

METİN MADENCİLİĞİNİN TALEP PLANLAMADAKİ ROLÜNÜN İNCELENMESİ

Erkan İŞIKLI^{1*}

¹İstanbul Teknik Üniversitesi, İşletme Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul

ORCID No: <http://orcid.org/0000-0002-8319-8782>

| Anahtar Kelimeler | Öz |
|--|--|
| İş analitiği, Metin madenciliği. Talep planlama. Yapılandırılabilir ürün | Yapılandırılabilir ürünler, önceden tanımlanmış ürün özelliklerinin çeşitli kombinasyonlarıdır. Bu ürünlerin imalatçıları için uzun dönem kârlılık ve müşteri memnuniyeti açısından en parlak ürün çeşitlerini belirlemek, özellikle stratejik bir öneme sahiptir. Çok sayıda ve derinlikte öznitelige sahip olan yapılandırılabilir ürünler, teorik olarak milyonlarca farklı yoldan oluşturulabilir. Bu yüzden talep tahmin veya stok modelleri gibi işlemsel modellerde kullanılacak ürün özelliği sayısının uygun şekilde azaltılması değerli bir çabadır. Bu çalışma, işlemsel modelleri "boyutsallık lanetine" maruz bırakabilen geleneksel yaklaşımların nitel yöntemlerle melezlenmesi sayesinde nasıl daha kolay ve doğru öngörülerde bulunulabileceğini göstermeyi amaçlamaktadır. Yapılandırılabilir bir ürün için müşteri değerlendirmeleri kullanılan mevcut çalışmada, büyük miktarda kritik bilgi kaybetmeden ürün özelliği sayısını azaltıp talep planlama sürecinin nasıl desteklenebileceği tartışılmıştır. Bu amaçla çeşitli metin madenciliği tekniklerine başvurulmuş, elde edilen sonuçlar ve ele alınan yapılandırılabilir ürünün talebini nicel yöntemler kullanarak modelleyen bir referans çalışmanın sonuçları karşılaştırılmıştır. |

INVESTIGATING THE ROLE OF TEXT MINING ON DEMAND PLANNING

| Keywords | Abstract |
|---|---|
| Business analytics, Text mining, Demand planning, Configurable products | A configurable product is defined as a combination of predefined product attributes. Identifying the most promising variants of such products in terms of long-term profitability and customer satisfaction has become a central issue for their manufacturers. In theory, configurable products that have numerous attributes with too many levels can be created in millions of different ways. Hence, properly reducing the number of product attributes to be considered in an operational model (e.g., demand forecasting model, inventory model) is therefore worth the effort. This study intends to demonstrate how easier and more accurate predictions can be made with the incorporation of qualitative methods into traditional approaches, which may cause operational models suffer from the "curse of dimensionality." Using customer reviews for a configurable product, the present study discusses how to support the demand planning process by reducing the number of product attributes without losing too much critical information. Various text mining techniques were thus employed, and the results obtained are then compared with those from another study that modeled the demand of the configurable product in consideration using quantitative methods. |

| | |
|-----------------------------|------------------------------|
| Araştırma Makalesi | Research Article |
| Başvuru Tarihi : 18.09.2020 | Submission Date : 18.09.2020 |
| Kabul Tarihi : 28.06.2021 | Accepted Date : 28.06.2021 |

*Sorumlu yazar; e-posta : isiklie@itu.edu.tr

1. Giriş

Hangi ürün çeşidinin sunulacağına karar vermek, şirketlerin özellikle satış ve pazarlama departmanları için her zaman zor bir problem olmuştur, çünkü bu karar, şirket içindeki diğer her işlevi etkileyebilir. Günümüzün küreselleşmiş ekonomilerinde ürün çeşitliliği yönetilirken genellikle iki ateş arasında kalınmaktadır. Tüketiciler alışverişe çıktıklarında, tam olarak akıllarındaki ürünü bulmaya isteklidirler; bu yüzden çeşitliliğe önem verirler. Bu çeşitlilik arama davranışı, tüketici isteklerine karşılık verebilmek için şirketlerin özelleştirilebilir ürünleri daha fazla sunmaya çalışmasıyla sonuçlanır. Ancak çok fazla çeşitlilik sunmak, talep değişkenliğini, üretim işlemlerinin karmaşıklığını, tahmin hatalarını, envanteri ve ürün noksanlıklarını artırabilir. Dahası, bütün bunlar kısmen veya birlikte pazar uyumsuzluğu maliyetlerinde artışa yol açabilir (Fisher, 1997; Ramdas, 2003). Çok sayıda seçenek arasından seçim yapmak tüketiciler için, kaç değişik ürün seçeneği sunacaklarına karar vermekse perakendeciler ve üreticiler için zordur. Yüksek düzeydeki çeşitliliğin şirketlere pazarda rekabet avantajı sağlayabileceği gösterilmiş olsa da (Berger, Cunningham ve Drumwright, 2007), bu çıkarım üzerinde hâlâ bir fikir birliği yoktur (Trattner, Hvam, Forza ve Herbert-Hansen, 2019). Bu nedenle, sabit maliyetlerini karşılamak, pazarda hayatta kalmak ve diğerleriyle rekabet etmek için talebin yeterince büyük bir kısmını ele geçirirken artan çeşitlilikten kaynaklanan maliyetleri düşürmek, şirketler için kritik öneme sahiptir. Dolayısıyla şirketler, gelir, üretim maliyetleri, daha doğru talep tahminleri ve tüketici sadakatinin uzun vadeli etkileri açısından çeşitliliği yönetmenin yollarını aramaktadırlar.

Tüm imalatçı firmalar, hangi ürün çeşidini oluşturacaklarını belirlemede birbirleriyle yarışan hedeflerle karşı karşıya kalsalar da bu sorun, yapılandırılabilir ürün imalatçıları için özellikle daha zorlayıcıdır. Çok sayıda özellik içeren bu karmaşık ürünler için tüketicinin oluşturulabileceği tüm ürün çeşitlerini içeren konfigürasyon uzayı fazlasıyla büyüktür. Bilgisayarlar ve otomobiller gibi yapılandırılabilir ürünler, önceden tasarlanmış bileşenlerin bir kombinasyonu olarak tanımlanır. Günümüzde tüketiciler, çeşitli platformlarda sunulan satış yapılandırıcıları yardımıyla ilgilendikleri ürünün akıllarındaki çeşidini yaratırlar. Bir yapılandırılabilir ürün imalatçısının sunabileceği muhtemel yapılandırma sayısı, mevcut

ürün özelliklerinin bir fonksiyonu olarak katlanarak artar. Örneğin, her birinin iki opsiyonu olan 10 ürün özelliğine sahip bir ürün, teorik olarak, 1024 (yani 2^{10}) farklı şekilde özelleştirebilir; ikili opsiyon sayısı 20 olduğunda, bu sayı 1.048.576 (yani 2^{20}) olmaktadır. Ancak bireysel mağazalar ve yerel pazarlardaki sınırlı envanter kanalı kapasitesi nedeniyle, olası yapılandırmaların belli bir anda yalnızca küçük bir kısmı kullanılabilir. Buna rağmen 2006 yılında Pontiac, sadece 300.000 araç satmayı bekliyorken, o zamanlarda ABD'deki sürücü sayısından (yaklaşık 200 milyon kişi) daha fazla yapılandırma sunabildiklerinin reklamını yapmıştır (Ferguson ve Donndelinger, 2010).

Bir yapılandırılabilir ürünün talebini tahmin etmek çok karmaşık bir iş olabilir. Talep planlama süreci her sektörde zorlu olabilir ancak yapılandırılabilir ürünlerle çalışırken geleneksel modeller, kestirilecek parametrelerin bolluğundan ve boyutsallık lanetinden dolayı zarar görebilir. Açıklayıcı değişken olarak yalnızca ikili olarak kodlanmış ürün özelliklerini kullanan basit bir doğrusal bağlanım modeli kurularak böyle bir ürünün talebini tahmin etmek istendiğinde, testin gücü göz önünde bulundurulmazsa, 10 özelliğe sahip bir ürün için en az 70, 20 özelliğe sahip bir ürün içinse en az 120 gözleme ihtiyaç duyulacaktır (Tabachnick ve Fidell, 2019). Büyük veri çağında bu sayıda gözlem toplanması ilk bakışta zor bir iş gibi görünmeyebilir, ancak gerçekte durum farklıdır. İlk olarak, bu çok basit bir model olduğundan isabetli tahmin gücü düşük çıkacaktır. İkincisi, böyle bir model ürünün satışa sunulduğu yer ve zamandan bağımsız olarak talebi tahmin etmeye çalışmaktadır. Açıklayıcı değişken sayısını artırmak ve daha karmaşık bir model tercih etmekse çoklu doğrusallık gibi sorunlar ortaya çıkacaktır. Dolayısıyla gözlem sayısının belirgin bir biçimde artırılması gerekecektir. Örneğin, çoklu doğrusallık sorununu aşmak için adimsal regresyon kullanılması halinde, yukarıda bahsedilen ilk durumda 400, ikinci durumda 800 gözleme ihtiyaç duyulacaktır (Tabachnick ve Fidell, 2019). Ancak gözlem sayısının yaklaşık altı kat artmasına yol açacak bu model bile talebi doğru tahmin edebilecek kadar iyi değildir çünkü zamansallık (örneğin, mevsimsellik) veya bölgesellik (örneğin, satış yapılan şehir) gibi etkiler hâlâ göz önünde bulundurulmamaktadır. Bu etkilerin hesaba katılması için gözlem sayısının onlarca kat artırılması gerekecektir ancak bir ürün çeşidinin başka bir ürün çeşidiyle aynı veya farklı yerlerde/zamanlarda satışa sunulmuş olup

olmaması yine de göz önünde bulundurulmamış olacaktır. Son olarak, açıklayıcı değişkenlerin ikili olarak kodlanması dolayısıyla yararlanılabilecek alternatif istatistiksel modeller de sayıca azdır ve hepsi kendine özgü kısıtlar getirmektedir. Örneğin, Temel Bileşen Analizi veya Kısmî En Küçük Kareler gibi yöntemlerin bu tipte açıklayıcı değişkenlerle kullanılması önerilmemektedir. Özetle, kritik öneme sahip büyük miktarda bilgi kaybetmeden işlemsel modellere dahil edilecek ürün özelliklerinin sayısını azaltmanın bir yolunu bulmak gerekmektedir. Özel olarak, şu iki temel soruya yanıt bulmak çok önemlidir:

