

Tavuk Sürüsü Optimizasyon Algoritmasıyla Amazon Müşteri Yorumları Üzerinde Duygu Analizi

Nagihan Yağmur¹

Özet

İnternet kullanımının artışıyla birlikte çeşitli platformlarda duygusal ve fikrinsel ifadelerin paylaşılmasının önemini vurgulanmakta ve bu bağlamda metin madenciliği ve makine öğrenmesi tekniklerinin duygu analizi üzerindeki etkileri incelenmektedir. İnternet üzerindeki kullanıcı yorumları, ürünler, filmler ve hizmetler hakkında geri bildirim sağlamak amacıyla yaygın bir şekilde kullanılmaktadır. Bu çalışmanın odak noktası, UCI Machine Learning Repository'den elde edilen 1000 İngilizce Amazon müşteri yorumları üzerinde Tavuk Sürüsü Optimizasyon Algoritması (TSO) kullanarak duygu analizi gerçekleştirmektir. Kelime çantası (Bag of Words) özniteliği ve metasezgisel algoritma kullanılarak yapılan sınıflandırma işlemi, doğrusal ve karesel modellerle gerçekleştirilmiştir. Çalışmanın sonuçları, önerilen yöntemin her iki model için de yüksek başarı gösterdiğini ve TSO algoritmasının metin madenciliği çalışmalarında etkili bir araç olabileceğini göstermektedir. Gelecekteki çalışmalarda, daha fazla öznitelik kullanımının daha iyi sonuçlara yol açabileceği ve TSO algoritmasının metin madenciliği alanında daha sık kullanılacağı düşünülmektedir.

1. Giriş

İnternet kullanımının son yıllarda artması, insanlara duygu ve düşüncelerini farklı platformlarda ifade etme konusunda geniş bir imkan sunmaktadır. İnternet, bilgi keşfi ve erişim konusunda kullanıcılara büyük ölçüde katkıda bulunmaktadır. Tatil veya otel rezervasyonu yapmak, bir ürün hakkındaki duygularını diğer potansiyel kullanıcılara iletmek, film izlemeden önce yapılan yorumları okumak gibi faaliyetler öncesinde, insanlar genellikle başkalarının düşüncelerini öğrenmek amacıyla interneti ilk

1 Kütahya Dumlupınar Üniversitesi, nagihan.yagmur@dpu.edu.tr, 0000-0002-6407-4338

başvurdukları kaynak olarak kullanmaktadırlar. İnsanlar, internet aracılığıyla aldıkları hizmetlerle ilgili deneyimlerini paylaşarak, olumlu veya olumsuz düşüncelerini diğerleriyle internet ortamında paylaşma eğilimindedirler. Kullanıcılar tarafından yapılan bu yorumlardan elde edilen faydalı bilgiler, hem tüketiciler hem de üreticiler için büyük önem taşımaktadır. Bir ürün satın almayı düşünen bir kişinin tüm yorumları okuması zaman alıcı bir süreç olduğundan, ürünle ilgili bilgileri çıkarmak için metin madenciliği ve makine öğrenmesi yöntemleri önemli bir rol oynamaktadır. Metin madenciliği, yapılandırılmamış metinlerden anlamlı bilgiler çıkarma sürecini ifade eder. Bu bağlamda, müşteri etkileşimlerini kayıt altına almak, kullanıcıların yorum ve görüşlerini incelemek için çeşitli duygu analizi yöntemlerinin geliştirilmesi büyük bir öneme sahiptir. Duygu analizi, bir metnin duygu içerip içermediğini ve içeriyorsa bu duygunun olumlu veya olumsuz olup olmadığını belirleme sürecini ifade eder [1]. Bu veri analizi yöntemi, metindeki kelimeleri ve kelime öbeklerini kullanarak metnin duygusunu çıkarmaya odaklanır. Duygu analizi problemine dair literatürde birçok çalışma bulunmaktadır. Farklı makine öğrenmesi tekniklerinin uygulandığı bir duygu analizi çalışmasında, IMDB platformundan alınan film yorumları olumlu ve olumsuz olarak sınıflandırılmıştır [2]. Başka bir çalışmada ise Türkçe dökümanlardaki sık geçen kelimeler (stop words), 2 ve 3-gram'lar, istatistik ve dilbilgisi nitelikleri kullanılarak özellik vektörleri çıkarılmıştır [3]. Başka bir metin sınıflandırma çalışmasında ise Anadolu Ajansı'ndan alınan 2600 haber verisi ile ekonomi, sağlık, spor ve politika kategorilerindeki makaleler sınıflandırılmıştır. Bu çalışmada KNN ve naive Bayes sınıflandırma algoritmaları, bit frekansı ve TF-IDF ağırlıklandırma yöntemleri kullanılarak test edilmiş ve sonuçlar karşılaştırılmıştır. Bu karşılaştırmada en iyi başarı TF-IDF kelime ağırlıklandırma metodunu kullanan naive Bayes modeli tarafından elde edilmiştir [4]. Twitter platformunda "Torku" kelimesi ile ilgili paylaşımların incelenerek, kullanıcıların duygusal tepkileri TextBlob ve VaderSentiment kütüphaneleri kullanılarak yapılan bir çalışmada etiketlenmiş veri seti üzerinde dört farklı makine öğrenmesi modeli eğitilerek, duygu ağırlıkları belirlenmiştir [5]. Bu modeller daha sonra karşılaştırılarak, "Torku" kelimesiyle ilgili genel görüşleri en iyi şekilde ifade eden model tespit edilmeye çalışılmıştır. Bu süreç sonucunda, kullanıcıların "Torku" hakkındaki duygusal eğilimleri oranlarıyla birlikte ortaya çıkarılmıştır. Destek vektör makineleri, naive Bayes ve yapay sinir ağları kullanılarak Arapça metin sınıflandırılması yapılan bir çalışmada 1400 Arapça dökümandan veri seti oluşturulmuş ve en yüksek performans yapay sinir ağından elde edilmiştir. Başka bir çalışmada makine öğrenmesi ve kelime dizilimi (lexical sequence) kullanılarak metinlerin polaritesini otomatik bir şekilde ifade eden

