

**T.C.
SAKARYA ÜNİVERSİTESİ
İŞLETME ENSTİTÜSÜ**

**TÜKETİCİ YORUMLARININ FAYDA DÜZEYİNİN
TAHMİNLENMESİNE YÖNELİK BİR ARAŞTIRMA:
MAKİNE ÖĞRENMESİ ALGORİTMALARININ
KARŞILAŞTIRILMASI**

**YÜKSEK LİSANS TEZİ
Oğuzhan ARI**

Enstitü Anabilim Dalı : Yönetim Bilişim Sistemleri

Tez Danışmanı: Doç. Dr. Adem AKBIYIK

MAYIS – 2022

Oğuzhan ARI tarafından hazırlanan ‘‘Tüketici Yorumlarının Fayda Düzeyinin Tahminlenmesine Yönelik Bir Araştırma: Makine Öğrenmesi Algoritmalarının Karşılaştırılması’’ başlıklı bu tez, 30/05/2022 tarihinde Sakarya Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yapılan Tez Savunma Sınavı sonucunda başarılı bulunarak, jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Adem AKBIYIK
Sakarya Üniversitesi

Jüri Üyeleri: Dr. Öğr. Üyesi Naciye Güliz UĞUR
Sakarya Üniversitesi

Dr. Öğr. Üyesi Anıl Zekiye GÖKER
Doğuş Üniversitesi

 SAKARYA ÜNİVERSİTESİ	T.C.		Sayfa : 1/1
	SAKARYA ÜNİVERSİTESİ		
	İŞLETME ENSTİTÜSÜ		
	TEZ SAVUNULABİLİRLİK VE ORJİNALLİK BEYAN FORMU		
Öğrencinin			
Adı Soyadı	:	Oğuzhan ARI	
Öğrenci Numarası	:	Y199054001	
Enstitü Anabilim Dalı	:	Yönetim Bilişim Sistemleri	
Enstitü Bilim Dalı	:	Yönetim Bilişim Sistemleri	
Programı	:	<input checked="" type="checkbox"/> YÜKSEK LİSANS <input type="checkbox"/> DOKTORA	
Tezin Başlığı	:	Tüketici Yorumlarının Fayda Düzeyinin Tahminlenmesine Yönelik Bir Araştırma: Makine Öğrenmesi Algoritmalarının Karşılaştırılması	
Benzerlik Oranı	:	% 2	
<p>Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen tez çalışmasının benzerlik oranının herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.</p>			
			17 / 05 / 2022 İmza Öğrenci
<p>Sakarya Üniversitesi İşletme Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen öğrenciye ait tez çalışması ile ilgili gerekli düzenleme tarafımda yapılmış olup, yeniden değerlendirilmek üzere@sakarya.edu.tr adresine yüklenmiştir.</p>			
Bilgilerinize arz ederim.			
		 / / 20.... İmza Danışman
Uygundur			
		Danışman Unvanı / Adı-Soyadı: Tarih: 17 / 05 / 2022 İmza:	
<input type="checkbox"/> KABUL EDİLMİŞTİR <input type="checkbox"/> REDDEDİLMİŞTİR		Enstitü Birim Sorumlusu Onayı	
EYK Tarih ve No:			

ÖNSÖZ

Bu tezin yazılması aşamasında, sadece tez çalışmamı değil, bir akademisyen olarak yetişirken de beni titizlikle takip eden danışmanım ve mentorum kıymetli hocam Doç. Dr. Adem AKBIYIK'a değerli katkı ve emekleri için içten teşekkürlerimi ve saygılarımı sunuyorum. İdareci ve akademisyen kimliğini aşarak bana her konuda dostça yaklaşan, her zaman dinleyen ve yardımcı olan kıymetli hocam Dr. Öğr. Üyesi Anıl Z. GÖKER'e çok teşekkür ediyorum. Lisans öğrencisi olduğum süreçten başlayarak ilgisini ve alakasını hiçbir zaman eksik etmeyen kıymetli hocalarım Dr. Öğr. Üyesi Naciye Güliz UĞUR ve Dr. Öğr. Üyesi Tuğba KOÇ'a çok teşekkür ediyorum. Her zaman sorularımı cevaplayan, desteklerini benden esirgemeyen hocalarım Arş. Gör. Bahadır AKTAŞ ve Arş. Gör. Kübra ÇALIŞKAN'a katkılarından dolayı çok teşekkür ediyorum. Bugünlere gelmem için desteklerini hiçbir zaman esirgemeyen, her zaman kahrımı çeken ve derdime ortak olan babam Osman ARI, annem Tülay ARI ve sevgili kardeşim Zeynep Begüm ARI'ya teşekkürlerimi ve minnetlerimi sunarım. Biricik kedimiz, her zaman enerjimizi yükselten Ares'e sevgilerimi sunarım. Son olarak, her zaman yanımda olan, her derdime koşan, desteğini her an yanımda ve arkamda hissettiğim biricik hayat arkadaşım Beyza YILDIRIM'a teşekkürü bir borç bilirim.

Oğuzhan ARI

30.05.2022

İÇİNDEKİLER

KISALTMALAR	iv
TABLO LİSTESİ	v
ŞEKİL LİSTESİ	vi
ÖZET	viii
ABSTRACT	ix
GİRİŞ	1
BÖLÜM 1: KAVRAMSAL ÇERÇEVE	4
1.1. Ağızdan Ağıza İletişim	4
1.1.1. Nihai Organik Tüketici Etkisi Modeli	5
1.1.2. Doğrusal Pazarlamacı Etki Modeli	6
1.1.3. Ortak Çevre Modeli	7
1.2. Kullanıcı Yorumları ve Yıldız Derecelendirme	9
1.3. Web 2.0	10
1.4. Faydalı Tüketici Değerlendirmesi	13
1.5. Önceki Çalışmalar	16
BÖLÜM 2: VERİ, ALGORİTMA VE MAKİNE ÖĞRENİMİ	21
2.1. Algoritma	21
2.2. Veri	24
2.2.1. Veri Değişken türleri	24
2.2.1.1. Kategorik değişkenler	24
2.2.1.2. Nümerik Değişkenler	24
2.2.2. Veri Ön İşleme	25
2.2.2.1. Veri Temizleme	25
2.2.2.2. Veri Dönüştürme	27
2.3. Makine Öğrenimi	27
2.3.1. Denetimli Öğrenme	28
2.3.1.1. Regresyon	29
2.3.1.2. Naive Bayes	32

2.3.1.3. Logistik Regresyon	33
2.3.1.4. Karar Ağacı	34
2.3.1.5. Rassal Orman	36
2.3.1.6. Gradyan Arttırma Sınıflandırıcısı ve XGB Sınıflandırıcısı	37
2.4. Denetimsiz Öğrenme ve Denetimsiz Öğrenme Türleri.....	38
2.5. Pekiştirmeli Öğrenme.....	39
2.6. Makine Öğrenmesi Modelini Sınamak	39
2.6.1. Bias ve Varyans.....	39
2.6.2. Aşırı Öğrenme	39
2.6.3. Az Öğrenme	41
2.6.4. Hata matrisi, Doğruluk, Duyarlılık ve Kesinlik.....	42
2.6.5. Doğruluk (Accuracy)	43
2.6.6. Hassasiyet (Recall).....	43
2.6.7. Kesinlik (Precision).....	44
2.6.8. Özgüllük (Specificity).....	44
2.6.9. ROC-AUC Eğrisi	45
2.6.10. F1 Skoru	46
2.7. Duygu Analizi ve BERT	46
2.8. Python	48
BÖLÜM 3: MODEL, EĞİTİM VE ANALİZ.....	51
3.1. Çalışma Modeli, Öznitelik Seçimi, Veri Alma Süreçleri.....	51
3.1.1. Çalışma Modeli ve Uygulama.....	51
3.1.1.1. Çalışma Modeli	51
3.1.1.2. Öznitelik Seçimi.....	54
3.1.1.3. Veri Elde Etme	54
3.2. Veri Ön İşleme	57
3.3. Model Eğitimi için Rastgele Veri Seçimi	59
3.4. Modellerin Eğitimi	59
3.4.1. Model Eğitimleri ve Değerlendirme	59

3.4.1.1. Naive Bayes Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları.....	65
3.4.1.2. Logistik Regresyon Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları.....	66
3.4.1.3. Karar Aęacı Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları.....	68
3.4.1.4. Rassal Orman Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları.....	69
3.4.1.5. Gradyan Arttırma Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları.....	71
3.4.1.6. XGBoost Tahmin Hata Matrisi Deęerlendirmesi ve Öznitelik Aęrlıkları	72
3.5. Algoritmaların Özniteliklerin Deęerlendirilmesi.....	74
3.6. Farklı Kategorilerinden Tüketici Deęerlendirmeleri Üzerinde Algoritma Başarıları.....	76
SONUÇ.....	81
KAYNAKÇA.....	85
EKLER.....	94
ÖZGEÇMİŞ.....	101

KISALTMALAR

AAİ	: Ağızdan Ağıza İletişim
E-AAİ	: Elektronik Ağızdan Ağıza İletişim
WWW	: World Wide Web

TABLO LİSTESİ

Tablo 1: Önceki Çalışmalar	17
Tablo 2: Pozitif Tüketici Değerlendirmesi-Satış Miktarı.....	30
Tablo 3: Negatif Tüketici Değerlendirmesi-Satış Miktarı	31
Tablo 4: Hata Matrisi	42
Tablo 5: Literatür Taraması.....	54
Tablo 6: Yorumların Yıldız Derecelendirmelerine Göre Dağılımı.....	55
Tablo 7: 1 ve 5 Yıldız Derecelendirmeli Yorum Sayısı.....	56
Tablo 8: Nitelik Özellikleri	57
Tablo 9: İki veya Daha Fazla Defa Faydalı İşaretlenen Değerlendirmelerin Özellikleri	57
Tablo 10: Bir Kez Faydalı İşaretlenen Değerlendirmelerin Özellikleri.....	57
Tablo 11: Özniteliklerin Özellikleri	58
Tablo 12: Faydalı Faydalı Değil Etiket Dağılımı	59
Tablo 13: Algoritma Öznitelik Ağırlıkları	74
Tablo 14: Test Verisinin Kategorilere Göre Dağılımı	76
Tablo 15: Yeni Tarihli Tüketici Değerlendirmesi.....	79

ŞEKİL LİSTESİ

Şekil 1: Nihai Organik Tüketici Etkisi Modeli.....	5
Şekil 2: Doğrusal Pazarlamacı Etki Modeli.....	6
Şekil 3: Ortak Çevre Modeli.....	7
Şekil 4: Web 1.0 Kullanıcı Etkileşimi	11
Şekil 5: Web 2.0 Kullanıcı Etkileşimi	12
Şekil 6: Web 1.0 ve 2.0 Karşılaştırması	12
Şekil 7: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri	14
Şekil 8: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri	14
Şekil 9: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri	15
Şekil 10: Değerlendirme Faydalı Etiketleme.....	15
Şekil 11: Değerlendirme Faydalı Etiketleme.....	16
Şekil 12: Soyut Makine Akış Diyagramı.....	22
Şekil 13: Özyineleme Algoritması Akış Diyagramı.....	23
Şekil 14: Tek Bağımsız Değişkenli Regresyon Denklemi	29
Şekil 15: Pozitif Tüketici Değerlendirmesi-Satış Miktarı Grafiği.....	31
Şekil 16: Negatif Tüketici Değerlendirmesi-Satış Miktarı Grafiği	32
Şekil 17: Bayes Formülü	33
Şekil 18: Logistik Fonksiyonu.....	34
Şekil 19: Karar Ağacı Örneği	35
Şekil 20: Karar Ağacı Örneği	36
Şekil 21: Ortalama Hata Karesi	37
Şekil 22: Az Öğrenme Örneği	41
Şekil 23: Doğruluk Formülü.....	43
Şekil 24: Hassasiyet Formülü	44
Şekil 25: Kesinlik Formülü.....	44
Şekil 26: Özgüllük Formülü	45
Şekil 27: Örnek ROC-AUC Eğrisi	45
Şekil 28: F1 Formülü	46
Şekil 29: BERT, OpenAI GPT ve ELMo	47
Şekil 30: Çalışma Modeli	51
Şekil 31: Veri Edinme Süreci	52
Şekil 32: Veri Ön İşleme Süreci	53

Şekil 33: Öznitelik Çıkarımı.....	56
Şekil 34: Algoritmalara Göre Doğruluk Oranları.....	60
Şekil 35: Algoritmalara Göre ROC-AUC Oranı.....	61
Şekil 36: ROC Eğrisi.....	62
Şekil 37: Algoritmalara Göre F1 Skorları.....	63
Şekil 38: Algoritmaların Doğruluk Oranı, ROC-AUC ve F1 Skoru.....	64
Şekil 39: Naive Bayes Tahmin Hata Matrisi.....	65
Şekil 40: Naive Bayes Öznitelik Ağırlıkları.....	66
Şekil 41: Logistik Regresyon Tahmin Hata Matrisi.....	67
Şekil 42: Logistik Regresyon Öznitelik Ağırlıkları.....	67
Şekil 43: Karar Ağacı Tahmin Hata Matrisi.....	68
Şekil 44: Karar Ağacı Öznitelik Ağırlıkları.....	69
Şekil 45: Rassal Orman Tahmin Hata Matrisi.....	70
Şekil 46: Rassal Orman Öznitelik Ağırlıkları.....	70
Şekil 47: Gradyan Arttırma Tahmin Hata Matrisi.....	71
Şekil 48: Gradyan Arttırma Öznitelik Ağırlıkları.....	72
Şekil 49: XGBoost Tahmin Hata Matrisi.....	73
Şekil 50: XGBoost Öznitelik Ağırlıkları.....	73
Şekil 51: Algoritmaların Ortak Öznitelik Ağırlığı.....	75
Şekil 52: Ağaç Temelli Algoritma Öznitelik Ağırlıkları.....	75
Şekil 53: Farklı Kategoriler Üzerinde Doğru Tahmin Oranı.....	77
Şekil 54: Farklı Kategoriler Üzerinde Doğru-Yanlış Tahmin Sayısı.....	78
Şekil 55: Geçmiş Tarihli Tüketici Değerlendirmesinin Güncel Durumu.....	80

Tezin Başlığı: Tüketici Yorumlarının Fayda Düzeyinin Tahminlenmesine Yönelik Bir Araştırma: Makine Öğrenmesi Algoritmalarının Karşılaştırılması	
Tezin Yazarı: Oğuzhan ARI	Danışman: Doç. Dr. Adem AKBIYIK
Kabul Tarihi: 30/05/2022	Sayfa Sayısı: ix(ön kısım) + 84(tez) + 6(ek)
Anabilim Dalı: Yönetim Bilişim Sistemleri	
<p>Web 2.0'ın gelişmesiyle birlikte tüketiciler arasında çevrimiçi paylaşım platformlarında bilgi alışverişi yaygınlaşmış, tüketiciler arasındaki bilgi alışverişi sınırı neredeyse kalkmıştır. Tüketiciler, Web 2.0 ve teknolojileri öncesinde ancak kendi şehirlerinde, kendi mahallelerinde olan insanlar ile bilgi ve fikir alışverişinde bulunabilirken, bu iletişim türü yavaşça internet platformlarına taşınmış ve Elektronik Ağızdan Ağıza İletişim kavramını ortaya çıkarmıştır. Bu sayede tüketiciler, bir ürün veya hizmet hakkında, sadece kendi çevresinde değil dünyanın her yerinde bulunan tüketicilerle bilgi alışverişinde bulunabilmektedir. Gelişen teknolojiler ve tüketicilerin internet kullanım alışkanlarının da gelişmesiyle, tüketicinin faydasına olacak bilgiye ulaşmak da zorlaşmaktadır. Çevrimiçi alışveriş platformları, faydalı bilginin tespiti ve ön plana çıkması için çeşitli yöntemler sunsa da bir tüketici yorum ve değerlendirmesinin faydalı olmasının belirlenmesi, diğer tüketicilerin onayına bağlıdır.</p> <p>Makine öğrenmesi alanında yaşanan gelişmeler, her geçen gün bir başka işin makinelere devredilmesine olanak sağlıyor. Bu sayede, insanların hiç müdahalesi olmadan ya da çok az müdahalesi olarak tekrarlı işler veya kontrol süreçleri otomatik bir şekilde ilerleyebilmektedir. Makine öğrenmesinin bir dalı olan denetimli öğrenme, belirli etiket veya sınıflara sahip veri gruplarının özniteliklerini öğrenerek, etiketi veya sınıfı belli olmayan varlıkların etiketlenmesinde ya da sınıflandırılmasında kullanılabilir.</p> <p>Çalışmanın ilk bölümü ağızdan ağıza iletişim, Web 1.0 ve Web 2.0, faydalı tüketici değerlendirmesi ile ilgili genel bilgiler içermektedir. İkinci bölüm ise, algoritmanın ne olduğu, verinin ne olduğu, veri ön işleme süreçleri, makine öğrenmesi, denetimli öğrenme ve alt başlıkları, denetimsiz öğrenme, pekiştirmeli öğrenme ve makine öğrenmesi algoritmalarını sınamak ile ilgili bilgiler içermektedir. Çalışmanın üçüncü bölümü, araştırma modeli, veri alma süreçleri, veri ön işleme, öznitelik çıkarımı, modellerin eğitilmesi ve test edilmesini içermektedir. Çalışmada 6 farklı denetimli makine öğrenmesi algoritması, çevrimiçi alışveriş platformunda bir ürüne yazılmış tüketici değerlendirmelerinden elde edilen özniteliklerle eğitilerek, yeni bir kullanıcı değerlendirmesinin, tüketiciler tarafından faydalı olarak işaretlenip işaretlenmeyeceği tahmin edilmeye çalışılmıştır. Algoritmalar, doğruluk, ROC eğrisi altında kalan alan ve F1 skorlarına göre sınanmış, farklı ürün kategorilerinden gelen tüketici değerlendirmeleri ile harici olarak da sınanmıştır. Bu çalışmanın, tüketicileri, diğer tüketicilere faydalı olacak bir tüketici değerlendirmesinin, diğer tüketicilerin onayına sunulmaksızın tespit edilebileceği bilgisiyle, yeni bir tüketici değerlendirmesi hazırlanması konusunda teşvik etmesi beklenmektedir.</p>	
Anahtar Kelimeler: Makine Öğrenmesi, Ağızdan Ağıza İletişim, Faydalı Tüketici Değerlendirmesi, Denetimli Öğrenme	

Title of Thesis: A Study on Estimating the Usefulness Level of Consumer Reviews: Comparison of Machine Learning Algorithms	
Author of Thesis: Oğuzhan ARI	Supervisor: Assoc. Prof. Adem AKBIYIK
Accepted Date: 30/05/2022	NP: ix(pre text) + 84(main body) + 6(App.)
Department: Management Information Systems	
<p>With the development of Web 2.0, information exchange among consumers on online sharing platforms has become widespread, and the limit of information exchange between consumers has almost disappeared. While consumers could only exchange information and ideas with people in their own cities and neighborhoods before Web 2.0 and technologies, this type of communication slowly moved to internet platforms and revealed the concept of Electronic Word of Mouth Communication. In this way, consumers can exchange information about a product or service not only with consumers in their own neighborhood but also all over the world. With the developing technologies and the development of consumers' internet usage habits, it becomes difficult to reach information that will benefit the consumer. While online shopping platforms offer a variety of methods for identifying and highlighting useful information, determining whether a consumer review and evaluation is useful depends on the approval of other consumers.</p> <p>Developments in the field of machine learning enable the transfer of another job to machines day by day. In this way, repetitive work or control processes can proceed automatically with little or no human intervention. Supervised learning, which is a branch of machine learning, can be used to label or classify entities with unknown labels or classes by learning the attributes of data groups with certain labels or classes.</p> <p>The first part of the study includes general information about word of mouth, Web 1.0 and Web 2.0, useful consumer evaluation. The second part includes information about what the algorithm is, what the data is, data preprocessing processes, machine learning, supervised learning and its sub-titles, unsupervised learning, reinforcement learning and testing machine learning algorithms. The third part of the study includes the research model, data acquisition processes, data preprocessing, feature extraction, training and testing of models. In the study, 6 different supervised machine learning algorithms were trained with the attributes obtained from consumer evaluations written on a product in the online shopping platform, and it was tried to predict whether a new user evaluation would be marked as useful by consumers. Algorithms were tested for accuracy, area under the ROC curve and F1 scores, and were also externally tested with consumer reviews from different product categories. It is expected that this study will encourage consumers to prepare a new consumer assessment with the knowledge that a consumer assessment that will be beneficial to other consumers can be determined without the approval of other consumers.</p>	
Keywords: Machine Learning, Word of Mouth, Useful Consumer Evaluation, Supervised Learning	

GİRİŞ

Web 2.0'ın ortaya çıkmasıyla birlikte, insanlar ağızdan ağıza iletişim kavramını çevrimiçi ortama taşıyarak yaşadıkları mahalle, şehir gibi sınırları aşarak dünyanın neresinde olduğu fark etmeksizin gelişen sosyal medya platformları ve çevrimiçi paylaşım siteleriyle birlikte bir ürün ya da hizmetle ilgili deneyimlerini birbirleriyle paylaşabilmektedir. Teknolojinin gelişmesi ve yaygınlaşmasıyla birlikte internet üzerinde kendi deneyimlerini paylaşan kullanıcıların sayısı da artmaktadır. Kullanıcıların %66'sı, internet üzerinde şahsen tanımadıkları ancak ürünü daha önce deneyimlediklerine inandıkları kişilerin yorumlarını güvenilir bulduklarını belirtmişlerdir (Nielsen, 2015b). Kullanıcıların bir ürün veya hizmet konusunda bilgi arayışına girdiklerine, araştırma süreçlerine faydalı kullanıcı deneyimlerinin ön plana çıkması için sosyal paylaşım platformları ve çevrimiçi alışveriş siteleri, kullanıcıların faydalı buldukları kullanıcı değerlendirmelerini “faydalı” ya da “yararlı” olarak işaretleyebilecekleri birçok özellik sunmaktadır.

Çalışmanın Amacı: Denetimli sınıflandırma algoritmalarıyla, kullanıcı değerlendirmeleri incelenecek olup, kullanıcıların faydalı buldukları değerlendirmelerin özelliklerinin algoritmalar tarafından öğrenilerek, kullanıcıların yeni hazırlamış oldukları kullanıcı değerlendirmelerinin yayımlandığı anda diğer tüketiciler tarafından faydalı bulunup bulunmayacağı hakkında bir tahminde bulunmasını sağlamak ve algoritmaların faydalı kullanıcı yorumlarını inceleyip öğrenirken, kullanıcı değerlendirmesinin hangi özelliklerine daha fazla ağırlık vererek tahmin sürecinde etkili kullanıcı değerlendirmesi özelliklerinin ortaya koymaktır.

Çalışmanın problemi ve soruları: Bu çalışmanın araştırma problemi “çevrimiçi alışveriş platformlarında paylaşılan yeni bir tüketici yorum ve değerlendirmesinin faydalılık düzeyi, diğer tüketicilerinin algıları olmaksızın makine öğrenme algoritmaları aracılığıyla tahmin edilebilir mi?” olarak belirlenmiştir. Bu soruyu yanıtlamak adına cevap aranacak alt sorular;

Araştırma Sorusu 1 - Faydalı tüketici değerlendirmesi tanımlanabilir mi?

Araştırma Sorusu 2 - Faydalı tüketici değerlendirme tahminimde hangi makine öğrenmesi algoritmasından yararlanılabilir?

Araştırma Sorusu 3 - Faydalı tüketici değerlendirmesi tahminlime farklı ürün kategorilerinde nasıl çalışır?

Çalışmanın Önemi: Çevrimiçi alışveriş platformlarının mevcut değerlendirme filtreleme kriterleri içinde, yeni hazırlanmış tüketici değerlendirmelerinin belirli filtreler seçilmeden kullanıcıların karşısına çıkmaması, yeni bir değerlendirme hazırlamayı planlayan kullanıcıların motivasyonuna olumsuz etkisi olabilir. Kullanıcılara, hazırladıkları değerlendirmenin algoritmalar tarafından faydalı olarak belirlenmesiyle birlikte diğer kullanıcılar öncelikli olarak gösterileceğinin bilinmesi, kullanıcıların yeni bir değerlendirme hazırlama konusunda teşvik edebilir.

Ayrıca, faydalı kullanıcı yorumlarının niteliklerinin, makine öğrenmesi algoritmaları tarafından öğrenilerek, hangi özniteliklerin bir tüketici değerlendirmesinin “faydalı” olarak etiketlenmesinde rol oynadığı ortaya koyularak, bir tüketici değerlendirmesi hazırlanırken, diğer tüketiciler tarafından faydalı bulunması için, değerlendirmenin hangi özelliklere önem verilmesi gerektiği ortaya koyulabilir.

Çalışmanın Yöntemi: Çevrimiçi alışveriş platformları üzerinde bulunan tüketici değerlendirmelerinin, kazıyıcı uygulamalar tarafından alınıp veri ön işleme süreçlerinden geçirerek, bir tüketici değerlendirmesine ait, yıldız derecelendirmesi, duygu etiketi, kelime sayısı ve fotoğraflı bir derecelendirme olup olmamasıyla ilgili elde edilen veri setiyle 6 farklı makine öğrenmesi algoritmasının veri seti üzerindeki başarısının değerlendirmesi, yapılan değerlendirmenin ardından algoritmaların hangi özneliğe daha fazla ağırlık verdiğinin ortaya koyulması ve aynı çevrimiçi alışveriş platformunda, farklı kategorilerdeki ürünlere yapılan diğer tüketici değerlendirmelerinden, diğer tüketiciler tarafından faydalı bulunan kullanıcı yorumlarının, algoritmayı test etmek amacıyla kullanılması. Çalışmanın değerlendirme kapsamında, farklı kategorilerden alınan, diğer tüketiciler tarafından faydalı olarak işaretlenen tüketici değerlendirmeleri üzerinden, eğitilmiş algoritmaların farklı kategorilerde göstereceği faydalı yorum tahminlime başarısı test edilecektir.

Çalışmanın Kısıtları: Veri setinde “faydalı” olarak etiketlenen tüketici değerlendirmelerinin, “faydalı değil” olarak etiketlenen tüketici değerlendirmelerinden sayıca az olmasından ötürü oluşan dengesiz veri seti, öğrenme seti üzerindeki kısıtlardan birisidir. Diğer diller üzerinde yapılan tüketici değerlendirmesi temelli çalışmalarında

kullanılan başlıca özniteliklerden birisi olan “okunabilirlik”, Türkçe dili için güncel ve doğruluğu test edilmiş bir araç olmamasından ötürü öznitelik olarak kullanılamamıştır. Ayrıca, hazırlanan veri setinde kullanılan öznitelikler ve literatürdeki çalışmalarda kullanılan diğer özniteliklerin bu çalışma kapsamında kullanılmaması, inceleme metinlerinin kullanılan kelimelerin anlam ve birliktelik analizi açısından değil, sadece kelime sayısı bakımından değerlendirilmesi, araştırmanın kısıtlarındandır.

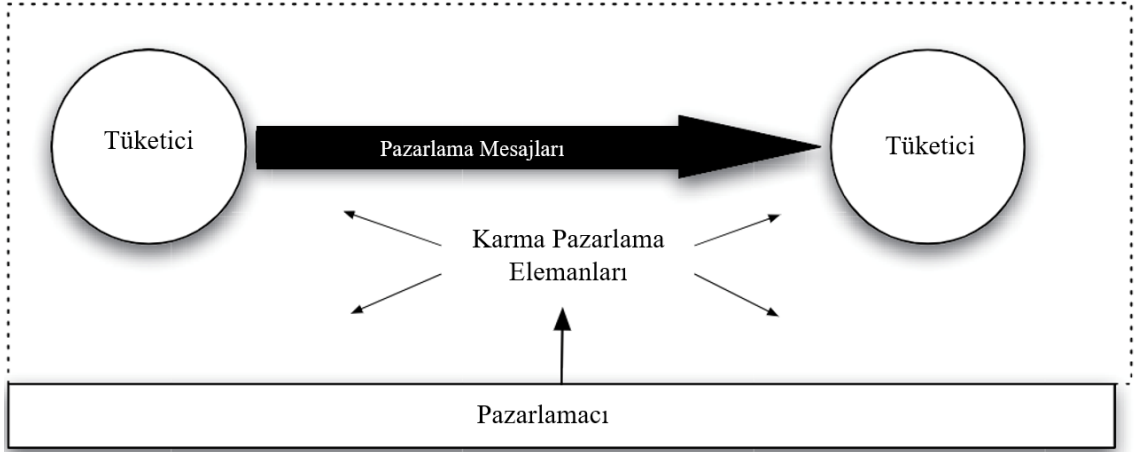
BÖLÜM 1: KAVRAMSAL ÇERÇEVE

1.1. Ağızdan Ağıza İletişim

Ağızdan ağıza iletişim (AAİ), pazarlama ve satışlar üzerindeki etkisi uzun yıllardır araştırma konusu olmuştur. Buttle, (2008) AAİ'nin, 2400 yıldır insanlar arasındaki etkisi düşünülse de ancak 1940'lardaki savaş sonrası dönemden itibaren pazar araştırmalarının konusu olduğunu ifade etmiştir. Genel olarak, AAİ'nin tüketiciden-tüketicie ilişkisi incelenmesine ve çalışmaların da bu şekilde yön verilmesine rağmen işe alma gibi alanlara da etkisi ortaya koyulmuştur (Buttle, 1998). Pozitif bir AAİ'nin deneyimle alakalı; etkilenilmiş veya kazanılmış bir tecrübeyi başkalarına tavsiye etmesi olarak tanımlanabilirken negatif bir AAİ, ürünün/deneyimin kötülenmesi, memnuniyetsiz deneyimlerin paylaşılması, dedikodu ve kişisel problemlerin paylaşılması olarak tanımlanabilir (E. W. Anderson, 1998b). Web 2.0'ın gelişmesiyle birlikte kullanıcıların da etkileşimli olarak deneyimlerini paylaşabilmesi olanak sağlamış ve bu noktadan itibaren internet kişilerarası bilgi alışverişinin en önemli araçlarından birisi olmuştur (Riegner, 2007). AAİ, gelişen teknolojiyle birlikte, marka oluşturma süreçlerine, yeni müşterilerin kazanımı, kazanılmış müşteriye elde tutma ve yeni ürün ve hizmetler geliştirme gibi konularda da büyük öneme sahip olmuştur (Dellarocas, 2003). Bu gelişmelerle birlikte kullanıcıların deneyimlerini birbirleriyle çevrimiçi topluluklar aracılığıyla paylaşabilmekte ve etkileşimde bulunabilmektedir (T. Zhou, 2011). Çevrimiçi topluluklar arasında gerçekleşen veya direkt olarak kullanıcıların çevrimiçi ortamlarda deneyim ve tecrübelerini paylaşmasıyla gerçekleşen ağızdan ağıza iletişimi, elektronik ağızdan ağıza iletişim (e-AAİ) olarak ifade edilir (Lee ve Youn, 2009).

Kozinets vd., (2010b) çevrimiçi toplulukların oluşturduğu e-AAİ'nin tüketiciler üzerindeki etkisini incelemek için yaptıkları çalışmada, topluluk etkileşimlerini üç ana model üzerinde toplamıştır. Bunları, "nihai organik tüketici etkisi modeli", "doğrusal pazarlamacı etki modeli" ve "ortak çevre modeli" olarak ifade edilir.

1.1.1. Nihai Organik Tüketici Etkisi Modeli

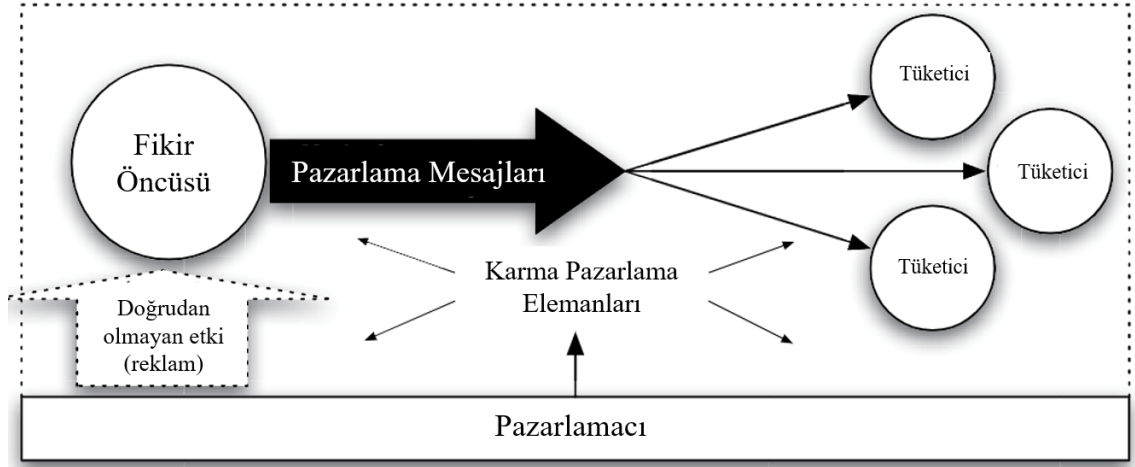


Şekil 1: Nihai Organik Tüketici Etkisi Modeli

Kaynak: Kozinets, R. v., de Valck, K., Wojnicki, A. C., ve Wilner, S. J. S. (2010b). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89. <https://doi.org/10.1509/JMKG.74.2.71>

Kozinets vd. (2010) bu modeli, tüketiciden tüketiciye arada başka hiçbir aktör olmaması sebebiyle “organik” olarak adlandırmaktadır. Bu modelin, diğer modellere göre en büyük farkı, direkt iletişimin olması ve deneyimin diğer aktörler tarafından müdahalesi bulunmamaktadır. Ürünü deneyimleyen bir “tüketici” bu deneyimini direkt olarak diğer “tüketici” ile paylaşmaktadır. Bu paylaşımı gerçekleştirirken, ürünü üreten firmanın veya pazarlamasını yapan diğer kaynakların bu etkileşime doğrudan veya dolaylı olarak müdahalesi bulunmamaktadır. İletişim, doğrudan “tüketiciden” – “tüketiciye” olarak “organik” şekilde gerçekleşmektedir.

1.1.2. Doğrusal Pazarlamacı Etki Modeli

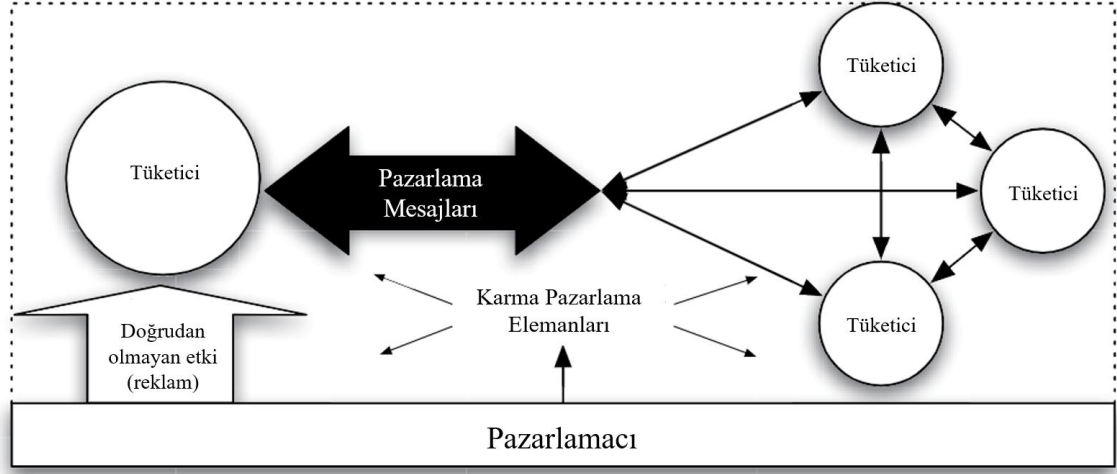


Şekil 2: Doğrusal Pazarlamacı Etki Modeli

Kaynak: Kozinets, R. v., de Valck, K., Wojnicki, A. C., ve Wilner, S. J. S. (2010b). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89. <https://doi.org/10.1509/JMKG.74.2.71>

Toplum mühendisliğinin pazarlamada kullanımıyla birlikte geliştirilen bu modelde, etkileşimin başlatıldığı taraf direkt olarak ürünün tüketicisi değil, pazarlamacılar tarafından çeşitli promosyonlar/reklam anlaşmaları ile etkilenmiş “fikir öncüsü” vardır. Fikir öncüsü, ürün hakkındaki fikirleri potansiyel “tüketiciler” ile paylaşır. Böylece, ürün hakkında dolaylı olarak reklam yapılarak potansiyel tüketicilere ulaşılmaya çalışılması yerine, fikir öncüsü ya da kanaat önderi kabul edilen kişiler aracılığıyla direkt olarak pazarlaması gerçekleştirilir (Kozinets vd., 2010). Günümüzde oldukça yaygın bir şekilde kullanılan bu pazarlama yöntemi, hedef kullanıcı grubunun belirlenmesinden sonra bu grup üstünde etkisi olan kişiler Aİİ aracı olarak kullanılarak ürünün pazarlaması gerçekleştirilir. Nihai organik tüketici etkisi modelinden farklı olarak, bu iletişim tarafların direkt olarak kendileriyle değil, pazarlamacıların etkisi altında gerçekleşir.