1. Stok ve talep modelleri gibi işlemsel modeller oluşturulmadan önce, elde edilen önemli bilgiden mümkün olduğunca ödün vermeden, milyonlarca bile olsa ürün yapılandırılmaları hakkındaki çok sayıda kararın daha az sayıda teknik özelliğe indirilmesinin en etkili yolu nedir?
2. Tüketicilerin yapılandırmalar hakkındaki algılarına değer katan ürün özellikleri etkili bir şekilde nasıl saptanır?

Bir otomobil üreticisinin yaşam döngüsü kârlarının iyileştirilmesini sağlayacak "doğru" ürün yapılandırmaları yelpazesini belirleme problemini ele almış olan Işıklı (2012), bu ürünlerin dış çeşitliliğini analiz eden bir çerçeve önermiştir. Çalışmanın temel amacı, birincil ve ikincil talebi tahmin etmek ve yapılandırılabilir ürün yelpazesini optimize etmek için işlemsel modeller oluşturmadan önce göz önünde bulundurulması gereken ürün çeşidi sayısını azaltmanın bir yolunu sunmaktır. Büyük bir otomotiv orijinal ekipman üreticisinden elde edilen, birden fazla segmenti ve ABD pazarındaki araçları kapsayan çoklu veri setlerinin kullanıldığı bu çalışmada, dış çeşitlilik hakkında, özellikle veri kaynakları sınırlı olduğunda, anlamlı ve uygulanabilir bilgilerin elde edilmesini sağlayacak bir dizi yöntem ve teknik önerilmiştir. Önerilen çerçevede, tüm mümkün ürün yapılandırmalarını içeren uzay çok büyük olabileceğinden, ürün çeşitlerinin yeniden tanımlanabilmesini sağlayacak boyutların belirlenmesi için her bir araç özelliğinin nispi karışım oranları kullanılarak temel bileşen analizi uygulanmaktadır. Nispi karışım oranı, belirli bir ürün özelliğinin, belirli bir zaman aralığında mevcut olan ürünlerin yüzde kaçında var olduğunu göstermektedir. Bu şekilde ürün özellikleri arasındaki etkileşimin dikkate alınmasının yanı sıra, benzer özelliklere sahip yapılandırmaların

kümelenebilirliği sağlanarak talep modelinde kestirilecek parametre sayısının da azaltılmış olacağı düşünülmüştür. Orijinal yapılandırmalardan sayıca daha az olan bu yeni oluşumlar "ürün varlıkları" ve "benzerlik kümeleri" olarak anılmıştır. Talep modellemesi aşamasında, orijinal yapılandırmalar yerine bu ürün varlıkları kullanılmıştır.

Işıklı (2012), ilgili problemde nitel ve nicel yöntemlerin bütünleşik kullanımını önermiş olsa da herhangi bir nitel yöntemle dair uygulama sonucu paylaşmamıştır. Daha sonra Işıklı ve Ketenci (2020), Işıklı (2012)'de ele alınan ürün için belli bir zaman aralığında yapılmış olan çevrimiçi tüketici yorumlarını sade ve basit metin madenciliği teknikleri kullanılarak incelemiştir. Mevcut çalışmada, söz konusu ürünün talep planlama sürecinin daha iyi yönetilebilmesi için Işıklı (2012)'nin önerdiği çerçevenin "Nitel Yöntemler" modülünün nasıl çalıştırılabileceği ayrıntılı olarak açıklanmıştır. Bu hedefle sunulan metodolojide, sonuç-odaklı metin madenciliği tekniklerine odaklanılmış ve Işıklı ve Ketenci (2020)'de sunulan yaklaşım yeni kaynaklar ve teknikler kullanılarak genişletilmiş ve zenginleştirilmiştir. Mevcut çalışmanın öncelikli amacı, büyük miktarda kritik bilgi kaybetmeden ürün özelliği sayısını azaltıp talep planlamada kullanılan nicel yöntemlerin nitel yöntemlerle nasıl desteklenebileceğini göstermektir. Bu hedef doğrultusunda elde edilen sonuçlar, incelenen yapılandırılabilir ürünün talebini nicel yöntemler kullanarak modelleyen Işıklı (2012)'deki sonuçlarla karşılaştırılmıştır. İlgili literatürde daha ziyade geleneksel yaklaşımlarla analitik çözümler sunulmakta, melez yaklaşımlara pek sıklıkla rastlanmamaktadır. Mevcut çalışmanın, yapılandırılabilir ürünlerin talep planlama süreçlerinde faydalanılacak nicel yöntemlerin "boyutsallık lanetine" maruz kalmamasını sağlayacak esnek bir veri ön işleme yaklaşımı sunarak söz konusu boşluğu doldurması beklenmektedir.

Bu çalışmanın akışı şu şekildedir: İkinci bölüm, metin madenciliği üzerine kısa bir inceleme sunmaktadır ve işlemsel modellerin oluşturulmasında bir araç olarak kullanımının yararlarını tartışmaktadır; üçüncü bölümde, kullanılan metin madenciliği teknikleri ve önerilen metodolojik çerçeveye bunun Işıklı (2012)'deki ana çerçeveye nasıl oturtulduğu açıklanmaktadır; dördüncü bölüm, uygulama sonuçlarını sunmaktadır; elde edilen sonuçlar ve gelecekteki

araştırmalar için olası yollar hakkında bir tartışmayla çalışma sona ermektedir.

2. İlgili Literatür

Müşterilerin istek ve ihtiyaçlarını anlamak, iş dünyasının en önemli amaçlarından biri olarak sayılabilir. İnternetin günümüzde son derece yaygınlaşmış olması, müşterilerin satış sonrası düşüncelerini anlama imkânını artırmasının yanı sıra, bu durumu karmaşıklştırmıştır. Önceleri çoğunlukla anket çalışmalarıyla yetinmek durumunda kalan pek çok işletmenin, şimdilerde ellerinde müşterileriyle ilgili analiz etmekte zorluk yaşayabilecekleri boyutta bir veri kaynağı bulunmaktadır. Son yıllardaki teknolojik gelişmelerle birlikte, zengin ve çeşitli kaynaklardan gelen ham verinin kıymetli parçaların en uygun şekilde çıkarmak için çözümler sunan veri madenciliği alanı da genişlemiş ve iş dünyasının ilgisini çekmiştir. Veri madenciliğinin etkili kullanımı, günümüz işletmelerinin başarısı üzerinde önemli bir rol oynamaktadır. Şirketler, halihazırdaki müşterileri ve potansiyel müşterileri hakkında daha fazla bilgi edinmek ve onları memnun etmek için veri madenciliği tekniklerinden sıkça yararlanmaktadır.

Yeni teknolojilerin ortaya çıkması, verilerin hem hacmini hem de çeşitliliğini günden güne artırmaktadır. Uluslararası Veri Kurumu (*İng. kısalt.*, IDC), 2025 yılına kadar dünyada yaratılan ve çoğaltılan dijital veri miktarının, 2011'deki değerin neredeyse 90 katı olan 163 zettabayt (1 zettabayt = 1 trilyon gigabayt) bulacağını öngörmektedir (King, 2017). İşletmeler için önemli bilgiler çoğunlukla metin biçiminde saklandığından, metinsel veriler dijital çağda özel bir öneme sahiptir (Shilakes ve Tylman, 1998). Metin madenciliği, veri kaynağı metinler olan bir veri madenciliği uygulamasıdır. Temel olarak, bilgisayar zekâsı kullanarak müşteri çağrı merkezine gelen aramaların ses-metin çevirisi, gazete ve dergi yazıları, profesyonel raporlar (Linoff ve Berry, 2010), tüketici yorumları, tivitler ve sinema filmi incelemeleri gibi yapılandırılmamış verileri yapılandırılmış verilere dönüştüren metin madenciliği teknikleri, bir şirketin iç veri tabanlarında ve dış kaynaklarında bulunan zengin metin bilgilerini analiz ederek bunlardan faydalı bilgiler elde etmeyi amaçlar. Yapılandırılmamış büyük verilerin çağında, ileri teknolojiyi kullanarak, bilgi değeri çok fazla olan yapılandırılmamış verileri çözmek acil bir sorundur (Li, Gao, Xiang, Shi ve Li, 2018).