algoritmalar geliştirilmiştir [6]. Yapılan başka bir çalışmada www.hepsiburada.com adlı web sitesinden alınan tüketici yorumları üzerinden çalışılmıştır [7]. Yorumların morfolojik analizinin yapılmasından sonra, sözcük türleri ve aldıkları ekler belirlenmiştir. Metinlerden negatif veya pozitif anlamlar taşıyan bilginin çıkarılması için, ürün özelliğini belirten kelimeler ile bunları niteleyen sıfatlar önceden tespit edilmiştir. Yorum cümleleri için cümle dizilişine ve Türkçe dilbilgisi kuralları göre ağaç yapısı oluşturulmuştur. İstenilen özelliklerin tespit edilmesi için oluşturulan bu ağaç yapısı üzerinde, Derinlik Öncelikli Arama (DFS) algoritmasıyla arama yaparak sonuca ulaşan bir yazılım geliştirilmiştir. Bir araştırmada, Twitter kullanıcılarının İyi Tarım, organik tarım ve sürdürülebilir tarım konularındaki duygu ve düşünceleri duygu analizi yöntemi kullanılarak incelenmiştir [8]. Toplam 15,984 tweet, tarımla ilgili üç farklı hashtag grubunda metin madenciliği tekniğiyle değerlendirilmiştir. Duygu analizi sonuçları, kullanıcılar arasında bu tarım teknikleri hakkında anlamlı farklılıklar olduğunu ortaya koymuştur. İyi Tarım uygulamasının Twitter kullanıcıları arasında en popüler ve güvenilir tarım uygulaması olduğu belirlenmiştir. Müşteri yorumlarının polarizasyonunu hedefleyen bir analizin yapıldığı çalışmada Destek Vektör Makineleri, Naive Bayes, Maksimum Entropi, C4.5 ve Yapay Sinir Ağları gibi makine öğrenimi algoritmaları kullanıldığı bir çalışmada ve performansları karşılaştırılmıştır [9]. Başka bir çalışmada ise yapay sinir ağları kullanılarak Bitcoin fiyatları ve sosyal medyadaki beklentilerin analizi gerçekleştirilmiştir [10]. Toplam 2,819,784 tweet kullanılarak sınıflandırma algoritmaları uygulanmıştır. Yapılan değerlendirmeye göre, Yapay Sinir Ağları'nın metin sınıflandırma konusunda %90'lık bir doğruluk oranına sahip olduğu belirlenmiştir. Avrupa'daki otellerin incelendiği ve otel yorumlarının sınıflandırıldığı bir çalışmada 12,500 otel yorumu olumlu ve olumsuz kategorilere ayrılmıştır. [11]. Metinler TF-IDF matrisi ile temsil edilmiş ve Rastgele Orman, Naive Bayes, DVM, Karar Ağaçları gibi algoritmalar kullanılarak sınıflandırılmıştır. Amazon'dan alınan ürün yorumları veri seti kullanılarak makine öğrenimi metodlarıyla duygu analizi sınıflandırma performansları analiz edilmiştir başka bir çalışmada 12,500 otel yorumu olumlu ve olumsuz kategorilere ayrılmıştır. [12]. Bu çalışmanın amacı, yorumlarda duygu bulunup bulunmadığını tespit etmek ve duygu içeriyorsa bu duygunun olumlu, olumsuz veya tarafsız olma durumunu belirlemektir.