1.1.3. Ortak Çevre Modeli



Şekil 3: Ortak Çevre Modeli

Kaynak: Kozinets, R. v., de Valck, K., Wojnicki, A. C., ve Wilner, S. J. S. (2010b). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89. <https://doi.org/10.1509/JMKG.74.2.71>

Ortak çevre modeli, her ne kadar doğrusal pazarlamacı etki modeline benzese de iki temel fark bulunur. Birinci fark, iletişim tek yönlü olarak değil iki taraflı olarak gerçekleşir. Yani “tüketiciden” – “tüketiciye” ileri ve geri yönlü bilgi akışı sağlanmaktadır. İkinci temel fark ise, doğrusal pazarlamacı modelinin aksine bu modelde “tüketici” ürünü/hizmeti kullanarak direkt olarak etkilenen taraf olmaktadır. Edindiği deneyim ve tecrübeleri, diğer tüketicileri ilk ağızdan anlatarak bu konu hakkındaki deneyimin taraflar arasında yayılmasını sağlar. Kozinets vd. (2010), bu ilişkiyi toplumsal (communal) ilişki olarak tanımlar ve pazarlamacıların, pazarlama stratejilerini bu toplumsal ilişki üzerine tasarladıklarını ifade eder. Bu ilişki günümüzde, çevrimiçi alışveriş sitelerinde ürünlerin altında bulunan yorumlarda, kullanıcı deneyimlerinin paylaşıldığı ürün odaklı (product dedicated) forumlarda rahatça gözlemlenebilmektedir. E-AAİ'nin pazarlama stratejisinin parçası olarak görülmesi sadece bu çalışmayla sınırlı değildir. AAİ'yle birlikte pazarlamacıların, e-AAİ'nin pazarlama stratejisi olarak kullanımında nasıl fayda sağlayabileceğini ifade eden çalışmayla birlikte (Kozinets et al., 2010b), e-AAİ'nin pazarlama stratejisi olarak kullanımında yaşanan zorlukların da incelendiği çalışma da mevcuttur (Litvin vd., 2008). Trusov vd., (2009) tarafından yapılan çalışma ile, e-AAİ'nin geleneksel pazarlama yöntemlerine göre daha uzun süreli etkiye sahip olduğunu ortaya konulmuştur. E-AAİ, sadece bir ürünün satın alınması ile sınırlı değildir, ayrıca

aynı ürün türünde ikame ürünlerin satın alınması konusunda da etkisi mevcuttur. E-AAİ'nin ikame ürünler arasındaki satışlara etkisi inceleyen Zhu ve Zhang, (2010) pazarlama stratejilerinde e-AAİ'nin de göz önüne alınması gerektiğini vurgulamıştır.

E-AAİ'nin analiz edilmesi ve kullanılması konusunda yapılan çalışmalar sadece teorik olanlar ile sınırlı değildir. Duan vd., (2008) yapmış oldukları çalışma ile, çevrimiçi kullanıcı yorumlarının filmlerin kâr marjı üzerindeki etkisini incelemiş, çevrimiçi kullanıcı derecelendirmelerinin film satışları üzerinde bir etkisi olup olmadığını bulamamış ancak, kullanıcı yorumları ile film satışları arasında anlamlı pozitif bir ilişki olduğunu ortaya koymuştur. Tüketiciler, %83 oranla birebir tanıdıkları insanların yorumlarını güvenilir bulurken, çevrimiçi ortamlarda yapılan yorumlar için bu oran %66'dır (Nielsen, 2015a). E-AAİ'de paydaşlar sadece tüketiciler değildir. Chen ve Xie, (2008) yaptığı çalışmada, bir ürünün maliyeti/fiyatı yüksek ise bu ürünle ilgili acemi/bilgisiz kullanıcıların daha fazla olduğunu ortaya koymuş, ürünün maliyeti/fiyatı düşük olduğunda ve ürün için herhangi bir uzmanlık gerekmediğinde çevrimiçi alışveriş sitesi tarafından sağlanan bilginin, diğer tüketiciler tarafından sağlanan bilgiler ile tamamlayıcı olarak hareket ettiğini ancak ürünün maliyeti/fiyatı yüksek olduğunda çevrimiçi alışveriş sitesinin sağlamış olduğu bilgiler, genel olarak acemi/bilgisiz tüketiciler için edinilen ürün bilgisini azaldığını vurgulamıştır. AAİ'nin kullanıldığı bir başka pazarlama yöntemi ise müşteri sadakatini yönetimidir. Sadık müşteri, bir ürün satın alacağı zaman öncelikli olarak belirli bir markaya ilgi duyan, o marka konusunda pozitif önyargıları olan ve o markadan en az iki kez alışveriş yapmış müşteriler olarak tanımlanır (Griffin, 1995).

E-AAİ'nin müşteri sadakatini sağlamak, müşteri sadakatini arttırmak ve müşteri sadakatini bir pazarlama yöntemi olarak kullanımıyla ilgili çalışmalar mevcuttur. İşletmelerden – Müşteriye (B2C) ilişkilerde müşteri sadakatini yönetimi için 8 faktör belirleyen Srinivasan vd., (2002) müşteri sadakatini yönetiminin iki büyük getirisi olduğunu vurgulamışlardır. Bu getiriler gelirlerdeki artış ve AAİ tanıtımıdır. Bir markaya sadık olan müşteriler, o markanın ürünlerine diğer markaların ürünlerine nazaran öncelik tanımakta ve onları tercih etmektedir. Ayrıca bu müşteriler, ürünler konusundaki deneyimlerini ve tecrübelerini diğer insanlarla paylaşmakta ve markanın/ürünün, markaya herhangi bir maliyeti olmadan markanın/ürünün tanıtımını yapmaktadır. Sadık müşterilerin AAİ yoluyla marka/ürün tanıtımı yapmaları için en büyük etkenlerden birisi

ürün/hizmet tatminidir (Hennig-Thurau vd., 2002). Üründen/hizmetten memnun kalan müşteriler deneyimlerini çevresiyle veya çevrimiçi platformlar üzerinden diğer potansiyel tüketiciler ile paylaşırlar.

1.2. Kullanıcı Yorumları ve Yıldız Derecelendirme

Kullanıcı yorumları, kişilerin kendi deneyimleriyle birlikte oluşturmuş oldukları ürün değerlendirmeleridir ve günümüzde en yaygın e-AAİ türüdür (Pelsmacker vd., 2018). Kullanıcıların almış oldukları ürün veya hizmet konusunda, ürün veya hizmeti satın alma sürecinde, satın aldıktan sonra ve deneyimlerken yaşadıkları olumlu ya da olumsuz durumları diğer kullanıcılarla çevrimiçi ortamlarda paylaşımları ile kullanıcı yorumları ortaya çıkmaktadır. Kullanıcı yorumları, ayrıca kullanıcı tarafından oluşturulmuş içerik (user generated content) olarak da bilinmektedir (Daradkeh, 2021).

Potansiyel müşteriler, bir ürün veya hizmeti almayı düşünen kişiler, satın almaya karar verme sürecinde birçok kaynaktan araştırma yapmaktadır. Bu araştırma sürecinde karşlarına diğer kullanıcıların oluşturmuş olduğu binlerce yorumlar çıkmaktadır. Bu yorumlar arasında kullanıcıların kendi deneyimleri, uzman kullanıcıların fikirleri ve otomatik olarak oluşturulmuş ürün bilgileri de yer almaktadır. Her bir yorum, potansiyel müşteriye değer katma potansiyelindedir (Mudambi ve Schuff, 2010).

Mudambi ve Schuff, (2010) çevrimiçi müşteri incelemelerini şirket veya üçüncü taraf web sitelerinde yayınlanan, akran tarafından oluşturulan ürün değerlendirmeleri olarak tanımlamışlardır. İnternetin yaygınlaşması ve tüketicilerin düşüncelerini paylaşabileceği farklı platformların ortaya çıkmasıyla birlikte faydalı bilgiye ulaşma süreci de zorlu bir hâl almıştır. Platformlar, kullanıcılarına kullanıcı yorumlarını filtreleme ve gruplandırma için çeşitli araçlar sunmaktadır. Bu araçlardan birisi kullanıcıların yazdıkları yorumlarla birlikte ürüne/hizmete bir derece verebilmesidir. Yıldız derecelendirmesi olarak da karşımıza çıkan bu kavramda, 1 yıldız tüketicilerin aldıkları ürün/hizmet açısından memnuniyetsizliklerini ve tüketicinin olumsuz fikrini yansıtırken, 5 yıldız ise tüketicilerin aldıkları ürün/hizmet açısından yaşadıkları memnuniyeti ve tüketicinin olumlu fikirlerini yansıtmaktadır (Krestel ve Dokoohaki, 2011). Tüketiciler tarafından oluşturulmuş içeriklere verilen yıldız derecelendirmesinin, içeriğin okuyuculara sağlayacağı faydaya olan ilişkisi konusunda birçok çalışma vardır. Park vd., (2007) yapmış olduğu çalışma ile çevrimiçi tüketici incelemelerinin kalitesinin tüketici satın

alma eğilimleri üzerinde pozitif etkisi olduğunu, bağlılığı düşük olan tüketicilerin, tüketici incelemelerin kalitesine nazaran sayısı ile ilgilendiğini, tam tersi olarak yüksek bağlılığı bulunan tüketicilerin ise ancak tüketici inceleme kalitelerinin kalitesi yüksek olduğunda tüketici incelemelerinin sayısı ile ilgilendiğini ve son olarak inceleme sayısı ile birlikte satın alma eğiliminin de arttığını ortaya koymuştur. Anderson, (1998a) tüketicilerin yaşamış olduğu deneyimleri ifade etme konusunda olumsuz deneyimlerin, olumlu deneyimlere göre daha çok kelime ve yorum ile ifade edildiğini ortaya koyarken, Leskovec vd., (2007) ortalama yorumların (1-5 yıldız derecelendirmesi ölçütünde 3 yıldız derecelendirmesi alan yorumlar), uç nokta olarak ifade edilen 1 ve 5 yıldız derecelendirmeli yorumlardan daha az faydalı olduğunu belirtmiştir. Yorumların sınıflandırılması ürün/hizmet kalitesinin ölçülmesi için yıldız derecelendirmesi ve yapılan incelemelerin ortalama yıldız derecelendirmesi bir ölçüt olarak kullanılmasına karşın, Çakar ve Akbıyık, (2018) bu tip derecelendirmelerin, üründen/hizmetten bağımsız olarak, kargo firmasının hizmet kalitesi ve ürünün güncel fiyatında bulunan artışlar/azalışlar gibi konulardan da etkilenmesi sebebiyle, ürünün genel kalitesinin değerlendirilmesinde ölçüt olarak tek başına yeterli olmayacağını vurgulamıştır.

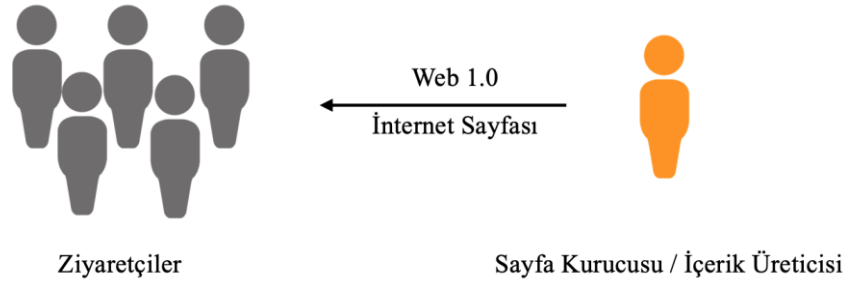
1.3. Web 2.0

Web 1.0 yada WWW, 1990'lu yıllarda CERN'de bir grup araştırmacı tarafından geliştirilmiş, 1994 yılının Eylül ayında gerçekleşen WWW Konsorsiyumu ile Amerika Birleşik Devletleri'nden Massachusetts Teknoloji Enstitüsü, Fransa'dan Ulusal Dijital Bilim ve Teknoloji Araştırma Enstitüsü ve Japonya'dan Keio Üniversitesi'nden gelen bilim insanlarıyla WWW için protokoller tartışılmış, bu protokollerin internete erişmek isteyen bütün cihazlar için standart olması üzerinde anlaşılmıştır (Berners-Lee, 1998).

Web 1.0, statik (sabit/değişken olmayan) internet sayfaları oluşturulabilmesine, sayfa yöneticilerinin oluşturmuş oldukları içeriklerin ağ üzerinden herhangi bir yerden erişilebilir olmasını sağlamıştır. Web 1.0, salt okunur (read-only), sadece sayfa yöneticilerinin kontrolüyle içerikleri değişebilen, ziyaretçilerin veya okuyucuların herhangi bir katkıda bulunmadığı internet sayfalarıdır (Choudhury, 2014). Web 1.0 ile oluşturulmuş internet sayfaları, günümüz internet sayfalarına göre kıyaslandığında basit bir tasarıma sahip, metin içeren sayfalardır.

Web 1.0'ın karakteristik özellikleri;

- Salt okunur sayfalara sahip olması,
- Statik sayfaların basit Hiper-Metin İşleme Dili (HTML) oluşturulmuş olması,
- Bir içeriğin çevrimiçi ortama yerleştirilebilmesine ve internet bağlantısı olan herhangi bir yerden içeriğe erişilebilmesi imkânı sunmasıdır (Choudhury, 2014).

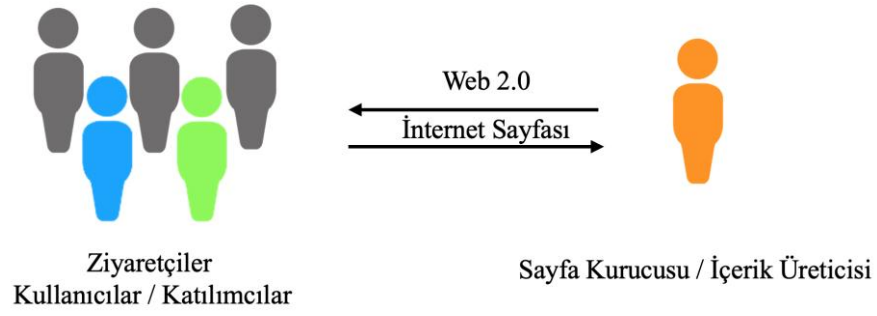


Şekil 4: Web 1.0 Kullanıcı Etkileşimi

Kaynak: Yazar.

Şekil 4, Web 1.0’ın paydaşları olarak ziyaretçiler ve internet sayfasını hazırlayan içerik üreticileri arasındaki ilişkiyi göstermektedir. Web 1.0 ile internete erişen, hazırlanmış olan internet sayfalarını kullanan kişiler, sadece internet sayfalarını ziyaret edebildikleri için “Ziyaretçi” olarak adlandırılmıştır. Ziyaretçiler, Web 1.0 teknolojisinin limitlerinden ötürü, internet sayfası üzerinde herhangi bir düzenleme, ekleme veya çıkarma işlemi gerçekleştiremez. Sayfalar statiktir, bu sayfalarda sadece ziyaretçi olarak bulunurlar.

Web 2.0’ın ortaya çıkmasıyla birlikte, salt okunur internet sayfalarından, okuma-yazma (read-write) internet sayfalarına dönüşümün önü açılmıştır. Web 2.0’la birlikte sadece sayfa sahipleri değil sayfayı ziyaret eden tüketiciler de internet sayfalarının içeriğine katkıda bulunabilme imkânı sahip olmuştur.



Şekil 5: Web 2.0 Kullanıcı Etkileşimi

Kaynak: Yazar.

Şekil 5, Web 2.0’da paydaşları olarak ziyaretçiler, kullanıcılar, katılımcılar ve internet sayfasını hazırlayan içerik üreticileri arasındaki ilişkiyi göstermektedir. Web 1.0’da sadece ziyaretçi olarak, statik internet sayfalarına erişen kullanıcılar, Web 2.0 ile birlikte kullanıcı ve katılımcılar olarak da rol almaya başlamışlardır. Kullanıcı katkılarıyla oluşan forumlar, forumu oluşturanlar tarafından herhangi bir içerik eklenmeden, tamamen kullanıcıların içeriklere yaklaşımları ve katılımcıların içerik üretmesiyle birlikte internet sayfaları dinamik bir hâl almış, sayfaların içerikleri diğer paydaşlarla birlikte şekillendirilmeye başlanmıştır.

Bu gelişme, sosyal ağların gelişiminin önün de açmıştır. Bir sosyal ağ, kullanıcılarına kolayca kendisine ait profil oluşturmaya, oluşturduğu profili diğer sosyal ağ kullanıcılarıyla paylaşabilmesine ve bu profilde metin olarak kendi düşüncelerini paylaşmasının yanı sıra fotoğraf, video ve müzik paylaşabileceklerine, kullanıcıların ilgi alanlarına göre topluluklar bulabilmelerine imkân veren çevrimiçi internet sayfalarıdır (Barsky ve Purdon, 2006).

	Statik İnternet Sayfaları Oluşturabilme	Dinamik İnternet Sayfaları Oluşturabilme	Metin Paylaşabilme	Medya (Fotoğraf, Video, Müzik) Paylaşabilme	Kullanıcıların İçerik Ekleyebilmesi
Web 1.0	✓	×	✓	×	×
Web 2.0	✓	✓	✓	✓	✓

Şekil 6: Web 1.0 ve 2.0 Karşılaştırması

Kaynak: Yazar.

Web 2.0'la birlikte gelişen sosyal ağlar, kullanıcıların ağızdan ağıza iletişim alışkanlıklarını çevrimiçi ortama taşımıştır. Ağızdan ağıza iletişimin, elektronik ağızdan ağıza iletişime dönüşümünde temel değişiklik gerçekleşmiştir; bilgi, yerel sınırları aşır internetin olduğu her yerde erişilebilir hale gelmiştir, geleneksel yollarda ancak kişi ve çevresiyle sınırlı olan bu etkileşim elektronik ağızdan ağıza iletişim ile aynı zevkleri paylaşan, aynı sorunlara çözüm arayan, aynı ürünle ilgilenen kişileri herhangi bir sınırlandırma olmadan fikir alışverişinde bulunmasına olanak sağlamaktadır (Sahelices-Pinto vd., 2014).

1.4. Faydalı Tüketici Değerlendirmesi

Bir ürün veya hizmeti kullanan, kullandıkları bu hizmet sonrasındaki deneyimlerini ürün veya hizmeti edindikleri çevrimiçi alışveriş sitesinde veya sosyal medya platformları aracılığıyla kişisel deneyimlerini paylaşmasıyla birlikte tüketici değerlendirmeleri ortaya çıkmaktadır. Bir tüketici, yeni bir ürün satın almak ya da hizmet deneyimlemek istediğinde, fiyat, uzun süreli kullanım tecrübesi, kargo seçenekleri, satış sonrası destek ve benzeri konularda ürün ya da hizmeti sağlayan firmanın hazırlamış olduğu bilgilendirme metinlerinden daha çok, ürün ya da hizmeti deneyimlemiş kullanıcıların hazırlamış oldukları tüketici değerlendirmelerini incelemeyi tercih etmektedir (Hennig-Thurau ve Walsh, 2014). Günümüzde çevrimiçi alışveriş siteleri, tüketici değerlendirmelerin ürünlerin yanında yayımlanmasına olanak sağlayan tasarımları kullanmaktadır. Tüketiciler inceledikleri ürün hakkındaki tüketici değerlendirmelerine de ürünlerin sayfalarında ulaşabilmekte, alışveriş sitesinin sağlamış olduğu filtreleme kriterlerine göre tüketici değerlendirmelerini filtreleyebilmektedir.

Müşteri Yorumları

★★★★★ 4,9 / 5 yıldız

30 genel puan

5 Yıldız ██████████ %95

4 yıldız ██████████ %5

3 yıldız ██████████ %0

2 yıldız ██████████ %0

1 yıldız ██████████ %0

▼ Puanlar nasıl hesaplanır?

En çok beğenilen müşteri yorumları

En yeni

Bu ürünü incele

Düşüncelerinizi diğer müşteriler ile paylaşın

Müşteri yorumu yazın

Şekil 7: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri

Kaynak: Amazon.com Tüketici Değerlendirmeleri Filtreleme Ekranı.

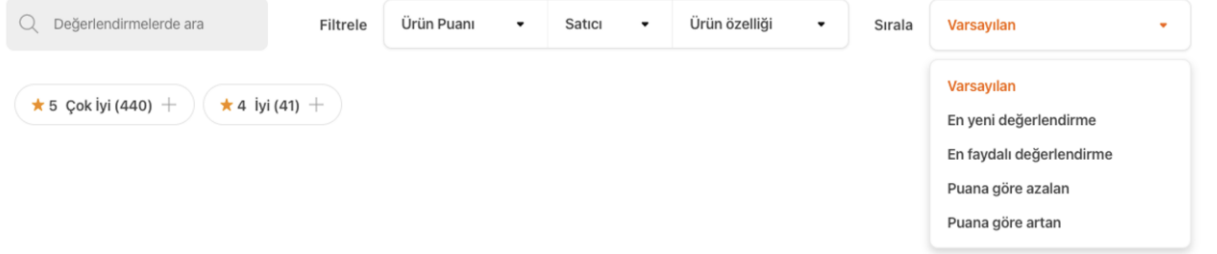
Bazı çevrimiçi alışveriş siteleri, yıldız derecelendirmesi dağılımları, genel puan (tüketici değerlendirmelerinin ortalama yıldız derecelendirmesi) gibi bilgileri sağlayıp, ayrıca kullanıcılara yorumları istedikleri şekilde filtreleme için de filtreleme kriterleri sunmaktadırlar.



Şekil 8: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri

Kaynak: Trendyol.com Tüketici Değerlendirmeleri Filtreleme Ekranı.

Tüketici değerlendirmelerinin filtrelenebilmesi için “yeniden eskiye” ve “eskiden yeniye” gibi filtreler uygulanmasına karşın, “en yeni değerlendirme”, “en faydalı değerlendirme” ve “en faydalı değerlendirme” kriterlere göre de filtrelenebilir.



Şekil 9: Çevrimiçi Tüketici Değerlendirmesi Filtrelemeleri

Kaynak: Hepsiburada.com Tüketici Değerlendirmeleri Filtreleme Ekranı.

Tüketici değerlendirmelerinin, diğer tüketiciler tarafından “Bu değerlendirme faydalı mı?” şeklinde sorulan sorularla tüketici değerlendirmesinin, diğer tüketiciler tarafından “faydalı” veya “faydalı değil” olarak işaretlenmesine olanak sağlamaktadır.



Şekil 10: Değerlendirme Faydalı Etiketleme

Kaynak: Hepsiburada.com Tüketici Değerlendirme Ekranı.

Bazı çevrimiçi alışveriş platformlarında ise, kullanıcıların sadece “Faydalı” olarak işaretlenmesine imkân verilmektedir. Tüketiciler faydalı buldukları değerlendirmeleri “Faydalı” olarak işaretleyebilmelerine olanak sağlanmaktadır. Tüketiciler tarafından faydalı bulunan inceleme, “... kişi bunu faydalı buldu” şeklinde etiketlenmektedir.



18 kişi bunu faydalı buldu

Faydalı

Şekil 11: Değerlendirme Faydalı Etiketleme

Kaynak: Amazon.com Tüketici Değerlendirme Ekranı.

Bir tüketici değerlendirmesinin, diğer tüketiciler tarafından faydalı bulunmasını belirleyen birden fazla kriter vardır. Tüketicilerin değerlendirmelerini uzun tutarak, birden fazla konudan bahsettiklerinde, söz konusu değerlendirmenin diğer tüketiciler tarafından faydalı bulunma ihtimalinin azaldığının ortaya koyulduğu çalışmalar olduğu gibi (Z. Liu ve Park, 2015), inceleme metnin uzunluğu, inceleme metnini hazırlayan kullanıcının belirlemiş olduğu yıldız derecelendirmesinin ve diğer kullanıcıların o değerlendirmeyi faydalı bulmasının, bir başka kullanıcının değerlendirmeyi faydalı bulmasında etkili olduğunu ortaya koyan çalışmalar da (Pelsmacker vd., 2018) bulunmaktadır. Tüketici değerlendirmelerinin, diğer tüketiciler tarafından faydalı bulunmasına etkili birden fazla kriter vardır.

Ürünü almak için araştırma yapan kullanıcıların %83'ü şahsen tanıdıkları kişilerin, ürünler hakkında yaptıkları değerlendirmelere güvendiklerini belirtmesine karşın (AAİ), %66 ile internet ortamında yapılan tüketici değerlendirmelere güvendiklerini belirtmişlerdir (e-AAİ) (Nielsen, 2015b). Faydalı tüketici değerlendirilmeleri, ürün satın alma sürecinde bulunan tüketiciler için önem arz etmektedir.

1.5. Önceki Çalışmalar

Tüketici/kullanıcı değerlendirmeleri üzerinde birçok çalışma yapılmıştır. Yapılmış çalışmaların amaçları, kullandıkları veri seti, kullanılan analiz/makine öğrenmesi yöntemi ve çalışmanın yapılış amacı Tablo 1'de verilmiştir.

Tablo 1: Önceki Çalışmalar

Referans	Veri Kaynağı	Kullanılan Veriler	Amaç	Model	Sonuç
Turney, 2002	Epinions	Kullanıcı Değerlendirmeleri, “faydalı”-“faydalı değil” etiketleri, duygu skorları	Tüketici değerlendirmelerini içerdikleri anlamsal yönelimlere “faydalı” ya da “faydalı değil” olarak sınıflandırmak.	Denetimsiz Öğrenim	Algoritma tüm veriler üzerinde %74, araba verileri üzerinde %84, film verileri üzerinde %66 başarıya ulaşmıştır.
Alsharif, 2022	Yelp	Yıldız Derecelendirmesi, Doğrulanmış Satın Alma Bilgisi, Değerlendirme Başlığı, Değerlendirici Numarası, Ürün Numarası, Şehir Adı ve Değerlendirme Metni	Sahte Tüketici Değerlendirmelerinin Yinelemeli Sinir Ağları aracılığıyla tespit edilmesi.	Yinelemeli Sinir Ağları	LSTM, Sahte Tüketici Değerlendirmelerini %98 başarı oranıyla tespit etmiştir.
Hmoud Al-Adhaileh ve Wasellah Alsaade, 2022	Amazon Mechanical Truckers	Özgünlük, Analitik Düşünme, Yıldız Derecelendirmesi, Kelime Sayısı, Kişi Zamirleri, Olumsuz Kelimeler, Ret Kelimeleri, Olumlu Kelimeler, Olumsuz Kelimeler	Tüketicileri yanıltma amacıyla yazılmış değerlendirmelerin tespiti ve tüketicilerin satın alma süreçlerine katkısının ortaya koyulması.	Yinelemeli Sinir Ağları	Yinelemeli Sinir Ağları ile Sahte Tüketici Değerlendirmelerini ve Faydalı Tüketici Değerlendirmeleri %84 başarı oranıyla ayrıştırılmıştır.

Kumar vd., 2022	Amazon Mechanical Truckers	Değerlendirme sayısı, Spam Değerlendirme Skoru, Yıldız Derecelendirme Sapması, Değerlendirme Metni Uzunluğu, Kullanıcı Üyelik Süresi, Değerlendirme Yazılma Zamanı, Kelime Yoğunluğu, Duygu Skoru, Pozitif Kelimelerin Oranı	Sahte Tüketici değerlendirmelerinin tespiti üzerinden “tüketici merkezli” ve “değerlendirme merkezli” özniteliklerin belirlenmesi ve veri ön işlemenin, tespit modellerindeki öneminin araştırılması.	XGBoost – LSTM – ANN – Rassal Ağaçlar	XGBoost %88 F1 skoru ile en yüksek tahmin başarısını gösterirken, diğer algoritmalar %79-%75 arasında değişen başarı göstermiştir. Model başarısını arttırmak için Çoğunluk Oyu kuralı uygulanarak model F1 skoru %90’a ulaşmıştır.
Pelsmacker vd., 2018	Çevrimiçi Alışveriş Sitesi	Metin Değeri, Yıldız Derecelendirmesi, Faydalı İşaretlenmesi, İlişki Değeri	Metin Değeri, Yıldız derecelendirmesi ve faydalı olarak işaretlenmiş olmasının, okuyucu ve E-AAİ üzerindeki etkisinin araştırması.	-	İnceleme metninin uzunluğu, söz konusu ürüne ait tüketici değerlendirmeleri için çok önemli olduğu, ancak inceleme metni içeren değerlendirmeler için Yıldız Derecelendirmesinin etkisi oldukça azdır. Tüketici değerlendirme metni içeren değerlendirmelerin, Yıldız derecelendirmelerinin göz ardı edilebileceği önerilmektedir.

Li, 2021	Çevrimiçi Turizm Sitesi	Değerlendirme Metinlerinin Kelime Vektörleri	Tüketici Değerlendirmelerinin birden fazla model ile işlenerek faydalı tüketici değerlendirmelerinin araştırılması.	Multimodel Öznitelik Birleştirme	Hem metin hem görsel içeren değerlendirmelerinin, sadece metin veya sadece görsel içeren değerlendirmelere göre faydalılığı daha kolay bir şekilde tahmin edilebileceği ortaya koyulmuştur.
Lai vd., 2021	Çevrimiçi Turizm Sitesi	Otel Yıldız Derecelendirmesi, Müşteri Duygu Analizi, Değerlendirici Uzmanlığı, Değerlendirme Metni, Görsel Detay İçeren Değerlendirme	Büyük Veri analizi ile Duygu Analizi aracılığıyla Tüketici Değerlendirmelerinin müşteri duyarlılığı ile çevrimiçi otel puanları arasındaki ilişkiyi e-AAİ bağlamında müşteri güdülerini açısından araştırmayı ve inceleme özelliklerinin düzenleyici etkilerini daha fazla tanımlamaya çalışılması.	Baidu AI Duygu Analizi, LTP	Sonuçlar, müşteri duyarlılığı ile çevrimiçi otel puanları arasındaki ilişkinin asimetrik olduğunu ve olumsuz bir duyarlılık puanının, olumlu duyarlılık puanına kıyasla çevrimiçi otel puanlarında daha büyük bir düşüşe neden olacağını göstermektedir. Bu arada, yorumcu düzeyi ve görsel incelemeler, müşteri duyarlılığı ile çevrimiçi otel puanları arasındaki ilişki üzerinde düzenleyici etkilere sahiptir. Ayrıca, farklı duygu analizi yöntemleriyle çıkan iki farklı duygu puanı türü, bu çalışmanın sonuçlarını doğrulamaktadır.

Krestel ve Dokoohaki, 2015	Çevrimiçi Alışveriş Sitesi	Tüketici Değerlendirmesi, Yıldız Derecelendirmesi, Değerlendirme Metni Uzunluğu	Verilerin elde edildiği alışveriş platformu kendi içerisinde bir sıralama (ranking) sistemine sahiptir. Değerlendirmeleri kapsam ya da duygu skorlarına göre optimize ederek, tüketiciler için ürün incelemesi açısından daha optimize bir sıralama sistemi önermek.	LDA/LM	Modelin bulguları, söz konusu alışveriş sitesinin içinde bulunan sisteme göre %30 daha iyi bir performans sergilemiştir.
----------------------------	----------------------------	---	--	--------	--

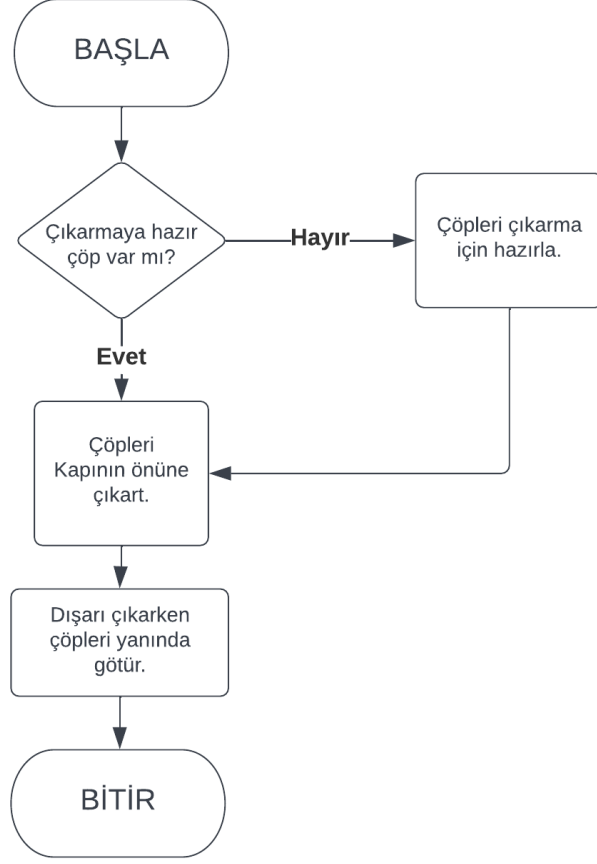
Önceki çalışmalar incelediğinde, çalışmaların değerlendirme yazıldıktan sonraki süreçleri ele aldıkları görülmektedir. Genel olarak çalışmalar, yazılmış olan bir tüketici değerlendirmesinin, “sahte mi yoksa gerçek mi?” ya da “faydalı olması üzerindeki etkenler nelerdir?” şeklinde araştırma sorusu sormuş ve bu sorulara cevap aranmıştır. Yapılan çalışmalarda, istenilen şartları sağlayan tüketici değerlendirmelerin tespiti için önceden elde edilmiş tüketici değerlendirmelerinin bir makine öğrenme algoritması tarafından öğrenilmesine odaklı çalışılmıştır. Makine öğrenmesi algoritmaları denetimli ve denetimsiz öğrenme algoritma olmasının yanında, derin öğrenme ve yinelemeli sinir ağları da kullanılmıştır.

Bu çalışmada, bir çevrimiçi alışveriş sitesinde yazılmış olan tüketici değerlendirmelerinin, denetimli makine öğrenmesi algoritmalarıyla öğrenilerek tüketiciler tarafından hazırlanan bir değerlendirmenin, diğer tüketiciler tarafından faydalı bulunup bulunmayacağına değerlendirme yazıldığı an tahmin edilmesi, böylece faydalı olarak tahmin edilen tüketici değerlendirmesinin direkt olarak değerlendirme okuyan tüketicilerin karşısına yeni hazırlanmış ve algoritmalar tarafından faydalı olarak tahmin edilmiş tüketici değerlendirmesini sunularak 1) değerlendirme yazan tüketicilerin, yazdıkları tüketici değerlendirmesinin algoritma tarafından takip edilen öznitelikler sınırları içerisinde faydalı olarak tahmin edilerek yorum görünürlüğü arttırılacak bilgisinin değerlendirme yazacak tüketicilere verilerek tüketicilerin motivasyonunun arttırılması, 2) bir ürün yada hizmet hakkında bilgi arayan tüketicilerin karşısına yeni tarihli tüketici değerlendirmeleri çıkartarak satın alma süreçlerine güncel değerlendirmeler ile fayda sağlaması amaçlanmaktadır.