Müşteri değerlendirmeleri ve kullanıcı yorumları da metin şeklinde veriler olduğundan, bunların analizi için metin madenciliği oldukça yararlı bir araçtır. Az masraflı ve etkili pazarlama stratejilerinden biri olan ağızdan ağıza pazarlama, çevrimiçi yorumlardan fazlaca etkilenmektedir. Bu yorumların analiz edilmesi, müşterilerin tatmin ve memnuniyet düzeyini anlamak için en verimli yollardan biridir. Bunun yanında, metin kaynakları genellikle bir gazete sütunu, bir kitap veya bir akademik makalede karşılaşılabilecek doğal dilde metinler olduğundan, içerdikleri bilgiler makine ve bilgisayarların doğrudan yorumlaması için anlaşılmalıdır. Bu nedenle, bilgi çıkarımı ve bilgi keşfi için bu tür zengin verileri araştırmak önemli bir değere sahiptir.

Tüketiciler bir ürünü fiziksel olarak inceleyemediğinde, satın alma kararı vermeden önce diğer tüketicilerin o ürünle ilgili deneyimleri hakkında bilgi sahibi olmak ister. İnternet üzerinde yapılan tüketici yorumları, ürünle ilgili güçlü bir bilgi kaynağı olarak, tüketicilerin karar alma sürecinde hayati bir rol oynamaktadır (Xia ve Bechwati, 2008). BrightLocal tarafından yayınlanan "Local Consumer Review Survey" (Yerel Tüketici Değerlendirme Anketi) başlıklı rapora göre, tüketicilerin %82'si yerel işletmeler hakkındaki çevrimiçi tüketici yorumlarını okumaktadır; 18-54 yaş aralığındaki kişilerin %52'siyse her zaman yorumları okuduğunu belirtmiştir (Murphy, 2019). Podium (2017)'ye göre, tüketicilerin %93'ü çevrimiçi yorumların satın alma kararlarını etkilediğini onaylamaktadır. Ayrıca, tüketicilerin üçte ikisinden fazlasının, daha iyi bir deneyim elde edeceklerini bildiklerinde, aynı ürün veya hizmete daha fazla para vermeye hazır oldukları belirtilmiştir. Tüketici yorumları, şirketlerin bir tüketici segmentindeki heterojenliğin derecesini belirleme ve tüketicilerinin algısı ve davranışı hakkında iç görü kazanmaları için iyi birer kaynaktır. Ye, Law ve Gu (2008), çevrimiçi tüketici yorumlarının otellerin işletme performansı üzerinde belirgin bir etkisi olduğunu ortaya koymuştur. Bu yüzden bir nevi yeni-nesil sosyal ağızdan-ağıza pazarlama bileşeni olarak düşünülebilir (Lowenstein, 2014).

Metin madenciliği katlanarak büyüyen İnternet tabanlı bilgilerle gelişmektedir ve şimdiye kadar müşteri ilişkileri yönetiminden (Lee ve Hu, 2004) finansa (Chan ve Chong, 2017), insan kaynakları yönetiminden (Fareri, Fantoni, Chiarello, Coli ve Binda, 2020) satış tahminine (Chong, Li, Ngai, Ch'ng ve Lee, 2016) kadar çeşitli alanlarda uygulanmıştır.

Mevcut çalışma tüketici değerlendirmelerinden işlemsel modellerde nasıl faydalanılabileceğini incelediğinden, bu bölümde ilgili literatürün teknik kısmından ziyade uygulamalarına odaklanılmıştır.

İnternet'in ve akıllı telefonların yaygınlaşması, müşterilerin şirketlere seslerini daha rahat duyurmaya başlamasını, işletmelerinse müşteri şikayetlerini daha kolaylıkla yönetmelerini sağlamıştır. Seyahat planı yapan kişilerin yaklaşık %95'inin, başka turistler ve gezginler tarafından yapılan çevrimiçi yorumları okuduğu belirtilmektedir (Maurer ve Schaich, 2011). Bu yüzden turizm ve otelcilik sektöründe metin madenciliği uygulamalarına sıkça rastlanmaktadır. Lee ve Hu (2004), çevrimiçi şikayetleri analiz ederek şikayetlerin niteliğini açıklamayı, şikayetlerde en sık kullanılan anahtar sözcükleri bulmayı ve bu şikayetlerin örüntüsünü anlamayı amaçlamıştır. Berezina, Bilgihan, Cobanoglu ve Fevzi (2015), otel müşterilerinin İnternet üzerinden yaptıkları yorumları kullanarak, aldıkları hizmetten tatmin olup olmadıklarını analiz etmiştir. Yazarlar, hangi değişkenlerin müşterilerde hayal kırıklığına yol açtığını, hangilerinin ele alınan otelin başkalarına tavsiye edilmesinde rol oynadığını belirlemeye çalışmıştır. Turgutlu ve Işıklı (2016), duygu analizi (sentiment analysis) ve anahtar sözcük çıkarımıyla günlük turistik planlar için bir tavsiye sistemi önermiştir. Wong ve Qi (2017), bir turistik yer olarak Macau hakkında yapılan ziyaretçi yorumlarının yıllar içindeki değişimini metin madenciliği tekniklerini kullanarak incelemiştir. Hong ve Park (2019), bir havayolu şirketinin pazarlama performansını belirlemek amacıyla kullanıcı yorumlarını incelemiştir. Temel olarak, çevrimiçi yorumları toplayarak başlıca anahtar sözcükleri belirleyen yazarlar, metin kümeleme yöntemiyle bu anahtar sözcüklerin anlamlarına ulaşmışlardır; *koltuklar*, *kabin* ve *uçuş personeli* sözcüklerinin müşteri tatmini üzerinde etkisi olduğunu bulmuş, *uçuş sınıfı* sözcüğününse böyle bir etkisi olduğuna rastlamamışlardır.

Ürünlerin daha akıllı hale gelmesiyle, ürün özelliklerine bağlı olarak çeşitli servis anlayışları da ortaya çıkmaya başlamıştır (Song, Yoon, Lee ve Park, 2017). Şikâyet yönetimi üzerine çeşitli sektörlerde uygulamaları görülen metin madenciliğinden, ürün ve hizmet tasarımı alanlarında da yararlanılmıştır. Song ve diğ. (2017), kullanılan hizmetlerle ilgili hizmet belgelerinin açıklamalarını veri olarak kullanmış ve değiştirilmiş biçimsel kavram analiziyle eş sözcük analizi gibi metin madenciliği tekniklerini uygulamıştır.

Hizmet sektöründeki kadar olmasa da metin madenciliği teknikleri imalat sektöründe de kullanılmaktadır. Abrahams, Fan, Jiao, Wang ve Zhang (2014), o zamana kadarki metin madenciliği çalışmalarını sentezleyerek ürün hatası tespiti için bütünleşik bir metin analitiği çerçevesi önermiştir. Söz konusu çerçeve sosyal medyadaki zengin içeriği etkin bir şekilde değerlendirerek, buradan elde edilen bilgiyi ürün hatası tespiti için kurulacak bir modelin girdisi olarak kullanmayı önermektedir. Hem otomotiv hem de tüketici elektroniği sektörlerinde uygulama yaparak bu çerçevenin geçerliliğini göstermiş olan yazarlar, ürün hatalarını metinden türetilen nicel ölçümlerle ilişkilendirmek amacıyla Temel Bileşen Analizi ve Lojistik Regresyon tekniklerinden faydalanmıştır. Otomotiv sektöründeki uygulamalarında, öncelikle yaklaşık 200 anahtar sözcük belirlemiş ve bu sözcüklerin her yorum başlığında kaç kez tekrarladığını bulmuşlardır. Sonrasında, her marka için en sık tekrarlanan şikâyet sözcüklerini içeren başlıkları bir araya getirmişler ve metinleri "kusurlu" ve "kusurlu değil" şeklinde sınıflandırmışlardır. Model doğruluğunun, içeriğe özgü özniteliklerin (anlamsal faktörler, farklı terimler ve ürün özellikleri) içeriğe bağlı (biçem, sosyal ve duyarlılık) özniteliklere eklendikçe geliştiğini ve genel duygu özniteliklerinin, kusurların varlığını öngörmeye önemsiz olduğunu göstermişlerdir.

Li ve diğ. (2018), büyük miktarda kaza ve hata analizi raporu verisine dayanarak demiryolu güvenliği sistemi için metin madenciliği tekniklerinin ve bunlar için gerekli teknolojilerin önemini analiz etmiş; bu alanda metin analizi uygulamalarının kilit bir rol oynadığını göstermiştir. Terim temsili ve kodlama kısmında terim doküman matrisi ve sözcük torbası (bag-of-words) yaklaşımlarını, tahmine dayalı analiz içinse konu modellemesi ve gizli Dirichlet tahsisi (*Ing.*, latent Dirichlet allocation) yöntemlerinden yararlanmışlardır.