Bu çalışmada, analizde kullanılan veri kümesi UCI Machine Learning Repository'nin açık erişimde bulunan sitesinden indirilmiş 1000 adet İngilizce Amazon müşteri yorumlarını içermektedir [13]. Alınan yorumlar üzerinde kelime çantası (Bag of Words (BoW)) öznitelikleri ile metasezgisel bir algoritma olan Tavuk Sürüsü Optimizasyon Algoritması (TSO) kullanılarak

doğrusal ve karesel formda modelleme ile sınıflandırma işlemi yapılarak sınıflandırıcıların performansları kıyaslanmıştır. Sonuç olarak önerilen yöntemin her iki model için de başarılı sonuç gösterdiği görülmüştür.

2. Yöntem

Metin Önışleme, Öznitelik Çıkarımı, Özellik Seçimi, Veri Madenciliği ve Değerlendirme adımlarından oluşmaktadır [14].

2.1. Ön İşlemler

Metinler üzerinde ön işleme yapmak, metin analizi süreçleri için gerekli olan bir adımdır. Metinleri analiz edebilmek ve metinsel verilerin sayısallaştırılmasını kolaylaştırmak için verilerin tutarlı hâle getirilmesi işlemleri ön işleme aşaması olarak adlandırılmaktadır. Metinlerin analizinde veri setinin toplanmasının ardından bir sonraki aşama metin ön işleme aşamasıdır. Bu aşamanın amacı, metni sayısallaştırmak için uygun bir formata getirebilmektir. Aşamalar genel olarak, metnin kelimelere ayrılması, metin içindeki kelimelerin büyük/küçük harf dönüşümlerinin sağlanması, metin içinde gereksiz noktalama işareti veya rakamların temizlenmesidir. Daha sonra metin içinde anlamsız olarak tabir edilen durak kelimelerin (stop words) çıkarılması sağlanarak metnin gereksiz kelimelerden temizlenmesi sağlanmaktadır. Sonraki aşamada kelimelerdeki eklerin (zaman, iyelik vb.) çıkarılarak kök bulma işlemleri yapılmaktadır. Böylece metinler, sayısallaştırılarak, analiz edilmeye uygun formata dönüştürülmüş olur.

Tablo 1’de UCI Machine Learning Repository’den alınan 1000 adet etiketli müşteri yorumlarını içeren veri setinden ilk 10 yorum gösterilmiştir.

Tablo 1. Veri setinden alınan tüketici yorumları örneği.

Sıra	Yorumlar	Sonuçlar
1	So there is no way for me to plug it in here in the US unless I go by a converter.	0
2	Good case, Excellent value.	1
3	Great for the jawbone.	1
4	Tied to charger for conversations lasting more than 45 minutes. MAJOR PROBLEMS!!	0
5	The mic is great.	1
6	I have to jiggle the plug to get it to line up right to get decent volume.	0
7	If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one.	0
8	If you are Razr owner...you must have this!	1
9	Needless to say, I wasted my money.	0
10	What a waste of money and time!.	0

Metin ön işleme sürecinde, veri seti üzerinde ilk olarak metinler kelimelere ayrılmıştır. Daha sonra cümleler noktalama işaretlerinden temizlenmiştir. Kök indirgeme işlemi yapılarak her bir kelime köklerine indirgenmiştir. En son aşamada da metin içinde bulunan durak kelimeler metinden atılmıştır. Tablo 1’de görülen ilk 10 cümle için ön işleme sonucu Tablo 2’de verilmiştir.

Tablo 2. Ön işleme sonucu ilk 10 cümle.

Sıra	Yorumlar						
1	wai	plug	unless	convert			
2	good	case	excel	valu			
3	great	jawbon					
4	charger	convers	last	minut	major	problem	
5	mic	great					
6	jiggly	plug	get	line	right	decent	volum
7	sever	dozen	hundr	contact	imagin	fun	send
8	razr	owner	thi				
9	needless	sai	wast	monei			
10	wast	monei	time				

2.2. Öznitelik Çıkarımı

Öznitelik çıkarımı, veri setinden ilgili özniteliklerin seçilerek ilgisiz özniteliklerin veri setinden ayrılmasını içerir. Öznitelik seçiminde amaç, ele alınan veri setini en iyi ifade eden alt kümeyi oluşturmaktır [15], [16]. Yani öznitelik çıkarımı sayesinde, veri setinde ilgili olmayan veriler çıkarılmış olup, veri kümesi sayısı azaltılır. Böylece daha az veriyle algoritmalar daha hızlı çalışır ve elde edilen anlamlı veriler ile modelleme başarısı da artar.

Literatürde bir çok öznitelik seçim yöntemi bulunmaktadır [17]. Fisher Skor, Ki- Kare Testi, BoW, Fasttext, Word2Vec, Terim Sıklığı – Döküman Sıklığı (TF-IDF), Glove, LSA, N-gram bunlardan bazılarıdır. Bu çalışmada öznitelik vektör oluşturmada BoW yöntemi kullanılmıştır. BoW’da metin içerisindeki tüm kelimelerin dökümandaki kullanım sıklığı hesaplanarak bir çantada toplanır [18].

Tablo 3. Negatif cümlelerdeki 10 kelime için BoW özniteliğinin çıkarımı.