BÖLÜM 2: VERİ, ALGORİTMA VE MAKİNE ÖĞRENİMİ

2.1. Algoritma

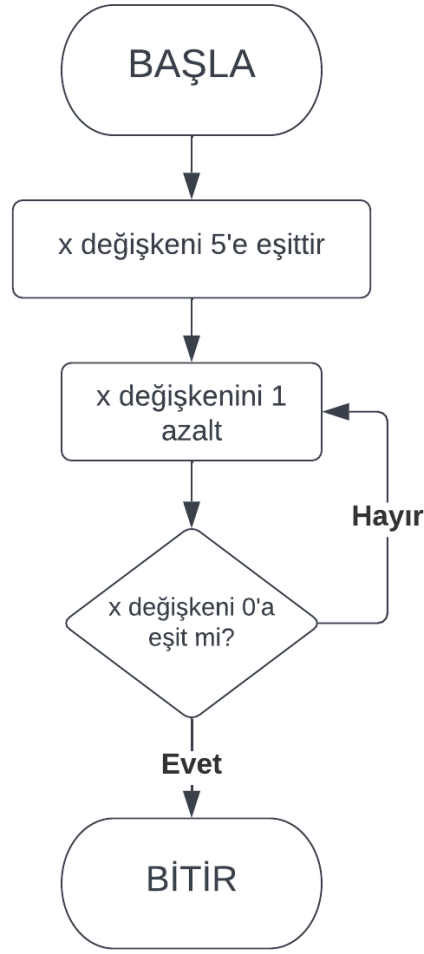
Algoritma, belirli bir problem kümesini çözmek için tasarlanmış, sonlu, bilgisayara uygulanabilir emir dizileridir (Hill, 2016; Moschovakis, 2001). Bu çalışma için algoritmalar, aynı seti üzerinde farklı veri işleyiş yaklaşımlarının değerlendirilmesi için büyük önem arz etmektedir. Bir algoritma, birden fazla değişken üzerinde işlem yapıp bir değişken üretebileceği gibi, bir değişken üzerinde gerçekleştireceği işlemler ile birden fazla değişken de üretebilir. Algoritmanın adımları belli ve sonludur. Nereden başlayacağı, hangi durum karşısında ne yapacağı, başlangıcı ve sonu bellidir. Algoritmalar matematiksel olarak gösterilebileceği gibi, programlama dilleri aracılığıyla kodlarla da ifade edilebilmektedir. Algoritmalar doğrudan matematiksel hesaplamalar gerçekleştirebileceği gibi, yapısal olmayan veriler (metinler, görseller) üzerinde de çalışabilir. Bilgisayar tabanlı algoritmalarda, girdinin ve çıktının türü ne olursa gerçekleştirdiği tüm işlemler matematiksel hesaplamalara dayanır. Soyut Makine (Abstract Machine) ve Özyineleme (Recursion), algoritmanın en temel iki kavramlarından birisidir. Soyut Makine, girdilerin ve çıktıların tanımlanmasından sonra, çıktı üretebilmek için girdilerin ne şekilde işleneceğini ya da hangi işlemlerden geçmesi gerektiğini ifade eden dizilimler bütünüyle oluşmaktadır. Özyineleme, algoritmanın belirli işlemleri gerçekleştirdiği süreçte kendi kendini yinelemesidir. Akış diyagramı, başlangıcı ve bitişi olan, belirli bir işlemi gerçekleştiren algoritmaların grafiksel olarak görselleştirilmesine imkân sağlar (Chapin, 2003).



Şekil 12: Soyut Makine Akış Diyagramı

Kaynak: Yazar.

Şekil 12, bir soyut makinenin akış diyagramını ifade etmektedir. Algoritmanın bir başlangıç ve bitişi vardır. Belirli bir durumlarda karşılaştığında, o durumlarla ilgili yapılacaklar ve karar mekanizmaları mevcuttur. Özyineleme, standart bir algoritma olmasına karşın, yapılan işlemi gerçekleştirirken algoritmanın kendi kendisini tekrar etmesidir. Özyineleme sırasında algoritmanın tasarımında, algoritmanın özyineleme aşamalarını gerçekleştirirken, algoritmanın ulaşması gereken hedefe doğru ilerlediğini bir tasarım olmalıdır. Örneğin, algoritmanın başlangıç sayısı 5, bitiş değeri 10 iken, algoritma söz konusu sayıdan her seferinde 1 azaltıyorsa, hiçbir zaman 10'a ulaşamayacağı için, algoritma sonsuz döngüye girebilir.



Şekil 13: Özyineleme Algoritması Akış Diyagramı

Kaynak: Yazar.

Şekil 13, bir özyineleme algoritmasının akış şemasını göstermektedir. Algoritma, tanımlanan x değerini 1 azaltmaktadır. Algoritmanın içerisindeki koşul, x değerinin sıfıra eşit olup olmadığını kontrol etmektedir. Eğer x değişkeni sıfıra eşit değilse, yapmış olduğu işlemi, x değişkeni sıfıra eşit olana kadar tekrar etmesi beklenmektedir. Peşi sıra yazılabilecek daha uzun bir akış diyagramı ile aynı sonuca ulaşılabılırken, özyineleme sayesinde aynı işlem istenilen koşul sağlanana kadar tekrarlanmaktadır.

2.2. Veri

2.2.1. Veri Değişken türleri

Veri değişken türleri, verilerin sınıflandırılması ve işlenmesi konusunda önem taşımaktadır. Verilerin türlerinden bağlı olarak, bazı değerler, bazı türler için anlamsız olmasına karşın bazı türler için bir değer ifade etmektedir. Bu sebepten ötürü makine öğrenmeleri algoritmaları için verinin ne olduğu, nasıl saklanacağı ve nasıl işleneceğinin belirlenmesi önem taşımaktadır.

2.2.1.1 Kategorik değişkenler

Kategorik değişkenler, sayısal olarak ifade edilmeyen belirli grupları aralında ilişki olacak veya olmayacak şekilde ifade eden değişkenlerdir (Jones vd., 2014). Aralarında bağlantı ve ilişki bulunmayan, sadece kategorik ifade etmek için kullanılan kategorik değişkenlere, nominal kategorik değişkenler denir. Nominal kategorik değişkenler sadece o özelliğe sahip olanların hangi kategoriye ait olduğunu ifade eder. Kategoriler arasında bir azlık – çokluk, altlık – üstlük durumu bulunmaz. Göz rengi, medeni durum, evcil hayvana sahip olma, cinsiyet ve benzeri değişkenler kategorik değişkene birer örnektir. Göz renginin yeşil ya da kahverengi olması, kişiler arasında herhangi bir sıralama yapılmasına olanak sağlamamaktadır (Gürbüz ve Şahin, 2015). Aralarında bağlantı ve ilişki bulunan, sıralama yapılabilen kategorik değişkenlere ise ordinal kategorik değişkenler denir. Ordinal kategorik değişkenler, numerik değişken olmamalarına rağmen aralarında bir sıralama ve anlam vardır. Ordinal değişkenlere örnek olarak, son mezun olunan eğitim kurumu, maaşların belirli aralıklarla ifade edilmesi, şehirlerin popülasyonlarının veya yüz ölçümlerinin “büyük”, “orta” ve “küçük” gibi ifade edilmesi ordinal değişkenlere birer örnektir. Ordinal değişkenler arasında sıralama yapılabilir, en düşük, en az, en fazla, en yüksek gibi kavramları ifade edecek kategoriler bulunabilir (J. A. Anderson, 1984; Gürbüz ve Şahin, 2015).

2.2.1.2. Nümerik Değişkenler

Nümerik değişkenler, Oran (Ratio) Ölçeği ve Aralık (Interval) Ölçeği olmak üzere iki ana gruba ayrılır. Oran ölçeği, bireylerin aylık geliri, kilosu, boyları ve nesnelerin, uzunluğu, genişliği ve hızı gibi değişkenleri ifade etmek için kullanılır. Oran ölçeği değişkenlerinde, 0 (sıfır) anlamsızdır (Gürbüz ve Şahin, 2015). Sıfır ancak başlangıç noktası olarak

kullanılır ve 0 değeri alan değişkenler yokluk bildirir (Harker ve Vargas, 1987). Aralık Ölçeği değişkenlerinde herhangi bir başlangıç noktası bulunmamaktadır. Sıcaklık gibi 0'ın (sıfır) anlamlı olduğu değerler aralık ölçeği olarak geçer (Gürbüz ve Şahin, 2015).

2.2.2. Veri Ön İşleme

Veriler bulunduğu sisteme girilirken, o sistemden alınırken ya da sistemin veriyi yansıtması konusunda hatalar meydana gelebilir. Hatalar meydana gelme dahi, verilerde eksiklik olabilir. Veriler direkt olarak bir algoritma tarafından işlem görülemeyecek durumda olabilir ya da verilerin direkt olarak kendisi değil ancak bu verilerden elde edilecek ikincil verilerle işlem yapılacak olabilir. Veri ön işleme, hatalı, eksik ya da yanlış veriden kurtulmak için gerçekleştirilebileceği gibi, verilerin algoritmalar için uygun hale getirilmesi amacıyla da veri ön işleme yapılabilir.

2.2.2.1. Veri Temizleme

Veri temizleme, verinin kalitesini arttırmak, veri seti içerisindeki düzensizliği gidermek ve tutarlılığı arttırmak için gerçekleştirilen bir dizi işlemde oluşmaktadır (Hellerstein, 2008). Veri temizleme, veri setini belirli kurallar çerçevesinde standartlaştırma ihtiyacı hissedildiğinde de kullanılmaktadır. Veri, tek kaynaktan geldiğinde, gelen verinin yanlış yazılmış olması, eksik veya girilmemiş olması, yanlış veri tipinin girilmiş olması karşılaşılan hatalar iken, veri birden çok kaynaktan geldiği durumlarda, veriler arasında standart olmaması, veri kaynakların yanlış birleştirilmesi gibi sebeplerden ötürü veri temizleme gereklilik haline gelmektedir. (Rahm ve Do, 2000).

Veri setinde oluşan yanlışlıklar, tek kaynaktan gelen ve birçok kaynağın birleştirilmesi ile oluşturan veri setlerinde farklılık göstermektedir. Tek kaynaktan gelen veri setlerinde, verilerin benzersiz olmaması, veri tutarlılığının sağlanamaması, veri girişlerinde yaşanan yazım yanlışlarının ya da eksik verilerin olması, aynı verilerin birden fazla kez girilmesi ve çelişkili değerlerin olması karşılaşılan sorunlardır. Birden fazla kaynaktan oluşan veri setlerinde (Veri Ambarları, Küresel Web Sitelerinin birleştirilmiş Veri Tabanları), veri kaynaklarının özelinde, tek kaynaktan oluşan setlerde karşılaşılan sorunlara ek olarak, isimlendirme karmaşası, yapısal karmaşıklar, tutarsız birleştirme ve tutarsız zamanlama sorunları görülebilir. İsimlendirme karmaşası, örneğin bir veri tabanında kişi isimleri “isim” olarak ifade edilirken, başka bir veri kaynağında “ad” olarak ifade edilebilir.

Yapısal karmaşıklar, bir veri tabanında, ad ve soyad ayrı ayrı sütunlarda ifade edilirken, bir diğerinde ad ve soyad aynı sütünde ifade edilebilir, bu veri tabanları arasında yapısal karmaşalara sebep olur, bir veri tabanının verileri birleştirilmesi veya diğer veri tabanında ad ve soyadın ayrılması gerekebilir. Tutarsız birleştirme, birden fazla veri tabanından alınan veriler birleştirirken, veri tabanlarında tutulan kayıtların sütunlarına dikkat edilmeden yapılan birleştirmelerdir, örneğin, bir veri tabanında isim – soyisim – cinsiyet – yaş olarak tutulan veriler diğer veri tabanında isim – soyisim – yaş – cinsiyet olarak bulunabilir, bu iki tablo birleştirirken bu tarz sıralamalara dikkat edilmemesi tutarsız birleştirmeye sebep olur. Tutarsız zamanlama, veri kaynakları farklı zamanlarda oluşturulmuş ve güncelleniyor olabilir, bu tarz veri tabanlarında, yeni kullanıcı kayıtları sistem tarafından 24 saatte bir, satış verileri haftalık olarak güncelleniyor ve düzenleniyor olarak görülebilir, bu tarz zaman farklılıkları olan veri tabanları birleştirilirken zamanlara dikkat edilmemesi veri kaybına ya da veri tekrarına yol açabilir (Chu vd., 2016; Hellerstein, 2008; Rahm ve Do, 2000).

Veri temizleme işlemi birkaç aşamadan oluşur. Verinin analizi, veri dönüşüm işlemi için iş akışı (work flow) ve dönüştürülecek/birleştirilecek veriler için eşleme kurallarının belirlenmesi, doğrulama, dönüştürme ve temizlenmiş verilerin geri yüklenmesi aşamaları, veri temizlemenin temel aşamalarıdır. Verinin analizi, eksik, yanlış ve uyumsuz verilerin tespiti için gereklidir. Verinin tamamı ya da veriden alınan örnek üzerinde yapılan analiz ile verilerin üst verisi elde edilir ve detaylı olarak raporlanır. Veri dönüşümü aşamasında, verinin alınmış olduğu veri kaynaklarının sayısı arttığında veri homojen ve dağınık olduğunda veri dönüşümü işleminden önce verinin eşlemesi için analizi yapılır. Verinin, çıkarılması, dönüştürülmesi ve yüklenmesi ETL (Extraction, Transforming, Loading) olarak geçmektedir. Veri, veri kaynaklarından çıkarılıp hazırlanan şemalara göre düzenlemeleri ve temizlemeleri yapılır. Düzenlemelerde, eksik veri içeren kayıtlar veri setinin boyutuna göre veri setinden çıkartılır veya ortalama değerler ile doldurulur. Dönüşüm işlemi gerçekleştirildikten sonra verilerin yükleme işlemi gerçekleştirilir (Rahm ve Do, 2000).

2.2.2.2. Veri Dönüştürme

Veri dönüştürme, veri temizlemeden farklı olarak eldeki verinin azalması değil, verinin bir başka veriye dönüşmesidir (Heer vd., 2015). Veri dönüştürmede, hazır bulunan veri

setindeki veriler bir başka veri tipine veya aynı veri tipinin bir başka gösterimine dönüşür. Örneğin tarihler, dönüştürmeden sonra sadece nümerik olarak gün şeklinde gösterilebilir. Müşteri memnuniyetleri 1 ve 5 arasında yıldız derecelendirmesiyle ifade edilebilirken, sadece 1 yıldız ve 5 yıldız verilen yorumların yıldız dereceleri, 1 yıldız yani hiç memnun değilim 0'a, 5 yıldız yani çok memnunum ise 1'e dönüştürülerek veri setinde bulunan müşteri yıldız derecelendirmeleri 1-5 ile değil 0-1 olarak ifade edilebilir.

2.3. Makine Öğrenimi

Turing, (1950) sormuş olduğu "Makineler düşünebilir mi?" sorusuyla günümüz kadar uzanan bir değişimin öncüsü olmuştur. Makinelerin, insanların gerçekleştirmiş olduğu işleri daha hızlı, daha az hata ve daha az emek ile gerçekleştirmesi makine öğrenmesi alanının temel hedeflerinden birisidir (Schmidhuber, 2015). Makine öğrenmesi, tarımda üretimin verimliliğinin artırılması için (Liakos vd., 2018), sağlıkta hastalıkların tespiti, sınıflandırması ve kişilerin genetik yatkınlığının incelenip gelecekte sahip olabilecekleri hastalıkların tespiti için (Beam ve Kohane, 2018; Bi vd., 2019; Ravi vd., 2017), hukukta karar vericilere destek sistemlerinin tasarımı için (Surden, 2014) ve müşterinin kaybını tahminleme (Vafeiadis vd., 2015), müşterinin satın alma eğilimlerinin tahmin edilmesinde (Martínez vd., 2020), müşterilerin marka bağlılığının incelenmesinde (Aluri et al., 2018) ve benzeri birçok çalışma ve alanda kullanılmakta ve kullanım alanları genişlemektedir.

Makine öğrenmesi, denetimli ve denetimsiz öğrenme olarak iki temel gruba ayrılmaktadır. Denetimli öğrenme, algoritmaya verilen veri setinde her x değerine karşılık gelen y değeri bulunmaktadır. Algoritmadan x ve y değerleri arasındaki ilişkiyi öğrenmesi ve daha önce karşılaşmadığı yeni x değerleri için bir y değeri tahmin etmesi istenir. Algoritmanın bu tahminleri gerçekleştirirken ne kadar doğru ve ne isabetli tahminlerde bulunduğunu ölçmek için birçok başarı metriği bulunmaktadır. Belirli başlı denetimli öğrenme algoritmaları, Naive Bayes, Logistik Regresyon, Karar Ağaçları (Decision Tree), Rastgele Ağaçlar (Random Forest), Gradient Boosting Classifier ve XGBClassifier olarak gösterilebilir. Denetimli öğrenme algoritmalarının günlük hayatta kullanım örnekleri; spam mesaj tespiti, hastalık tespiti, banka dolandırıcılığı tespiti ve benzeri, veriler ve sonuçlarının etiketlendiği yerler olarak gösterilebilir. Denetimsiz öğrenme, algoritmanın almış olduğu x değerleri için belirli bir y değeri sonuçları yoktur. Algoritmanın x değerleri arasında ilişkiyi tespit edip birbirine benzeyen x değerlerini

gruplandırması ve sınıflandırması beklenir. Denetimli öğrenmeden farkı, daha önce belirli bir y olmadığı için denetimsiz öğrenme algoritmalarının çıktılarının tahmin edilmesi zordur. Denetimsiz öğrenme algoritmalarının yapmış olduğu tahminlerde, algoritmaların belirli bir sonuç üzerine çalışmamasından ötürü başarı sınanması yapılması zordur (Lecun vd., 2015; Pedregosa vd., 2011; Schmidhuber, 2015).

2.3.1. Denetimli Öğrenme

Cunningham vd. (2008), göre, denetimli öğrenme, bir dizi X girdisinin, bir dizi Y çıktısına göre eşlenmesini öğrenimi ve bu öğrenmenin daha önce görünmemiş girdiler üzerinde gerçekleştirilerek potansiyel çıktılarının tahminlimesidir. Denetimli Öğrenmenin istatistiksel olarak dayandığı temeller, Risk Azaltma, Ampirik Risk Azaltma ve Risk Sınırlama olarak ifade edilebilir. Denetimli öğrenmenin, denetimsiz öğrenmeye göre farkı, girdilerin sahip olacağı potansiyel çıktıların önceden belirlenmiş olmasıdır. Denetimli öğrenmenin tercih edildiği durumlar; potansiyel çıktılar (veya hedef çıktılar) belirli olduğu veya veri setinin istenilen etiketlere (label) göre sınıflandırılması istenildiği zamanlarda kullanılmaktadır. Denetimli öğrenmenin Ensemble öğrenme kavramı altında 2 temel yaklaşımı bulunmaktadır; Torbalama (Bagging) ve Yükseltme (Boosting). Torbalama için, algoritma asıl veri setinden n miktarda örneklem alıp, bu örneklemeler üzerinde çalışıp sonuç alan temel algoritmanın elde etmiş olduğu tahminler birleştirilerek algoritmanın nihai sonuçları elde edilir (Cunningham vd., 2008). Rastgele Ağaçlar (Random Forest) bir torbalama algoritmasıdır. Yükseltme (Boosting) ise, örneklemeler torbalamada olduğu gibi bir kere baştan seçilmek yerine, her seferinde yeniden oluşturulur. Oluşturulan sınıflandırıcı algoritması, tüm veri seti üstünde test edilir. Her yeni oluşturulan algoritma, bir önceki algoritmanın iyi bir şekilde öğrenemediği verilere odaklanır. AdaBoost, GBM ve XGBM birer yükseltme algoritmasıdır. Denetimli öğrenmenin tahminleme için kullanıldığı birden fazla alan vardır. Altın Yavuz ve Yavuz (2021), yüzey su kalitesinin sınıflandırılması konusunda denetimli öğrenmeyi kullanırken, Martín vd., (2018) turistlerin yapmış hazırlamış oldukları incelemelerin duygu analizi için denetimli öğrenmeden faydalanmışlardır. Denetimli öğrenme, diyabet araştırmalarında kullanıldığı gibi (Kavakiotis vd., 2017), kredi risk faktörlerinin hesaplanmasında da kullanılmıştır (Khashman, 2010).

2.3.1.1. Regresyon

Regresyon, bağımlı bir değişkenin, bağımsız bir değişken ile ilişkisinin tahmini için kullanılan istatistiksel bir yöntemdir. Bu yöntemde iki temel değişken vardır. Değişkenlerden, bağımsız değişken, denklem içerisinde herhangi bir değişken ile pozitif veya negatif anlamda bağlantısı bulunmayan değişkendir. Bağımlı değişken ise, bağımsız değişken/değişkenlere bağlı olarak değişen, bağımsız değişken ile arasında pozitif veya negatif bir ilişki bulunan değişkendir. Pozitif ilişki, bağımsız değişken niceliksel olarak artış gösterirken, bağımlı değişkenin de belirli bir oranda artış göstermesi olarak ifade edilebilirken, negatif ilişki ise pozitif ilişkiye ters olarak bağımlı değişken ile aralarında negatif yönlü bir ilişki olması ve bağımlı değişken artarken bağımsız değişkenin azalması olarak ifade edilebilir.

$$y = mx + b$$

Şekil 14: Tek Bağımsız Değişkenli Regresyon Denklemi

Kaynak: Ünver, Ö., & Gamgam, H. (2008). Uygulamalı temel istatistik yöntemler.

Şekil 14, tek bağımsız değişkenli regresyon denklemini ifade etmektedir.

y , bağımlı değişkeni ifade etmektedir. Bağımlı değişken, eşitliğin diğer tarafındaki her değişkenden etkilenmektedir.

x , bağımsız değişkeni ifade etmektedir. x üzerindeki her artış veya azalış, y üzerinde etkili olmaktadır. Etkinin pozitif ya da negatif yönlü ve ne oranda olacağına m değişkeni ile belirlenmektedir.

m , bağımsız değişkenin katsayısını ifade etmektedir. Örneğin, m 'in 0.1 olduğu bir durumda, x üzerinde yaşanan 10 birim artış, y 'ye 1 birim olarak yansıtacaktır. m 'in 0 olduğu bir durumda, x üzerindeki artış veya azalışların y üzerinde hiçbir etkisi olmaz.

b , sabit değişkeni ifade etmektedir. Sabit değişken, diğer tüm bağımsız değişkenlerin sıfıra eşit olduğu durumda, y değişkeninin alacağı değerdir.

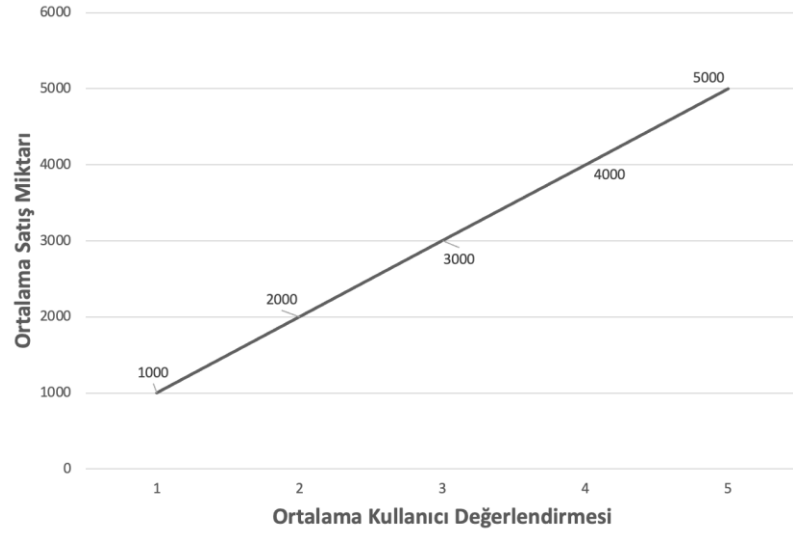
Bu regresyon modeli ayrıca lineer regresyon modeli olarak da bilinmektedir. Lineer regresyon modelleri, tahminlerini doğrusal olarak gerçekleştirdiği için veri setinde bulunan aykırı (outlier) veriler modelin başarısını etkilemekle birlikte, aykırı verilerin tahmini lineer regresyon modelleriyle zordur (Freund vd., 2006).

Tablo 2: Pozitif Tüketici Değerlendirmesi-Satış Miktarı

Ortalama Kullanıcı Değerlendirmesi	Ortalama Satış Miktarı
5	5000
4	4000
3	3000
2	2000
1	1000

Kaynak: Yazar.

Tablo 2, Pozitif bir ilişkiyi ifade etmektedir. Bağımsız değişken olan “Satış Miktarı”, “Tüketici Değerlendirmesi”ne bağlıdır. Kullanıcıların 1 ile 5 arasında yapmış oldukları değerlendirmelere göre, ürünlerin satış miktarı görünmektedir. Ürünlerin ortalama satış miktarı, daha önce ürünü almış kullanıcıların yapmış oldukları değerlendirmelerle ilişkilidir. Ürünü daha önce alan kullanıcıların 5 üzerinde 5 verdiği ürünün satışları en fazla iken 5 üzerinden 1 vermiş oldukları ürünün satışları en azdır. Grafikselsel olarak görünümü Şekil 15’teki gibidir.



Şekil 15: Pozitif Tüketici Değerlendirmesi-Satış Miktarı Grafiği

Kaynak: Yazar.

Ortalama Tüketici Değerlendirmesi ile Ortalama Satış Miktarı arasındaki bu pozitif ilişkiye “pozitif korelasyon” denir. Korelasyon, değişkenler arasındaki ilişkiyi ifade ederken, pozitif ve negatif ilişkinin türünü ifade etmektedir. Model, ortalama satış miktarı ile ortalama tüketici değerlendirme arasındaki ilişkiyi matematiksel olarak analiz ettikten sonra gelecek verilerin tatmini için kullanılabilmesi gibi, bağımlı değişken ile bağımsız değişkenin arasındaki ilişkinin ortaya koyulması için de kullanılmaktadır.

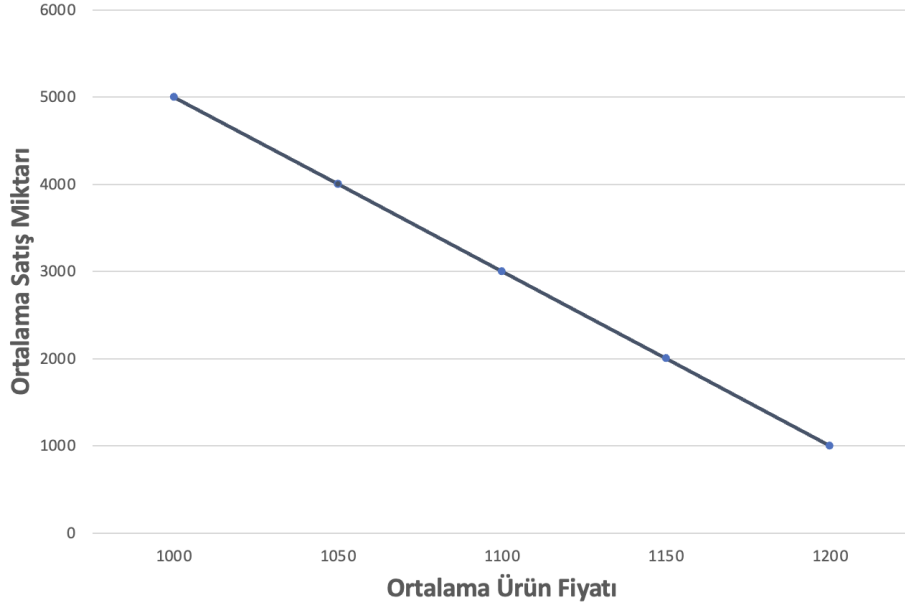
Negatif bir ilişkiye örnek olarak, Tablo 3 gösterilebilir.

Tablo 3: Negatif Tüketici Değerlendirmesi-Satış Miktarı

Ortalama Ürün Fiyatı	Ortalama Satış Miktarı
1000	5000
1050	4000
1100	3000
1150	2000
1200	1000

Kaynak: Yazar.

Birbirine ikame olan beş ürünün ortalama fiyatı ve bu ürünlerin ortalama satış miktarını gösteren bu şekilde, ikame ürünlerden fiyatı yüksek olanın, ortalama satış oranının da diğerlerine göre daha düşük olduğu görülmektedir. Ürünlerin ortalama fiyatları ile ortalama satış miktarlarında negatif korelasyon görülmektedir. Ürünün fiyatı arttıkça, satışı azalmaktadır.



Şekil 16: Negatif Tüketici Değerlendirmesi-Satış Miktarı Grafiği

Kaynak: Yazar.

Regresyon başlıca doğrusal olan, doğrusal olmayan ve çoklu (multicollinearity) regresyon gibi çeşitli regresyon türlerine sahiptir. Logistik regresyon bu türlerden birisidir.

2.3.1.2. Naive Bayes

Bayesian veri analizi, Naive Bayes'in temelini dayandırdığı olmasıyla birlikte, Bayesian analizlerin karmaşıklığına karşın Naive Bayes analizinin efektif olduğu durumlar oldukça yaygındır (Cunningham vd., 2008). Naive Bayes, veri setinin içindeki her bir değişkenin diğer değişkenlerle birlikteliği oranını, tüm veri setinde bulunan değişkenlere oranıyla kıyaslanmasıyla elde edilen oranı verir. Naive Bayes hesaplamasından sonra ortaya çıkan bu oran, sınıflandırma yapılması için algoritma tarafından işlenir.

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

Şekil 17: Bayes Formülü

Kaynak: Cunningham, P., Cord, M., & Delany, S. J. (2008). *Machine learning techniques for multimedia*.

$P(A/B)$: B olayının gerçekleştiği konusunun bilindiği evrende A olayının gerçekleşme olasılığı.

$P(B/A)$: A olayının gerçekleştiği konusunun bilindiği evrende B olayının gerçekleşme olasılığı.

$P(A)$: A olayının olma/gerçekleşme olasılığı.

$P(B)$: B olayının olma/gerçekleşme olasılığı.

Naive Bayes ve diğer denetimli öğrenme algoritmalarını metinler üzerinde kıyaslayan Aggarwal ve Kaur (2013), Naive Bayes'in farklı veri tiplerinde diğer sınıflandırma algoritmalarına göre, gereken işlem gücü açısından daha az maliyetli ve uygulanma açısından daha kolay olduğunu öne sürmüştür. Yine aynı çalışmada Aggarwal ve Kaur (2013), her ne kadar çalışma konusunda verimli olsa ilk kez karşıladığı kelimelerle ilgili Naive Bayes'in performans sorunları yaşadığını belirtmişlerdir.

2.3.1.3. Logistik Regresyon

Logistik regresyon, kategorik verilerin sınıflandırılmasında kullanılan güçlü bir sınıflandırma algoritmasıdır. Logistik regresyon, genel olarak Logistik fonksiyonunu baz alarak çalışmaktadır. Lineer regresyona karşın, Logistik regresyonun çıktıları 0 ile 1 arasındadır. (Belyadi ve Haghighat, 2021).

$$\text{Lojistik fonksiyonu} = \frac{1}{1 - e^{-x}}$$

Şekil 18: Logistik Fonksiyonu

Kaynak: Belyadi, H., & Haghghat, A. (2021). Machine Learning Guide for Oil and Gas Using Python: A Step-By-Step Breakdown with Data, Algorithms, Codes, and Applications. In *Machine Learning Guide for Oil and Gas Using Python*. Gulf Professional Publishing.

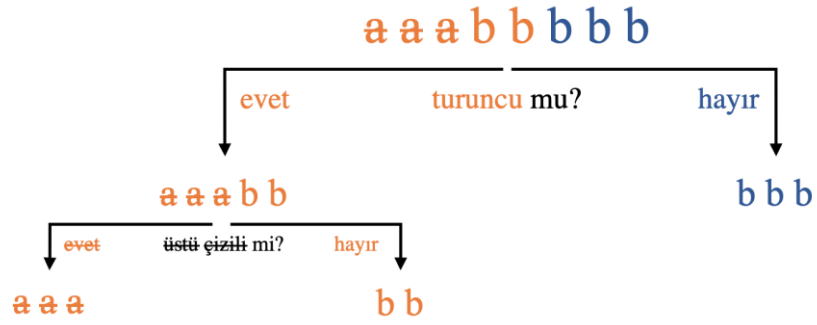
Kategorik verilerin nominal veriler ve ordinal veriler olmak üzere iki türü vardır. Nominal veriler, birbirine karşın herhangi bir üstünlük ifade etmeyen, gruplandırma ve sınıflandırma için kullanılan verilerdir. Örneğin, göz rengi, medeni hal, cinsiyet ve favori renk gibi özellikler nominal verilerdir. Bu kategorilerin diğer kategorilerle herhangi bir ast üst ilişkisi bulunmamaktadır. Buna karşın, ordinal veriler, kategorik olmasına rağmen veriler arasında ast üst ilişkisi bulunmaktadır. Öğrenim durumu, alınan maaşın kategorik olarak ifade gibi, veriler arasında bir ast ve üst ilişkisi vardır. Örneğin, Lisans öğrenimi, kategorik olarak Yüksek lisans öğreniminin altında olmasına karşın lise öğreniminin üstündedir (Gürbüz ve Şahin, 2014).

Zhou vd., (2020) yapmış oldukları çalışmada, kategorik veriler ve Logistik regresyon kullanarak Covid-19 ile ölümlerin risk faktörlerini analiz etmişlerdir. Yapmış oldukları analiz ile kişilerin daha önce hangi hastalığa sahip olmasının, Covid-19 ile birleşmesinin ölümle sonuçlanacağını ortaya koymaya çalışmışlardır. Çocukluk yaşta yaşanan istismarların yetişkin ölümleri üzerindeki etkisini inceleyen Felitti vd., (1998) demografik özelliklerin kümülatif etkisinin analizi için Logistik regresyon kullanmıştır. Young vd., (1993) uyku esnasındaki solunum bozuklarındaki risk faktörlerinin analizinde Logistik regresyon kullanmıştır.

2.3.1.4. Karar Ağacı

Karar Ağacı, denetimli öğrenme algoritmasıdır. Daha önce belirli olan çıktılara göre tasarlanmış ağaç yapısı üzerinden verilerin kategorilendirilmesi (sınıflandırılması) gerçekleştirilir. Karar ağacı, ağaç yapısı kullanarak verileri belirli kurallara göre sınıflandırma işlemi gerçekleştirdikten sonra alt dallarda sınıflandırmaya ve verileri ayrı

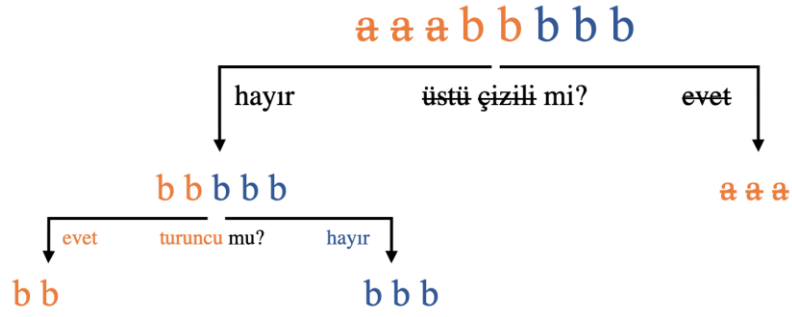
ayrı kategorilere ayırmaya devam edebilir. Karar ağacı, bir veri kümesinin özellik sınıflarını, her sınıfta bir özellik olacak şekilde bölmek için özyinelemeli olarak uygulanan, ayırt edici işlevlerden veya karar kurallarından oluşan hiyerarşik bir modeldir. Veri kümelerinin yönetiminde ve sınıflandırılmasında oldukça faydalı olan bu yöntemde en üstte bulunan düğüm (node) başlangıç yapılacağı değişkeni ifade eder. Ağacın dallarından ayrılan diğer düğümler belirli karar şartları içerir, bu şartlara göre düğümler diğer alt düğümlere ayrılır. Başlangıç seviyesi, seviye 0 olarak ifade edilir. Seviye sıfırda sadece bir düğüm bulunur. Seviye 1’de başlangıç seviyesindeki düğüme göre şartları sağlayan diğer düğümler bulunur, bunlar ebeveyn (parent) düğüm olarak da ifade edilir. Bu düğümlerin altında bulunan seviye 2 ve diğer seviyedeki düğümler çocuk (child) olarak ifade edilir (Myles vd., 2004).