Kullanıcı tarafından oluşturulan içeriğin talep planlamadaki önemini kabul eden araştırmaların sayısı da giderek artmaktadır. Linoff ve Berry (2010)'un bildirdiğine göre, Ford Motor Şirketi, arama motorlarında yapılan aramaların trendinin gecikmeli etkisini gözetererek, zaman serisi tabanlı kısa vadeli satış tahminlerini iyileştirmiştir. Duygu analizi ve sinir ağları kullanan Chong ve diğ. (2016), İnternet üzerinden yapılan tüketici değerlendirmelerinin, çevrimiçi tanıtım stratejilerinin, tüketici değerlendirmelerindeki duyguların ve bu göstergeler arasındaki etkileşimin

ürün satışlarını tahmin etmedeki önemini araştırmıştır. Cui, Gallino, Moreno ve Zhang (2018), satış tahminlerini iyileştirmek için bir giyim perakendecisinin satış tahminlerini, satış, reklam ve promosyon verilerini, kamuya açık sosyal medya bilgileriyle bir araya getirerek iyileştirmiştir. Benzer şekilde Boone, Ganeshan, Hicks ve Sanders (2018), tahmin hatalarını azaltmada müşteri sorgularını dikkate almanın değerini göstermiştir.

3. Metodoloji

3.1 Metin Madenciliği Sürecine Genel Bakış

Genel olarak bir metinden yararlı bilgi çıkarılması süreci üç aşamaya ayrılır: Metin Ön İşleme, Metin Sınıflandırma, Metin Son İşleme. Birinci aşamadaki işler, öznitelik çıkarma, öznitelik indirgeme ve belge temsilidir. Öznitelik çıkarma esnasında, belgenin karmaşıklığını azaltmak ve sınıflandırma aşamasındaki işleri kolaylaştırmak amacıyla belirteleme (tokenization), filtreleme veya gereksiz sözcükleri eleme, kök çözümleme (lemmatization) ve köküne ayırma (stemming) faaliyetleri gerçekleştirilir. Öznitelik indirgeme işlemleri genel olarak iki grupta toplanır: öznitelik seçme, öznitelik dönüştürme. Öznitelik seçme esnasında, verilerin daha iyi anlaşılmasıyla orijinal fiziksel anlamı eşdeğer bir şekilde temsil edebilen bir özellik alt kümesinin seçilir. Buradaki ana amaç, eğitim veri setini daha küçük boyutta oluşturmak için boyutsallık lanetini azaltmaktır çünkü bu sayede daha düşük hesaplama süresi gerekecektir. Öznitelik dönüştürme esnasında, orijinal öznitelik kümesini dönüştürerek veya eşleyerek yeni ve daha küçük bir öznitelik kümesi oluşturulur. Burada sıkça kullanılan yöntemler, Gizli Anlamsal Endeks Analizi (*Ing.*, Latent Semantic Index Analysis), doğrusal diskriminant analizi ve genelleştirilmiş tekil değer ayrıştırma (*Ing.*, Singular Value Decomposition). Birinci aşamada son olarak, belge temsili yapılır. Burada öznitelikler ham metin verilerinden çıkarıldıktan sonra, daha verimli bir şekilde sınıflandırma yapmak için eldeki tüm belgeler birim uzunluğa sahip olacak şekilde standardize edilir. Literatürde bu amaçla en çok kullanılan üç model vardır: vektör uzayı yöntemi, olasılıklı modeller ve çıkarsama ağı modeli. Bunların arasında en yaygın kullanılan vektör uzayı yöntemidir. Belge kümesi $B = \{b_1, b_2, \dots, b_B\}$, sözcük dağarcığı kümesi $D = \{d_1, d_2, \dots, d_D\}$ ve b belgesi için terim vektörü $\vec{t}_b = (f_b(d_1), f_b(d_2), \dots, f_b(d_D))$ kullanılarak belgeler nümerik vektörlere dönüştürülür. Burada, $i = 1, 2, \dots, D$ için d_i ile B kümesindeki özgün sözcükler veya terimler, $f_b(d_i)$

ilse b belgesinde d_i sözcüğünün bulunma sayısı gösterilmektedir. Vektör uzayı yönteminde, bir öznitelik ağırlığı iki yolla hesaplanabilir: mantıksal model, TF-IDF yaklaşımı. Mantıksal model, b_j belgesinde bulunan ($j = 1, 2, \dots, B$) her d_i ($i = 1, 2, \dots, D$) sözcüğüne pozitif bir a_{ij} ağırlığı atar. Aksi durumlarda $a_{ij} = 0$ olarak belirlenir. TF-IDF yaklaşımında her terim, verilen bir cümlede kaç kez geçtiğine göre ağırlıklandırılır. TF, Terim Sıklığı; IDF, Ters-Belge Sıklığı için kullanılan İngilizce kısaltmalardır. Genel olarak, $TF - IDF = TF \times IDF$ şeklinde ifade edilir; öyle ki, burada TF, verilen bir belgedeki belli bir sözcüğün sıklığı ve IDF, [toplam belge / o sözcüğü içeren belge sayısı] oranının iki tabanına göre logaritmasıdır (Turgutlu ve Işıklı, 2016). Matematiksel olarak, bir d_i sözcüğünün b_j belgesindeki terim ağırlığı, $TF - IDF(d_i, b_j)$, aşağıdaki gibi hesaplanır:

$$TF - IDF(d_i, b_j) = TF(d_i, b_j) \cdot \log\left(\frac{|B|}{TF(d_i, B)}\right) \quad (1)$$

Burada $TF(d_i, b_j)$, d_i sözcüğünün b_j belgesinde görülme sıklığını; $|B|$, B kümesinin eleman sayısını; $TF(d_i, B)$ ise, d_i sözcüğünün kaç belgede geçtiğini gösterir. TF, sözcük torbasına çok benzer. Vektör uzayı modelinde metinlerin vektörler olarak klasik yapılandırılmış temsili sözcük torbası olarak bilinir. Matematikte kümeler tekrarlanan elemanlar içermez ancak yine de bir elemanın birden çok bulunduğu yapılara rastlanır. Bu yapılara matematikte çoklu küme (*Ing.*, multiset), doğal dil işleme alanında torba adı verilir. Sözcük torbası yaklaşımında, işlenen belgelerden türetilen öznitelikler, belgelerde geçen sözcüklere dayanmaktadır. Böylece her belge bir torbayla temsil edilir. Belgede geçen tüm sözcükler bu çoklu kümede bulunur. Belgede bulunmayan sözcüklerse torbanın elemanı değildir. Torbalar, makine öğrenmesinde daha yaygın bir veri temsili olan vektörlere dönüştürülebilir. Diğer yandan IDF, belirli bir sözcüğün belli bir belgede bir bütün olarak ne sıklıkla geçtiğini ölçer.

Metin sınıflandırma aşamasında, temel olarak üç kümeden yararlanır: eğitim için kullanılacak belgeleri içeren E kümesi, sınıf etiketlerini içeren S kümesi, test için kullanılacak belgeleri içeren T kümesi. E' deki her b_k ($k = 1, 2, \dots, E$) belgesi, S' den bir s_l ($l = 1, 2, \dots, S$) sınıf etiketiyle etiketlenir; ancak T' deki hiçbir belge etiketlenmez. Bu aşamada birincil amaç, E kümesini kullanarak metin belgelerindeki öznitelikleri hedef sınıf etiketlerinden biriyle ilişkilendirecek bir metin sınıflandırıcı oluşturmaktır. Sınıflandırma modeli

eğitildiğinde, T kümesindeki belgelerin sınıf etiketlerini tahmin edebilecektir. Başka bir deyişle, bir sınıflandırma algoritmasının amacı, $f(b) = s$ olacak şekilde $B = E \cup T$ kümesinden S 'ye bir fonksiyon tanımlamaktır. İlgili literatürde, bu aşamada kullanılabilecek çeşitli sınıflandırma algoritmaları mevcuttur ancak alan sınırlaması nedeniyle aşağıda bunlardan sadece birkaçına ait kısa bir özet verilmiştir. Bu konuyla ilgili daha fazla bilgi için Žižka, Dařena ve Svoboda (2019) veya Jo (2019)'a başvurulabilir.