Sıra	1	2	3	4	5	6	7	8	9	10
Kelimeler	wai	plug	unless	convert	charger	convers	last	minut	major	problem
Sayı	5	12	3	1	12	4	11	5	2	15

Tablo 4. Pozitif cümlelerdeki 10 kelime için BoW özniteliğinin çıkarımı.

Sıra	1	2	3	4	5	6	7	8	9	10
Kelimeler	good	case	excel	value	great	jawbon	mic	razr	owner	thi
Sayı	61	23	28	5	92	3	2	3	1	105

Tablo 4'e göre, "good" kelimesi pozitif cümlelerde 61 kere, Tablo 3'e göre "problem" kelimesi negatif cümlelerde 15 kere geçmiştir.

Tablo 5'te veri setindeki 10 yorum için, her bir yorumun pozitif ve negatif olmak üzere 1x2 boyutunda Bag of Words'ü oluşturulmuştur.

Tablo 5. Verisetindeki ilk 10 cümle için BoW değerleri.

Sentences Row	Positive BoW (PBoW)	Negative BoW (NBoW)	Result
1	5	21	0
2	117	26	1
3	95	5	1
4	23	49	0
5	94	7	1
6	30	48	0
7	8	16	0
8	109	99	1
9	6	42	0
10	14	54	0

2.3. TSO Yönteminin Duygu Analizi Problemine Uygulanması

Tavukların, horozların ve civcivlerin yiyecek arayışındaki davranışlarını modellemek amacıyla geliştirilen doğa tabanlı bir algoritma olan TSO, 2014 yılında Xianbing Meng, Yu Liu, Xiaozhi Gao ve Hengzhen Zhang tarafından literatüre kazandırılmıştır [19], [20]. Bu algoritma, tavukları sürü halindeki doğal davranışlarına dayanarak kategorize eden bir modeli temsil etmektedir.

HSO algoritması, sürüde horozlar, tavuklar ve civcivlerden oluşan gruplara dayanmaktadır [19], [20]. Horozlar, en yüksek uygunluk değerine sahip bireyler olarak belirlenir ve aynı zamanda sürüde lider olarak hareket ederler. Civcivler ise sürüdeki en düşük uygunluğa sahip bireyler olarak tanımlanır. Geriye kalan bireyler ise tavuk olarak adlandırılır. Her alt gruptaki tavuklar, yiyecek arayışında lider konumundaki horozları takip ederler. Arada sırada, diğer tavuklardan yiyecek çalma durumları da yaşanabilir. Civcivlerin, diğer bireyler tarafından bulunan yiyecekleri rastgele çalabileceği ve her civcivin yiyecek arama sırasında annesini takip ettiği varsayılır. Sürü içindeki hiyerarşik düzen, rastgele seçilen tavuklara, rastgele seçilen civcivlerin annesi rolü verilerek korunur [19], [20].

Pozisyon değiştirme işlemleri, yiyecek arama sürecinde önemli bir rol oynar. Bu işlemler, horozlar, tavuklar ve civcivler için farklı şekillerde gerçekleştirilir. TSO algoritmasının hiyerarşik yapısı, Şekil 2’de gösterilmiştir.

Algoritmada kullanılan denklem 2, denklem 1’e bağlı olarak horozların konum değiştirme formülünü ifade etmektedir. Bu formül, horozların en iyi uygunluk değerlerine dayanarak yiyeceğe ulaşma stratejilerini belirler. Horozlar, en yüksek uygunluk değerlerine sahip oldukları için, yiyeceğe ulaşma konusunda öncelikli konumlandırılırlar ve bu durum daha geniş bir alanı kapsamalarını sağlar [21].

$$\sigma = \begin{cases} 1, & \text{if } pfit_i \leq pfit_k, \\ \exp\left(\frac{pfit_k - pfit_i}{|pfit_i| + \epsilon}\right), & \text{otherwise} \end{cases} \quad k \in [1, N], k \neq i \quad [19], [21] \quad (1)$$

$$x_{i,j}^{t+1} = x_{i,j}^t * (1 + \text{Randn}(0, \sigma)) \quad [19], [21] \quad (2)$$

Denklemden, $x_{i,j}^t$, sembolü, i . üyenin t zamanındaki konum bilgisini temsil eder. Horoz grubundan rastgele seçilmiş bir horoz, k ile gösterilir ve uygunluk değeri $pfit$ ’e sifıra bölme probleminden kaçınmak amacıyla sabit bir değer (ϵ) ile düzeltilir [19], [21]. $r1$ ve $r2$ ise sürüden seçilen rastgele bir tavuğu veya horozu temsil eder.