Şekil 19: Karar Ağacı Örneği

Kaynak: Yazar.

Şekil 19, karar ağacı örneği içermektedir. Ağaç algoritması entropi gibi çeşitli matematiksel denklemlere ile ağacın gruplandırılmalarına karar verir. Her alt seviyede, üst seviyede ayrıştığı gruptan verilerden arındırılmış alt kümeler oluşmaktadır. Örnekte, üstü çizili turuncu, turuncu ve mavi renkte harfler bulunmaktadır. Algoritma öncelikli olarak renklerine göre ayırım yapmıştır. Eldeki veri setinin özellikleri belirgin olduğu için ilk seviye farklı bir şekilde değişebilir. Şekil 19’de algoritmanın aynı veri setinde başka bir yaklaşımını görmekteyiz. Uygulamanın seviye 1’de ilk dağılımını nasıl yapacağı veri setinin boyutu ve çeşitliliğine göre değişmektedir.



Şekil 20: Karar Ağacı Örneği

Kaynak: Yazar.

Örnekteki veri setinde eğer üstü çizili olmayan turuncu a harfleri de olsaydı daha farklı bir yaklaşım izlemesi gerebilirdi. Karar ağacının en büyük faydalarından birisi algoritmanın hazırlamış olduğu karar ağacını ve karar verme şartlarını, karar ağacı hazırlandıktan görsel olarak görebilmemizdir. Şekil 19 ve Şekil 20'deki görsellerde bulunan evet ve hayır sekmeleri, karar ağacındaki karar şartlarını ifade etmektedir.

2.3.1.5. Rassal Orman

Rassal Orman (Random Forest/Rastgele Orman/Rassal Ağaçlar), birden fazla karar ağacının farklı alt kümeler üzerinden oluşturulması ile elde edilir. Karar ağacı, veri seti üzerinde aşırı öğrenme gerçekleştirebilmesi sebebiyle çok fazla öznelik içeren veri setlerinde tercih edilmemektedir. Rassal Orman, birden fazla öznelik içinden birden fazla alt karar ağacı geliştirerek, hem algoritmanın belirli bir öznelik üstünde aşırı öğrenme gerçekleştirmesini engeller hem de hangi özneliğin daha önemli olduğunu, algoritmanın daha iyi tahminler yapabilmesi için gerekli olduğunu ortaya çıkartır. Rassal orman sadece bir öznelik için karar ağacı oluşturup onu test edebileceği gibi, birkaç farklı özneliği birlikte test edebilir. Rassal ormanın birden fazla karar ağacını üzerinde öğrenip gerçekleştirip bu karar ağaçları üstünden nihai algoritmayı oluşturmasına kolektif (ensemble) öğrenim denir (Biau ve Scornet, 2016; Myles vd., 2004). Kolektif öğrenimi, Rassal Orman algoritması gibi, algoritmanın kendi yapısında buluna bileneceği gibi, algoritmanın tasarımında birden fazla algoritmanın kullanılmasına da kolektif öğrenim denir.

2.3.1.6. Gradyan Arttırma Sınıflandırıcısı ve XGB Sınıflandırıcısı

Gradyan Arttırma Sınıflandırıcısı, Rassal Orman gibi Karar Ağacı algoritmasını birçok defa birçok alt küme üzerinden çalıştırır. Gradyan Arttırma Sınıflandırıcısı, çalıştırmış olduğu Karar Ağaçlarının ortalamasını sonuç olarak sunar. Çalıştırma aşamasında birden fazla metot kullanır. Bu metotlardan birisi kayıp fonksiyonudur. Kayıp Fonksiyonu ya da ortalama hata karesi (mean squared error), tahminlerin gerçek değerlerden ne kadar uzak olduğunun karesi toplamı ile hesaplanır.

$$\text{Ortalama Hata Karesi} = \frac{1}{n} \sum_{i=1}^n (g - t)^2$$

Şekil 21: Ortalama Hata Karesi

Kaynak: Yazar.

Ortalama hata karesinde;

n, veri setinde bulunan toplam veri sayısını,

g, karşılık gelen gerçek değeri,

t, tahmin edilen değeri ifade etmektedir.

Ortalama hata karesi, gerçek değer ile tahmin değerinin farkının karesini aldığından ötürü negatif olamaz. Ortalama hata karesi, gerçek değer ile tahmin değerinin karesini aldığından ötürü özellikle aykırı değerlerden çok etkilenir. Gradyan Arttırma Sınıflandırıcısı, oluşturmuş olduğu karar ağacının, ortalama hata karesine göre bir diğer karar ağacı oluşturulur ve belirlenmiş tekrar sayısı kadar ağaç, bir öncekinin ortalama hata karesine göre oluşturulur. Gradyan Arttırma Sınıflandırıcısı, sadece son oluşturulan ağacı baz almaz. Nihai sonuç, oluşturulan tüm ağaçların ortalamasıdır (Natekin ve Knoll, 2013).

XGBClassifier, Gradyan Arttırma Sınıflandırıcısının, L1 regülarizasyon ile düzenlenmiş halidir. Gradyan Arttırma Sınıflandırıcısı'nın, XGB Sınıflandırıcısı'na göre aşırı öğrenme yapması daha olasıdır (Ran vd., 2017).

2.4. Denetimsiz Öğrenme ve Denetimsiz Öğrenme Türleri

Ghahramani, (2003) denetimsiz öğrenmeyi diğer öğrenme türlerinden karşılaştırırken, denetimli öğrenmeyi, verilmiş olan bir dizi x girdisinin beklenen bir y çıktısı olduğunu ve $x - y$ değerleri arasındaki ilişkiyi matematiksel olarak ortaya koyup daha önce karşılaşmadığı x girdileri için isabetli tahmin yapması beklendiğini, pekiştirmeli öğrenmeyi ise algoritmaya verilmiş bir dizi x girdisi için tahminler yaparken, yaptığı tahminlere göre “ödül” ya da “ceza” olarak cezadan kaçınacak ve sürekli ödüle öncelik vererek tahminler yaptığını algoritma tipleri olarak tanımlamıştır. Denetimsiz öğrenmede ise herhangi bir beklenen y çıktısı ya da ödül – ceza sistemi bulunmamaktadır. Belirli ve beklenen y çıktıları olmadan ya da yapılan her tahmin konusunda bir geri dönüş almadan algoritmanın nasıl bir tahminleme yapacağı belirli olmayabilir ancak, denetimsiz öğrenme, gürültü olarak görünen veriler arasındaki ilişkileri ortaya koymak ve çıktıları hakkında önceden tahminde bulunulamayan veri setlerinde görülmeyen bağlantıları ortaya koymak için kullanılmaktadır (Sathya ve Abraham, 2013).

Denetimsiz öğrenmenin iki temel türü vardır; kümeleme (clustering) ve boyutsal azaltma (dimensionality reduction). Kümeleme, k-means algoritması gibi algoritmaların çalışma biçimidir. K-means adında bulunan k harfi, algoritmanın veri setini kaç kümeye ayıracağını ifade etmektedir. Algoritma çalışmaya başlamadan önce belirlenmiş küme sayısına göre veri setindeki verileri kümelere ayırmaya çalışır. Bunu algoritmanın temel kullanım alanları, müşteri alışveriş alışkanlıklarını gruplandırma, ürün pazar özelliklerini gruplandırma gibi daha önce herhangi bir etiketi veya sayısal karşılığı bulunmayan setler üzerinde kümeleme yapması için kullanılır. Bir diğer denetimsiz öğrenme yöntemi olan boyut azaltma, veri setinde bir verinin bağlı olduğu birden fazla boyutu, o veri için önemli tek boyuta indirgeyerek veri görselleştirme ve diğer makine öğrenmesi algoritmaları için öznitelik için verileri kümeler. Örneğin, müşterinin zaman ve konuma bağlı verileri bulunan bir sette müşterileri sadece zaman veya sadece konum olacak şekilde kümeleneğine boyut azaltma denir (Ghahramani, 2003; Sathya ve Abraham, 2013).

2.5. Pekiştirmeli Öğrenme

Pekiştirmeli Öğrenme, (Reinforcement Learning) bir makine öğrenmesi türü olarak, denetimli ve denetimsiz öğrenmeden farklı olarak algoritmanın yapmış olduğu tahminler için bir hedef belirlenir ve algoritmanın bu hedef için tahminler yapması istenir. Yapılan tahminler hedefe yakınsa algoritma ödüllendirilir, hedefe uzaksa algoritma cezalandırılır. Bu ödül – ceza sisteminde algoritmanın hangi özneliklere değer vereceği algoritma tarafından belirlenir (Ghahramani, 2003; Sathya ve Abraham, 2013). Pekiştirmeli öğrenme algoritmalarının dezavantajlarından birisi çok fazla işlemci gücü gerektirmesi ve öğrenme sürelerinin görece olarak daha uzun sürmesi olarak gösterilebilir.

2.6. Makine Öğrenmesi Modelini Sınamak

2.6.1. Bias ve Varyans

Bias, modelin yapmış olduğu tahminlerin, gerçek değerler ile arasındaki farkı ifade eder. Varyans, modelin yapmış olduğu tahminlerin dağılımını ifade eder. Bias ve Varyans, denetimli öğrenme modellerinin başarısının sınanması için kullanılan yöntemlerden birisidir. İyi bir denetimli öğrenme algoritmasının tahminlerinde düşük bias ve düşük varyans olması beklenir (Breiman, 1996).

Bias'ın yüksek olması ancak varyans'ın düşük olması modelin eğitim verisi üstünde yeterince öğrenemediğini yani az öğrenme gerçekleştiğini gösterir. Bias'ın düşük olması ancak varyans'ın yüksek olması ise modelin eğitim verisi ezberlediğini yani aşırı öğrenme gerçekleştiğini gösterir (Friedman, 1997).

2.6.2. Aşırı Öğrenme

Aşırı öğrenme, istatistiksel öğrenme süreçlerinde ortaya çıkan, modelin eğitim veri seti üzerindeki verileri aşırı öğrenip, buradan elde ettiği sonuçları ezberlemesi ve bu eğitim seti üstünde gerçekleştirilen testlerde görece çok iyi sonuçlar alıp daha önce karşılaşmadığı veriler ile test edildiğinde, eğitim verisi üzerinde göstermiş olduğu başarıyı gösterememesi olarak ifade edilebilir (Ying, 2019). Aşırı öğrenmede model düşük bias ancak yüksek varyans gösterir. Yapılan tahminler, eğitim setine oldukça yakın olduğu için bias düşük olur. Ancak model, aşırı öğrenme gerçekleştirdiği için varyansı yüksek olur. Aşırı öğrenme, veri setinin azlığından veya modelin karmaşıklığından ötürü gerçekleşebilir. Veri setinin azlığıyla birlikte, setin içerisinde

bulunan veriler arasında bulunan gürültünün, veri setinin geneline olan etkisi artmaktadır. Gürültü, veri kaynağından hatalı veri gelmesi veya veri toplama sırasında araştırmacının yapmış olduğu hatalardan kaynaklı olarak veri setinin, veri kaynağını hatalı ifade etmesidir (Kalapanidas vd., 2003).

Aşırı öğrenmeden kaçınmak için;

- 1- Veri setinin boyutunu arttırmak,
- 2- Öğrenimi erken sonlandırmak,
- 3- Algoritmada bulunan öznitelik sayısını düşürmek,
- 4- Algoritma üzerinde regülarizasyon (düzenleme) yapmak,

Uygulanabilmektedir.

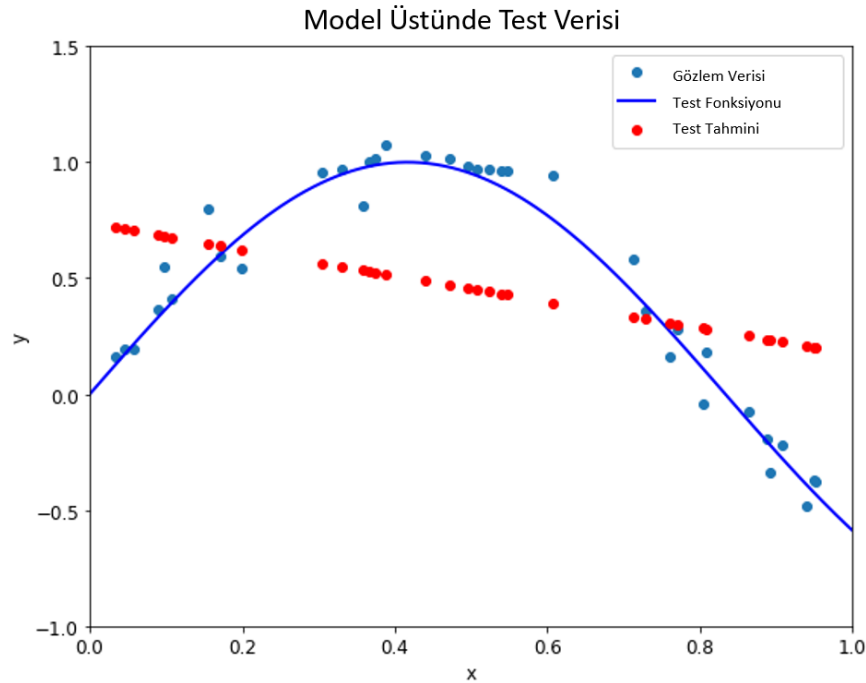
Öznitelik, gözlemlenmiş olan bir varlık ya da olguya ait, diğer varlık ya da olgularla ayrışabilen ve ölçümü yapılabilen niteliktir (Sarle, 1995). Veri setinin boyutunu arttırmak, veri setinin gürültüsünü azaltabilir. Makine öğrenmesinin belirli bir seviyeye geldikten sonra öğrenimini sonlandırmak, aynı veri seti üzerinde öğrenmeye devam ederek, öğrenimin aşırı öğrenmeye dönüşmesinin önüne geçebilir. Makine öğrenmesi algoritmasının öğrenme noktalarından olan özniteliklerin sayısını azaltmak, veri seti üzerinden yaşanabilecek aşırı öğrenmeyi azaltır. Makine öğrenme algoritmasının düzenlenmesi, algoritmanın özniteliklere vermiş olduğu ağırlığın düzenlenmesiyle birlikte aşırı öğrenmenin önüne geçilebilir (Cook ve Ranstam, 2016; Sarle, 1995; Ying, 2019).

Özniteliklerin ağırlığı, makine öğrenmesi algoritması içinde, bu özniteliklerde gerçekleşen değişimlerin sonuca etkisini ifade eder. Regülarizasyon (düzenleme) yapmak için kullanılan yöntemlerden birisi ise L1 ve L2 regülarizasyondur. L1 regülarizasyon Ridge regülarizasyon olarak geçerken, L2 LASSO regülarizasyon olarak ifade edilir. Ridge ve LASSO regülarizasyon, özniteliklerin ağırlıkları üzerinde düzenlemeler yapar. Bu düzenlemeler ile, sonuca etkisi az olan özniteliklerin ağırlığını azaltırken, sonuca etkisi yüksek olan özniteliklerin ağırlığını artırır. Ridge ve LASSO regülarizasyon arasındaki en temel fark, Ridge, etkisinin az olduğunu tespit ettiği özniteliklerin ağırlığını azaltırken, LASSO bu özniteliklerin ağırlığını sıfıra eşitler ve sonuca etkisini sıfırlar. Özellikle özniteliklerin çok olduğu ve öznitelikler arasında seçim yapılması gereken

durumlarda LASSO, sonuca etkisi daha düşük olan özneliklerin ağırlığını sıfırlayarak çıkartır (Melkumova ve Shatskikh, 2017; Ying, 2019).

2.6.3. Az Öğrenme

Az öğrenme, aşırı öğrenmenin tam tersi olarak, algoritmanın eğitim setini öğrenememesidir. Şekil 22’de görülebileceği üzere, mavi noktalar eğitim setini, kırmızı noktalar eğitim seti üzerindeki tahminleri ifade ederken, mavi çizgi olması gereken fonksiyonu ifade etmektedir. Az öğrenme, veri setine göre daha basit şekilde tasarlanmış algoritmalarda görülebilir. Algoritma, veri setindeki kalıp göremez/öğrenemez (Koehrsen, W., 2018).



Şekil 22: Az Öğrenme Örneği

Kaynak: Koehrsen, W. (2018, January 28). *Overfitting vs. Underfitting: A Complete Example*. <https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765>

Az öğrenmede modelin bias'ı oldukça yüksektir. Modelin tahminleri ile eğitim setinin farkı fazla olduğu için modelin bias değeri oldukça yüksektir. Ancak, model az öğrenim gerçekleştirdiği için ve modelin yapısını öğrenemediğinden yapılan tahminlerin varyans değeri düşüktür. Aşırı öğrenmenin önüne geçmek için kullanılan öğrenimi erken durdurma, az öğrenmenin bir sebebi olabilir. Veri seti üzerinde yeterince öğrenme

gerçekleşmeden sonlanan eğitim nedeniyle, model eğitim setini yeterince öğrenemez (Jabbar ve Khan, 2015).

2.6.4. Hata matrisi, Doğruluk, Duyarlılık ve Kesinlik

Hata Matrisi, denetimli öğrenme algoritmalarının başarısının sınıandığı ölçütlerden birisidir (Visa vd., 2011). Algoritmaların, ne kadar doğru tahmin ettiği kadar, neyi doğru tahmin ettiği önemlidir. Özellikle sağlık gibi hayati konuların söz konusu olduğu yerler için doğru tahminin önemi artmaktadır (Breiman, 1996; Friedman, 1997).

Tablo 4: Hata Matrisi

		Tahmin	
		Doğru	Yanlış
Gerçek	Doğru	a	b
	Yanlış	c	d

Kaynak: Yazar.

Tablo 4, Hata Matrisini göstermektedir. Hata Matrisi, gerçek ile tahmin arasındaki cevapların ilişkisini adlandırmaktadır.

a = Gerçekte doğru olan verilerin, algoritmanın doğru olarak tahminlimesidir. Algoritmanın, gerçekte “doğru” olan verileri “doğru” olarak tahmin etmesi, gerçek pozitif olarak ifade edilir. Gerçek Pozitif (GP), kanser hastası olan bir kişinin, kanser hastası olarak tespit edilmesi olarak gösterilebilir.

b = Gerçekte doğru olan verilerin, algoritmanın yanlış olarak tahminlimesidir. Algoritmanın, gerçekte “doğru” olan verileri “yanlış” olarak tahmin etmesi, sahte negatif olarak ifade edilir. Sahte Negatif (SN), kanser hastası olan bir kişinin, kanser hastası değil olarak tespit edilmesi olarak gösterilebilir. Yanlış negatif, Tip 2 hata olarak geçer.

c = Gerçekte yanlış olan verilerin, algoritmanın doğru olarak tahminlimesidir. Algoritmanın, gerçekte “yanlış” olan verileri “doğru” olarak tahmin etmesi, sahte pozitif

olarak ifade edilir. Sahte Pozitif (SP), kanser hastası olmayan bir kişinin, kanser hastası olarak tespit edilmesi olarak gösterilebilir. Yanlış pozitif, Tip 1 hata olarak geçer.

d = Gerçekte yanlış olan verilerin, algoritmanın yanlış olarak tahminlimesidir. Algoritmanın, gerçekte “yanlış” olan verileri “yanlış” olarak tahmin etmesi, gerçek negatif olarak ifade edilir. Gerçek Negatif (GN), kanser hastası olmayan bir kişinin, kanser hastası değil olarak tespit edilmesi olarak gösterilebilir. (Buckland ve Gey, 1994)

2.6.5. Doğruluk (Accuracy)

Doğruluk, algoritmanın veri seti üzerinde tahmin ettiği Gerçek Pozitif ve Gerçek Negatif tahminlerin sayısının, tüm veri setine oranıyla elde edilir. Yapılan tahminlerin ne kadarının Gerçek Pozitif veya ne kadarının Gerçek Negatif olduğu ile ilgili bilgi vermez. 100 kişilik bir grup içerisinde yapılan Gerçek Pozitif ve Gerçek Negatif tahminlerin sayısı 90 ise, bu algoritmanın doğruluk oranı $90/100 = \%90$ olarak hesaplanır.

$$Doğruluk = \frac{GP + GN}{GP + GN + SP + SN}$$

Şekil 23: Doğruluk Formülü

2.6.6. Hassasiyet (Recall)

Hassasiyet, algoritmanın veri seti üzerinde tahmin ettiği Gerçek Pozitif tahminlerin sayısının, Gerçek Pozitif ve Sahte Negatif tahminlere oranını ifade eder. Algoritmanın, hatalı tahminlerden kaçınılması gereken durumlarda hassasiyet oranına bakılır (Friedman, 1997). 100 kişilik bir grup içerisinde Algoritmanın, Gerçek Pozitiflerin, yani gerçek kanser hastalarının, kanser hastası olarak doğru tahminlerinin sayısı 20, Sahte Negatiflerin, yani gerçek kanser hastalarının hasta değil olarak yapılan tahmin sayısı ise 30 ise, bu algoritmanın hassasiyet oranı $20/50 = \%40$ olarak hesaplanır.

$$Hassasiyet = \frac{GP}{GP + SN}$$

Şekil 24: Hassasiyet Formülü

2.6.7. Kesinlik (Precision)

Kesinlik, algoritmanın veri seti üzerinde tahmin ettiği Gerçek Pozitif tahminlerin sayısının, Gerçek Pozitif ve Sahte Pozitif tahminlere oranını ifade eder. Algoritmanın, “doğru” olarak tahmin ettiği değerlerin ne kadarının gerçek “doğru” olduğunun anlaşılması gerektiği durumlarda kesinlik oranına bakılır (Friedman, 1997). Algoritmanın, Gerçek pozitiflerin, yani gerçek kanser hastalarının, kanser hastası olarak tahminlerinin sayısı 40, Sahte Pozitiflerin, yani aslında kanser hastası olmayan kişilerin hasta olarak tahminleme sayısı 10 ise, bu algoritmanın kesinlik oranı $40/50 = \%80$ olarak hesaplanır.

$$Kesinlik = \frac{GP}{GP + SP}$$

Şekil 25: Kesinlik Formülü

2.6.8. Özgüllük (Specificity)

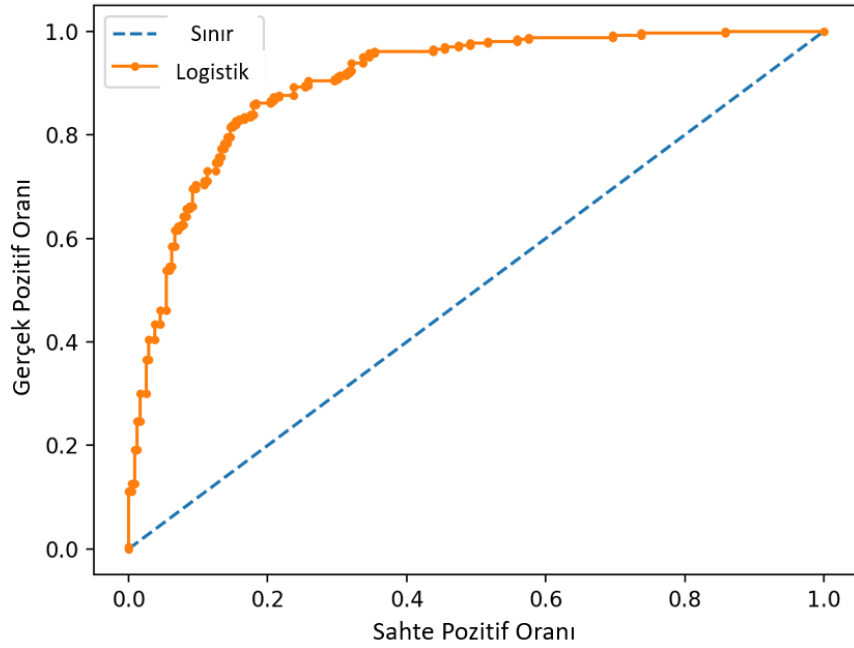
Özgüllük, algoritmanın veri seti üzerinde tahmin ettiği Gerçek Negatif tahminlerin sayısının, Gerçek Negatif ve Sahte Pozitif tahminlerin oranını ifade eder. Algoritmanın, “yanlış” olarak tahmin etmesi gereken verilerin ne kadarını doğru olarak “yanlış” olarak tahmin ettiğini gösterir. Algoritmanın, Gerçek Negatiflerin, yani kanser hastası olmayan kişilerin, kanser hastası değil olarak tahminleme sayısı 30, Sahte Pozitiflerin, yani kanser hastası olmayan kişilerin, kanser hastası olarak tahminleme sayısı 10 ise, bu algoritmanın özgüllük oranı $30/40 = \%75$ olarak hesaplanır.

$$\text{Özgüllük} = \frac{GN}{GN + SP}$$

Şekil 26: Özgüllük Formülü

2.6.9. ROC-AUC Eğrisi

ROC eğrisinde, Sahte Pozitiflerin Oranı (FPR) X ekseninde bulunurken, Gerçek Pozitiflerin Oranı (TPR) Y ekseninde bulunur. Sahte Pozitiflerin Oranı, $(1 - \text{Özgüllük})$, Gerçek Pozitiflerin Oranı ise direkt olarak Hassasiyet'e eşittir. Bu eğrinin altında kalan alan büyüdükçe, algoritmanın sınıflandırma başarısı da artmaktadır. Kanser örneğinde, algoritmanın ROC eğrisinde altta kalan alan büyüdükçe, kanser hastalarını hasta, kanser hastası olmayan kişileri hasta değil olarak sınıflandırma başarısı artmaktadır (Buckland ve Gey, 1994).



Şekil 27: Örnek ROC-AUC Eğrisi

Kaynak: Brownlee, J. (2018, August 31). *How to Use ROC Curves and Precision-Recall Curves for Classification in Python*. <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>

Brownlee, (2018) yapmış olduđu çalışmada, hazır veri setleri üzerinden lojistik regresyon algoritmasıyla makine öğrenmesi gerçekleştirmiş ve bu algoritmanın başarısını ROC-AUC Eğrisi üzerinden göstermiştir. Sahte Pozitiflerin Oranı (FPR) X ekseninde bulunurken, Gerçek Pozitiflerin Oranı (TPR) Y ekseninde bulunmaktadır. Turuncu çizgi, algoritmanın almış olduđu skorları ifade ederken, mavi çizgi algoritmanın hiç öğrenme göstermeseydi alacağı değerleri ifade etmektedir. Turuncu çizginin altında kalan kısım Algoritmanın başarısını ifade etmektedir.

2.6.10. F1 Skoru

F1 Skoru, modelleri değerlendirmek için kullanılan başarı metriklerinden birisidir. F1 Skoru, Kesinlik ve Hassasiyet değerlerinin harmonik ortalamasıyla hesaplanmaktadır (Fujino vd., 2008).

$$F_1 = 2 * \frac{\text{kesinlik} * \text{hassasiyet}}{\text{kesinlik} + \text{hassasiyet}}$$

Şekil 28: F1 Formülü

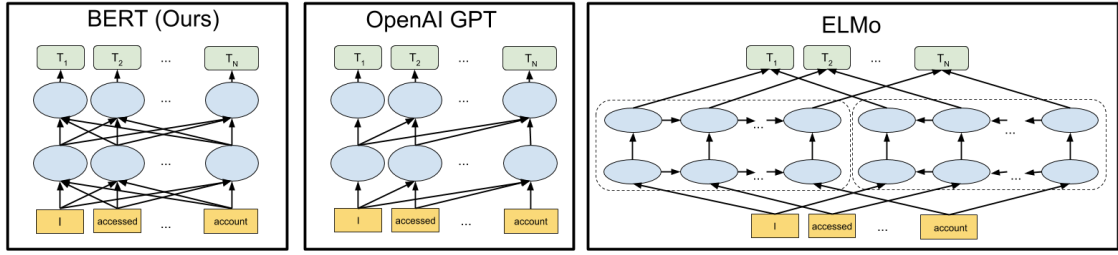
Kaynak: Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1), 1–28. <https://doi.org/10.1186/S12880-015-0068-X/TABLES/5>

F1 Skoru, özellikle veri setinin dengeli dağılmadığı durumlarda modellerin başarısının sınanması için kullanılmaktadır. Kesinlik ve Hassasiyet değerlerinin sadece ortalaması alınması algoritmanın dengesiz dağılan veri setinde nasıl bir başarı gösterdiğini ortaya koyma konusunda yetersiz kalacaktır. Harmonik ortalamasının alınmasıyla birlikte, Kesinlik değeri 1 olan, Hassasiyet değeri 0,5 olan bir algoritmanın normal ortalaması 0,75 olarak çıkabilirken harmonik ortalaması 0,66 olarak karşımıza çıkmaktadır.

2.7. Duygu Analizi ve BERT

Duygu analizi ya da düşünce madenciliği, kişilerin haberlerde, kişisel sayfalarında, sosyal paylaşım sitelerinde, çevrimiçin alışveriş platformlarında yapmış oldukları paylaşımların içerdikleri duygu bakımından analizine verilen isimdir (B. Liu, 2015). 2000’li yıllardan

sonra Web 2.0'ın gelişimiyle birlikte en çok gelişim gösteren doğal dil işleme alanı, duygu analizidir (Zhang et al., 2018). Duygu analizi kullanılan yapıya (framework) göre farklılık gösterse de temel kullanım alanı girdilerin olumlu ya da olumsuz olarak etiketlenmesidir. Farklı yapılarda içerdiği olumlu ve olumsuz duygulara ek olarak bir de nötr duygu içerip içermediğine göre de etiketlenebilse de bazı yapılarda nötr duygular bulunmamaktadır. Duygu analizi için kullanılacak modeller, kullanıcıların kendi etiketlemesini ve makine öğrenimini gerçekleştirilip oluşturulabilirken, modelin eğitimi için günümüzde Google tarafından geliştirilen BERT (Bidirectional Encoder Representations from Transformers) yaygın olarak kullanılmaktadır (Devlin ve Chang, 2018). BERT'i diğer doğal işleme algoritmalarından ayıran en temel özellik, cümleleri baştan sona ilişkilendirerek çift yönlü kodlama yapabilesidir.



Şekil 29: BERT, OpenAI GPT ve ELMo

Kaynak: Devlin, J., & Chang, M.-W. (2018, October 2). Open Sourcing BERT: *State-of-the-Art Pre-training for Natural Language Processing*.

Devlin ve Chang (2018), hazırlamış oldukları görselde, BERT'in diğer dönüştürücülerle (transformers) yapısal olarak kıyaslanmasını göstermişlerdir. BERT, derin bir şekilde çift yönlü olmasına karşın, OpenAI tarafından geliştirilen GPT-2 algoritması tek yönlüdür. ELMo ise "yüzeysel" olarak çift yönlüdür. GPT-3'ün OpenAI tarafından tanıtılmasından sonra, BERT ile GPT-3 arasındaki farkı inceleyen Topal vd., (2021) GPT-3'ün 175 milyar parametre üzerinde tahmin yaptığını ancak BERT'in derin çift yönlü ilk denetimsiz öğrenme algoritması olarak günümüzde hala etkili denetimsiz doğal dil işleme algoritmalarından birisi olarak bilinmektedir. Hugging Face gibi girişimler, BERT ile herkesin kendi veri setiyle birer algoritma geliştirebilmesine olanak sağlamaktadır.

Duygu analizini, büyük şirketler kullanıcılarının memnuniyet ve şikayetlerinin analizi için kullanmaktadır. Geleneksel yöntemlerde kullanıcıların memnuniyeti, kullanıcıların şikayetleri ve benzeri analizler için kullanıcıların birebir iletişime geçilmesi, anket

çalışması yapılması, daha detaylı konularda odak grup görüşmesi yapılması gerekebilirdi. Günümüzde tüketiciler, ürün ve hizmetler konusundaki memnuniyet ve şikayetlerini sosyal ağlar, kullanıcı forumları, kişisel ağ günlükleri üzerinden paylaşabilmektedir. Kullanıcılara gönderilecek bir dizi soru içeren anket gibi veri toplama yöntemlerine göre çevrimiçi ortamlarda yapılan paylaşımların analiz edilmesi kullanıcılarla ulaşma, soruları yöneltmesi ve cevaplanması gibi süreçleri içermediği için veriye ulaşım daha hızlıdır. Ancak, erişilen bilgiler direkt olarak kullanıcıların ürün ya da hizmet hakkındaki olumlu ya da olumsuz deneyimlerini içerdiği için, analiz edilmek istenen belirli konu hakkında istenen bilgileri içermeyebilir.

Bir markanın sunduğu ürün veya hizmet konusunda kullanıcılarının genel memnuniyetinin araştırılması için marka adının geçtiği paylaşımlar toplanarak, belirli veri ön işleme süreçlerinden sonra duygu analizi yöntemleriyle birlikte kullanıcıların memnuniyetleri hakkında genel bilgi sahibi olunabilir.

2.8. Python

Python, ilk sürümü 1991 yılında yayınlanmış, kodların okunabilirliğini artırma hedefiyle geliştirilmiş olmasına karşın, kullanıcılara yazmış oldukları kavramları daha az kod ile ifade etme imkânı soran kodlama dilidir. Python, açık kaynak kodlu bir programlama dili olup kullanıcıların kullandıkları kodların içeriğini görmesine izin vermekte, tüketiciler isterlerse kendi yazmış oldukları kütüphaneleri (packages) çeşitli siteler aracılığıyla açık kaynak olarak kullanıma sunabilmektedir (*History of Python*, 2022). Python, bir yorumlayıcı (interpreted) bir programlama dilidir. Python ile yazılmış program kodları öncelikle binary olarak dönüştürülür ve daha sonra Python tarafından yaratılan sanal çekirdek üzerinde çalıştırılır (Bogdanchikov vd., 2013). Geleneksel programlama dilleri kodları bir derleyici (compiler) tarafından işlenir. Python, yorumlayıcı bir programlama dili olduğu için, programdan yapılması istenen işler daha az kod ile ifade edilebilir. Ancak, arka planda yazılan kodların yorumlanması makine tarafından gerçekleştirildiği için çalışma süresi ve kaynak tüketimi geleneksel programlama dillere göre fazladır. Genel kullanıcıların kullanım alanlarına hitap eden birçok veri yükleme, veri manipülasyon etme ve veri dönüştürme kütüphaneleri olmasına karşın, spesifik grupların kullanımı için de yazılmış Astro.py gibi, astronomi ile ilgilenen kullanıcıların kullanabileceği, kendi içerisinde astronomik dönüşümler yapılmasını kolaylaştıracak,

sanal gözlem evi tabloları oluşturulmasına imkân veren araçlar içeren kütüphaneler de bulunmaktadır (Robitaille vd., 2013).