- **Naive Bayes Sınıflandırıcısı:** Bayesgil sonsal (posterior) olasılık dağılımına dayalı bir olasılıksal sınıflandırıcıdır. Belge ön işleme aşamasında, vektör uzayı modelindeki belgelere karşılık gelen her öznitelik, belirli bir özneliğin o belgedeki varlığına veya yokluğuna bağlı olarak bir veya sıfır değerini alır. Özniteliklerin birbirinden bağımsız olduğu varsayımına dayanır. İki sınıflı bir problem için $p(s = i)$: eğitim kümesindeki herhangi bir belgenin i . sınıfta olma olasılığı, $p(d_j | s = i)$: bir d_j özneliğin (sözcüğün) i . sınıfta olma olasılığı olmak üzere, belli bir özneliğin sınıfı aşağıdaki gibi tahmin edilir (Zhang ve Jatowt, 2020):

$$\hat{y} = \underset{i \in \{0,1\}}{\operatorname{argmax}} p(s = i) \prod_{j=1}^k p(d_j | s = i) \quad (2)$$

- **Karar Ağaçları ve Rassal Orman:** Karar ağaçları hiyerarşik yapılardır ve kök, iç düğümler ve yaprak düğümler gibi üç tür düğümden oluşur. Bir karar ağacı oluşturmak, genel olarak, özyinelemeli bir bölüntüleme prosedürüdür; tüm eğitim verilerinin dahil edilmesiyle kök düğümden başlar ve bir bölme ölçütünü müteakip defalar optimize ederek yaprak düğümlerinde sona erer. Karar ağacı oluşturulduktan sonra, daha önce kullanılmamış bir örnek, kökten yaprak düğüme kadarki en iyi yoldan akarak test edilebilir. Test için kullanılan örneğin son adımdaki sınıf etiketi, ulaşılan yaprak düğüme bağlı olarak belirlenir (Yan, Zhang, Lin, Yang ve Luo, 2020). İki sınıflı problemlerin çözümünde pek sık kullanılsa da karar ağaçları sınıflandırma işinde genelde çok kötü bir performans sergilemez. Rassal orman, torbalama tipi bir topluluk öğrenme algoritmasıdır. Rastgele orman, her biri eğitim verilerinin rastgele bir alt kümesine göre eğitilen ve özniteliklerin yalnızca bazı rastgele alt kümelerini kullanmasına izin verilen karar ağaçları topluluğudur. Rasgele seçimde koordinasyon yoktur; yani, belirli bir gözlem

veya öznitelik rastgele olarak tüm ağaçlara bağlanabildiği gibi, ağaçların hiçbirine bağlanmayabilir veya bazılarına bağlanırken bazılarına bağlanmayabilir. Bir gözlem için nihai sınıflandırma puanı, tüm ağaçlardan elde edilen puanların ortalamasıdır. Rastgele ormanın karar ağacına göre temel avantajı, aşırı uyma sorununu ortadan kaldırması ve sınıflandırma doğruluğunu artırmasıdır (Cady, 2017).

- **Lojistik Regresyon:** Son zamanlarda önem kazanmış ve kullanımı büyük ölçüde artmış bir denetimli sınıflandırma algoritmasıdır. Lojistik fonksiyon (sigmoid eğri) yardımıyla gözlemleri iki sınıfa ayırmak amacıyla kullanılır. Derlemdeki tüm sözcükler birer açıklayıcı değişken olarak görev yapar ve bir belge/kayıt/cümle verildiğinde bunun eldeki iki sınıftan hangisine gireceğini aşağıdaki ilişki sayesinde tahmin eder (Shah, Patel, Sanghvi ve Shah, 2020):

$$\ln\left(\frac{p}{1-p}\right) = \beta^T X \quad (3)$$

Burada p : ilgilenilen kaydın birinci sınıfta olma olasılığı, X : açıklayıcı değişkenleri içeren bilgi matrisi, β : parametre vektörüdür.

- **Doğrusal Destek Vektör Makinesi:** Temel olarak, bir hiperdüzlem kullanarak iki sınıfı birbirinden ayırmayı amaçlar. Farklı sınıflardaki öğeleri ayıracak böyle bir hiperdüzlem (veya iki boyutlu uzayda düz bir çizgi) yerleştirmenin birkaç yolu vardır. Belirli bir eğitim kümesindeki örnekler için bu türden bir sınır sonsuz sayıda oluşturulabilir. O halde amaç, gelecekte olası yanlış sınıflandırmaları mümkün olduğunca en aza indirmek için eğitim örneklerine dayalı olarak optimal bir çözüm bulunup bulunmadığını araştırmaktır. Bu nedenle optimum hiperdüzlem, noktaların mesafesinin minimum değerinin bu düzlemden mümkün olduğu kadar büyük olacağı şekildedir. İlk olarak, iki sınıfı ayıran en geniş bölgeyi (kayışı) bulmak gerekir. Destek vektörleri adı verilen ve uygun şekilde seçilmiş örnekler, bu bölgenin kenarlarını belirler. Daha sonra optimum sınır (hiperdüzlem), bölgenin merkezine yerleştirilir. Tüm olası durumlar bilinmediğinde (teorik olarak sonsuz sayıda olabilir), eğitim kümesinde mevcut olmayan durumların gelecekteki sınıflandırmaları için mutlak hatasızlık garanti edilemeyebilir; ancak sınıra çok yakın olan durumların bile doğru şekilde sınıflandırılma şansı artar (Luo, 2021; Žižka ve diğ., 2019).

- Oylama Sınıflandırıcısı: s adet sınıfın olduğu durumda, her sınıf için ikili bir senaryo üretir. Böylelikle s -sınıflı tek bir problem yerine iki sınıflı s adet problem oluşturulmuş olur. Her problem, bir sınıfı diğerlerinden ayıştırmaktan sorumlu olan bir ikili sınıflandırıcı tarafından yapılır. Bunun devamında, bir sınıfın tüm örneklerinin olumlu, geri kalan sınıfların örneklerininse olumsuz olarak değerlendirilmesiyle tüm eğitim verileri öğrenme aşamasında kullanılır. Test aşamasındaysa, bilinmeyen tüm örnekler tüm sınıflandırıcıların değerlendirmesine sunulur (Yan ve diğ., 2020).
- En Yakın Ağırlık Merkezi (Rocchio Sınıflandırıcısı): Çoğunlukla belge sınıflandırması için kullanılan örnek tabanlı bir öğrenme algoritmasıdır. Eminağaoğlu (2020)'nin aktardığına göre, Rocchio sınıflandırıcısı, en basit haliyle, "temsilci vektörlerin (ağırlık merkezlerinin) temsil edilen sınıfa atanan eğitim belgelerinin ortalama vektörü tarafından verildiği, sınıf sınırlarını tanımlamak için her sınıfın temsili vektörlerini hesaplayan, ağırlık merkezi tabanlı bir sınıflandırıcı" (Manning, Raghavan ve Schütze, 2009) olarak tanımlanabilir. B_s , s sınıfındaki tüm belgeleri, $\theta(b)$ ise b belgesinin vektör uzayını göstermek üzere, s sınıfının ağırlık merkezi aşağıdaki gibi hesaplanır:

$$\mu_s = \frac{1}{|B_s|} \sum_{b \in B_s} \theta(b) \quad (4)$$

Eğitim aşamasında elde bulunan belgeler kullanılarak her sınıf için ağırlık merkezi hesaplanır; bunlar her sınıfın temsilcisi olarak belirlenir.

Yeni bir belge veya kayıt, yakınlık derecesi Öklid mesafesi veya kosinüs benzerliğiyle ölçülerek, en yakın küme merkezine ait sınıfa atanır.

Son aşamada, sınıflandırıcının performansı doğruluk, hassasiyet, duyarlılık veya F_1 skoru gibi ölçülerle değerlendirilir. Öncelikle, kolaylık sağlanması için iki sınıflı bir problemle ilgileniyor olduğumuzu varsayalım: Var, Yok. Performans ölçülerini matematiksel olarak ifade edebilmek için dört temel bileşene ihtiyaç vardır: gerçek pozitifler (GP), gerçek negatifler (GN), yanlış pozitifler (YP), yanlış negatifler (YN). GP, gerçek değeri Var olup Var sınıfına atanan belgelerin sayısını; GN, gerçek değeri Yok olup Yok sınıfına atanan belgelerin sayısını; YP, gerçek değeri Yok olduğu halde Var sınıfına atanan belgelerin sayısını; YN, gerçek değeri

Var olduğu halde Yok sınıfına atanan belgelerin sayısını gösterir. Bu durumda,

$$\text{Doğruluk oranı} = \frac{GP+GN}{GP+GN+YP+YN}, \text{Hassasiyet} = \frac{GP}{GP+YP},$$

$$\text{Duyarlılık} = \frac{GP}{GP+YN}, F_1 \text{ skoru} = \frac{2(\text{Hassasiyet} \cdot \text{Duyarlılık})}{\text{Hassasiyet} + \text{Duyarlılık}}$$

olarak ifade edilir.

3.2 Metodolojik Çerçeve

Diğer alanların aksine pazarlama araştırması, öncelikle ürün seçimini etkileyen değişkenleri belirlemeye odaklanır. Bununla birlikte, bir ürün çeşidini tanımlayan niteliklerin sayısı nispeten büyük olduğunda, ürün özelliklerini işlemsel modellere dahil etmek zorlu bir iştir. Bu zorluk, her şeyden önce, basitçe mevcut ürün özelliklerinin farklı kombinasyonları olarak tanımlanabilen yapılandırılabilir ürünlerin talebinin tahmin edilmesinde daha belirgindir. İki olası yaklaşım olan Lancaster yaklaşımı ve kesikli seçim modelleri, çeşitli nedenlerden dolayı her zaman çalışmayabilir. Birincisi, çok fazla ürün özelliği var olduğunda, tüketiciler, milyonlarca senaryo arasından arama yapmak zorunda kalabilirler. Bu patlamanın üstesinden gelinse bile, mevcut ürün çeşitlerinin sayısına oranla pazarda bulunan görece az sayıdaki potansiyel tüketici göz önünde bulundurulduğunda, şirketlerin her bir yapılandırma için en az bir işlem gözlemlemesi imkânsız olur. İkincisi, istatistik modeller değişim ve tekrarı sevdiğinden, sadece bir kez gözlemlenmiş yapılandırmalar da işleri fazla kolaylaştıramayacaktır. Dahası, ürün özelliklerini tahmin değişkenleri olarak kullanan herhangi bir istatistiksel model, çoklu doğrusallık veya boyutsallık lanetinden mustarip olabilir. Bu nedenle uygulamacılar, olası engelleri önlemek için mevcut verileri önceden işlemek ve yeniden yapılandırmak için alternatif yollar araştırmalıdır. Işıklı (2012), bu konuları ele almak üzere yapılandırılabilir ürünlerin talep planlaması ve bunlardaki dış çeşitliliğin yönetilmesi için nitel ve nicel verilerin bütünleşik kullanımını önermekte ve bunun için bir çerçeve sunmaktadır. Mevcut çalışmada önerilen ve görsel anlatımı Şekil 1'de verilmiş olan metodolojiye, Işıklı (2012)'deki çerçevenin birkaç modülünden biri olarak hizmet edebilir ve aşağıdaki gibi özetlenebilir:

Aşama 1 - Metin çekme, ön işleme ve temsil: Müşteri değerlendirmeleri birden fazla kaynaktan toplanır ve *derlem* adı verilen bir yapıya dönüştürülür. Noktalama işaretleri, özel karakterler ve gereksiz

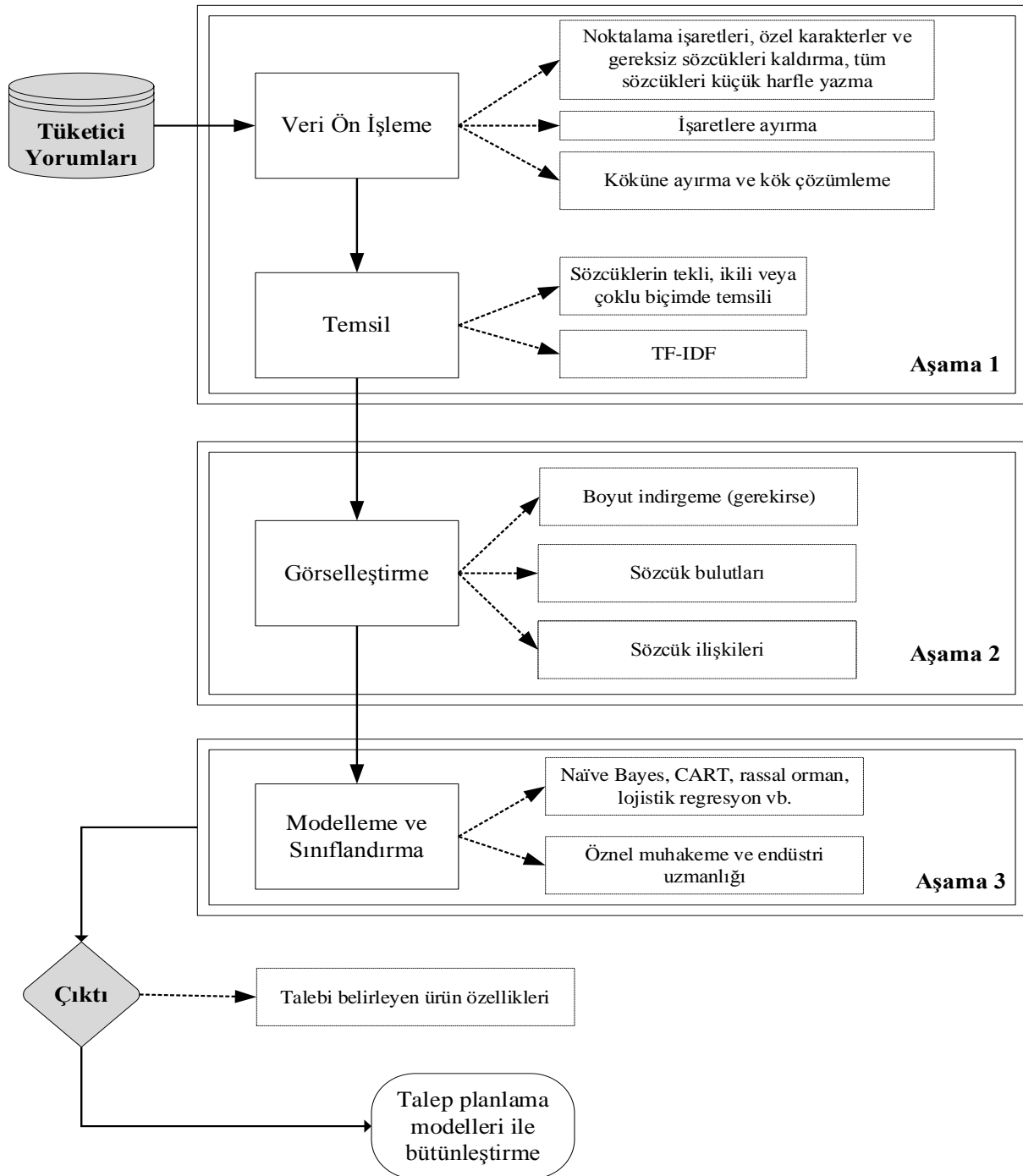
sözcükler kaldırılır ve büyük harfler küçük harflere dönüştürülür. Referans sözlük göz önüne alınarak yazım yanlışları düzeltilir. Derlemdeki her sözcük daha sonra belirteçlere (tokens) dönüştürülür. Bu belirteçler bilgilendirici ve bilgilendirici olmayan şekilde iki gruba ayrılır. Birinci grupta isim, fiil, sıfat ve zarf türünde sözcükler bulunurken ikinci grupta edatlar, bağlaçlar, adılar gibi sözcük türleriyle kısaltmalar ve sayılar bulunur. Sözcüklerin temel bir formunu elde etmek ve Terim-Belge Matrisinde (*İng. kısalt.*, TF-IDF) anlamsal olarak benzer sözcükleri tekrarlamaktan kaçınmak için köküne ayırma ve/veya kök çözümüleme işlemleri yapılması gerekebilir. Veriler derleme dönüştürüldükten ve ön işleme adımları uygulandıktan sonra *TF-IDF*, sözcük torbası yaklaşımı kullanılarak oluşturulur. Ayrıca veriye konu modellemesi ve duygu analizi de yapılabilir. Her parçaya (cümleye), bulunduğu tüketici yorumunun, varsa, derecelendirmesi puan olarak atanabilir. Her parçanın anahtar sözcükleri, makine öğrenmesi kullanılarak üretilir ve en yaygın anahtar sözcükler *TF-IDF* sayesinde belirlenir. Özellikle tüketici yorumlarının çok fazla olduğu durumlarda, eş anlamlı sözcüklerin bir sonraki aşamaya geçmeden önce belirlenmesi yararlı olacaktır.

Aşama 2 - Görselleştirme: Derlemdeki en yaygın sözcüklerle bir sözcük bulutu oluşturulur. Sözcük bulutu, bir sözcüğü bir belgedeki sıklığına göre farklı boyutlarda görüntüleyen, iyi bilinen ve anlaşılabilir bir görselleştirme tekniğidir. Boyutu azaltmak için Gizli Anlamsal Endeks Analizi ve alternatif kümeleme algoritmaları kullanılabilir. Yeterli bilgi verildiğinde sözcük bulutu, yorumların

yapıldığı konumdan kaynaklı farklılıkları da ayırt edebilecek şekilde çizdirilebilir. Anahtar sözcükler arasındaki ilişkiler de belirlenebilir. Farklı terimleri karşılaştırmak için ortaklık bulutu ve karşılaştırma bulutu elde edilebilir. Sözcük veya sözcük grupları (ikili, üçlü vs.) arasındaki ilişkiler görselleştirilir. Görselleştirme aşamasına dahil edilecek sözcük veya sözcük gruplarının önceden belirlenmiş bir sayıda tekrarlanmış olması yönünde bir kısıt konularak bu koşulu sağlamayan terimler için farklı bir analiz yapılması düşünülebilir. Yine bu aşamada, hangi sözcüklerin sıklıkla birlikte kullanıldığını saptamak amacıyla, basit bir hiyerarşik kümeleme de yapılabilir.

Aşama 3 - Modelleme ve Sınıflandırma: Yüksek performans gösterdiği literatürde yaygın olarak kabul edilen Naive Bayes, Karar Ağaçları, Rastgele Orman, Lojistik Regresyon vb. sınıflandırıcıların yanı sıra öznel muhakeme ve endüstri uzmanlığına dayanarak talep modeline dahil edilmesi yararlı olacak ürün özellikleri belirlenir. Satıcının etkisi veya bölgesel farklılıklar gibi ek girdiler daha sonra modelin çekirdek girdilerine eklenebilir. Bu aşamada, klasik bir şikâyet yönetimi modelinde olduğu gibi belli sözcükleri kullanan müşterilerin tavsiye etme ya da etmeme durumunun ne olacağı önceden kestirilmeye çalışılmaz. Öncelikli amaç, tahminden ziyade sınıflandırmadır, zira elde edilecek sonuçlar daha sonra talep modelinde kullanılacak girdiler olacaktır.

Bu çalışmada araştırma ve yayın etiğine uyulmuştur.