$$S_1 = \exp\left(\frac{pfit_i - pfit_{r1}}{\text{abs}(pfit_i + \epsilon)}\right) \quad [19], [21] \quad (3)$$

$$S_2 = \exp(pfit_{r2} - pfit_i) \quad [19], [21] \quad (4)$$

S_2 , iki tavuğun uygunluk değerleri arasındaki farkı ifade eden bir ölçüdür; yani, tavukların konumları arasındaki mesafenin büyük olduğunu belirtir. Bu durum, tavukların diğer tavuklardan kolayca yiyecek çalamayacakları

anlamına gelir. S_2 'nin küçük olması, sürü içindeki bireyler arasındaki uzaklığın arttığını ve rekabetin azaldığını gösterir. S_1 formülü, S_2 'den farklı bir ölçüdür çünkü grup içinde rekabet söz konusudur. Eğer S_2 'nin değeri 0 ise, bu durum tavuğun kendi bölgesinde horozu takip ettiği anlamına gelir, yani yiyecek arama stratejisinin daha bağımsız olduğu bir durumu ifade eder. Tavukların hareketi, Denklem 5'te belirtilen bir formül ile ifade edilir.

$$x_{i,j}^{t+1} = x_{i,j}^t + S_1 * rand(x_{r_1,j}^t - x_{i,j}^t) + S_2 * rand(x_{r_2,j}^t - x_{i,j}^t) \quad [19], [21] \quad (5)$$

Bu formül, tavukların mevcut konumlarına ve sürü lideri olan horozun konumuna dayanarak, yiyecek arama sırasında izledikleri yolun matematiksel bir ifadesini sunar.

Yiyecek arama sürecindeki civcivlerin konumlandırılması, algoritmada Denklem 6 ile belirtilmiştir. Bu denklem, civcivlerin anneleri etrafında nasıl hareket edeceklerini ve yiyecek arama stratejilerini ifade eder. FL sembolü ise bir civcivin annesini ne kadar hızlı takip edeceğini gösteren temsili bir parametreyi temsil eder [21].

$$x_{i,j}^{t+1} = x_{i,j}^t + FL * (x_{mother,j}^t - x_{i,j}^t) \quad [19], [21] \quad (6)$$

Veriler, Model-1TSO doğrusal formu ve Model-2TSO karesel formu iki farklı formda modellenmiş ve test edilmiştir. Bu modellerin her biri, duygu analizi veri setini sınıflandırma konusundaki performanslarını belirlemek amacıyla çeşitli formüllerle ifade edilmiştir.

Tablo 6. TSO'nun uygulandığı modeller.

TSO uygulanan Model İsmi	Model Tipi
Model-1TSO	Çoklu Lineer Form
Model-2TSO	Çoklu Karesel Form

Tablo 6'da gösterildiği gibi, Model-1TSO doğrusal formu ve Model-2TSO karesel formu her birinin kendi içerisinde benzersiz bir matematiksel ifade ve modelleme şekline sahiptir. Bu modeller, TSO algoritmasının kullanıcı yorumlarını sınıflandırmak için farklı matematiksel formülleri nasıl kullandığını ve bu formüllerin hangi durumlarda daha etkili olduğunu anlamak amacıyla oluşturulmuştur.

Elde edilen negatif ve pozitif BoW öznitelikleri Denklem 7'de sunulmuştur.

$$Q = [Q_1, Q_2] \quad (7)$$

Denklem 7'ye göre, özellik vektörünün 1. sütunu (Q_1) pozitif BoW sayısını verirken, 2. sütun (Q_2) negatif Bow sayısını vermektedir.

Denklem 8'de, özniteliklerin kombinasyonu cinsinden uygulanan çoklu doğrusal regresyon modelinin denklemi sunulmaktadır. Bu denklem, duygu analizinde sınıflandırma probleminde kullanılan özellik vektörünün veri seti değişkenleriyle olan kombinasyonunu ifade eder. Bu bağlamda, NBoW ve PBoW değişkenleri, çoklu doğrusal regresyon modelinde yer alır problemi sınıflandırmak için kullanılan katsayılarla ilişkilendirilir.

$$y = Q_0 + Q_1 \text{PBoW} + Q_2 \text{NBoW} = Q_0 + \sum_{i=1}^k Q_i x_i \quad (8)$$

Duygu analizi değişkenlerinin kombinasyonu cinsinden uygulanan çoklu ikinci dereceden regresyon modelinin denklemi, denklem 9'da ifade edilmiştir.

$$y = Q_0 + Q_1 \text{PBoW} + Q_2 \text{NBoW} + Q_3 \text{PBoW}^2 + Q_4 \text{PBoW} \cdot \text{NBoW} + Q_5 \cdot \text{NBoW}^2 \quad (9)$$

Eşitlik 8 ve 9'daki değerler sırasıyla $0 \leq i \leq 2$ ve $0 < i < 5$ 'tir. Bu parametreler, Eşitlik 8 ve 9 için belirlenecek olan değerlerdir ve bunlar TSO kullanılarak belirlenecektir.

$$J(Q) = \frac{1}{N} \sum_{i=1}^N (Y - X_i Q^T)^2 \quad (10)$$

Denklem 10'da X , girdi kümesini yani kullanıcı yorumlarını, X_i ise i . cümle kaydındaki kayıtları temsil eder, cümle sayısı N , Y ise etiket değerlerini ifade eden çoklu doğrusal formdaki maliyet fonksiyonu örnek olarak verilmiştir.