Python içerisinde bulunan kütüphaneler aracılığıyla birçok işleme izin vermektedir. NumPy kütüphanesi sayesinde, bilimsel hesaplamalar yapma ve çok boyutlu matrislerin düzenlenmesine ve manipüle edilmesine imkân sağlamaktadır. Nümerik verilerin analizi için güçlü bir araç olan NumPy üzerinde yazılmış herhangi bir veri kümesi, herhangi bir dönüşüme ihtiyaç duymadan C veya Fortran gibi dillerde çalışabilir. pandas, yapısal veriler üzerinde çalışmayı kolaylaştıracak araçlar içeren bir veri analizi kütüphanesidir. Veri kümelerinin düzenlenmesi (manipüle edilmesi) ya da analiz edilmesi için yerleşik araçlara sahip kütüphane, NumPy ile beraber SQL gibi yapısal verilerin düzenlenmesi, analiz edilmesi, kümelere ayrılması, sütunları veya satırların ayrılması, sütunlar üzerinden yeni sütunların oluşturulmasına olanak sağlamaktadır. matplotlib, verilerin 2 boyutlu olarak görselleştirilmesine olanak sağlayan bir kütüphanedir (McKinney, 2017).

İçerdiği kütüphaneler ve yazım dilinin kolaylığı sayesinde Python, günümüzün popüler programlama dillerinden birisidir. An itibariyle güncel sürümü 3.10 olan Python, içerdiği scikit-learn kütüphanesi sayesinde birçok algoritmayı temel alan birçok makine öğrenmesi algoritması da sunmaktadır.

scikit-learn, Python kullanıcı toplulukları tarafından geliştirilen bir veri analiz kütüphanesidir. İçerdiği NumPy gibi kütüphaneler sayesinde, veri analizini kolaylaştıran scikit-learn, sınıflandırma, regresyon, kümeleme, boyutsal azaltma, model seçimi için yardım araçları ve veri ön işleme araçları içeren güçlü bir makine öğrenmesi kütüphanesidir. Sınıflandırmayı, objelerin hangi sınıflara ait olduğunun belirlemesi olarak tanımlayan scikit-learn, kullanım alanlarına örnek olarak istenmeyen (spam) elektronik postaların tespiti ve fotoğraf tanımlama verilmiştir. Regresyonu, objelerin sürekli değişkenlerine ait değerlerin tahmini olarak tanımlayan scikit-learn, hastaların ilaçlara verdikleri/verecekleri tepkinin tahmini ve hisse senetlerinin fiyatlandırılmasının tahmini gibi alanlarda kullanılmaktadır. Kümeleme ve boyutsal azaltma, denetimsiz öğrenme metodu olarak, müşteri bölümlenme, deney sonuçlarını gruplandırma veya belirli özelliklere göre bölümlenme için kullanılmaktadır. Model seçimi araçları ise, makine öğrenmesi algoritmalarında modellerde öznelilikler üzerinde ayarlama (tunning) yapılmasına ve model performansının artırılmasına olanak sağlamaktadır. Veri ön

işleme, metin gibi ham girdilerden belirli özelliklerin çıkarılması ve çıkarılan verilerin makine öğrenmesi algoritmalarına öznitelik olarak kullanılmasına olanak sağlamaktadır (Buitinck vd., 2013).

BÖLÜM 3: MODEL, EĞİTİM ve ANALİZ

3.1. Çalışma Modeli, Öznitelik Seçimi, Veri Alma Süreçleri

Bu çalışma için, Hepsiburada isimli alışveriş sitesi üzerinden Xiaomi Redmi Basic 2 TWS Bluetooth 5.0 adlı ürün seçilmiştir. Ürün, sınıflandırma olarak giriş seviyesi ürün olup hem erkek hem kadın tüketicilere hitap etmesi, alışveriş sitesinde en çok değerlendirilmiş ürünlerden birisi olması ve 5 yıldız derecelendirmesine sahip tüketici değerlendirmelerinin sayısının, 1 yıldız derecelendirmesine sahip tüketici değerlendirmelerinin sayısına oranının diğer ürünlere kıyasla daha yüksek olmasından ötürü bu ürün tercih edilmiştir. Ürüne ait Temmuz 2019 – Mart 2022 tarihinde yazılmış olan tüketici değerlendirmeleri alınmıştır. Sadece ürün yıldız derecelendirmeleri, değerlendirme metni ve faydalı bulunma sayısı veri setine dahil edilmiştir.

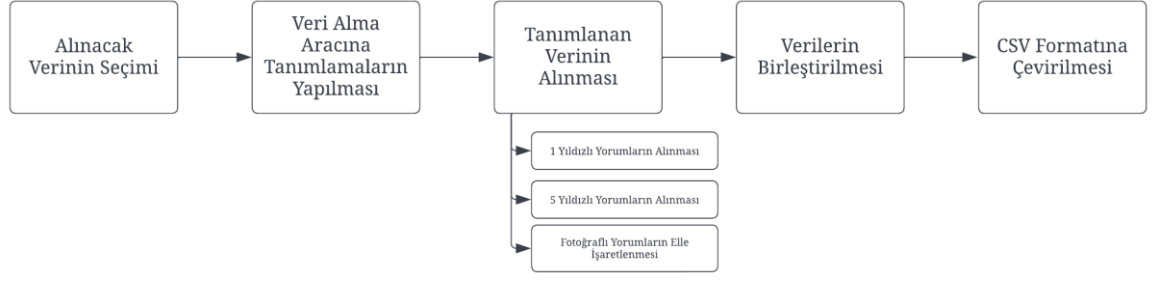
3.1.1. Çalışma Modeli ve Uygulama

3.1.1.1. Çalışma Modeli



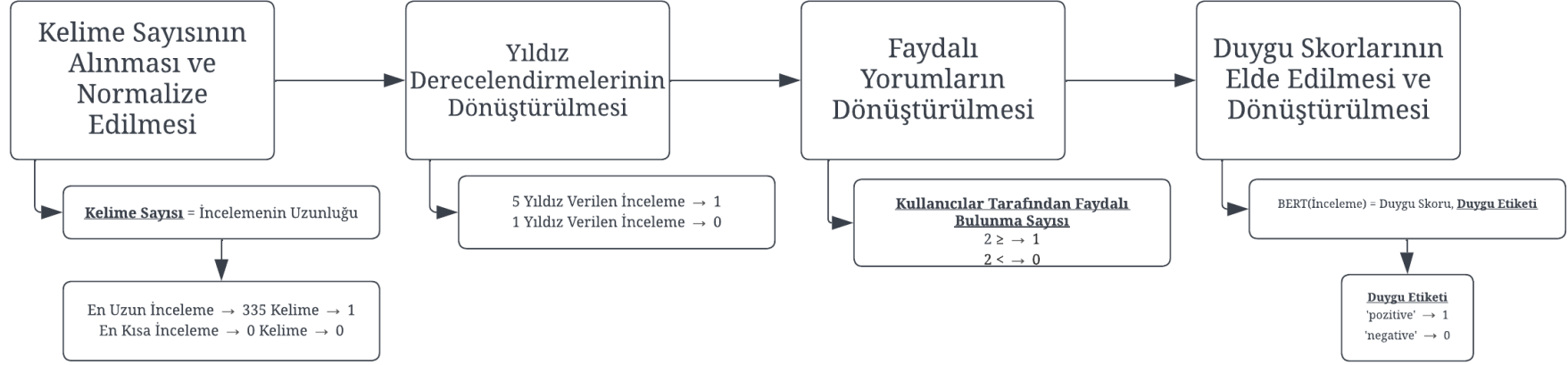
Şekil 30: Çalışma Modeli

Çalışma modeli, veri elde etme, veri ön işleme, modeller için rastgele veri seçimi, veri setlerinin eğitimi ve değerlendirme olarak planlanmıştır. Veri elde etme sürecinde, çevrimiçi alışveriş platformundan inceleme metni içeren tüketici değerlendirmeleri alınacaktır. Alınan tüketici değerlendirmeleri CSV (Virgüller Ayrılmış Değerler) dosya formatına dönüştürülerek veri alma süreci sonlandırılacaktır.



Şekil 31: Veri Edinme Süreci

Alınan incelemelerden oluşan veri seti, veri ön işleme sürecinden geçilerek makine öğrenimi modelleri için hazır hale getirilecektir.



Şekil 32: Veri Ön İşleme Süreci

Veri ön işleme sürecinde elde edilen makine öğrenimine hazır seti, rastgele biçimde ayırarak “eğitim” ve “test” setleri oluşturulacaktır. Oluşturulan bu setler üzerinde algoritma “eğitim” setiyle eğitilip, daha önce hiç karşılaşmadığı veriler olan “test” setleri üzerinden başarıları değerlendirilecektir.

3.1.1.2. Öznitelik Seçimi

Tablo 5’de literatürde tüketici değerlendirmeleri üzerinde çalışma yapan makalelerde kullanılan öznitelikler işaretlenmiştir. Bu çalışma için kullanılacak öznitelikler, birden fazla literatür çalışmasının incelenmesiyle belirlenmiştir.

Tablo 5: Literatür Taraması

Çalışma	Duygu Etiketi	İnceleme Metni Uzunluğu	Yıldız Derecelendirmesi	Fotoğraflı Değerlendirme
Turney, 2002	✓		✓	
Alsharif, 2022	✓	✓	✓	
Hmoud Al-Adhaileh ve Waselallah Alsaade, 2022	✓	✓	✓	
Kumar vd., 2022	✓	✓	✓	
Pelsmacker vd., 2018	✓	✓	✓	✓
Li, 2021		✓	✓	✓
Lai vd., 2021	✓	✓	✓	✓
Krestel ve Dokoohaki, 2015	✓	✓	✓	

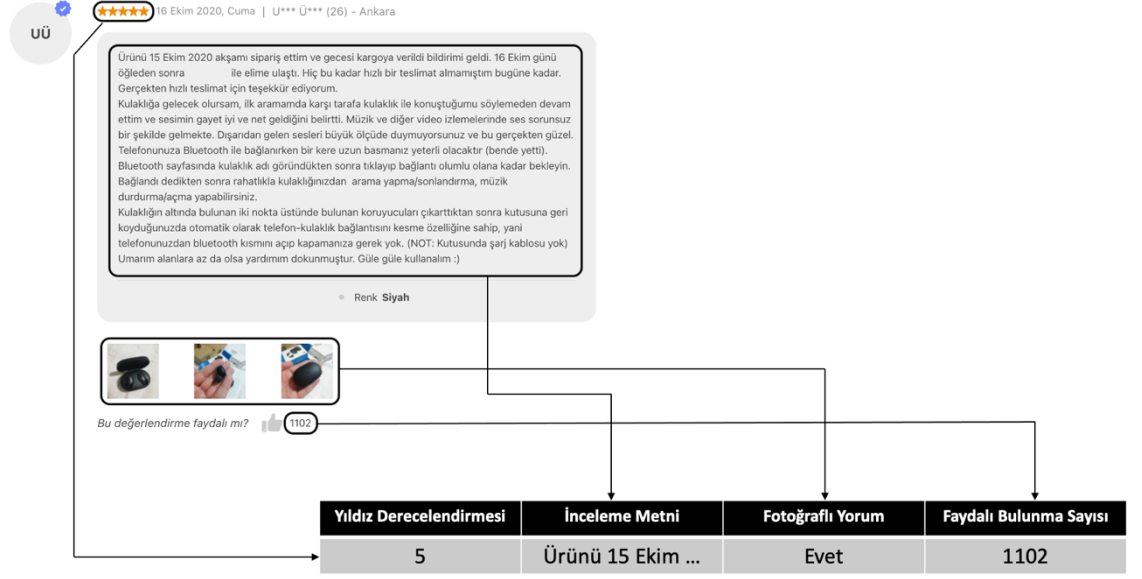
3.1.1.3. Veri Elde Etme

Algoritma eğitmek ve analiz etmek için kullanılacak veri, tüketiciler tarafından oluşturulmuş ürün incelemelerinden oluşturulacaktır. Kullanıcıların ürün için belirledikleri yıldız derecelendirmeleri sayılarının sayıca yakın olduğu ürünlerden birisi eğitim setine alınmak üzere seçilmiştir. Seçilen üründe, kullanıcıların ürün için oluşturdukları incelemelere verdikleri yıldız derecelendirmelerinin dağılımı Tablo 6’daki gibidir.

Tablo 6: Yorumların Yıldız Derecelendirmelerine Göre Dağılımı

Yıldız Derecelendirmesi	Yorum Sayısı	Toplam Yorum İçindeki Oranı
5	8675	%57,65
4	2526	%16,78
3	1245	%8,27
2	800	%5,31
1	1800	%11,96
Toplam	15046	%100

Seçilen ürünün, ürün inceleme sayfasında bulunan tüketici değerlendirmeleri, WebScrapier.io adlı Google Chrome Uzantısı aracılığı alınacaktır. Alınacak veriler ve bu verilerin veri setindeki temsil edilmiş biçimi Şekil 33'te gösterilmiştir. Verilerin elde edileceği çevrimiçi alışveriş sitesi, tüketici değerlendirmelerini nümerik olarak değil de görsel olarak ifade ettiği için veriler 3 parça halinde elde edilecektir. Tüketici değerlendirmeleri öncelikle “Sadece 5 Yıldız Derecelendirmesine Sahip Tüketici Değerlendirmeleri” şeklinde filtrelenerek, gösterilen veriler uzantı aracılığıyla alınıp Excel üzerinden, elde edilen verilere “Yıldız Derecelendirmesi” sütunu eklenecek ve bu filtreleme sonucu elde edilen tüm verilerin yıldız derecelendirmeleri “5” olarak girilecektir. Aynı işlem “Sadece 1 Yıldız Derecelendirmesine Sahip Tüketici Değerlendirmeleri” şeklinde uygulanan filtre ile gerçekleştirilecektir. Söz konusu tüketici değerlendirmelerinde herhangi bir inceleme metni içermeyen, sadece yıldız derecelendirmesinden oluşan değerlendirmeler de bulunmaktadır. Çevrimiçi alışveriş sitesinin sağladığı filtreleme özelliklerinin içerisinde “Fotoğraflı Değerlendirmeler” şeklinde bir filtre bulunmamakla birlikte, “Fotoğraflı Değerlendirmeler” için başka bir sayfaya yönlendirmektedir. Bu yönlendirilen sayfada geçen fotoğraflı değerlendirmeler, daha önce elde edilen 5 yıldız derecelendirmesine ve 1 yıldız derecelendirmesine sahip tüketici değerlendirmelerinde bulunup işaretlenecektir. Veri seti, sadece inceleme metni içeren tüketici değerlendirmelerinden oluşacaktır.



Şekil 33: Öznitelik Çıkarımı

Şekil 33, veri ön işleme sürecine girecek tüketici incelemelerinin veri setine nasıl yansıtıldığını göstermektedir. Yıldız derecelendirmesi incelemelerin yayımlanmış olduğu çevrimiçi alışveriş sitesinde görsel olarak ifade edilmesine karşın veri setine yıldız derecelendirmesi nümerik olarak eklenmiştir. İnceleme metni, herhangi bir işlemten geçmeden direkt veri setine eklenmiştir. İncelemenin fotoğraflı bir inceleme olup olmadığı, durumuna göre “Evet” ya da “Hayır” şeklinde veri setine eklenmiştir. Faydalı bulunma sayısı, herhangi bir işlemten geçmeden veri setine eklenmiştir. İnceleme metni içeren 5 yıldız derecelendirmesine ve 1 yıldız derecelendirmesine sahip yorumlardan oluşan veri setinin son hali Tablo 7’deki gibidir.

Tablo 7: 1 ve 5 Yıldız Derecelendirmeli Yorum Sayısı

Yıldız Derecelendirmesi	Yorum Sayısı	Toplam Yorum İçindeki Oranı
5	3749	%78,84
1	1006	%21,16
Toplam	4755	%100

Toplam incelemeler içerisinde 5 yıldız verilen yorum sayısı 8675 olmasına karşın, kullanıcıların inceleme metni yazdığı yorum sayısı 3749, oranı ise %43,21’dir. Toplam değerlendirme içerisinde yorum sayısı 1800 olmasına karşın, kullanıcıların inceleme metni yazdığı yorum sayısı 1006, oranı ise %55,88’dir. İnceleme metni içermeyen

yorumlar sadece yıldız derecelendirmesinden oluşmaktadır. *Yıldız Derecelendirmesi*, *İnceleme Metni*, *Fotoğraflı Yorum* ve *Faydalı Bulunma Sayısı* sütunlarından oluşan veri seti CSV formatına dönüştürülerek veri elde etme süreci sonuçlandırılır.

Veri ön işleme öncesinde elde edilen verilerin son hali Tablo 8’de verilmiştir.

Tablo 8: Nitelik Özellikleri

Sütun Adı	Veri Tipi	Veri Sayısı	Minimum Değer	Maksimum Değer	Standart Sapma	Ortalama Değer
İnceleme Metni Uzunluğu	integer	4755	1	336	20,99	17,5
Faydalı İşaretlenme	integer	4755	0	1097	20,27	1,14
Yıldız Derecelendirmesi	integer	4755	1	5	1,63	4,15
Fotoğraflı Yorum*	integer	4755	0	1	0,07	0,005

Fotoğraf içeren yorumlar 1, içermeyen yorumlar 0 ile kodlanmıştır.

Tüketiciler tarafından 2 veya daha fazla defa “Faydalı” işaretlenen tüketici değerlendirmelerinin genel özellikleri Tablo 9’da gösterilmiştir.

Tablo 9: İki veya Daha Fazla Defa Faydalı İşaretlenen Değerlendirmelerin Özellikleri

Sütun Adı	Veri Sayısı	Minimum Değer	Maksimum Değer	Standart Sapma	Ortalama Değer
İnceleme Metni Uzunluğu	267	1	336	46,51	43,85
Yıldız Derecelendirmesi	267	1	5	1,85	3,75
Fotoğraflı Yorum*	267	0	1	0,27	0,08

Tüketiciler tarafından 1 kez “Faydalı” olarak işaretlenen ya da hiç faydalı işaretlenmeyen tüketici değerlendirmelerinin genel özellikleri Tablo 10’da gösterilmiştir.

Tablo 10: Bir Kez Faydalı İşaretlenen Değerlendirmelerin Özellikleri

Sütun Adı	Veri Sayısı	Minimum Değer	Maksimum Değer	Standart Sapma	Ortalama Değer
İnceleme Metni Uzunluğu	4488	1	248	17,18	15,93
Yıldız Derecelendirmesi	4488	1	5	1,61	4,17
Fotoğraflı Yorum*	4488	0	1	0,0004	0,021

3.2. Veri Ön İşleme

Veri ön işleme sürecinde, Microsoft Visual Studio Code ve Python 3.10 ortamı kullanılarak elde edilen veriler pandas kütüphanesinin sağladığı fonksiyonlar sayesinde algoritmaların eğitimine uygun olacak şekilde dönüştürülecektir. Şekil 32’de gösterildiği şekilde,

- İnceleme Metni sütununda bulunan inceleme metinlerinden, metinlerin içerdiği kelime sayısı nümerik olarak çıkartılmıştır. Elde edilen kelime sayılarının

maksimum değeri 336 iken, minimum değeri 0'dır. Kelime sayıları normalize edilerek öğrenme setine eklenmiştir. Normalize etmek için, kullanılan formül;

$$\circ \text{ (Mevcut Değer – En Düşük Değer) / (En Yüksek Değer – En Düşük Değer)}$$

- Normalize edilmeden önce veri aralığı 1-336 arasında iken, normalize edildikten sonra veri seti 0 ile 1 arasındadır.
- 5 ve 1 olarak elde edilen yıldız derecelendirmeleri, 5 için 1, 1 için 0 olacak şekilde dönüştürülerek öğrenme setine eklenmiştir.
- Kullanıcıların, kendi yorumlarını da faydalı bulabilmelerinden ötürü, Tüketiciler tarafından, iki kere veya daha fazla kez faydalı bulunan yorumlar 1, 1 kez faydalı bulunan ya da hiç faydalı olarak işaretlenmeyen yorumlar ise 0 olarak şekilde dönüştürülerek öğrenme setine eklenmiştir. Böylece, tüketici kendi yorumunu faydalı bulmuş olsa bile, faydalı bulunma sayısı iki veya daha fazla olduğunda, başka bir tüketici tarafından faydalı bulunduğu öngörülebilir.
- İnceleme metinlerinin duygu etiketlerini çıkartmak için BERT üzerinden daha önce eğitilmiş model olan “savasy/bert-base-turkish-sentiment-cased” isimli model kullanılmıştır. Model, hem duygu etiketi (“Pozitif”, “Negatif”) hem de bu etikete ne kadar uyduğunu gösteren, 0 ile 1 arasında değişen bir değer döndürür. Modelin metin için vermiş olduğu duygu etiketleri, “Pozitif” için 1, “Negatif” için 0 olacak şekilde dönüştürülerek öğrenme setine eklenmiştir.

Öğrenme seti, yapılan ön işleme sonucunda Tablo 11’de görüldüğü gibi son halini almıştır.

Tablo 11: Özniteliklerin Özellikleri

Özniteliğin Adı	Veri Tipi	Veri Sayısı	Minimum Değer	Maksimum Değer
duygu_skoru	integer	4755	0	1
kelime_sayısı	integer	4755	0	1
yıldız_derecelendirmesi	integer	4755	0	1
fotoğraflı_yorum	integer	4755	0	1
faydalı_yorum	integer	4755	0	1

3.3. Model Eğitimi için Rastgele Veri Seçimi

Kullanılacak olan Logistik Regresyon, Naive Bayes, Karar Ağacı, Rassal Orman, Gradyan Arrtırma ve XGBoost algoritmalarının eğitimleri, oluşturulan öğrenme seti

üzerinden gerçekleştirilecektir. Oluşturulan öğrenme setini, sklearn kütüphanesinin train-test-split fonksiyonu aracılığıyla “X_öğrenme”, “y_öğrenme”, “X_test” ve “y_test” olarak veri setinin %75’i öğrenme için, %25’i test için olacak şekilde ayrılacaktır. X ile ifade edilen set içerisinde “duygu_skoru”, “kelime_sayısı”, “yıldız_derecelendirmesi” ve “fotoğraflı_yorum” sütunları bulunurken, Y ile ifade edile set sadece “faydalı_yorum” değerlerini içermektedir. Algoritmaların, verilen X_öğrenme setindeki öznitelikler ile y_öğrenme seti arasındaki ilişkiyi anlayıp, daha önce hiç görmediği test setinde verilen X değerlerine göre y değerlerini tutarlı şekilde tahmin etmesi beklenecektir.

3.4. Modellerin Eğitimi

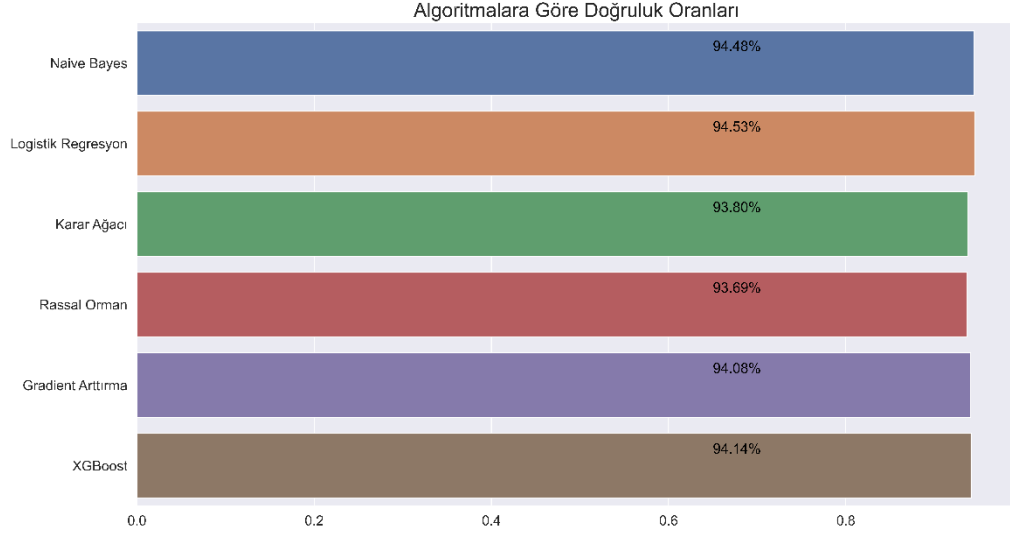
3.4.1. Model Eğitimleri ve Değerlendirme

Naive Bayes, Logistik Regresyon, Karar Ağacı, Rassal Orman, Gradyan Arttırma ve XGBoost algoritmaları oluşturulmuş eğitim seti ile eğitilerek bu modellerin başarıları karşılaştırılacaktır. Öğrenme setinde bulunan veriler, dengesiz biçimde dağılmaktadır. Toplam 4755 tüketici değerlendirmesi içerisinde, 267 adet tüketici değerlendirmesi “Faydalı Yorum” olarak etiketlendiği için, her “Faydalı Yorum” olarak etiketlenen tüketici değerlendirmesine karşılık 9 adet “Faydalı Yorum Değil” olarak etiketlenen tüketici değerlendirmesi bulunmaktadır.

Tablo 12: Faydalı Faydalı Değil Etiket Dağılımı

Etiket	Veri Sayısı	Oranı
Faydalı	267	%5,61
Faydalı Değil	4488	%94,39
Toplam	4755	%100

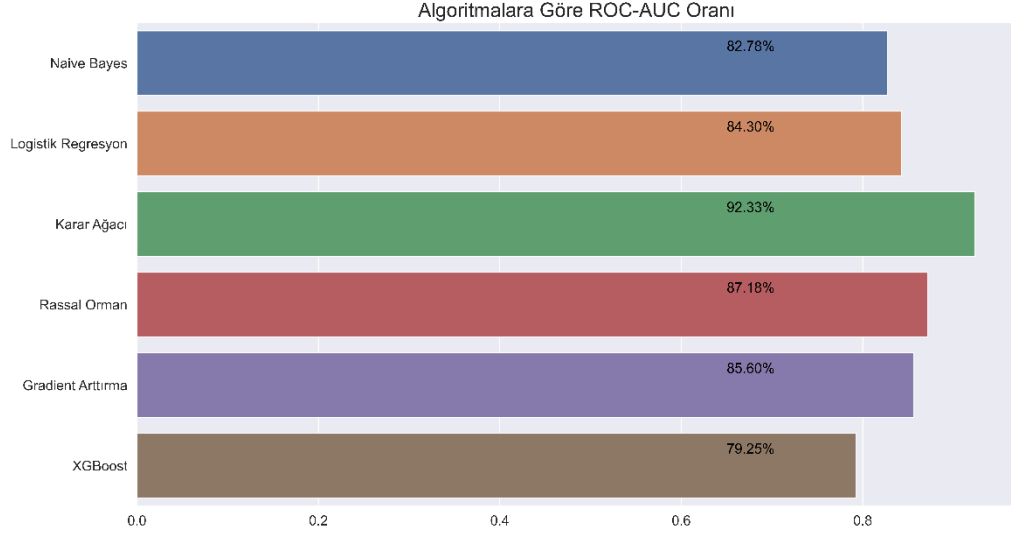
Algoritmalar eğitim setleri ile eğitildikten sonra daha önce hiç görmediği test seti üzerindeki doğruluk oranlarına göre başarı yüzdeleri Şekil 34’te gösterilmiştir.



Şekil 34: Algoritmalarla G6re Doęruluk Oranları

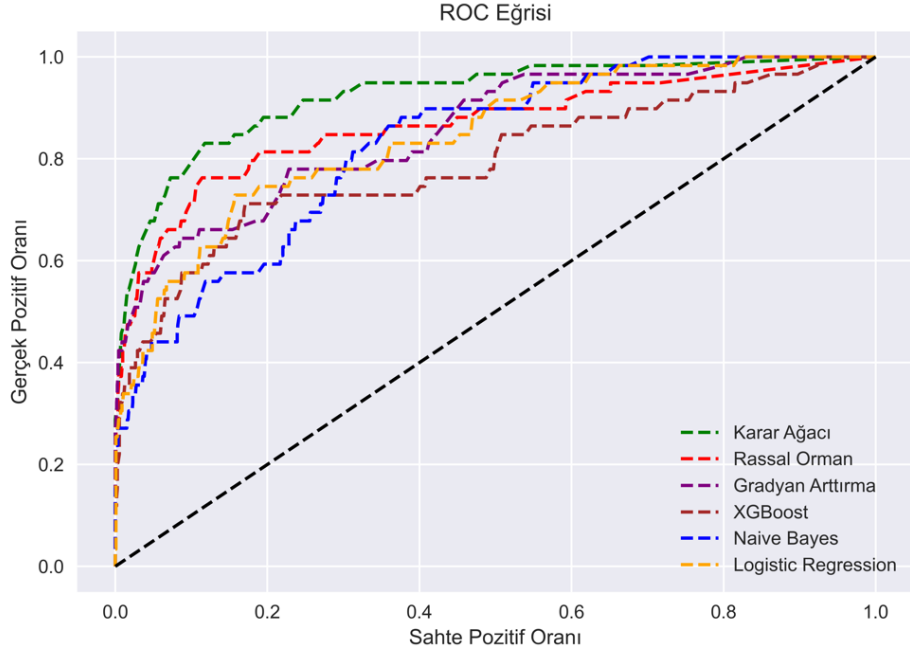
Doęruluk oranlarına g6re aęaç yapısına sahip Karar Aęacı, Rassal Orman, Gradyan Arttırma ve XGBoost sınıflandırma algoritmaları arasında, %94,14 ile XGBoost en y6ksek doęruluk oranına sahip olmasına karřın en d6ř6k doęruluk oranı %93,69 ile Rassal Orman algoritmasıdır. %94,53 ile Logistik Regresyon en y6ksek deęere sahip algoritma olmasına karřın en d6ř6k deęer %93,69 ile Gradyan Arttırma algoritmasına aittir. Doęruluk, algoritmanın yapmıř olduęu tahminlerin ne kadarının doęru olduęunu ifade eden oranı vermektedir. 6ęrenme seti dengesiz olduęu ve ‐Faydalı Deęil‐ etiketli yorumlar veri setinin %94,31’ini oluřturduęu iin, algoritmaların b6t6n t6keticilerde deęerlendirmelerini ‐Faydalı Deęil‐ olarak tahmin ettiklerinde elde edecekleri doęruluk oranı %94,31’dir.

ROC-AUC (Alıcı İřletim Karakteristięi – Eęrinin Altındaki Alan) oranları iin algoritmaların bařarı oranları Şekil 35’teki gibidir.



Şekil 35: Algoritmalarla G6re ROC-AUC Oranı

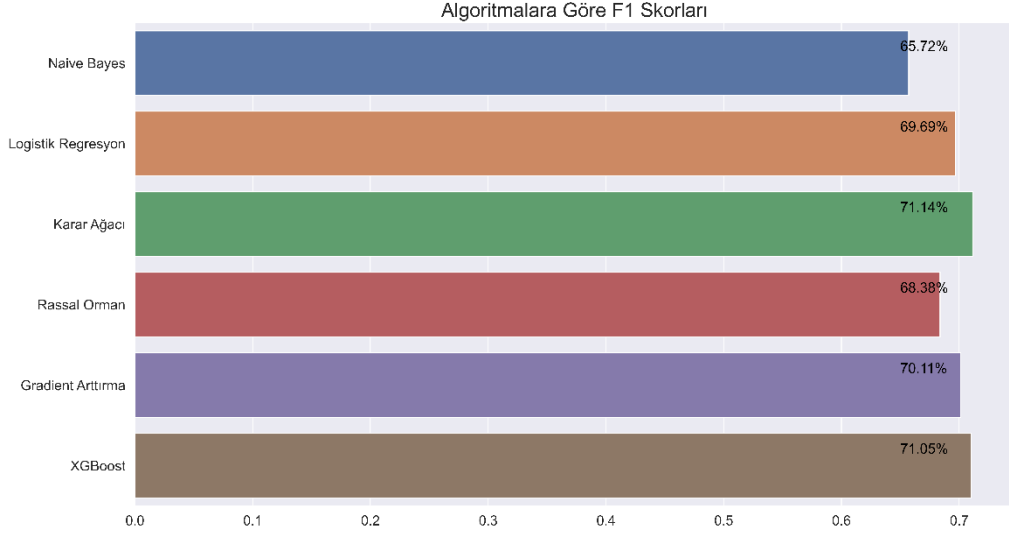
ROC – AUC oranı algoritmaların, test setinde bulunan deęerlendirmeleri “Faydalı” ve “Faydalı Deęil” olan etiketlerini ne kadar doęru bir şekilde tahmin edebildięini g6stermektedir. Aęaę yapısına sahip algoritmalar ierisinde en y6ksek ROC – AUC oranına sahip algoritma %92,33 ile Karar Aęacıdır. %79,25 ile XGBoost algoritması t6m algoritmalar arasında ve aęaę algoritmaları iinden d6ş6k ROC – AUC oranına sahip algoritmadır. Algoritmaların ROC – AUC oranlarına g6re g6sterimleri Şekil 36’daki gibidir.



Şekil 36: ROC Eğrisi

Algoritmaların sınıflandırma başarılarının artmasıyla birlikte, algoritmaya ait eğrilerin kapsadıkları alanlar da artış göstermektedir. Şekil 36'daki oranlara bağlı olarak, en çok alan kapsayan eğri Karar Ağacı algoritması, en az alan kapsayan algoritma ise XGBoost algoritmasıdır.

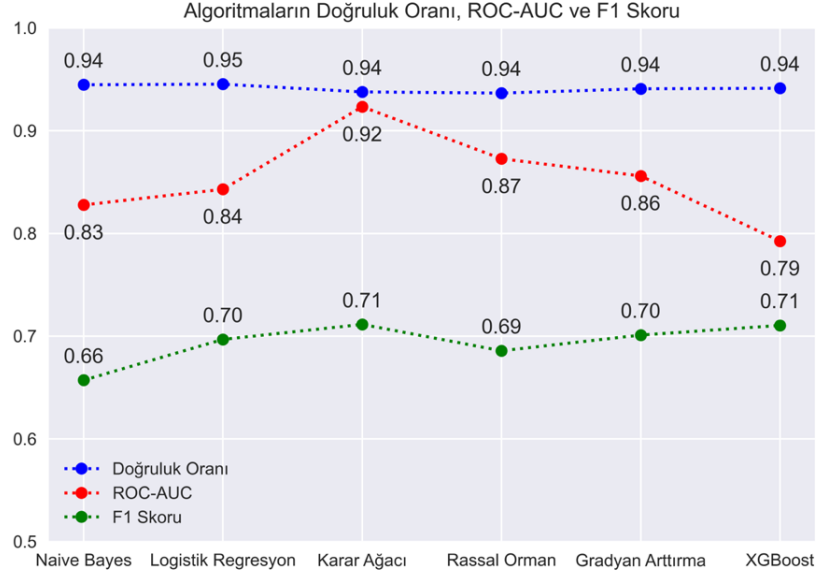
Algoritmalara ait F1 skor değerleri Şekil 37'deki gibidir.



Şekil 37: Algoritmalarla Gre F1 Skorları

F1 skoru aısından, Karar Ađacı ve XGBoost algoritmaları en yksek deęere sahip olmasına karřın, ROC – AUC deęerlerine gre en yksek bařarıya sahip algoritma Karar Ađacı olarak grnmektedir. Ađa yapısına sahip algoritmalar arasında en dřk ROC – AUC deęerlerine sahip algoritma XGBoost algoritmasına raęmen, en yksek F1 deęerine sahip iki algoritmadan birisi XGBoost algoritmasıdır. Algoritmaların sınıflandırma konusunda bařarısının algoritma bazlı incelenbilmesi iin her algoritmanın yapmış olduęu tahminlerin deęerlendirileceęi hata matrisine bakılmalıdır. Algoritmaların yapmış oldukları tahminlerin doęru sınıfa ait olması konusunda bařarısını gsterecektir.

Algoritmaların eęitimi elde edilen veri seti, tek bir rn iin yazılmış tketiciler deęerlendirmelerden elde edilmiştir. Elde edilen deęerlendirme zerinde algoritmaların gstermiş olduęu genel bařarı Şekil 38’de gsterilmiştir.