Şekil 1. Önerilen metodolojik çerçeve

4. Uygulama

Tüketicilere sunulan ürün özelliklerinin önemini belirlemek için bir çapraz kullanımlı araç (*İng. kısalt., CUV*) modeli incelenmiştir. Üretim geçmişi 1970'lere kadar uzanan bu araçlar, açık alan

kullanımlı araçlarla (*İng. kısalt., SUV*) binek araçların çapraz eşleşmesiyle oluşmuştur. Özellikle 2000'li yılların başından itibaren önemli ölçüde rağbet görmeye başlayan çapraz kullanımlı araçlar, geniş iç hacimlerinin yanı sıra düşük yakıt tüketimi sağladıkları için özel olarak çocuklu genç ailelere

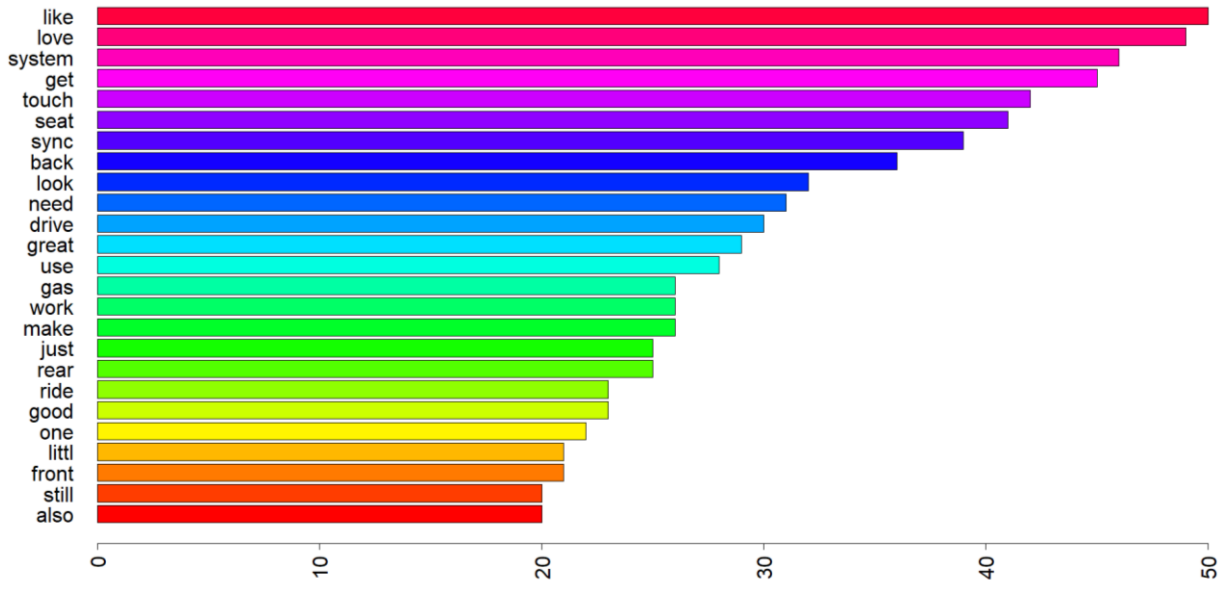
fazlasıyla çekici gelmektedir. Bu ürün için Cnet.com, Edmunds.com ve Kbb.com gibi otomobil bloglarında yapılan tüketici yorumları dikkate alınmıştır. Alexa.com'a göre, söz konusu bloglar ABD genelinde, ziyaretçi trafiği açısından, sırasıyla en gözde 111., 409. ve 294. İnternet siteleridir (28 Mayıs 2021 itibarıyla). Özellikle Cnet.com, uzman görüşlerine de yer verdiği için kullanıcılar tarafından yakından takip edilmektedir. Çekilen tüketici görüşleri Ağustos 2010 – Ağustos 2011 arasındaki dönemi kapsamaktadır. Bunun sebebi, Ağustos 2010 – Mart 2012 arasındaki dönemi göz önüne alan Işıklı (2012)'deki talep kestirimi sonuçlarıyla karşılaştırma yapma arzudur. Yorumların hepsi İngilizcedir ve ürünün çeşitli boyutlarda derecelendirmelerini içermektedir. Metinde birbirinin kopyası yorumlar veya derecelendirmelere dair eksik bilgi bulunmamaktadır. 492 cümle (satır) ve 842 özniteliğe (sütun) dönüştürülmüş toplam 66 tüketici yorumu toplanmıştır. Söz konusu öznitelikler, derlemede en az üç kez rastlanmış olan tek sözcüklerin yanı sıra en az iki kere tekrarlanmış olan ikili ve üçlü sözcük gruplarından oluşmaktadır. Bu sözcükler ürün özelliklerinin yanı sıra bunları betimleyen sıfatları da içermektedir.

Bu çalışmanın öncelikli amacı, hakkında olumlu ya da olumsuz düşünülen ürün özelliklerinin ortaya çıkarılmasıdır çünkü bitaraf kalınmış bir ürün özelliğinin talep tahminine etkisinin olmayacağı varsayılmıştır. Dolayısıyla, belirlenen bu gibi ürün özelliklerinin işlemsel modellere sokulmaması önerilmiştir. Buna ek olarak, anahtar sözcükler kullanılarak ürünü tavsiye eden yorumları, tavsiye etmeyenlerden ayıran sınıflandırma modelleri kurulmuştur. Uygulama için Python ve R yazılımlarından yararlanılarak çeşitli makine öğrenme algoritmaları ve doğal dil işleme araçları kullanılmıştır.

Sözcük torbası yaklaşımının ilk aşamasında, ham metin verisi vektörler halinde yazılarak metin derlemi oluşturulmuştur. Bunu takiben, metin ön işleme adımları uygulanmıştır. Her bir yorum etiketlendikten sonra cümlelere ayrılmıştır. Her cümle bir veya daha fazla öznitelikten oluşmaktadır (burada ürün özelliği kastedilmemektedir). Bir özelliğin tekrar sıklığının artmasıyla öneminin de artacağı varsayılmıştır. Yorumlar öncelikle "and", "the", "a", "an", "can", "could", "on", "at" gibi gereksiz sözcükler ve noktalama işaretlerinden

arındırılmıştır; ancak olumsuzlamalar için herhangi bir eylemde bulunulmamıştır. Sonrasında, sözcükler türlerine göre sınıflandırılmıştır. Ardından, İngiliz diline göre köklerine ayırma işlemi yapılmış ve tüm büyük harfler küçük yazılmıştır. Özellikle ikili sözcük gruplarına katabilecekleri anlamlar olduğundan, sayılar metinden çıkarılmamıştır. Daha sonra, isim türünde sözcüklerin frekansları bulunmuş, isim gruplarının belirlenmesi için birliktelik kural çıkarımı uygulanmış ve ürün özellikleri tanımlanmıştır. Devamında, birinci tip modeller için tarafsız cümleler bilgilendirici olmayan (0), diğerleri ise bilgilendirici (1) olarak işaretlenmiştir; ikinci tip modeller içinse ürünü tavsiye eden yorumlar (1), tavsiye etmeyen yorumlar (0) olarak işaretlenmiştir. Tarafsız cümleler, herhangi bir ürün özelliği hakkında olumlu veya olumsuz hiçbir bilgi içermeyen cümlelerdir. Son olarak hem görselleştirme hem de her sözcüğün sıklığını tutma açısından önemi olan Terim-Belge Matrisi (TF-IDF) oluşturulmuştur.