Bu araştırma, veri madenciliği parametre optimizasyonunu ele alarak TSO algoritmasını kullanarak model parametrelerini (Q) tahmin etmeyi hedeflemektedir. Çalışmanın temel odak noktası, duygu analizi ile tüketici yorumlarının sınıflandırması yaparak veriyi analiz etmektir.

Bu bağlamda, Tablo 7'de aşamaları verilen TSO algoritması kullanılarak elde edilen model parametreleri, tüketici yorumlarının sınıflandırılmasındaki etkinliği değerlendirmek amacıyla optimize edilmeye çalışılacaktır.

Tablo 7. Model-ITSO için TSO algoritmasının aşamaları.

Adım 1	Veri kümesi tanıtılır ve xtrain, ytrain, xtest ve ytest olarak ayrılmıştır.
Adım 2	Popülasyon büyüklüğü (N), maksimum iterasyon sayısı (T) tanımlanmıştır. (Çalışmada $N=50$ ve $T=100$ alınmıştır.)
Adım 3	Popülasyon büyüklüğü kadar bir Q matrisi üretilir. ($-5 < Q_0 < 5$ ve $-3 < Q_1, Q_2 \dots Q_N < 3$)
Adım 4	Amaç fonksiyonu Denklem 11 dikkate alınarak hesaplanır $J = \frac{1}{2} (Q_0 + \sum_{i=1}^k y_i - Q_i x_i)^2 \quad (11)$
Adım 5	En düşük hata değerine sahip amaç fonksiyonuna ait Q matrisi bulunur ve saklanır.
Adım 6	Tavukların uygunluk değerleri sıralanarak hiyerarşik bir düzen kurulur. Sürü daha sonra birçok gruba ayrılacaktır.
Adım 7	i 'nin değerine bağlı olarak a) Horozlar için çözüm kümesi denklem 2 kullanılarak güncellenir. b) Tavuklar için çözüm kümesi denklem 5 kullanılarak güncellenir. c) Yavru tavukların çözüm kümesi denklem 6 kullanılarak güncellenir.
Adım 8	Elde edilen sonuç saklanan sonuçtan daha üstünse, güncellenecektir.
Adım 9	Çift yer değiştirme mutasyonu uygulanacaktır.
Adım 10	Popülasyon optimize edilecektir.
Adım 11	Yineleme tamamlanana kadar 5. adıma geri dönün.

Tablo 8 ve Tablo 9'da gösterilen katsayılar, TSO algoritmasının doğrusal ve karesel modellere uygulanması sonucunda elde edilmiştir.

Tablo 8. Model-ITSO ile bulunan parametre ağırlıkları.

Parameter	Ağırlık
$1(Q_0)$	0.499
PBoW (Q_1)	1.699
NBoW (Q_2)	-1.313

Tablo 9. Model-2TSO ile bulunan parametre ağırlıkları.

Parametre	Ağırlık
1	-3.308
PBoW	0.530
NBoW	-0.286
PBoW²	2.483
PBoW x NBoW	1.663
NBoW²	-0.212

3. Deneysel Çalışma

Çalışma kapsamında 1000 adet veri seti %80 eğitim ve geri kalanı test verisi olarak ayrılmış ve 2 farklı deney gerçekleştirilmiştir.

3.1 Veri Seti

Çalışma kapsamında analiz edilen veri, UCI Machine Learning Repository'nin açık erişimde bulunan sitesinden indirilmiş 1000 adet İngilizce <https://www.amazon.com/> sitesinden alınan etiketli müşteri yorumlarını içermektedir. 500 olumlu, 500 olumsuz olmak üzere toplamda 1000 adet yorum ele alınmıştır. Bu yorumlar etiketli olup, olumlu ve olumsuz yorum şeklide 2 kategoridedir. Veri setinin %80'i tanesi eğitim seti, %20'si test seti olarak ayrılmıştır.

Bu çalışmada TSO algoritması kullanılarak Bag of Words özneliğiyle 2 farklı deney gerçekleştirilmiştir. Yapılan deneylerle İngilizce tüketici yorumlarını içeren ve <https://www.amazon.com/> 'dan alınan veri seti üzerinde duygu analizi ile sınıflandırma amaçlanmıştır. Sınıflandırıcıların performansını kıyaslamak için ROC analizi ölçütünden yararlanılmıştır. ROC analizinin amacı, çeşitli metodlar ile elde edilen sonuçların performansını karşılaştırabilmek için, sonuçların duyarlılık, özgüllük ve doğruluk olarak değerlendirilebilmesini sağlayan bir metottur [22]. Analizde kullanılan ROC parametreleri TP, TN, FP ve FN'dir. TP (Doğru Pozitif); olumlu olarak sınıflandırılmış bir yorumun, analiz sonucunda olumlu olarak tespit edilmesiyken, TN (Doğru Negatif); olumsuz olarak sınıflandırılmış bir yorumun, analiz sonucunda olumsuz olarak tespit edilmesidir. FP (Yanlış Pozitif); olumlu olarak tespit edilen bir yorumun analiz sonucu test sonuçlarının olumsuz olarak tespit edilmesiyken, FN (Yanlış Negatif);

olumsuz olarak tespit edilen yorumun analiz sonucu test sonuçlarının yanlış olarak yani olumlu olarak tespit edilmesidir.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