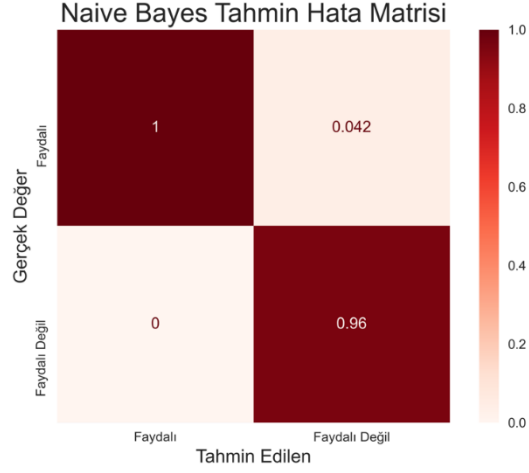


Şekil 38: Algoritmaların Doğruluk Oranı, ROC-AUC ve F1 Skoru

Eğitilen algoritmalar en az %94 doğruluk, %79 ROC eğrisi altında kalan alan oranı, %66 F1 skoru oranı almıştır. Algoritmaların göstermiş olduğu başarılar, algoritmanın eğitildiği değerlendirme setinin elde edildiği yerden alınan, ancak algoritmaların eğitiminde hiç kullanılmamış veri seti üstünde olmuştur. Algoritmalar, temel karar verme mekanizmaları ve yapılarının farklılıklarıyla birlikte ortalama %94 Doğruluk, %85 ROC eğrisi altında kalan alan oranı ve %69 F1 skoruna sahiptir. Genel oran üzerinde eğitim seti üzerinde en yüksek öğrenmeyi gösteren algoritma Karar Ağacı algoritması olmasına karşın, en düşük sonucu F1 skorunun oldukça düşük olması sebebiyle Naive Bayes algoritması vermiştir. ROC eğrisi altında kalan alan oranı ve F1 skoru, algoritmanın yapmış olduğu tahminlerin ne kadar doğru olduğunu ortaya koymaktadır. Eğer bir algoritma, bütün değerlendirmeleri “ faydalı değil” olarak tahmin etseydi, ulaşacağı doğruluk değeri %94 olurdu. Çünkü veri setindeki değerlendirmelerin %94’ü “ faydalı değil” etiketine sahiptir sadece %6’sı “ faydalı” etiketine sahiptir. Bu sebepten ötürü algoritmaların tahmin başarısının ölçülmesi için doğruluk tek başına yeterli değildir. Bahsi geçen senaryoya göre, doğruluk %94 olmasına karşın, ROC eğrisi altında kalan alan oranı ve F1 skoru %0 olarak karşımıza çıkacaktır. Algoritmaların eğitim sonucunda göstermiş oldukları başarı Şekil 38’de görünmektedir.

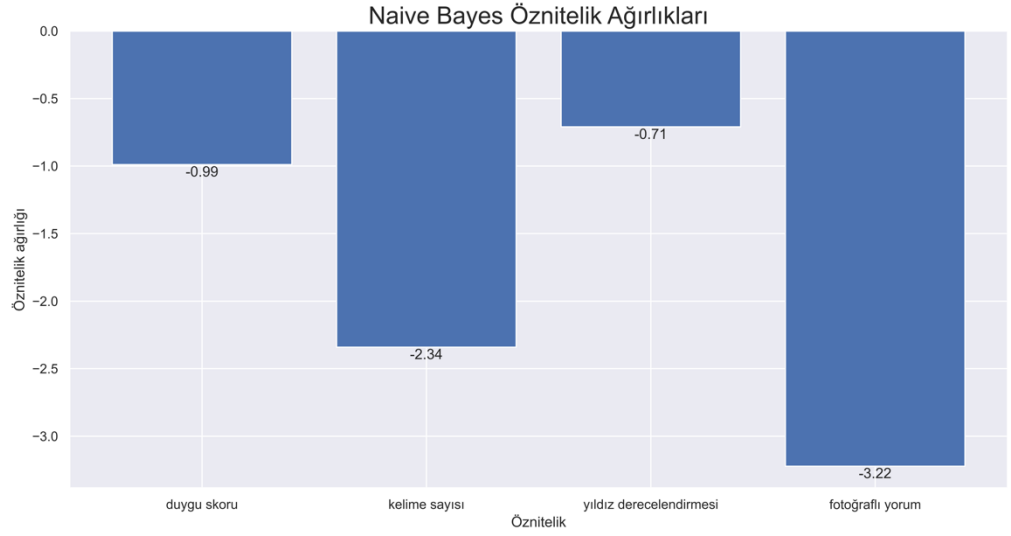
3.4.1.1. Naive Bayes Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

Naive Bayes algoritmasının doğruluk oranı %94,48, ROC eğrisinin altında kalan alan oranı %82,78 ve F1 skoru ise %65,72'dir. Algoritmanın hata matrisi Şekil 39'daki gibidir.



Şekil 39: Naive Bayes Tahmin Hata Matrisi

Naive Bayes algoritmasının, hiç görmediği test setindeki (algoritmanın eğitimi sırasında kullanılmayan set) tüketici değerlendirmeleri üzerinde yaptığı bütün “Faydalı” tahminleri doğru, yapmış olduğu “Faydalı Değil” tahminlerinin ise %96’sı doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin hiçbiri aslında “Faydalı Değil” olmamasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %4,2’si “Faydalı” etiketine sahiptir.

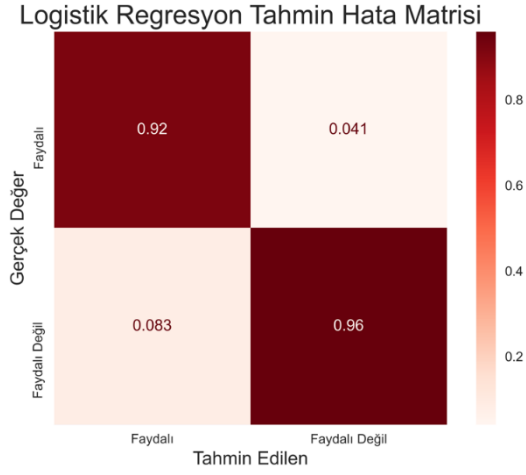


Şekil 40: Naive Bayes Öznitelik Ağırlıkları

Naive Bayes algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 40'taki gibidir. En yüksek ağırlığa 3,22 öznitelik katsayısıyla incelemenin fotoğraflı olma özneliği sahipken, en düşük öznitelik katsayısı olan 0,71'e incelemeye ait yıldız derecelendirmesi özneliği sahiptir.

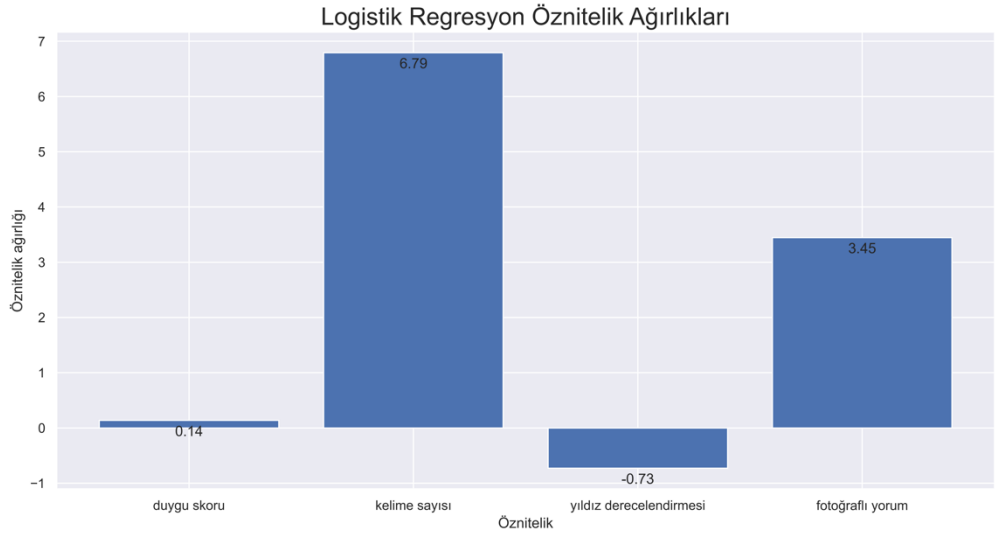
3.4.1.2. Logistik Regresyon Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

Logistik Regresyon algoritmasının doğruluk oranı %94,53, ROC eğrisinin altında kalan alan oranı %84,30 ve F1 skoru ise %69,69'dur. Algoritmanın hata matrisi Şekil 41'deki gibidir.



Şekil 41: Logistik Regresyon Tahmin Hata Matrisi

Logistik Regresyon algoritmasının, test setindeki tüketici değerlendirmeleri üzerinde yaptığı “Faydalı” tahminlerinin %96’sı, “Faydalı Değil” tahminlerinin ise %92’si doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin %8,3’ünün “Faydalı Değil” etiketine sahip olmasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %4,1’si “Faydalı” etiketine sahiptir.



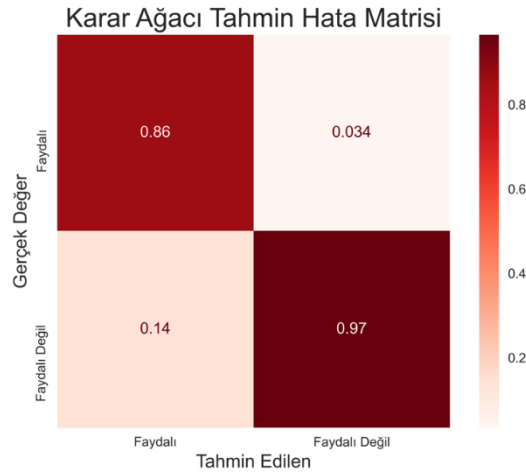
Şekil 42: Logistik Regresyon Öznitelik Ağırlıkları

Logistik Regresyon algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 42’deki gibidir. En yüksek ağırlığa 6,79 öznitelik katsayısıyla

inceleme metnin içerdiği kelime sayısı özniteliği sahipken, en düşük öznitelik katsayısı olan 0,14'e incelemeye duygu skoru (duygu etiketi) özniteliği sahiptir.

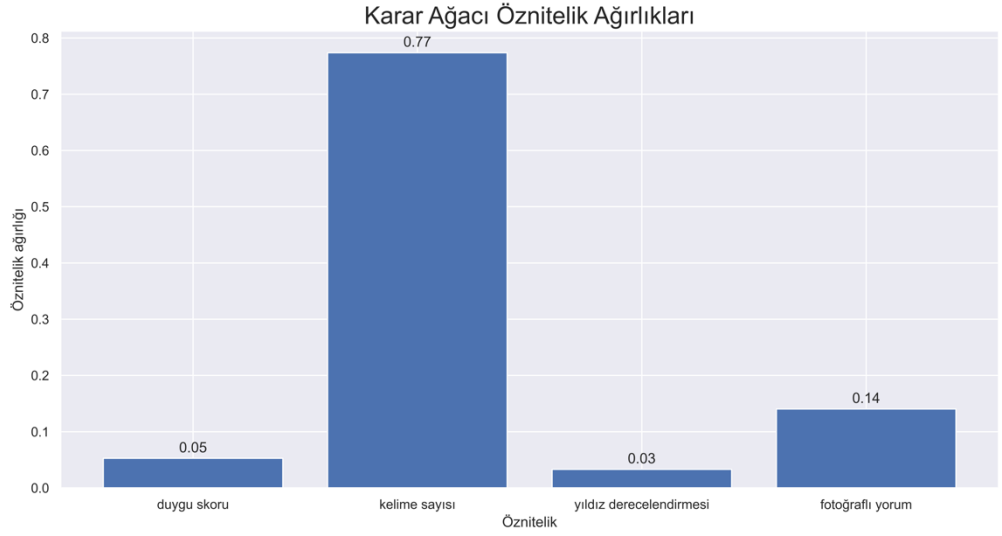
3.4.1.3. Karar Ağacı Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

Karar Ağacı algoritmasının doğruluk oranı %93,80, ROC eğrisinin altında kalan alan oranı %92,33 ve F1 skoru ise %71,14'tür. Algoritmanın hata matrisi Şekil 43'teki gibidir.



Şekil 43: Karar Ağacı Tahmin Hata Matrisi

Karar Ağacı algoritmasının, test setindeki tüketici değerlendirmeleri üzerinde yaptığı “Faydalı” tahminlerinin %97’si, “Faydalı Değil” tahminlerinin ise %86’sı doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin %14’ünün “Faydalı Değil” etiketine sahip olmasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %3,4’ü “Faydalı” etiketine sahiptir.

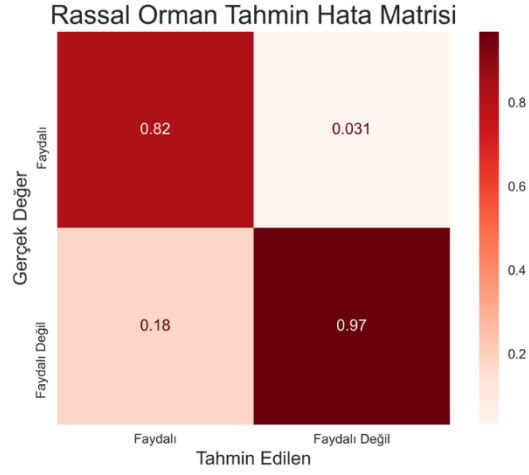


Şekil 44: Karar Ağacı Öznitelik Ağırlıkları

Karar Ağacı algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 44'teki gibidir. En yüksek ağırlığa 0,77 öznitelik katsayısıyla inceleme metnin içerdiği kelime sayısı özniteliği sahipken, en düşük öznitelik katsayısı olan 0,03 incelemeye ait yıldız derecelendirmesi özniteliği sahiptir.

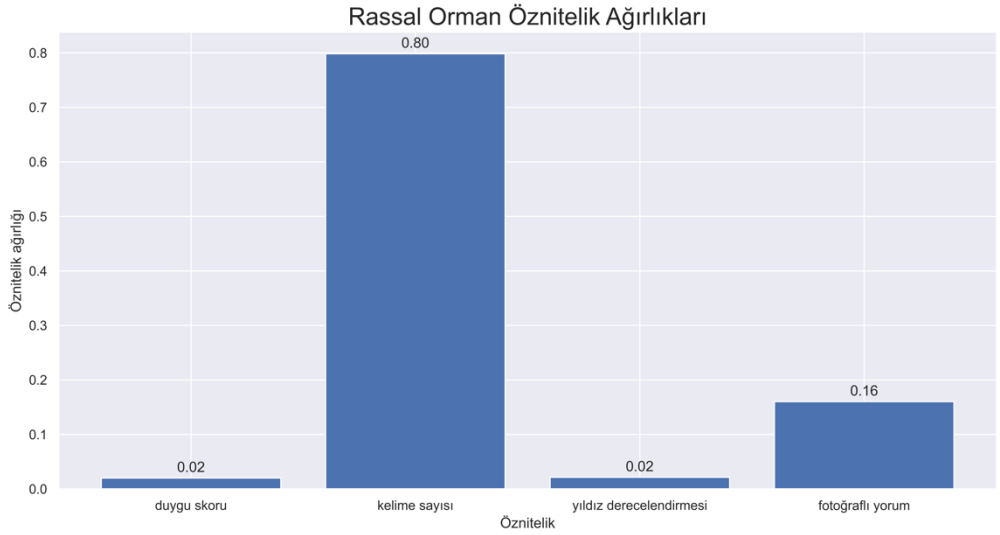
3.4.1.4. Rassal Orman Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

Rassal Orman algoritmasının doğruluk oranı %93,69, ROC eğrisinin altında kalan alan oranı %87,27 ve F1 skoru ise %68,58'dir. Algoritmanın hata matrisi Şekil 45'teki gibidir.



Şekil 45: Rassal Orman Tahmin Hata Matrisi

Rassal Orman algoritmasının, test setindeki tüketici değerlendirmeleri üzerinde yaptığı “Faydalı” tahminlerinin %97’si, “Faydalı Değil” tahminlerinin ise %82’si doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin %18’inin “Faydalı Değil” etiketine sahip olmasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %3,1’i “Faydalı” etiketine sahiptir.



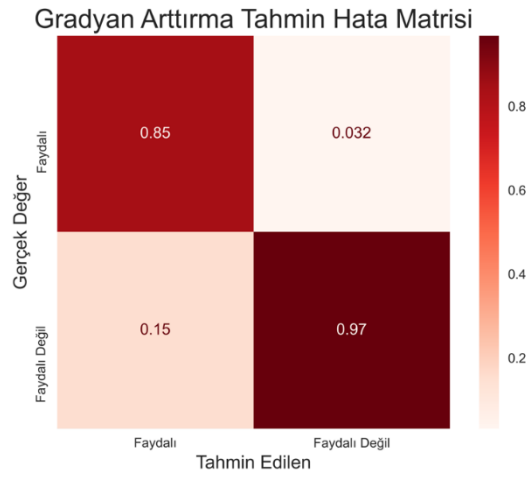
Şekil 46: Rassal Orman Öznitelik Ağırlıkları

Rassal Orman algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 46’daki gibidir. En yüksek ağırlığa 0,80 öznitelik katsayısıyla inceleme metnin içerdiği kelime sayısı özniteliği sahipken, en düşük öznitelik katsayısı

olan 0,02 incelemeye ait yıldız derecelendirmesi ve incelemeye ait duygu skoru (duygu etiketi) özniteliği sahiptir.

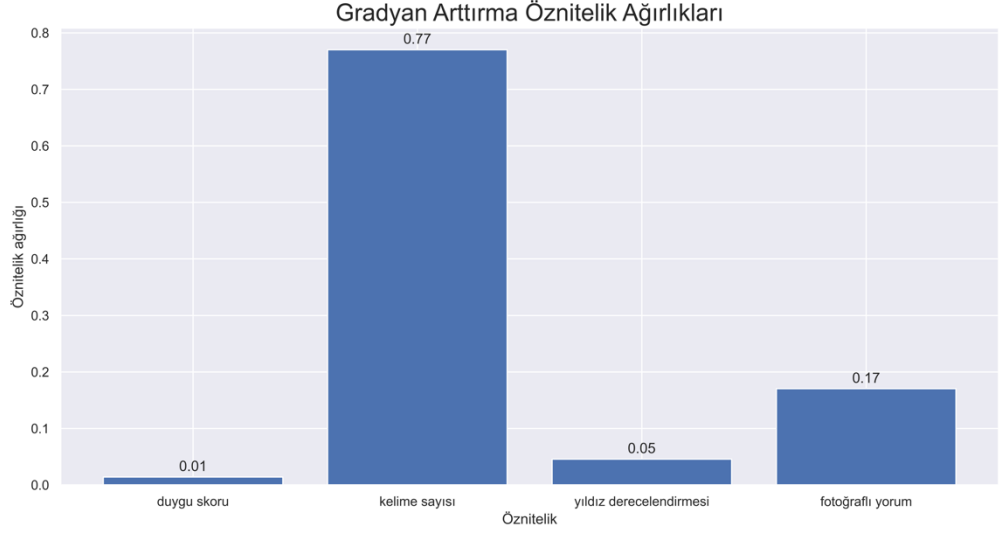
3.4.1.5. Gradyan Arttırma Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

Gradyan Arttırma algoritmasının doğruluk oranı %94,08, ROC eğrisinin altında kalan alan oranı %85,60 ve F1 skoru ise %70,11'dir. Algoritmanın hata matrisi Şekil 47'deki gibidir.



Şekil 47: Gradyan Arttırma Tahmin Hata Matrisi

Gradyan Arttırma algoritmasının, test setindeki tüketici değerlendirmeleri üzerinde yaptığı “Faydalı” tahminlerinin %97’si, “Faydalı Değil” tahminlerinin ise %85’i doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin %15’inin “Faydalı Değil” etiketine sahip olmasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %3,2’si “Faydalı” etiketine sahiptir.

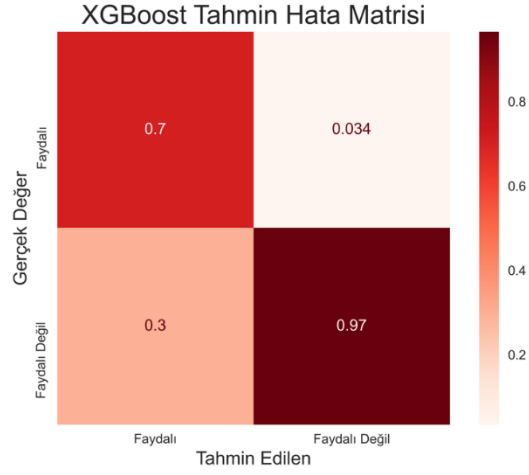


Şekil 48: Gradyan Arttırma Öznitelik Ağırlıkları

Gradyan Arttırma algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 48'deki gibidir. En yüksek ağırlığa 0,77 öznitelik katsayısıyla inceleme metnin içerdiği kelime sayısı özniteliği sahipken, en düşük öznitelik katsayısı olan 0,01 incelemeye ait duygu skoru (duygu etiketi) özniteliği sahiptir.

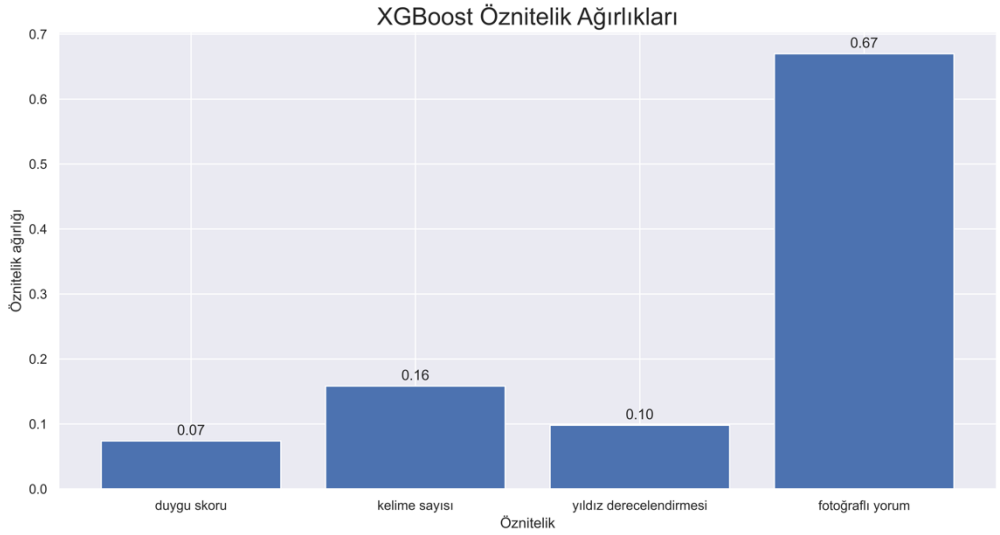
3.4.1.6. XGBoost Tahmin Hata Matrisi Değerlendirmesi ve Öznitelik Ağırlıkları

XGBoost algoritmasının doğruluk oranı %95,14, ROC eğrisinin altında kalan alan oranı %79,25 ve F1 skoru ise %71,05'tir. Algoritmanın hata matrisi Şekil 49'deki gibidir.



Şekil 49: XGBoost Tahmin Hata Matrisi

XGBoost algoritmasının, test setindeki tüketici değerlendirmeleri üzerinde yaptığı “Faydalı” tahminlerinin %97’si, yapmış olduğu “Faydalı Değil” tahminlerinin ise %70’i doğrudur. “Faydalı” olarak tahmin ettiği tüketici değerlendirmelerinin %30’unun “Faydalı Değil” etiketine sahip olmasına karşın, “Faydalı Değil” olarak tahmin ettiği tüketici değerlendirmelerinin %3,4’ü “Faydalı” etiketine sahiptir.



Şekil 50: XGBoost Öznitelik Ağırlıkları

XGBoost algoritmasının faydalı yorum tahmini için öznitelikler için belirlemiş olduğu ağırlıklar Şekil 50’teki gibidir. En yüksek ağırlığa 0,67 öznitelik katsayısıyla incelemenin

fotoğraflı olma özneliği sahipken, en düşük öznelik katsayısı olan 0,07 incelemeye ait duygu skoru (duygu etiketi) özneliği sahiptir.

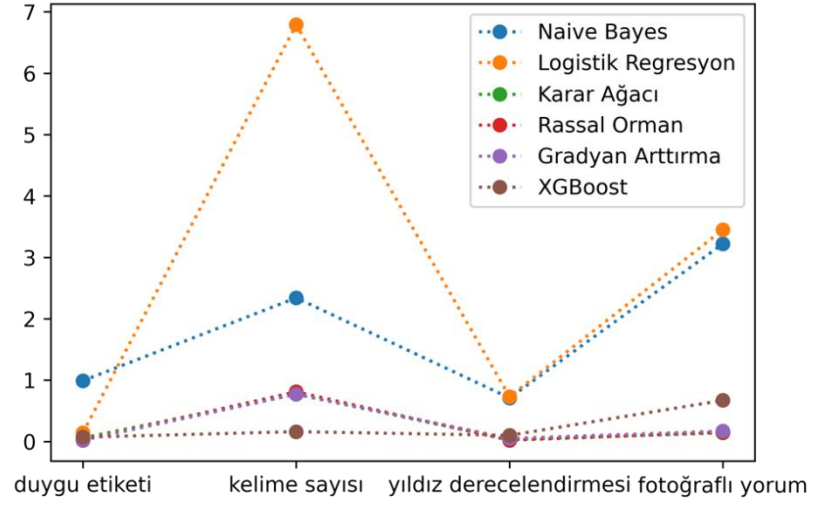
3.5. Algoritmaların Özneliklerin Değerlendirilmesi

Algoritmalar, eğitim süreçlerinden sonra faydalı kullanıcı değerlendirmeli eğitim sürecinde öznelik olarak almış oldukları duygu skorları (duygu etiketleri), kelime sayıları, yıldız derecelendirmeleri ve fotoğraflı yorum olup olmadığını belirten etiketler üzerinde, hangi özneliğin bir değerlendirmenin faydalı olup olmadığını tespiti konusunda bir denklem oluşturmuştur. Bu denklemde, her özneliğin bir kat sayı atanarak, bu özneliğin üzerindeki herhangi bir değişikliğin sonucu ne kadar değiştireceğini ifade etmektedir. Algoritma bazlı öznelik ağırlıkları Tablo 13 ve Şekil 51’de verilmiştir.

Tablo 13: Algoritma Öznelik Ağırlıkları

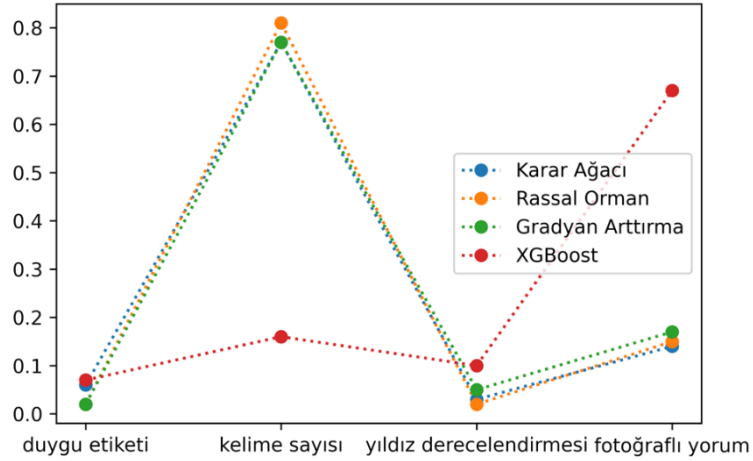
Algoritma	Duygu etiketi	Kelime sayısı	Yıldız derecelendirmesi	Fotoğraflı Yorum
Naive Bayes	0,99	2,34	0,71	3,22
Logistik Regresyon	0,14	6,79	0,73	3,45
Karar Ağacı	0,05	0,77	0,03	0,14
Rassal Orman	0,02	0,80	0,02	0,16
Gradyan Arttırma	0,01	0,77	0,05	0,17
XGBoost	0,07	0,16	0,10	0,67
Ortalama	<i>0,21</i>	<i>1,94</i>	<i>0,27</i>	<i>1,3</i>

Logistik Regresyon algoritması “yıldız derecelendirmesi”, Naive Bayes algoritması ise bütün öznelikleri negatif katsayıyla hesaplamıştır. Tablo 13 ve Şekil 51, hangi özneliğin algoritmanın karar verme aşamasında daha fazla ağırlığa sahip olduğunu ifade etmek için oluşturulduğundan, bütün öznelik ağırlıkları pozitif olarak değerlendirilmiştir.



Şekil 51: Algoritmaların Ortak Öznitelik Ağırlığı

Ağaç temelli algoritmaların öznitelikler için belirlemiş oldukları ağırlıklar 0,01 ile 0,80 arasında değişirken, lineer model olan Naive Bayes ve Logistik Regresyon algoritmalarının ağırlık aralıkları 0,14 ile 6,79 arasında değişmektedir. Sadece ağaç temelli algoritmaların ağırlık grafiği Şekil 52'deki gibidir.



Şekil 52: Ağaç Temelli Algoritma Öznitelik Ağırlıkları

Diğer ağaç temelli algoritmalar benzer ağırlık değerlerine sahipken, XGBoost L1 regülarizasyon işlemlerini öğrenme sırasında gerçekleştirdiği için diğer ağaç algoritmalarından farklı bir öznitelik ağırlığına sahiptir.

Algoritmaların test setleri üzerinde göstermiş oldukları başarı ve özniteliklerin ağırlıkları, çevrimiçi alışveriş platformlarında kullanıcıların faydalı buldukları tüketici değerlendirmeleri için, tüketici tarafından hazırlanan inceleme metninin kelime sayısı bakımından uzunluğu ve incelemenin görsel materyaller ile desteklenmesi, içeriği hazırlayan kullanıcıların pozitif veya negatif bir tecrübe aktarması ya da ürüne vermiş oldukları yıldız derecelendirmelerinden daha fazla önem arz etmektedir. Gradyan Arttırma algoritması için duygu etiketi 0,02 ağırlığına sahipken, kelime sayısı 0,77 ağırlığına sahiptir. Logistik Regresyon algoritmasında bu oran, 0,14'e 6,79'dur. Bütün algoritmalar içinde en yüksek ağırlığa sahip öznitelikler kelime sayısı ve değerlendirmenin fotoğraflı bir değerlendirme olmasıdır.

3.6. Farklı Kategorilerinden Tüketici Değerlendirmeleri Üzerinde Algoritma Başarıları

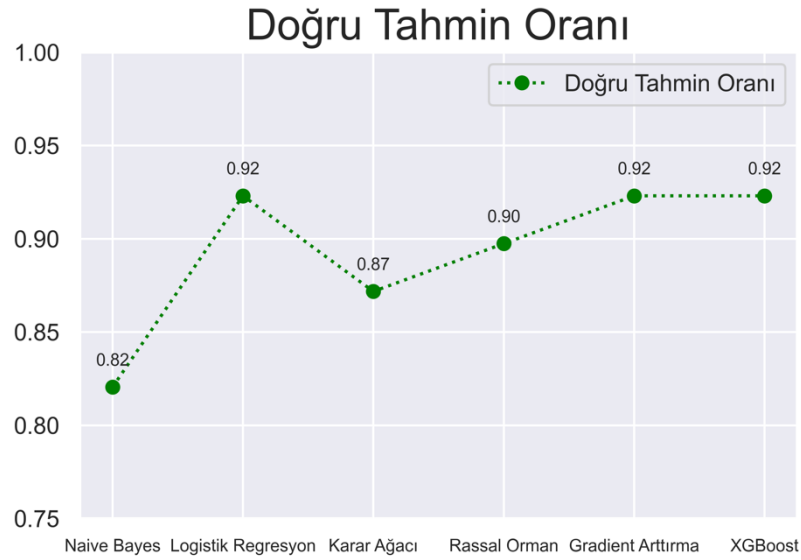
Algoritmaların, farklı ürün kategorilerindeki faydalı yorumları da tahmin başarısının ölçülmesi için EK-1'de değerlendirme metni, Tablo 14'de kategorileri ve sayıları verilen, 32'si fotoğraflı, 7'si fotoğrafsız, tamamı 5 yıldız derecelendirmesine sahip ve en az 2 defa faydalı bulunmuş 39 farklı tüketici değerlendirmesi elde edilmiş olup, tek ürün üzerinden elde edilen algoritmaların bu değerlendirmelerin "Faydalı" ya da "Faydalı Değil" olarak tahmin etmesi istenmiştir.

Tablo 14: Test Verisinin Kategorilere Göre Dağılımı

Ürün Ana Kategorisi	Kategoriye Ait Değerlendirme Sayısı
Elektronik	3
Moda	2
Beyaz Eşya Mutfak Ürünleri	3
Yapı Market, Bahçe, Oto	5
Süpermarket	2
Spor Outdoor Ürünleri	6
Kozmetik Kişisel Bakım	2
Telefon ve Aksesuarlar	8
Ev Elektronik Ürünleri	6
Hobi Oyun Konsolları	2
Toplam	39

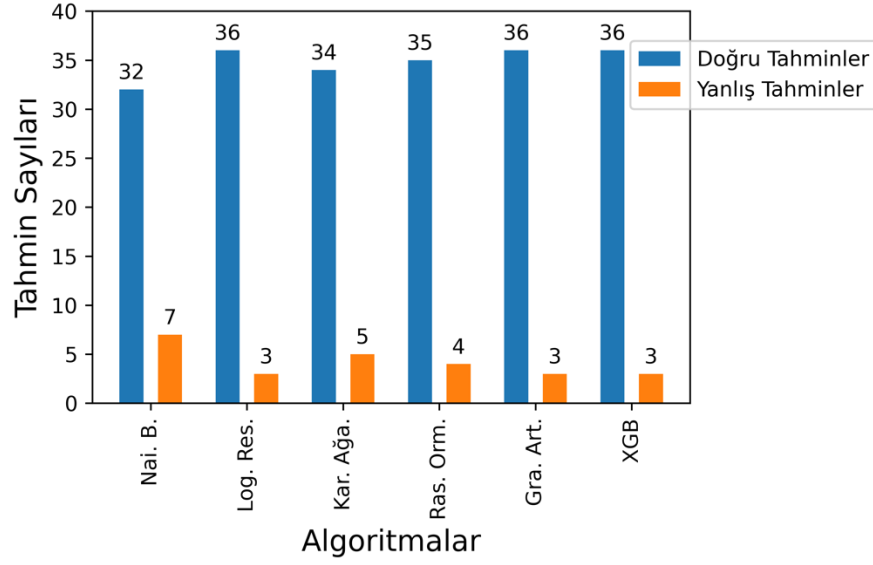
Elde edilen tüketici değerlendirmelerinin hepsi 5 yıldız derecelendirmesine sahip, en çok faydalı bulunan değerlendirmeler olup, 39 değerlendirmenin 32'si fotoğrafı değerlendirilmiştir. Değerlendirmelerin hepsi iki ve daha fazla kez faydalı bulunduğu için

çalışmamız kriterlerine göre hepsi “Faydalı” etiketine sahip değerlendirmelerdir. Değerlendirmelerin öznitelikleri çıkarıldıktan sonra algoritmaların yapmış olduğu tahminlerin başarısı Şekil 53’te görülebilir. Eğer bir algoritma, 39 değerlendirme tamamını “Faydalı” olarak tahmin edebilseydi, Şekil 53’te başarısı %100 olarak görülecekti.



Şekil 53: Farklı Kategoriler Üzerinde Doğru Tahmin Oranı

Algoritmalar arasında %82 ile Naive Bayes olmasına karşın, Logistik Regresyon, Gradyan Arttırma ve XGBoost algoritmaları %92 oran ile doğru tahmin etmiştir. Algoritmaların doğru ve yanlış tahmin ettiği değerlendirme sayıları Şekil 54’teki gibidir.



