journal of Computing. 4(1):MARET, 2019. - Vol. DOI 10.21108/INDOJC.2019.4.1.284..

Saeed An ensemble approach for spam detection in Arabic opinion texts. [Journal]. - [s.l.] : Journal of King Saud University - Computer and Information Sciences 34(1):1407–1416, 2019. - Vol. DOI 10.1016/j.jksuci.2019.10.002..

Saidani A semantic-based classification approach for an enhanced spam detection [Journal]. - [s.l.] : . Computers & Security 94(1):101716 , 2020. - Vol. DOI 10.1016/j.cose.2020.101716..

Sandulescu Detecting singleton review spammers using semantic similarity [Journal]. - [s.l.] : Proceedings of the 24th International Conference on World Wide Web. 971–976., 2015.

Shahariar Spam review detection using deep learning. [Journal]. - [s.l.] : In: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile, 2019.

Sharmin Spam Detection in social media employing machine learning tool for text mining [Journal]. - 2017.

Tandel Sayali Sunil A Survey on Text Mining Techniques [Journal]. - [s.l.] : 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS 2019) , 2019.

Watcharenwong Spam detection for closed Facebook groups [Journal]. - [s.l.] : In: 14th, 2017.

Yelp Veritabı [Online]. - Yelp Veritabanı. - http://odds.cs.stonybrook.edu/yelpzip-dataset/..

Zhu Learning to Identify Review Spam [Journal]. - [s.l.] : Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2011.