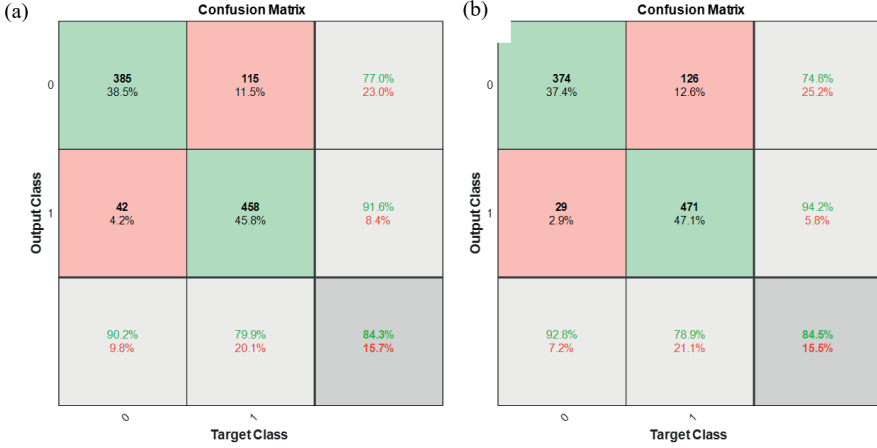
$$F1 - Score = 2 \frac{TP}{(2TP + FP + FN)} \quad (16)$$

Denklem 11-15 ROC analizi parametrelerinin formülasyonları verilmiştir. Bu parametreler incelendiğinde Accuracy parametresi, genel doğruluğu ifade ederek, sınıflandırılan yorumlardan yüzde kaçının önerilen yöntemle doğru sınıflandırıldığı bilgisidir. Sensitivity, gerçekte olumlu olarak etiketlenmiş yorumların yüzde kaçının önerilen yöntemle olumlu sınıflandırıldığı bilgisidir. Specificity parametresi, gerçekte olumsuz olarak etiketlenmiş yorumların yüzde kaçının önerilen yöntemle olumsuz sınıflandırıldığı bilgisidir. Precision, tüm olumlu etiketlenmiş yorumlardan kaçının doğru tahmin edildiği bilgisidir. F1-Score, Precision ve Sensitivity değerlerini harmonik ortalamasıdır. Yapılan deneylerde Sensitivity, Specificity, Precision, Accuracy, F1-Score değerleri bulunarak Tablo 10'da listelenmiştir.

Tablo 10. ROC performans analizleri.

Model Name	Accuracy	Precision	Recall	Specificity	F1-Score
Model-1TSO	0.843	0.799	0.916	0.770	0.854
Model-2TSO	0.845	0.928	0.748	0.942	0.828

2 farklı model için elde edilen konfüzyon matrisleri Şekil 1'de verilmiştir.



Şekil 1. Önerilen modeller için konfüzyon matrisleri; (a) Model-1TSO için konfüzyon matrisi, (b) Model-2TSO için konfüzyon matrisi.

Şekil 1 (a)'daki karışıklık matrisi değerlendirildiğinde Model-1TSO ile veri setinde toplamda 157 tüketici yorumunun doğru sınıflandırılmadığı görülmüştür. Şekil 1(b)'deki karışıklık matrisi değerlendirildiğinde Model-2TSO ile toplamda 155 tüketici yorumunun doğru sınıflandırılmadığı görülmüştür.

Bu çalışmada, tüketici yorumlarına ait bir veriseti üzerinde BoW öznitelikleri çıkarılarak öznitelikler arasındaki ilişkiyi daha etkili bir şekilde modelleyebilmek amacıyla doğrusal ve karesel senaryolarda çeşitli testler gerçekleştirilmiştir. Bu testlerde, TSO metodu, her iki senaryo üzerinde kullanılan farklı modellerle detaylı bir şekilde incelenmiştir. Yapılan araştırmada, her iki modelin de sınıflandırma başarısının yüksek olduğu sonucuna ulaşılmıştır. Bu sonuçlar, parametreler arasındaki ilişkinin tüketici yorumlarının sınıflandırılarak duygu analizi yapma problemine katkısının önemli olduğunu göstermektedir.