Şekil 54: Farklı Kategoriler Üzerinde Doğru-Yanlış Tahmin Sayısı

Algoritmalar, farklı kategorilerden alınmış olan yorumlar üstünde de en az %82 oranında doğru tahmin yapabilmektedir. En düşük sonucu veren Naive Bayes algoritması, öğrenim seti üzerinde de en düşük sonucu veren algoritmadır. Söz konusu algoritmanın veri seti ve öznitelikler üzerinde diğer algoritmalar kadar öğrenme gerçekleştirmediği ifade edebilir.

Bu çalışmadan makine öğrenmesi algoritmalarıyla birlikte faydalı yorumların tüketici onayına sunmaksızın tahminlenmesi için çevrimiçi alışveriş platformunda belirlenmiş bir ürüne 2019 – 2022 yılları arası yazılmış tüketici değerlendirmelerinin değerlendirme metni, yıldız derecelendirmesi ve faydalı bulunma sayısı özellikleri alınarak, söz konusu niteliklerden veri ön işleme sonucu elde edilen özniteliklerle hazırlanan makine öğrenimi setleriyle, bilinen denetimli öğrenme algoritmalarının altısı kullanılarak bir dizi öğrenme sürecinden geçirilerek algoritmaların, yeni tüketici değerlendirmelerinin, diğer tüketiciler tarafından faydalı bulunup bulunmayacağını tahmin edilmeye çalışılmıştır. Algoritmaların öğrenme setlerinde ve farklı kategorilere ait tüketici değerlendirmeler üzerindeki başarısı incelenmiştir. Algoritmaların başarısının sınanması ve yeni yorumları tahmin performansının ölçülmesi için kayıt altına alınan yeni tarihli tüketici değerlendirmesi Tablo 15’te görünmektedir.

Tablo 15: Yeni Tarihli Tüketici Değerlendirmesi

Yorum Tarihi	Yıldız Derecesi	Faydalı Bulunma Sayısı	İnceleme Metni
11/08/2021	1	0	Selamlar, hemen üründen bahsetmek istiyorum kargo falan bunlar boş muhabbet, bir kere malzeme kalitesi olarak kutu çok kötü... ...dürüst şekilde sizlere tecrübelerimi aktarmaya çalıştım, sağlıklı kalın.

Yorum, yazıldığı hafta içerisinde kayıt altına alınarak algoritmaların eğitim sonrasında sınanması için saklanmıştır. Naive Bayes algoritması hariç, tüm algoritmalar Tablo 15'te görünen tüketici değerlendirmesinin, diğer tüketiciler tarafından faydalı bulunacağı yönünde tahminde bulunmuştur. Tüketici değerlendirmesinin yazılmasından 9 ay sonra, tüketici değerlendirmesinin mevcut faydalı bulunma derecesi Şekil 55'teki gibidir.



★★★★★ 11 Ağustos 2021, Çarşamba | B**** D**** (32) - Trabzon

Selamlar , hemen üründen bahsetmek istiyorum kargo falan bunlar boş muhabbet , bir kere malzeme kalitesi olarak kutu çok kötü, ben sporda kullanmak için aldım iş görüyor ama böyle ne bileyim telefon görüşmesi için falan kullanımı hayal, bir arkadaşımı aradığımda test için tuvaletti misin dedi sesin çok derinden geliyor ve yankı yapıyor , sol kulaklığın şarjı sağ kulaklığa göre çok çabul bitiyor ses deseniz ben müzikte yapıyorum aynı zamanda hi-fi falan dinliyordum bir dönemler bu kulaklığı onlarla karşılaştırmayacağım elbette ama bass tizz mid allaha emanet , hayatta ucuz mal almam sizlerinde almasını tavsiye etmiyorum fakat , sporda böylebangir bangir müzik dinlemek için ideal kulaklık zaten çin malı herhalde kullanma kılavuzu falan çin , ne poşetleme var ne bi garanti belgesi , kadıköy bit pazarından aldık ürünü sanki (gerçi oranın paketlemesi daha iyi) suya dayanıklı mı diye soracaksınız :D terlediğim zaman çıkarıg siliyorum bişey olur diye , umarım salonda bi yerde kırılır da daha düzgün bir şey alırım , tavsiye etmiyorum ben şahsi görüşüm bunlar daha önce hiç bluetooth kulaklık kullanmamış arkadaşlar sanırım bu 5-10 yıldızları vermiş , satıcı ile bir derdim de yok ürün elime ulaştı yani , ucuz malın yahnisi yavan olur diyorum , param yok bunu alayım dersiniz anlarım , onun haricinde kulaklık kutusunun kapağı nefesle üflesen alttan açılır o kadar hassas , herkese iyi alışverişler diliyorum bir ürünü dürüst şekilde sizlere tecrübelerimi aktarmaya çalıştım , sağlıkla kalın.

Kullanıcı bu ürünü 'saticisından aldı. ● Renk **Siyah**

Bu değerlendirme faydalı mı? 10

Şekil 55: Geçmiş Tarihli Tüketici Değerlendirmesinin Güncel Durumu

Yazıldığı tarih itibariyle kayıt alınan, eğitim seti içerisinde bulunmayan ve algoritmalar tarafından, diğer kullanıcıların faydalı bulacağını yönünde tahmin edilen tüketici değerlendirmesi, tüketiciler tarafından faydalı olarak işaretlenmiştir.

SONUÇ

Araştırmaya genel bakış: Web 2.0 gelişmesiyle birlikte kullanıcıların çevrimiçi ortamda bilgi alışverişi yapmasına olanak sağlamıştır. İnternet kullanıcıları, yapmış oldukları paylaşımlar ile diğer kullanıcılarla aynı sosyal ortamı paylaşır gibi Ağızdan Ağıza İletişim gibi kavramları kullanmaktadır. Tüketiciler, çevrimiçi ortamda belirli hizmet veya ürün hakkında bilgi arayışındayken çeşitli platformlardan, platformların sunmuş olduğu imkanlar (inceleme metni, yıldız derecelendirmesi, inceleme fotoğrafı gibi) ve diğer kullanıcıların hazırladıkları incelemelerin filtrelenmesi için belirli filtreler (en faydalı bulunan değerlendirmelerine göre, kullanıcıların verdikleri yıldız derecelendirmelerine göre, tüketici değerlendirmesinin oluşturulma tarihine göre ve benzeri) sunmaktadır. Faydalı bir tüketici değerlendirmesi, diğer tüketiciler tarafından çevrimiçi alışveriş platformunun sunduğu imkanlara göre işaretlenerek ortaya çıkmaktadır. Bu çalışmada, hali hazırda tüketiciler tarafından ortaya çıkartılmış faydalı yorumlardan faydalanılarak, yeni bir değerlendirme tüketici onayına sunulmadan faydalı bir değerlendirme olup olmadığı tahminlenmeye çalışılacaktır.

Araştırmanın Bulguları: Bu bölümde araştırmanın temel sorusu ve araştırmanın temel sorusunu destekleyecek alt araştırma soruları, elde edilen bulgular ile yanıtlanmıştır.

Araştırma Sorusu 1 – Faydalı tüketici değerlendirmesi tanımlanabilir mi?

Bulgular, tüketici değerlendirmelerinin belirli özelliklerinin baskın olmasıyla birlikte diğer tüketiciler tarafından faydalı bulunan yorumların ortak özellikleri, algoritmaların öznitelik ağırlık analizinde ortaya koyulmuştur. Bir tüketici değerlendirmesinin, fotoğraflı veya fotoğrafsız olmasına bağlı olarak değişen kelime sayısı miktarı, bir tüketici değerlendirmesinin, diğer tüketiciler tarafından faydalı bulunabileceğini belirlemede en baskın öznitelikleridir (Bkz. Şekil 51). Tüketici değerlendirmelerinin faydalı bulunması konusunda, tüketicilerin değerlendirmeye yansıttıkları pozitif veya negatif duygunun ya da verilen yıldız derecelendirmesinin, görece olarak diğer özniteliklerden daha az öneme sahip olduğu görülmüştür. Bir değerlendirme metni varken, yıldız derecelendirmesinin Pelsmacker vd., (2018) yapmış olduğu çalışmaların bulgularını desteklemektedir. Çalışmada, değerlendirme metni içeren incelemelerde, kullanıcıların söz konusu değerlendirmeyi faydalı bulup bulmaması açısından yıldız

derecelendirmesinin neredeyse hiçbir etki olmadığını ortaya koymuştur. Baskın öznitelikler, kullanıcı değerlendirmelerinin makine öğrenmesi algoritmaları tarafından tanımlanmasında önemli etkiye sahiptir.

Araştırma Sorusu 2 – Faydalı tüketici değerlendirme tahminimde hangi makine öğrenmesi algoritmasından yararlanılabilir?

Bu çalışma için kullanılan makine öğrenim algoritmaları, denetimli öğrenim algoritmaları seçilmiş olup, 2 adet lineer, 4 adet lineer olmayan denetimli makine öğrenmesi algoritması seçilmiş, seçilen algoritmaların, bir üründen alınan tüketici değerlendirmeleri üzerinde öğrenim gerçekleştirip, test edilmiştir. Test bulgularında, en iyi performans gösteren denetimli öğrenme algoritmasının Karar Ağacı, en düşük performans gösteren algoritmanın ise Naive Bayes olduğu görülmüştür.

En iyi performans gösteren denetimli öğrenme algoritması olarak Karar Ağacı, %94 doğruluk, %92 ROC eğrisi altında kalan alan değeri ve %71 F1 skoruna ulaşmıştır (Bkz. Şekil 38). Tahmin Hata Matrisi, yapmış olduğu bütün “faydalı” tahminlerinin %86’sının, bütün “faydalı değil” tahminlerinin ise %97’sinin doğru olduğu görülmüştür (Bkz. Şekil 43).

Araştırma Sorusu 3 – Faydalı tüketici değerlendirmesi tahminlime farklı ürün kategorilerinde nasıl çalışır?

Elde edilen 10 farklı ana ürün kategorisine ait 39 farklı tüketici değerlendirmesi üzerinde yapılan testlerde, algoritmaların en yüksek %92, en düşük %82 olacak şekilde doğru tahmin edilmiştir. 39 değerlendirmenin en yüksek olarak 36’sı, en düşük olarak ise 32’si tek bir algoritma tarafından doğru tahmin edilmiştir. 39 değerlendirmenin 36’sı, Karar Ağacı, Gradyan Arttırma ve XGBoost algoritmaları tarafından doğru bir şekilde faydalı olarak tahmin edilmiştir (Bkz. Şekil 53 ve Şekil 54).

Bu çalışmayla birlikte, çevrimiçi alışveriş platformlarında paylaşılan yeni bir tüketici yorum ve değerlendirmesinin faydalılık düzeyi, diğer tüketicilerin algıları olmaksızın (Araştırma Sorusu 1) makine öğrenmesi algoritmaları aracılığıyla (Araştırma Sorusu 2) değerlendirmenin hangi kategoride olduğu fark etmeksizin tahmin edileceği (Araştırma Sorusu 3) ortaya koyulmuştur.

Bu çalışmayla, faydalı tüketici değerlendirmelerinin, diğer tüketicilerin algısı olmaksızın, değerlendirilmenin girildiği tarih itibariyle yazılan değerlendirme metninin içeriğine bakılmadan, sağlamış olduğu metnin uzunluğu ve değerlendirmenin ürün ya da hizmete ait fotoğraflar içerip içermemesiyle, değerlendirmeyi sonradan okuyacak tüketiciler tarafından tahmin edilebileceği ortaya koyulmuş olup, ürün ve hizmet açısından sürekli güncel faydalı değerlendirmelerin ortaya koyulması sağlanmıştır. Çalışmada hazırlanan uygulamanın ve faydalı tüketici değerlendirmeleri için ortaya koyulan yaklaşımın, değerlendirme hazırlayan tüketiciler açısından; diğer tüketicilerin satın alma süreçlerine faydalı olacak şekilde hazırlanmış oldukları tüketici değerlendirmelerinin ön plana çıkacağı bilgisiyle tüketicilerin değerlendirme hazırlama motivasyonlarına katkıda bulunması, değerlendirme okuyan tüketiciler açısından; ürün ya da hizmet konusunda bilgi arayan tüketicilerin, güncel tarihli tüketici değerlendirmeleri ile karşılaşması ve yıllara içinde değişen ürün hizmet özelliklerine yönelik güncel tarihli tüketici değerlendirmesiyle karşılaşması amaçlanmıştır.

Çalışmanın Kısıtları ve Gelecek Araştırmalar

Her çalışmada olduğu gibi, bu çalışmanın da yapıldığı zaman, dönem, teknoloji ve yaklaşımlar açısından kısıtları bulunmaktadır.

Çalışmada kullanılan veriler 02.07.2019 ile 01.03.2022 tarihleri arasında Hepsiburada adlı çevrimiçi alışveriş platformundan satılan Xiaomi Redmi Airdots Basics 2 TWS Bluetooth 5.0 Kulaklık ürününe ait sadece 1 ve 5 yıldız derecelendirmesine sahip tüketici değerlendirmeleri alınarak veri setine dahil edilmiştir. Tüketici değerlendirmelerine ait değerlendirme metni veri setinde sadece kelime sayısı ve duygu skoru olarak teslim edilmiş olup metin içerikleri (TF/IDF – Word2Vec) gibi yöntemler kullanılarak metinlerin okunabilirlik skorları, tekrar eden kelime oranları, pozitif – negatif kelime oranı gibi analizleri dahil edilmemiştir.

Tarih itibariyle alışveriş sitesinin sunmuş olduğu yeni değerlendirme yöntemleriyle hazırlanmış tüketici değerlendirmeleri sadece metin ve yıldız derecelendirmeleri olarak veri setine dahil edilmiş, tüketicilerin sağlamış olduğu tematik bilgiler veri setinde temsil edilmemiştir.

Bu çalışmada yalnızca değerlendirme metnlerinin duygu skorları, yıldız derecelendirmeleri, metinlerin kelime sayısı ve fotoğraflı değerlendirme olup olmaması

öznitelik olarak kullanılmıştır. Gelecek çalışmalarda, değerlendirme metinlerine okunabilirlik analizi, birliktelik analizi, pozitif – negatif kelime oranları, TF/IDF skorları gibi analizler gerçekleştirerek bir tüketici değerlendirmesinin diğer tüketiciler tarafından faydalı bulunmasında içeriğinin ne derece etkili olduğu araştırılabilir.

KAYNAKÇA

- Aggarwal, S., & Kaur, D. (2013). Naïve bayes classifier with various smoothing techniques for text documents. *International Journal of Computer Trends and Technology*, 873–876.
- Alsharif, N. (2022). Fake opinion detection in an e-commerce business based on a long-short memory algorithm. *Soft Computing*. <https://doi.org/10.1007/s00500-022-06806-5>
- Altın Yavuz, A., & Yavuz, H. S. (2021). Denetimli Makine Öğrenme Yöntemleri ile Yüzey Su Kalitesinin Sınıflandırılması. *Biyoloji Bilimleri Araştırma Dergisi*, 2, 142–155.
- Aluri, A., Price, B. S., & McIntyre, N. H. (2018). Using Machine Learning To Cocreate Value Through Dynamic Customer Engagement In A Brand Loyalty Program: *Https://Doi.Org/10.1177/1096348017753521*, 43(1), 78–100. <https://doi.org/10.1177/1096348017753521>
- Anderson, E. W. (1998a). Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1), 5–17. <https://doi.org/10.1177/109467059800100102>
- Anderson, E. W. (1998b). Customer Satisfaction and Word of Mouth. *Journal of Service Research*, 1(1), 5–17.
- Anderson, J. A. (1984). Regression and Ordered Categorical Variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(1), 1–22. <https://doi.org/10.1111/J.2517-6161.1984.TB01270.X>
- Barsky, E., & Purdon, M. (2006). Introducing Web 2.0: Social networking and social bookmarking for health librarians. *Journal of the Canadian Health Libraries Association*, 27. <https://doi.org/10.5596/c06-024>
- Beam, A. L., & Kohane, I. S. (2018). Big Data and Machine Learning in Health Care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/JAMA.2017.18391>
- Belyadi, H., & Haghghat, A. (2021). Machine Learning Guide for Oil and Gas Using Python: A Step-By-Step Breakdown with Data, Algorithms, Codes, and Applications. In *Machine Learning Guide for Oil and Gas Using Python*. Gulf Professional Publishing.
- Berners-Lee, T. (1998, May 7). *The World Wide Web: A very short personal history*. <https://www.w3.org/People/Berners-Lee/ShortHistory.html>
- Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, 188(12), 2222–2239. <https://doi.org/10.1093/AJE/KWZ189>

- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/S11749-016-0481-7/FIGURES/4>
- Bogdanchikov, A., Zhaparov, M., & Suliyev, R. (2013). Python to learn programming. *Journal of Physics: Conference Series*, 423(1), 012027. <https://doi.org/10.1088/1742-6596/423/1/012027>
- Breiman, L. (1996). *Bias, Variance, and Arcing Classifiers*. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.30.8572>
- Brownlee, J. (2018, August 31). *How to Use ROC Curves and Precision-Recall Curves for Classification in Python*. <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>
- Buckland, M., & Gey, F. (1994). The Relationship between Recall and Precision. *Journal of The American Society for Information Science*, 45(1), 12–19. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1)
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Müller, A. C., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). *API design for machine learning software: experiences from the scikit-learn project*. <https://doi.org/10.48550/arxiv.1309.0238>
- Buttle, F. A. (1998). Journal of Strategic Marketing Word of mouth: understanding and managing referral marketing Word of mouth: understanding and managing referral marketing. *Journal of Strategic Marketing*, 6(3), 241–254. <https://doi.org/10.1080/096525498346658>
- Çakar, E. N., & Akbıyık, A. (2018). Hızlı Tüketim Mallarına Yönelik Tüketici Yorumlarında Odak Sorunu: Ürün mü, Satış Hizmeti mi Değerlendiriliyor? *AJIT-e: Online Academic Journal of Information Technology*, 9(33), 147–158. <https://doi.org/10.5824/1309-1581.2018.3.009.X>
- Chapin, N. (2003). Flowchart. In *Encyclopedia of Computer Science* (pp. 714–716).
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491. <https://doi.org/10.1287/MNSC.1070.0810>
- Choudhury, N. (2014). World wide web and its journey from web 1.0 to web 4.0. *International Journal of Computer Science and Information Technologies*, 5(6). <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.666.6445&rep=rep1&type=pdf>
- Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. *Proceedings of the ACM SIGMOD International Conference on Management of Data, 26-June-2016*, 2201–2206. <https://doi.org/10.1145/2882903.2912574>

- Cook, J. A., & Ranstam, J. (2016). Overfitting. *British Journal of Surgery*, 103(13), 1814–1814. <https://doi.org/10.1002/BJS.10244>
- Cunningham, P., Cord, M., & Delany, S. J. (2008a). *Machine Learning Techniques for Multimedia*.
- Cunningham, P., Cord, M., & Delany, S. J. (2008b). *Machine learning techniques for multimedia*.
- Daradkeh, M. K. (2021). Exploring the Usefulness of User-Generated Content for Business Intelligence in Innovation. *International Journal of Enterprise Information Systems*, 17(2), 44–70. <https://doi.org/10.4018/IJEIS.2021040103>
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407–1424. <https://doi.org/10.1287/MNSC.49.10.1407.17308>
- Devlin, J., & Chang, M.-W. (2018, October 2). *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*. <https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>
- Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4), 1007–1016. <https://doi.org/10.1016/J.DSS.2008.04.001>
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The adverse childhood experiences (ACE) study. *American Journal of Preventive Medicine*, 14(4), 245–258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- Freund, R. J., Wilson, W. J., & Sa, Ping. (2006). *Regression analysis* (2nd ed.). Elsevier Academic Press.
- Friedman, J. H. (1997). On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality. In *Data Mining and Knowledge Discovery* (Vol. 1). Kluwer Academic Publishers.
- Fujino, A., Isozaki, H., & Suzuki, J. (2008). Multi-label Text Categorization with Model Combination based on F1-score Maximization. *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*. <https://aclanthology.org/I08-2116>
- Ghahramani, Z. (2003). Unsupervised Learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3176, 72–112. https://doi.org/10.1007/978-3-540-28650-9_5
- Griffin, J. (1995). *Customer loyalty*. <http://altfeldinc.com/pdfs/Customer%20Loyalty.pdf>
- Gürbüz, S., & Şahin, F. (2014). Sosyal bilimlerde araştırma yöntemleri. In *academia.edu* (2nd ed.).

https://www.academia.edu/download/63615871/Guvenilirlik_5_Yerli20200613-38326-6bfqj9.pdf

- Gürbüz, S., & Şahin, F. (2015). Sosyal bilimlerde araştırma yöntemleri. In *academia.edu* (1st ed.). Seçkin Yayıncılık.
- Harker, P. T., & Vargas, L. G. (1987). The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process. *Http://Dx.Doi.Org/10.1287/Mnsc.33.11.1383*, 33(1), 1383–1403. <https://doi.org/10.1287/MNSC.33.11.1383>
- Heer, J., Hellerstein, J. M., & Kandel, S. (2015). Predictive interaction for data transformation. *CIDR 2015*.
- Hellerstein, J. M. (2008). Quantitative data cleaning for large databases. *Computer Science Journal*, 1(1).
- Hennig-Thurau, T., Gwinner, K. P., & Gremler, D. D. (2002). *Understanding Relationship Marketing Outcomes An Integration of Relational Benefits and Relationship Quality*.
- Hennig-Thurau, T., & Walsh, G. (2014). Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet. *Https://Doi.Org/10.1080/10864415.2003.11044293*, 8(2), 51–74. <https://doi.org/10.1080/10864415.2003.11044293>
- Hill, R. K. (2016). What an algorithm is? *Philos. Technol*, 29, 35–59. <https://doi.org/10.1007/s13347-014-0184-5>
- History of Python*. (2022, February 11). <https://www.geeksforgeeks.org/history-of-python/>
- Hmoud Al-Adhaileh, M., & Waselallah Alsaade, F. (2022). Detecting and Analysing Fake Opinions Using Artificial Intelligence Algorithms. *Intelligent Automation & Soft Computing*, 32(1). <https://doi.org/10.32604/iasc.2022.021225>
- Jabbar, H. K., & Khan, R. Z. (2015). *Methods to Avoid Over-Fitting and Under-Fitting in Supervised Machine Learning (Comparative Study)*. 163–172. https://doi.org/10.3850/978-981-09-5247-1_017
- Jones, M., Hockey, R., Mishra, G. D., & Dobson, A. (2014). Visualising and modelling changes in categorical variables in longitudinal studies. *BMC Medical Research Methodology*, 14(1), 1–8. <https://doi.org/10.1186/1471-2288-14-32/FIGURES/4>
- Kalapanidas, E., Avouris, N., Craciun, M., & Neagu, D. (2003). Machine Learning algorithms: a study on noise. *Ist Balcan Conference in Informatics*. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.9256>
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research.

Computational and Structural Biotechnology Journal, 15, 104–116.
<https://doi.org/10.1016/J.CSBJ.2016.12.005>

- Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, 37(9), 6233–6239. <https://doi.org/10.1016/J.ESWA.2010.02.101>
- Koehrsen, W. (2018, January 28). *Overfitting vs. Underfitting: A Complete Example*. <https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765>
- Kozinets, R. v, de Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010a). Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities. *Journal of Marketing*, 74, 71–89.
- Kozinets, R. v., de Valck, K., Wojnicki, A. C., & Wilner, S. J. S. (2010b). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing*, 74(2), 71–89. <https://doi.org/10.1509/JMKG.74.2.71>
- Krestel, R., & Dokoohaki, N. (2011). Diversifying product review rankings: Getting the full picture. *Proceedings - 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011, 1*, 138–145. <https://doi.org/10.1109/WI-IAT.2011.33>
- Krestel, R., & Dokoohaki, N. (2015). Diversifying customer review rankings. *Neural Networks*, 66, 36–45. <https://doi.org/10.1016/J.NEUNET.2015.02.008>
- Kumar, A., Gopal, R. D., Shankar, R., & Tan, K. H. (2022). Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering. *Decision Support Systems*, 155. <https://doi.org/10.1016/J.DSS.2021.113728>
- Lai, X., Wang, F., & Wang, X. (2021). Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics. *International Journal of Contemporary Hospitality Management*, 33(6), 2137–2156. <https://doi.org/10.1108/IJCHM-07-2020-0708/FULL/PDF>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature* 2015 521:7553, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, M., & Youn, S. (2009). Electronic word of mouth (eWOM). *International Journal of Advertising*, 28(3), 473–499. <https://doi.org/10.2501/S0265048709200709>
- Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1). <https://doi.org/10.1145/1232722.1232727>
- Li, M. (2021). Research on Extraction of Useful Tourism Online Reviews Based on Multimodal Feature Fusion. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(5). <https://doi.org/10.1145/3453694>

- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors* 2018, Vol. 18, Page 2674, 18(8), 2674. <https://doi.org/10.3390/S18082674>
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468. <https://doi.org/10.1016/J.TOURMAN.2007.05.011>
- Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, 1–367. <https://doi.org/10.1017/CBO9781139084789>
- Liu, Z., & Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140–151. <https://doi.org/10.1016/j.tourman.2014.09.020>
- Martín, C. A., Aguilar, R. M., Torres, J. M., & Díaz, S. (2018). Automatic detection of sentiments in tourist reviews by using long short-term memory recurrent neural networks. *30th European Modeling and Simulation Symposium, EMSS 2018*, 356–360.
- Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. *European Journal of Operational Research*, 281(3), 588–596. <https://doi.org/10.1016/J.EJOR.2018.04.034>
- McKinney, W. (2017). *Python for data analysis : data wrangling with pandas, NumPy, and IPython* (Vol. 2). O'Reilly.
- Melkumova, L. E., & Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering*, 201, 746–755. <https://doi.org/10.1016/J.PROENG.2017.09.615>
- Moschovakis, Y. N. (2001). What is an algorithm? *Mathematics Unlimited — 2001 and Beyond*, 919–936. https://doi.org/10.1007/978-3-642-56478-9_46
- Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. In *MIS Quarterly: Management Information Systems* (Vol. 34, Issue 1, pp. 185–200). University of Minnesota. <https://doi.org/10.2307/20721420>
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*, 18(6), 275–285. <https://doi.org/10.1002/CEM.873>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7(DEC), 21. <https://doi.org/10.3389/FNBOT.2013.00021/BIBTEX>

- Nielsen. (2015a). *Global Trust in Advertising*. <https://www.nielsen.com/eu/en/insights/report/2015/global-trust-in-advertising-2015/>
- Nielsen. (2015b, September 28). *Global Trust in Advertising*. <https://www.nielsen.com/eu/en/insights/report/2015/global-trust-in-advertising-2015/>
- Park, D. H., Lee, J., & Han, I. (2007). The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International Journal of Electronic Commerce*, 11(4), 125–148. <https://doi.org/10.2753/JEC1086-4415110405>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pelsmacker, P. de, Dens, N., & Kolomiiets, A. (2018). The impact of text valence, star rating and rated usefulness in online reviews. *International Journal of Advertising The Review of Marketing Communications*. <https://doi.org/10.1080/02650487.2018.1424792>
- Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, 23(4). <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.98.8661>
- Ran, Y., Liu, S.-Y., Yu, X., Liang, Y., Huang, L., & Wang, Z. (2017). Link prediction in social network based on local information and attributes of nodes. *Journal of Physics: Conference Series*, 887(1), 012043. <https://doi.org/10.1088/1742-6596/887/1/012043>
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Riegner, C. (2007). Word of Mouth on the Web: The Impact of Web 2.0 on Consumer Purchase Decisions. *Journal of Advertising Research*, 47(4), 436–447. <https://doi.org/10.2501/S0021849907070456>
- Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Günther, H. M., ... Streicher, O. (2013). Astropy: A community Python package for astronomy. *Astronomy & Astrophysics*, 558, A33. <https://doi.org/10.1051/0004-6361/201322068>

- Sahelices-Pinto, C., Rodríguez-Santos, C., Sahelices-pinto, C., & Rodríguez-santos, C. (2014). E-WoM and 2.0 Opinion Leaders. *Journal of Food Products Marketing*, 20, 244–261. <https://doi.org/10.1080/10454446.2012.732549>
- Sarle, W. S. (1995). Stopped Training and Other Remedies for Overfitting. *Proceedings of The 27th Symposium on The Interface of Computing Science And Statistics*. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.3920>
- Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, 2. <https://doi.org/10.14569/IJARAI.2013.020206>
- Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/J.NEUNET.2014.09.003>
- Srinivasan, S. S., Anderson, R., & Ponnayolu, K. (2002). Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, 78(1), 41–50. [https://doi.org/10.1016/S0022-4359\(01\)00065-3](https://doi.org/10.1016/S0022-4359(01)00065-3)
- Surden, H. (2014). Machine Learning and Law. *Washington Law Review*, 89. <https://heinonline.org/HOL/Page?handle=hein.journals/washlr89&id=94&div=&collection=>
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1), 1–28. <https://doi.org/10.1186/S12880-015-0068-X/TABLES/5>
- Topal, M. O., Bas, A., & van Heerden, I. (2021). *Exploring Transformers in Natural Language Generation: GPT, BERT, and XLNet*. <https://towardsdatascience.com/Transformers-141e32e69591>
- Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing*, 73(5), 90–102. <https://doi.org/10.1509/JMKG.73.5.90>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/MIND/LIX.236.433>
- Turney, P. D. (2002). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. <https://doi.org/10.48550/arxiv.cs/0212032>
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9. <https://doi.org/10.1016/J.SIMPAT.2015.03.003>
- Visa, S., Ramsay, B., Ralescu, A., & VanDerKnaap, E. (2011, January 1). Confusion Matrix-Based Feature Selection.

Proceedings of the Twentysecond Midwest Artificial Intelligence and Cognitive Science Conference. <https://openworks.wooster.edu/facpub/88>

- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168(2), 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Young, T., Palta, M., Dempsey, J., Skatrud, J., Weber, S., & Badr, S. (1993). The Occurrence of Sleep-Disordered Breathing among Middle-Aged Adults. *New England Journal of Medicine*, 328(17), 1230–1235. <https://doi.org/10.1056/NEJM199304293281704>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/WIDM.1253>
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229), 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- Zhou, T. (2011). Understanding online community user participation: A social influence perspective. *Internet Research*, 21(1), 67–81. <https://doi.org/10.1108/106622411111104884>
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on Sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133–148. <https://doi.org/10.1509/JMKG.74.2.133>

EKLER

EK 1: Farklı Kategorilerden Tüketici Değerlendirmeleri

Değerlendirme Numarası	Değerlendirme Metni
#1	<p>Kulaklığı yaklaşık 6-7 ay önce almayı düşünüyordum 10 dan fazla inceleme video su izledim herkes şiddetle öneriyordu kırmızı rengi ile siyah arasında kalmıştım ama sonra siyahın daha iyi olacağına karar verdim ve siyah aldım beklediğim gibi kalitesi mükemmel. Ses beklediğimden iyi baseup özelliği de başlı müziklerde çok fark ettiriyor güzel bir özellik. Pasif görültü engelliyor kulaklık kulağı sardığı için dışardaki sesi içeri almıyor, yüksek sesle müzik dinlediğinde de dışarı vermiyor. İçerisinden çıkan type-c kablo ve aux kabloda kaliteli onu bile basit yapmamışlar yani Fiyat konusunda sürekli inip çıkıyordu ben hepsiburadanın kampanyasıyla baya uyguna aldım, almak isteyen düşünmeden alabilir ben sizin yerinize fazlasıyla inceledim ve düşündüm fiyatını fazlasıyla hakeden bir ürün</p>
#2	<p>Yıllarca titizlikle kullandığım kulak içi kulaklığım arızalanınca yenisini alanya karar verdim. Ses konusunda çok hassas bir insanım. Biraz basslı ve yüksek sesli müzik dinlemeyi seviyorum. Fakat bütçem 100 lira ile sınırlı maalesef. Bunun için 10 gün boyunca sabahlara kadar bu fiyat aralığındaki tüm ürünlerin onlarca inceleme videosunu izleyip yüzlerce yorumu okuduktan sonra bu ürünü almaya karar verdim. Sonuç olarak bu fiyata nokta atışı bir tercih yapmışım. Ses yüksekliği yeterli seviyede. (telefonum s7 edge bu arada) Kulaklığın ses kalitesi harika. Basslar dolgun, derin ve dengeli. Tizler yeterli. Medium sesler, sanki bass ve tizlerin arasına hafiften gömülmüş gibi. 8d müzikleri gece sessizliğinde dinleyerek kulaklığınızı test etmenizi öneririm. Bayılacaksınız. Ses yalıtımı konusunda çok başarılı, kesinlikle yanınızdaki kişiyi rahatsız etmezsiniz, dış sesleri de kolay kolay duymazsınız. Ses yüksekliği ve kalitesi konusunda ses kaynağınızın kalitesi de çok önemli bunu da göz önünde bulundurun. Yukarıda yazdığım tüm güzel şeyler 100 lira altı bir ürün için yapılmıştır. Lütfen binlerce liralık üst segmentteki ürünlerle kıyaslamayalım. O segmentteki kulaklıkları da test ettim aradaki farkı biliyorum. Fakat cebinizdeki para bu aralıkta ise kesinlikle parasının tam karşılığı. Bence başarılı bir ürün olmuş. Tam bir f/p ürünü. Eksi olarak söyleyebileceğim yüksek seste dinlemiyorsanız kablo bir yere dokunduğunda ya da sürttüğünde sesi pıt pıt kulağınıza iletiyor ve rahatsız oluyorsunuz. Bir de kulak içi kauçukları biraz daha soft olabilirdi. Ürün hakkında söyleyebileceklerim bunlar. Bir de size tavsiyem olacak. Kulaklığınızın işi bittiğinde rulo şeklinde sararak kulaklık kutusuna kaoyup taşırsanız ömrü daha uzun olacaktır. Bir önceki kulaklığımı bu şekilde tam 7 yıll kullandım. Yürüyüş ve koşubuyduğum halde. Cebe yada çantaya rastgele atılan kulaklığın kablosu katlanıp kırılıyor ve bir süre sonra ses kayıpları başlıyor. Yardımcı olabildimse ne mutlu, kalın sağlıcakla.</p>
#3	<p>Bildiğiniz ssd gayet hızlı ve güzel paketlenmiş biçimde geldi. Ben laptopa taktım ve çok iyi sonuçlar aldım. Yanında usb 3.0 disk kutusu aldım hem klonlama için hem de sonrasında eski diski kullanmak için. Herkese tavsiye ederim. Heee bu arada disk klonlama yaptıktan sonra size ortalama 240gb ssd den 207 gb kalıyor. Geri kalanı sistem dosyaları alıyor.</p> <p>Yaptığım test değerlendirmeleri şöyle:</p> <p>Ssd öncesi</p> <p>Şifre ekranı 26 saniye</p> <p>Autocad 150 saniye</p> <p>Chrome 45 saniye</p> <p>3ds max 165 saniye</p> <p>Sketchup 66 saniye</p> <p>Lumion 100 saniye</p> <p>Kapanma 35 saniye</p> <p>Ssd sonrası</p> <p>Şifre ekranı 12 saniye</p> <p>Autocad 40 saniye</p>