4. Sonuçlar

Bu çalışma kapsamında UCI Machine Learning veri tabanından alınan <https://www.amazon.com/> deki 1000 adet İngilizce müşteri yorumlarını içeren veri seti üzerinde duygu analizi gerçekleştirilmiştir. Ele alınan tüketici yorumlarının %80'i eğitim geri kalan %20'si eğitim setine ayrılarak metasezgisel bir algoritma olan Tavuk Sürü Optimizasyon Algoritması kullanılarak çoklu doğrusal form ve karesel formda modelleme ile sınıflandırma işlemi yapılarak BoW özniteliğiyle sınıflandırıcıların performansları test

edilmiştir. Testler sonucunda TSO algoritmasının uygulandığı iki modelin de yüksek başarı gösterdiği görülmüştür.

Gelecek çalışmalarda, TSO algoritmasıyla daha iyi performansta modelleme ve sınıflandırma yapabilmek için daha çok öznelik kullanılmasıyla daha iyi sonuçların alınacağı ve TSO algoritmasının metin madenciliği çalışmalarında sıklıkla kullanılacağı düşünülmektedir.

REFERANSLAR

1. Liu B., Sentiment Analysis and Subjectivity, in *Handbook of Natural Language Processing*, 2010. [Online]. Available: <https://api.semanticscholar.org/CorpusID:13908471>
2. Pang B., Lee L., Vaithyanathan S., Thumbs up? Sentiment classification using machine learning techniques, arXiv preprint cs/0205070, 2002.
3. Amasyalı M.E., Diri B., Automatic Turkish text categorization in terms of author, genre and gender, in *International Conference on Application of Natural Language to Information Systems*, Springer, 2006, 221–226.
4. Pilavcılar İ.F., Metin madenciliği ile metin sınıflandırma, 2007.
5. Bahar M.A., Twitter verileri ile makine öğrenmesi kullanılarak duygu analizi: Torku örneği , Necmettin Erbakan Üniversitesi Sosyal Bilimler Enstitüsü, 2023.
6. Pajupuu H., Altrov R., Pajupuu J., Identifying polarity in different text types, *Folklore: Electronic Journal of Folklore*, 64, 125–142, 2016.
7. Ergün K., Metin madenciliği yöntemleri ile ürün yorumlarının otomatik değerlendirilmesi, Doktora, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya, 2012.
8. TEMİZHAN E., M.M., Tarım hakkında atılan tweetlerin duygu analizi değerlendirmesi, *Harran Tarım ve Gıda Bilimleri Dergisi*, 27 (3), 352–361, 2023.
9. Zubrinic K., Sjekavica T., Milicevic M., Obradovic I., A Comparison of Machine Learning Algorithms in Opinion Polarity Classification of Customer Reviews, *International Journal of Computers*, 3, 2018.
10. ÇILGIN C., Ceyda Ü., ALICI S., AKKOL E., GÖKŞEN Y., Metin Sınıflandırmada Yapay Sinir Ağları ile Bitcoin Fiyatları ve Sosyal Medyadaki Beklentilerin Analizi, *Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi*, 4 (1), 106–126, 2020.
11. Campos D., Silva R.R., Bernardino J., Text Mining in Hotel Reviews: Impact of Words Restriction in Text Classification., in *KDIR*, 2019, 442–449.
12. Göçgün Ö.F., Aytuğ O., Amazon ürün değerlendirmeleri üzerinde derin öğrenme/makine öğrenmesi tabanlı duygu analizi yapılması, *Avrupa Bilim ve Teknoloji Dergisi*, (24), 445–448, 2021.
13. Kotzias D., Denil M., De Freitas N., Smyth P., From group to individual labels using deep features, in *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2015, 597–606.
14. Zohar Y.E., Introduction to text mining, University of Illinois, 2002.

15. Forman G., An extensive empirical study of feature selection metrics for text classification., J. Mach. Learn. Res., 3 (Mar), 1289–1305, 2003.
16. Budak H., Özellik seçim yöntemleri ve yeni bir yaklaşım, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 22, 21–31, 2018.
17. Onan A., Korukoğlu S., Metin sınıflandırmada öznelik seçim yöntemlerinin değerlendirilmesi, Akademik Bilişim, 2016.
18. Aksu M.Ç., Karaman E., FastText ve Kelime Çantası Kelime Temsil Yöntemlerinin Turistik Mekanlar İçin Yapılan Türkçe İncelemeler Kullanılarak Karşılaştırılması, Avrupa Bilim ve Teknoloji Dergisi, (20), 311–320, 2020.
19. Meng X., Liu Y., Gao X., Zhang H., *A new bio-inspired algorithm: chicken swarm optimization*. Springer, 86–94.
20. Verma S., Sahu S.P., Sahu T.P., MCSO: Levy's Flight Guided Modified Chicken Swarm Optimization, IETE Journal of Research, 1–15, 2023.
21. Gullu M., Author Identification Based On The Ensemble Learning Approach Supported By Ooptimiazation – Based Feature Selection Methods.
22. Obuchowski N.A., ROC analysis, American Journal of Roentgenology, 184 (2), 364–372, 2005.