	<p>Chrome 6 saniye 3ds max 40 saniye Sketchup 35 saniye Lumion 37 saniye Kapanma 13 saniye</p>
#4	<p>İlk kez MARKA terlik alığım için siparişimi 3.kez tekrarlamak zorunda kaldım...ayak numaram normalde 37,5 ilk alışımında 37-38 numara olanını aldım fakat parmaklarım terliğin ucuna dayandı bu nedenle iade ettim.buçuklu numara giyiyorsanız bir numara büyük alabilirsiniz...2.siparişimde başka bir satıcıdan sipariş verdim 38-39 numarasını boyu tam oldu ancak bu defa da ürünün orjinallüğinden şüphe ettim iade etmek zorunda kaldım.Bu son siparişimi bugün teslim aldım...kesinlikle orjinal dokusu kalitesi her şeyi kusursuz.fotoğraflarımı ekleyeceğim.tek yıldızı paketlemeden dolayı kırdım.crocs poşetinde bile değildi sadece kargo poşetine koyulmuştu ve etiketlerinin yarısı kopmuştu...daha özenli olmalıydı.</p>
#5	<p>Kumaş kalitesi süper beklediğim üstünde bir alışveriş oldu bundan sonra merkezimiz burası eşime aldım pişman olmicamı düşünüyorum renge bayıldım</p>
#6	<p>KARGO sayesinde ertesi gün geldi.Tavalar Aliminyum döküm Bazı yorumlar da arkadaşlar saf döküm olduğunu zannederek almış.Bu fiyata hakiki döküm olmaz ve üstelik ben kaliteli bir markanın gerçek döküm ürünlerini kullanıyorum memnunum fakat bazı yiyecekleri ne kadar ısıtırsan ısıt ve yağ ekle yapıştırıyor mesela patates,müçver vs..Eti harika mühürlüyor orası ayrı.yani diyeceğim böyle sağlam kaliteli yiyecekleri yapıştırmayan bir tava arıyordum.indirime de denk gelince harika oldu.Boyutları ideal.Teşekkürler FIRMA.</p>
#7	<p>Hızlı kargo güzel paketleme 1 güne elime ulaştı. Güzel şık ve kibar bir ürün. Şarjı dolu vaziyette geldi. Çalışmasında bir sıkıntı yok biraz ses çıkarıyor o da normal bence. Suyu orta derecede akıtıyor ne çok az ne çok fazla bence bir sıkıntısı yok güzel ürün. Almak isteyenlere tavsiye ederim</p>
#8	<p>Ürünü daha öncede kullanmıstım. 1 filtre 1 ay kadar ömrü oluyor. Sadece icme suyu olarak 5 yetişkin kullanıyor Damacanadan kurtuldum. Diğer tezgah altı filtrelerin maliyeti ve uygulaması kapladığı yer ve yatırım maliyeti acısından mantikli gelmemiştir.</p>
#9	<p>Ürünler eksiksiz ve fiyatına göre çok uygun. Paketler kontrol edilerek konulmuş. Numaraların kesiminde hiçbir sıkıntısı yoktu. Kargoda biraz hırpalanmış, daha iyi paketlenmeli.Tavsiye ederim.</p>
#10	<p>kapı fitillerine sıkarsanız fitillerin ömrünü uzatır ve yumuşatmış olursunuz. kapı menteşelerinize sıkıldığınızda kapılarınız çok daha sessiz açılıp kapanır. cam çerçevesine içten sıkıldığınızda camlarınız çok daha kolay ve gıcirtısız çalışır hale gelir. emniyet kemerine sıkıldığınızda kemeriniz otomatik ve seri şekilde kendiliğinden toplanır. kontak anahtarı, kapıdaki ve bagajdaki anahtar yuvasına sıkıldığınızda anahtarınız şıkır şıkır yuvaya oturur ve problemsiz çalışır. kaputu kaldıran hidroliğe, kaputu ve bagajı kilitleyen mekanizmaya kullandığınızda eskiye nazaran daha sessiz ve rahat çalışır. elektrikli koltuk, elektrikli katlanan aynalar, sunroof, çalışırken tutaklık yapan tüm hareketli parçalar için çok etkili bir çözümdür. araç içinden gelen trim sesleri için de ilaç gibidir. gıcirtıyı, titreşimi, rahatsız edici tüm sesleri anında keser. torpido için kullanılması önerilirse de ben bugüne kadar herhangi bir zararını görmedim. oldukça parlak yapmakla birlikte toz itici özelliği de bulunmakta. karantina günlerinde canı sıkılan ve arabasının bakımını yapmaktan keyif alanlar için arabasına level atlatacak özellikte mucizevi bir spreidir kendileri. elektrik aksamina bir zarar vermemekle birlikte kullandığınız yüzeyleri besleyip koruyarak kullanım ömrünü uzatmaktadır. kullanım, kullandırın efendim.</p>
#11	<p>MARKA bildiğimiz ve güvenle kullandığımız bir marka. Çalışma/oyun odamız için aldık. Anahtarı hep açık konumda bırakarak kumandada yer alan on/off tuşunu kullanabilirsiniz. Ampülün kendi rengi sıcak beyaz/gün ışığı. Işıması ayarlanabilen bir ampül (dim.) ve ayarını kumanda üzerinden yapabiliyorsunuz. Tüm renkleri güzel, 3 farklı modu mevcut. Flash, renkler arasında direkt geçiş sağlıyor otomatik olarak. Strobe, renkler arasında seri bir geçiş sağlayarak disko topu efekti veriyor :D. Smooth ise renkler arasında yumuşak geçişler sağlayarak ambiyans yaratıyor. Modları kullanırken dim özelliği kullanılmıyor çünkü dim özelliğini modlar belirliyor. Direkt renk seçimi yaptığınızda ise dim özelliğini kullanabilirsiniz.</p>

	Kapatmadan önce W'ye tıklarsanız ampül orijinal rengindeyken açılacaktır. Herhangi bir moddayken kapatırsanız açtığınızda ampül son kaldığı renkte ışıyacaktır. Kumandayı ampüle doğru tutarak kullanmanız gerekmekte, aksi taktirde kumandayı algılamayabilir. Kumanda üzerinde pil ile gelmekte.
#12	İçine koydukları şeker için GÖNDERİCİ firmasına teşekkürler. Küçük ama gülümseten bir hediye. Aynı zamanda kapağı kırılmasın diye kesilmiş rolikten kapak yapmaları çok mantıklı bir hareket olmuş. Bunların dışında ürünü gıcırdayan kapı kolunda denedim. Çok iyi performans gösterdi. Ses sifıra indi diyebilirim.
#13	Ürün anlatılandan ve beklentilerden bir basamak daha ileri seviyede. Su sızdırmazlığı ve kolay uygulanımı sayesinde çok pratik. Üretici ve tedarikçilerin eline sağlık.
#14	MARKA çok uygun fiyat avantajıyla KARGO kargo nun jet hızıyla sağlam ambalaj paketlemesiyle elime ulaşan ürünü için tesekkür ederim.Dogacak olan bebegimiz için tavsiye üzerine aldığım bir ürün. Fiyat ve kargo çok iyi görselleri laylastim umarim yardimci olabilmışimdir. Almak ksteyenler gerçekten gönül rahatlığıyla tercih edebilir. Tesekkurler.
#15	Ürün kendisini kanıtlamış bir ürün kalitesine laf diyemem. Ben taze ebeveynlere tavsiyede bulunayım bu kremi her alt degistiklerinde kullanacaklar ki çocuk pisik olmasın, pisik olduktan sonra kullanmanın bir anlamı yok onun için ayrı ürün almaları gerek. SATICI isimli satıcıdan aldım, sağlam paketleme için ayrı ileri kullanma tarihli ürün gönderdiği için ayrı tesekkurler. Sunduğu kampanya içinde saolsun.
#16	merhaba arkadaşlar. öncelikle hepimiz buradaki yorumları kontrole istinaden malzeme alıyoruz. bu sebepten daha faydalı olacak yorumlar yapmalıyız diye düşünüyorum. kargo süresi v.s abartı olmadıkça çok önemli değil ki o siparişe özgü bir aksaklık olabilir. bu siparişe ilgili olarak ise malzeme bildiğim bir malzeme ama o kadar kötü yorum okudum ki tereddütü aldım. malzeme piyasadakilerle aynı herhangi bir sorun yaşamadım. evet biraz rahatsız olabilir bu malzeme hatası değil. piyasadakilerde böyle. daha iyisi 80 tl den başlıyor 250-300 tl ye çıkıyor. kullanım amacına bağlı olarak parasına göre kaliteli ve iş görür..
#17	Gördüğüm bu model sandalyelerden daha geniş.Rahat mı derseniz bu modeller bana çok rahat gelmemiştir.Yemeginizi yerken biraz dinlenirken işinizi görür Belki oturma yerine ince bir minder atarsanız çok uzun süre keyifle oturabilirsiniz.Kaliteli iskelet kaliteli kumaş dengeli sağlam bir sandalye.Kendi emsalleriyle kıyaslandığında 5 yıldız.Fiyat olarakda çok uygun.Tavsiye ederim.
#18	ürün güzel anlatıldığı gibi paketlemesi gayet yeterli. FİRMA kendisinden aldım. MARKA aynı tip tabureli 4-6 kişilik versiyonu 350 tl. pratik bir ürün kurulum gerektirmiyor. ayakları açıp kilitlemeniz yeterli olacaktır. kamp piknik kullanımı için ideal.çok ağır değil ama denge problemi de yok bence. boyunu uzatıp kısaltabiliyorsunuz. katlanır sandalye ile muhteşem bir takım olacaktır. Kesinlikle tavsiye ediyorum. şu koronavirüs günleri geçip ülkece özgürlüğümüze ve sağlığınıza kavuştuğumuzda bol bol kullanacağınız inşallah. şimdilik Evde Kal Türkiye
#19	Öncelikle sorunsuz teslimata teşekkürler. Fiyatına göre gayet güzel bir masa. Taburelerde sağlam görünüyor. Gelir gelmez kurdum. Ortasında şemsiye için delikte mevcut. Tek sorun taşıma kulpunun ve bağlı olduğu çerçevesinin taşırken esnemesi. O alanın metali daha sağlam olsaydı uzun ömürlü olurdu.
#20	Ürünü sepetteki ürünlerimle indirimli olarak aldım, fiyat performans olarak oldukça iyi, yeterli bir aydınlatma seviyesi var. Bildiğimiz fenerlerden daha etkili ancak tabii ki km'lerce aydınlatacak bir ürün değil. Kutusu vs yok, bir patpat'a sarılı geldi. 3 adet AAA (ince) pil ile çalışıyor, içinden pil çıkmıyor. Herhangi bir kalitesizlik hissetmedim ama sürekli zoom out/in yapılırsa ucundaki bu mekanizma ne kadar dayanır göreceğiz. Garanti beklentiniz olmasın zira garanti namına herhangi bir belge yok. Amacını fazlası ile görüyor, evde çektiğim resimlerden ışiksiz bir ortamdaki aydınlatma seviyelerini görebilirsiniz.
#21	Ürün iyi hoş da, boyutu çok küçükmüş, ben resimde daha büyük sanmıştım, açıklamalarda yazıyormuş , 10,5cm olarak, ama gerçekten ufak, yani yorum resminde yanına bildik bir kalem de koyuyorum ki kafanızda daha rahat şekillensin. güzel paketlenmiş bir şekilde elime ulaştı. yanındaki kalem daha uzun bu arada, paketten belli olmayabiliyor. Ben aldığımda 2km menzilli yazıyordu ve diğer yorumlarda da bunu görebilirsiniz, öyle olmadığı için de düzeltmişler ve uzun menzilli yapmışlar. 2km değil ama 25 metre menzili vardır herhalde. daha ötesini beklemeyin. görünce beklemezsiniz zaten. neden 5 yıldız peki? uygun fiyatla işini

	yapıyor, fiyatı da çok uygundu o yüzden aldım. yoksa fiyat performans olarak çok tavsiye etmem. ama bu haliyle 5 yıldız. ayrıca güzel paketlenme ve hızlı teslimat. daha ne olsun.
#22	Cok cok cok cok başarılı. Öylesine umudum yoktu ki ayaklarım için. O toynaga dönen ayaklarım normal insanlarınki gibi oldu. Şahane.
#23	Uyguladıktan 4. günde soyulmalar başladı çok sorunlu bir ayaktı bizimkisi süper sonuç aldık yaklaşık yavaş yavaş soyulduğu için 1 hafta 10 gün sonrasında tertemiz oldu. Kesinlikle tavsiye ediyorum
#24	Selam Canlar. Öncelikle ben bir ürün alırken önce çok satılanlara göre sıralıyorum. Ürünün özelliğinden çok yorumlara ve puan ortalamasına dikkat ediyorum. Bu ürününde yorumlarını epey okudum. İlk defa kablosuz kulaklık aldım. 25 Kasımda sipariş vermiştim. Bugün (11 Aralık) elime ulaştı. Sanıyorum Efsana Cuma muhabbeti nedeniyle gecikmeler oldu. Ürüne gelirsem. Gelir gelmez tek kulaklık ile denedim. Müzik kalitesi gerçekten muazzam. Hatta dalgınlıkla son ses açtım. Kulağım patlayacak sandım aniden çıkardım. Dışarıya dahi çok ses veriyor. Orta derece sesle yaklaşık 2 saat müzik dinledim. Telefonun şarjı da fazla gitmedi. Herkes çok beğendi. Hatta bir arkadaşım da bugün kendine sipariş verdi. Bunca zaman boşuna kablolu kulaklık kullanmışım bunu anladım. Hareket durumum rahatladı. Kulağa da güzel oturuyor. Müzik kalitesi süper. Gelelim konuşmaya. Akşam iş çıkışı kalabalık sayılabilecek bir kafeye gittim. Deneme için annemi aradım. Sesimi iyi alabildiğini söyledi. Ama etraf kalabalık geliyor dedi. 5 dakika kadar konuştuk. Sıkıntı veya anlaşılama olmadı. Yani kalabalık bir ortamda da konuşulabiliyor. Anlaşıyor. Ev ortamında keyif halinde konuşmalar yapılabilir. Özetle iyiki almışım diyorum. Alacaklara da tavsiye ediyorum. Umarım yorumum faydalı olur da birileri de benim gibi yorumlara güvenerek alıp iyi ki almışım der. Saygılar arkadaşlar...
#25	Başta orijinal olup olmadığından şüpheliydim fakat %100 orijinal test yaptım. Kulağa takılan silikon kısmını çıkarın orada çok küçük beyaz renkle yazılmış bir ID kodu bulunuyor. Bu şekilde orijinal olup olmadığını anlayabilirsiniz. Gözünüz kapalı alabilirsiniz çok kaliteli ve güçlü sesi var fakat kafamda bere varken telefonla konuştum ve sesim buğulu geliyormuş biraz. Ama sorun değil zaten beni kimse aramıyor...
#26	Oldukça şık, sade ve fonksiyonel bir ürün. Eşim ve kendim için dün sipariş ettik bugün ulaştı. Şu an şarj ediyorlar ama bir yandan kendi bilekliğimi eşleştirdim. Bir güncelleme yaptı ve ürün otomatik olarak Türkçe oldu. İlk izlenim oldukça iyi oldu. Günlük yürüyüşleri ve kalp atışlarının takip özelliği istediğim özelliklerdi ama uyku takibi özelliği sonradan öğrendiğim bir güzellik oldu. Su geçirmemesi de benim gibi duşta veya denizde saati çıkarmayanlar için çok faydalı bir özellik. Bir süre kullandıktan sonra da ilave görüşlerimi aktarmak isterim. Ben de pek çok video seyrettikten ve okuduktan sonra satın aldım. Şimdilik sahip olduğu özelliklerin bana yeteceğini düşünüyorum. İnceleme yazılarında aktarıldığı gibi aslında bir bileklikten çok akıllı bir saat sahibi oluyorsunuz. Aynı segmentte bu fiyat aralığındaki ürünlere göre şarjının uzun gitmesi sanıyorum en büyük artısı. Alacak olan arkadaşlara şimdiden iyi günlerde, sağlıklı kullanmalarını diliyorum. İyi günler.
#27	Merhaba...Açıkçası ürünü alırken çok tedirgindim. Çok araştırdım en sonunda buradaki ürüne karar verdim. Gece saat 2 de siparişi verdim ve sabah 11 de evime geldi. Evime gelen en hızlı ürün kargosu oldu bu da Ürünü kullanmadan önce herkes şarj ediyor ve öyle kullanıyor. Bu yanlış bence. Ben ürünü aldığımda %52 falandı galiba şarjı, ben saati kurcaladım sürekli. Ayarıydı şuyuydu buyuydu derken şarjı 0'a inene kadar kullandım. Yaklaşık şarjı 5 gün gitti o şekilde ve kapandı. Daha sonra şarjı fulledim, şu an şarjı %49. Bilekliği 16 gün önce şarj ettiğim ve şu an 49'da, gerisini siz düşünün. Bir defa bileklikle duşa girdim denemek için, su da geçirmiyor bunu da test etmiş oldum. Teşekkürler firmaya
#28	Sanırım dünyadaki en iyi akıllı telefon ÜRÜN oldu. Kamerası profesyonel kamerayı aratmıyor. Şarjı çok iyi gidiyor. Normal kullanımda 2 gün rahat rahat kullanabiliyorsunuz. 3. Gün öğlen saatlerinde şarj cihazını arıyorsunuz. Bu zamana kadar çıkan tüm MODEL modellerini kullandım ama 11 hepsinin ötesinde olmuş. Menü geçişleri ve internet akışı olağanüstü hızlı. İki sim kart takılması, iki hat ile arama yapabilme seçeneği eklenmesi büyük sürpriz oldu. Cam ve aliminyum malzeme ile plastik bir telefonu tutmak arasındaki farkı dibine kadar hissediyorsunuz. Çıkış fiyatı Türkiye şartlarımızda tabi ki herkesin alabilmesini engelliyor ama ödenen paranın hakkını da sonuna kadar veriyor ki belirli bir süre kullanıp satsanız bile, temiz kullandıysanız maddi olarak sizi yine üzüyor. Bu telefonlar altın gibi. İmkani olan hiç düşünmeden alsın. Almış olanlar güle güle kullansın. Almak isteyip de alamayanlara da nasip olmasını dilerim

#29	Arkadaşlar cihazı yaklaşık 1 haftadır kullanıyorum kargo elime 1 gün sonra ulaştı hepsijet tarafından cihaza MODEL dan geçiş yaptım ve kırmızı rengini aldım bir erkek olarak cesur bi renk gibi duruyor olabilir ama hep olumlu tepkiler aldım ve gerçekten çok şık duruyor . Ekran olarak handikaplı benim için tedirgin oluyordum ya iyi değilse diye gerçekten çok iyi hatta MODEL kullanan arkadaşımla yaptığımız karşılaştırmalarda düz MODEL her kulvarda daha önde ekran farkını hissetmiyorsunuz bile ki ön kamerası ondanda iyi 12 mp, yeterince büyük zaten MODEL ebatları kaba duruyor bu ideal ayrıca şarj konusunda test ettim dolu dolu 2 günü çıkartabilir çok yoğun oyun oynasanız bile 1 günü şarj etmeden çıkartabilirsiniz sadece arama ve mesajla 2,5- 3 günü zorlayabilir harbiden bu sefer yapmışlar evet fiyatı mâlesef yüksek ama muadillerine göre misliyle hak ediyor kalite ülkemizde ucuz olmuyor. Cihazı farklı çin cihazlarıyla kıyaslayanların daha önce bu markayı kullanmadığı çok bariz stabil 5 sene tak demeden çatır çatır kullanabileceğiniz sürekli güncelleme ile yenilenen bir cihaz istiyorsanız tercih edebilirsiniz .. Not:MODEL 2015 te almıştım, MODEL 2020 de, 2025 te görüşmek üzere
#30	5w'luk klasik adaptörden sonra ilaç gibi geldi. MODEL'imi 30 dakika gibi bir sürede yaklaşık %50 şarj etti. Priz başından kurtuldum gerçekten büyük rahatlık. Ürünü alırken Satıcının SATICI olmasına özellikle dikkat ederek aldım. KARGO kargo ile 1 günde geldi. Hemen kutunun üzerindeki ve adaptörün üzerindeki seri numarasını kontrol ettim ikisinde birbiriyle uyumlu ürünün orjinal MARKA olduğundan emin oldum. SATICI gerçekten güvenimi boşa çıkarmadı. Orjinallik açısından içinizde kuşku kalmasını istemiyorsanız kesinlikle Satıcının SATICI olmasına dikkat edin. SATICI aldığımız ürünü gönül rahatlığı ile kullanabilirsiniz.
#31	MODEL cihazım var. Kutusundan eski tip 5w şarj cihazı çıkmıştı ve 4 saate yakın şarj süresi olabiliyordu. Akıllılık yapıp sadece 20 w adaptörü aldım. Bir de usb yi typ c ye çeviren dönüştürücü aldım. Fakat adaptör 20w olmasına rağmen eski kabloyu dönüştürücü ile kullandığım için şarj süresi kısalmadı neredeyse aynı sürede şarj yapmaya devam etti. Koskoca Apple dönüştürücü ve eski kabloyla hızlı şarj deneyimine müsade eder mi ? Etmez... Orijinal type c kablosunu da alman lazım diyor resmen. Eee mecbur onu da aldım. Şarj süresi hiç dokunmazsanız: MODEL da : tamamen kapalıyken yarım saatte %50, 1 saat 50 dakikada %100 yaptı. MODEL de : tamamen kapalıyken yarım saatte %50, 2 saat 20 dakikada %100 yaptı. Batarya kapasiteleri farklı olduğu için süre değişkenlik gösteriyor. Ayarlardan : İyileştirilmiş pil şarjı aktifken bu değerler oluşuyor. Seçeneği Kapatırsanız belki daha hızlı olabilir. %80 e kadar hızlı şarj yapıp sonra yavaşlıyor. Hızlı şarj olurken Telefon çok az ısınıyor. Sonuç olarak hem kablo hem de 20w adaptör alırsanız hızlı şarjın faydasını görebilirsiniz... normalde siparişin ertesi günü ürün elimde olurdu ama hepsijet bu sefer biraz gecikti.. alacaksınız farklı satıcılardan değil SATICI alın. Diğerlerinin orijinal olup olmadığını bilemezsiniz..
#32	Ürünü başka siteden aldım. Bir hafta oldu, çok memnunum. Paket içerisinde 2 döner fırça (biri yedek, tek fırçalı makine bu arada), Hepa filtreli toz tankı, yine hepa filtreli 2 si 1 arada (suvetoz) toz tankı. 2 tane microfiber bez (biri süpürgevepaspas modu, diğeri paspas modu için, son olarak microfiber bezi taktığımız aparat. 2100 mpa çekiş gücü bulunmakta, 4 modu mevcut sessiz, standart, orta ve turbo. Tahmin edebildiğiniz gibi turboya doğru daha fazla ses yapmakta, şarj yemekte ve çekiş gücü yükselmektedir. Haritalama olayı çok güzel, yeneden adlandırma bölme birleştirme olayları çok pratik. Her yemekten sonra mutfağa gönderiyorum, 2 tur yaptırıyorum tertemiz yapıyor. Koltuk ünite yere sıfır değilse, eviniz düz ayaksa ve çok eşyanız yoksa ideal. Bizde eşya çok, 2 tane de çocuk var, çalıştırmadan önce mutlaka topluyoruz evi. Yoksa çorap kablo oyuncak takılıyor. Ev temizken bile çalıştırdık bir ton toz çıktı. Bazı parçaları almaması sizi hayal kırıklığına uğratmasın, sonuçta harita bazlı çalışıyor, önemli olan tozu toprağı topluyor. Yanına da hava temizleme cihazı aldım tozdan eser yok.
#33	Bir aydır kullanıyoruz, çok memnunuz. Kesinlikle tavsiye ediyorum. İlk birkaç çalışmasına göre yargıya varmayın. Evin haritasını çıkarıp tanıdıktan sonra gerçek performansını görüyorsunuz. İlk çalışmalarda karışık bir çalışmayla işi çok uzattı, bir şarjla bitiremedi temizliği ama 2-3 çalışma sonrası ve şimdilerde daha hızlı ve tekrar şarja dönmeden tüm evi bitiriyor. Evim 180 metrekare. Süpürmesi de silmesi de şahane. Tabii yerdeki lekeleri güç uygulayarak çıkarmıyor, sadece yüzeysel bir silme. Ama evdeki toz inanılmaz azaldı bir ayda. Oğlumda alerjik geniz akıntısı vardı, 20 gündür geniz açık. Eşimdeki kuru öksürük kesildi.

	<p>Her gün çalıştırıyorum bir avuç toz çıkıyor. Önceden haftada 1 süpürge açardık, hepimizin alerjik olmasına şaşmamalı ☺</p> <p>Önerim ilk çalışmada kenarları süpür özelliği ile çalıştırın, evi haritalasın sonra tam temizlik yaptırın.</p> <p>Odaları bizim mantığımızla oda oda ayırmıyor. Harita oluştuktan sonra siz app uygulaması üzerinden odaları ayırıp adlandırıp, oda oda temizlemesini sağlayabilirsiniz.</p>
#34	<p>urunu çok araştırarak aldım ama burdan değil bi mağazadan gelir gelmezde daha önce supurdugum haliyi supurdum ve çıkan toza inanmadım o kadar iyi emis gücü. inanılmaz sessiz olusu da çok iyi. kablosu çok uzun kolidora takmamla 140 metre kare evin her odasını dip kose supuruyor bu çok iyi.. icindeki baslıklar inanılmaz rahatlık sağlıyor her ture uygun baslık var.. toz haznesi çok pratik dolunca hop cope bu da çok güzel.. saptan kumandali olusu da çok iyi şimdi bunlar olumlu özellikler. gecelim olumsuz özelliklere 1. si supurge çok ağır ne kadar pratik olursanız olun kapılara köşelere takilip kalıyor ve hemen oyle kayıp gelmiyor bu baya zorluyor. 2. emiz gücü çok yüksek olduğu için normal supurge baslıgiyla supurmek çok zor asiri yoruyor kolumu dahi incilltim ve zorlandığım için verimli supuremedim. 3. turbo firca baslıgi var onu taktim onla daha kolay supurdum hem icindeki donen fircalar haliyi dovme islevi goruyor daha fazla toz alıyor fakat normalde sesi çok az çıkan supurgenin turbo baslıgi takınca cikardığı ses asiri ve rahatsız edici bu yuzden sessizlik ozelligi kullanışlı olmuyor. 4. filtresi hemen toz oluyor ve sürekli temizleyip yıkamak gerekiyor. 5. kablosu bi yere takılırsa otomatik kendini topluyor ve bu her seferinde kabloyu tekrar acmama neden oluyor.. iyi ve kötü yanları var evet ama fiyatına göre bence alınabilecek bir ürün ben kendimce sevip sevmediğim yönlerini yazdım alacak olanlar ona göre alsın.</p>
#35	<p>Fiyatına bakmaksızın ne alırsam alayım kılı kırk yaran insanlardanım almadan önce 2 ay araştırdım şikayetleri kadınlar kulübünde bu ürünü kullananları farklı sitelerdeki yorumları tek tek okudum</p> <p>Alalı 1 ay oldu gözlemlerim şöyle</p> <p>sessiz çalışma olayı en kısıt kademe de çok sessiz onun dışında normalinde de bir elektrikli süpürgeye daha sessiz zaten</p> <p>çektığı tozları un ufak yapıyor yumuşatıyor resmen benim yıkamadan gelen halılardan çıkan tozları görmeliydiniz</p> <p>turbo başlığı</p> <p>koldan ayarlama çok kullanışlı</p> <p>büyük ama hantal değil</p> <p>Alacak olanlara tavsiye ediyorum</p> <p>4 yıldır bu siteden alışveriş yapıyorum hb ailesine teşekkür ederim</p> <p>not: ben bir ürün alırken babam hep şunu der kızım yorumlara güvenme onların kendileri yazıyo der)</p> <p>sizde babam gibi düşünüyorsanız diye yazdım</p> <p>inşallah yazdıklarım yardımcı olur sizlere</p>
#36	<p>Almayı düşünen arkadaşlar lütfen içinizde ukde kalmasın. Ürün gayet kullanışlı bir ürün. Tavsiye ederiz. Günlük kullanımlarda çok başarılı bir kullanım sergiliyor. Uzun zamandır Philip Azur serilerini kullanıyoruz ve bu sefer yine yanıltmadı Philips. Ütü için çok ağır, şu haznesi yetersiz, rengi siyah olduğu için içini göremiyoruz gibi bahaneler yazılmış, ancak hiç de öyle değil. Günlük kullanım için gayet yeterlidir. Çok çabuk ısınıyor, buhar gücü ve anlık basınçlı buhar üretimi memnun ediyor. Kablo uzunluğunun diğerlerine göre fazla olması çok büyük kolaylık sağlıyor. Tavsiye ederiz. Memnun kalırsınız.</p>
#37	<p>Ütü almak için çok bi fikriniz yoksa, çeyizim için almak istiyorum ama karar veremiyorum diyosanız bu yorumu sizin için yazıyorum.</p> <p>En üst modellerden en alt modellere kadar fiyat/performans kıyaslaması yaptım. Buhar kazanlı mı alsam ? yoksa sadece buharlı ütü yeter mi diye düşünüp durdum. Bütün sitelerde olumlu olumsuz tüm yorumları okudum. Bu ürüne karar verdim.</p> <p>1- Ütü hepsi jet ile geldi. Ütünün kutusunun dışında bir koli daha vardı. İçinde korumak adına patpat vs. yoktu. Ama bi zarar görmemiş çok şükür. 1 günde Ankaraya teslim edildi.</p> <p>2-Ütüyü açtım açar açmaz kötü yorumlarda gördüğüm çizikler vs. var mı diye baktım gayet sağlamdı.</p> <p>3-Kablosu çok uzun gayet güzel ekstra uzatma çekme gerek kalmayacak.</p>

	<p>4-Hemen açıp denemek istedim. Sonra temizleyip çeyizime kaldırıcam çünkü. Su haznesi dedikleri kadar küçük değil. Babamın 4 gömleğini 1 pantolonunu ilk doldurduğum suyla ütüledim. Öyle çabucak bitmedi. Hemde otomatik buhar ayarındayken kullandım.</p> <p>5- Daha önce evde kullandığımız ütü tefal'in normal bi ütüsüydü.1 gömleği nerden baksanız 7-8 dk ütülüyorduk. Ama bu üründe 3 dk almadı ütölemek.</p> <p>6- Bahsedildiği kadar ağır bi ütü değil bende 50 kilo bi kızım sonuçta öyle bilek güzü kas gücü vs. gerek yok. Ütünün bence biraz ağır olması gerekiyor. Bu ağırlı sayesinde kırıksıklıları daha güzel eziyor.</p> <p>7. Hemen ısınması çok hoş kazanlı ütülerde gibi uzun süre beklemeye gerek kalmıyor.</p> <p>8-Kireç çözme özelliğine bakmak için hazneyi çıkarttım gerçekten kireç topluyormuş.İlk kullanımda bile hafif kireç tortusu vardı.</p> <p>Özet olarak almak isteyen varsa ve üst modellere o paraları vermek istemiyorsanız bence güzel bi ütü. Bakabilirsiniz.</p>
#38	<p>Ürünü 8 haziranda aldım. İyice bir kullanayım ona göre yorum yapayım dedi.</p> <p>Ürün ile oynadığım oyunlar:</p> <p>Ets2 (27 saat)</p> <p>Rocket League (28 saat)</p> <p>Kingdom Come Deliverance (67 saat)</p> <p>Assassins Creed 2 (10 saat)</p> <p>Assassins Creed Chronicles China (6 saat)</p> <p>Life is Strange (2 saat)</p> <p>Child of Light (8 saat)</p> <p>Gris (2 saat)</p> <p>Önce artılarından bahsedeyim</p> <p>Dokusu çok hoş. Ele oturuyor. Kaymıyor. Tak çalıştır. Herhangi bir driver kurmanıza gerek yok. Kablosu örgü kablo. Kolay kolay yıpranmaz. Bu da size uzun bir kullanım sunar. Kablosu yeterince uzun. Analoglarda parmağınızı koyduğunuz yerler oyuk yapılmış parmağınız oturuyor oraya. Üründe iki adet titreşim motoru var bu ayrı bir hava katıyor oynadığımız oyuna. Bilgisayar ürünü MARKA kolu olarak algılıyor. Ortada bulunan "P3" tuşuna 3 saniye basarsanız mod değiştiriyor. Hiç kullanmadım ama o da iyi bir özellik. R2 ve L2 tuşları biraz hassas. Örneğin bir yarış oyununda ne kadar basarsanız o kadar hızlanıyor araç. Bu da hoş. Bir de arkasında reset deliği var eğer kolda bir sıkıntı olursa bir kaç saniye basılı tutarak resetleyebilirsiniz.</p> <p>Şimdi gelelim eksilerine</p> <p>Öncelikle tuşlardaki semboller playstation kolundakine benziyor ama pc kolu MARKA kolu olarak algıladığı için siz oynarken MARKA koluymuş gibi oynamak zorunda kalıyorsunuz. O biraz kafa karışıklığı yapıyor ama ben çok az kol kullanmış biri olarak 1-2 saatte alıştım bir daha da oynarken tuşlara bakmadım. Muhtemelen sizin içinde aynıdır olur.</p> <p>Diğer eksi ise gerçekten can sıkıcı. Eğer kol çok titirse kolun pc ile bağlantısı kopuyor. Ben Rocket League oynarken yaşadım bunu birkaç kere. Sonra ayarlardan titremeyi kapatmak zorunda kaldım. Ama diğer oyunlarda yaşamadım. Çok uzun sürmeyen titreşimlerde sıkıntı olmuyor ama 2 saniye falan sürünce bağlantı kopuyor.</p> <p>Genel olarak bakacak olursak ben aldığıma pişman değilim. Tabi bu fiyata daha f/p bir ürün var mı bilmiyorum. Ama parasına değer bence. Artılar ve eksilerden aklıma gelenleri yazdım. Eğer almak isterseniz kolayca karar verebilirsiniz diye.</p>
#39	<p>Bir ürün satın alırken çok detaylı araştırma yaparak alırım bu kol ise tam bir F/P ürünü diyebilirim. Öncelikle kol güzel bir şekilde kargolanmış ve özel kutusu ile elime ulaştı kutu içeriğinde ayrıca sürücü CD'si de yer alıyordu. PC'ye takar takmaz çalıştı bunun haricinde internetten sürücüsünü de indirirseniz titreşim ayarları vs... hepsi sorunsuz çalışacaktır ve herhangi bir problem yaşamadım bağlantı kopma, oyundan atma gibi. Kol görünüş itibarı ile çok şık ki görünüşünden çok daha hafif emin olun bunun haricinde tuşları hafif sert gibi gelse de malzeme kalitesi iyi gibi çok az plastik kokusu bulunuyor. Ayrıca ortada ki P3 tuşuna 3 saniye basılı tutarsanız mod değiştiriyor ki orjinal MARKA kolu olarak algılıyor PC aynı zamanda. İndirimdeyken almıştım, profesyonel kola gerek yok dersanız daha iyisini bulamazsınız tavsiye ederim.</p>

ÖZGEÇMİŞ

Adı Soyadı: Oğuzhan ARI

Oğuzhan ARI, 2019 yılı temmuz ayında Sakarya Üniversitesi Yönetim Bilişim Sistemleri Bölümü'nden fakülte derecesi ile mezun oldu. 2020 – 2022 yılları arasında Doğuş Üniversitesi Yönetim Bilişim Sistemleri bölümünde Araştırma Görevlisi olarak çalıştı. 2022 Şubat ayı itibariyle Sakarya Üniversitesi Yönetim Bilişim Sistemleri bölümünde Araştırma Görevlisi olarak çalışmaktadır.

ÖĞRENİM DURUMU

Derece	Eğitim Birimi	Mezuniyet Yılı
Yüksek Lisans	Sakarya Üniversitesi/İşletme Enstitüsü/Yönetim Bilişim Sistemleri	Devam ediyor
Lisans	Sakarya Üniversitesi/İşletme Fakültesi/Yönetim Bilişim Sistemleri	2019
Lise	Manisa Hasan Türek Anadolu Listesi	2014

İŞ DENEYİMİ

Yıl	Yer	Görev
2022-Halen	Sakarya Üniversitesi	Araştırma Görevlisi
2020-2022	Doğuş Üniversitesi	Araştırma Görevlisi

YABANCI DİL

İngilizce

EŞERLER

1- Akbıyık, A. ve Arı, O. (2022). *Lojistik Regresyon ile Faydalı Müşteri Yorumlarını Tahminleme* . Journal of Research in Business , IMISC 2021 Special Issue , 15-32 . DOI: 10.54452/jrb.1024602

2- Halil İbrahim Cebeci, Mustafa Yılmaz, Oğuzhan Arı - Transformation of Made in China Perception in Turkey: An Evaluation Based on Sentiment Analysis - 2021