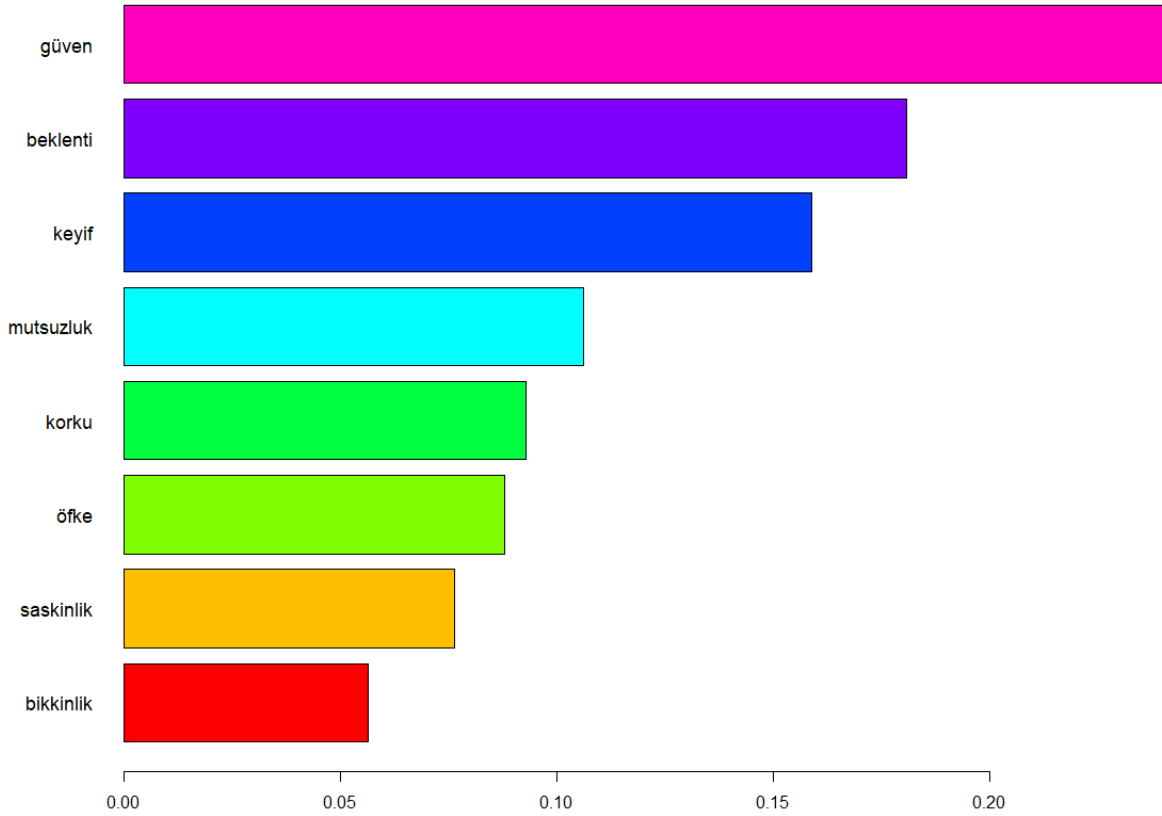
Ön işleme aşamasından sonra, yorumlarda en sık tekrarlanan yirmi beş sözcük Şekil 2'de verildiği gibi bulunmuştur. İlk bakışta, "system", "touch", "seat", "sync", "rear" ve "front" sözcüklerinin amacımız doğrultusunda kullanılabileceği fark edilmektedir. Bu sözcüklerden birincisi, ikincisi ve dördüncüsü aslında tek bir ürün özelliğini vurgulamaktadır. Araç kullanıcılarının sesli komutlar vererek telefon araması yapmalarını ve müziği kontrol etmelerini sağlayan bir araç içi iletişim ve eğlence sistemi olan senkronizasyon sistemi (sync system) tüketici yorumlarında sıklıkla geçirilmiştir. İkinci sözcük, dördüncü sözcüğün 2010 yılı sonrasında üretilen araçlardaki adıdır ve çoğunlukla bu sistemin yazılımına karşılık gelmektedir. Her ne kadar araştırma dönemindeki araçlardaki ürün özelliği dördüncü sözcükle pazarlanıyor olsa da tüketicilerin ikinci sözcüğü bu ürün özelliğini anarken hâlâ kullanabileceği göz ardı edilmemelidir zira bu gibi değişikliklerin tüketicinin kullandığı dilde karşılık bulması zaman alabilmektedir. Araç koltukları (üçüncü sözcük), arka görüş kamerası (beşinci sözcük) ve önden çekişlilik (altıncı sözcük) gibi ürün özelliklerinin de tüketici için önemli olduğuna işaret edilmektedir. Öne çıkan bu ürün özelliklerinden olumlu yönde bahsedilip bahsedilmediğini bilmesek de bir talep modelinde öncelikli olarak bunların göz önünde bulundurulması gerektiğini söylemek yanlış olmaz.



Şekil 2. Müşteri Yorumlarında En Sık Tekrarlanan Sözcükler

Şekil 2'nin talep planlama açısından yararı görece kısıtlıdır zira yorumlarda geçen fiiller ve sıfatlar, ürün özelliklerinden daha sık tekrarlandığı için ilk bakışta daha fazla dikkat çekmektedir. Bu yüzden ürünü tavsiye edenlerle tavsiye etmeyenlerin sıklıkla yineledikleri sözcükler, bir karşılaştırma bulutu olarak Şekil 3'te verilmiştir. Bu şeklin üst kısmında yer alan yeşil renkli sözcükler, müşterilerin beğendiği ve ürünü başkalarına tavsiye etmelerinde etkili olan ürün özellikleridir. Şeklin alt kısmında yer alan kırmızı renkli sözcüklerse, ürünü tavsiye etmeyen müşterilerin yorumlarında en sık geçen sözcüklerdir. İlk bakışta, aracın koltukları (seats), arka görüş kamerası (camera), iç tasarım (interior), dış tasarım (exterior), ortam paketi (ambient package), konfor paketi (comfort

package), tavan penceresi (roof ve sunroof) ve kör nokta bilgi sistemi (blis) gibi ürün özelliklerinin, aracın tavsiye edilmesinde etkili olduğu söylenebilir. Benzer şekilde, teknolojik özellikler (sync, GPS, control), önden çekişlilik (front) ve SEL (standart limitli sürüm) gibi ürün özelliklerinin aracın tavsiye edilmemesinde etkili olduğu görülmektedir. Bu aslında çok basit bir duygu analizidir. Standart bir duygu analizinde olduğu gibi sıfatlara pozitif veya negatif dereceler atanarak sözcüklerin skorları hesaplanmamıştır ancak müşterilerin hangi özelliklerle ilgili olumlu, hangisiyle ilgili olumsuz düşünceleri olduğunu kabaca görmek halihazırda mümkündür.



Şekil 5. Müşteri Yorumlarındaki Temel Duyguların Dağılımı

Son aşamada, veri setinin %70'i bir eğitim kümesi, geri kalanı bir test kümesi olarak ikiye ayrılmış ve ilk önce *TF-IDF*, sonrasında Sözcük Torbası yaklaşımları uygulanarak çeşitli sınıflandırma modelleri denenmiştir. *TF-IDF*, sadece nadir gözlemlendikleri için bazı sözcüklere daha fazla ağırlık verebilir (Hotho, Nürnberg ve Paaß, 2005). Ancak bu, söz konusu sözcüklerin duygu analizi için önemli olduğu anlamına gelmez. Sözcük torbası yaklaşımı bu yüzden denenmiştir. Sadece belirli bir cümle içinde geçen sözcüklerin sayısını sayan bu yaklaşımda, sözcükler bir dizide saklanmaktadır (Hotho ve diğ., 2005).

Birinci tip modellerde, cümleler tarafsız ve olumlu/olumsuz olarak sınıflandırılmıştır; diğerlerinde, ürünü tavsiye eden yorumların tavsiye etmeyenlerden ayrıştırılması amaçlanmıştır. Aynı tip modellerde aynı açıklayıcı değişken kümesi kullanılmıştır. Tüketicilere ait demografik bilgiler ve

kullanıcı istatistikleri bulunmadığından, sadece sözcüklerin sıklığı ve ağırlıklarını temel alan yaklaşımlar benimsenebilmiştir.

Hem *TF-IDF* hem de Sözcük Torbası yaklaşımlarında, lojistik regresyon, doğrusal destek vektör makinesi, en yakın ağırlık merkezi, doğrusal olmayan destek vektör makinesi, oylama sınıflandırıcısı ve rassal orman gibi farklı sınıflandırıcılar dikkate alınmıştır. Bu sınıflandırıcılar, uygulanması kolay ve nispeten iyi sınıflandırma sonuçlarına sahip oldukları için seçilmiştir. Her model için hem eğitim hem de test verisinde en isabetli sınıflandırma sonuçlarını üreten tekniklere dair bilgi Tablo 1'de özetlenmiştir. *TF-IDF* yaklaşımı kullanılan modellerde, Sözcük Torbası yaklaşımı kullanılanlara kıyasla, sınıflandırıcıların daha yüksek performans gösterdiği görülmektedir.

Tablo 1
Sınıflandırma algoritmalarının sonuçları

| Bağımlı Değişken | Yaklaşım | Sınıflandırma Algoritması | Doğruluk Oranı | |
|--------------------------------------|-----------------------------------|---------------------------|--------------------------------|--------------------------------|
| Tavsiye Edilme / Tavsiye Edilmeme | TF-IDF | Rassal Orman | %87,8 (Eğitim) %82,4 (Test) | |
| | | Lojistik Reg. | %89,8 (Eğitim) %88,2 (Test) | |
| | Sözcük Torbası | Rassal Orman | %81,6 (Eğitim) %70,6 (Test) | |
| | | En Yakın Ağırlık Merkezi | %98,0 (Eğitim) %82,4 (Test) | |
| | Tarafsız Yorum / Taraflı Yorum | TF-IDF | Rassal Orman | %75,0 (Eğitim) %64,8 (Test) |
| | | | Lojistik Reg. | %75,0 (Eğitim) %73,5 (Test) |
| Doğrusal Destek Vektör Makinesi | | | %60,7 (Eğitim) %53,5 (Test) | |
| Oylama Sınıflandırıcısı | | | %76,0 (Eğitim) %67,0 (Test) | |
| Sözcük Torbası | | Naive Bayes | %50,0 (Eğitim) %66,3 (Test) | |

Bu algoritmalarla göre, tüketicilerin dört çekişli konfigürasyonların, özellikle *Sınırlı* veya *SEL model* olanlarına değer verdiği bulunmuştur. Ayrıca *senkronizasyon sistemi*, *navigasyon merkezi*, *arka kamera* ve *kör nokta bilgi sistemi (blis)* gibi çoğu teknolojik özellikle de ilgilenmişlerdir. *Deri koltuklu* ve *birinci sınıf tekerlekli* konfigürasyonlar da beğenilmiştir.

Sonuç olarak, sınıflandırma algoritmalarından elde edilen bulguların ve görselleştirmenin ortak olarak işaret ettiği, tüketicilerin belirtilen ürüne olan algısını yansıtan ve bu yüzden talep planlamasında kullanılması yararlı olabilecek ürün özelliklerinin *senkronizasyon sistemi* ve *navigasyon sistemi* gibi teknolojik özelliklerle *dört çekişlilik* ve *sürüş* gibi yapısal özellikler olduğu söylenebilir.

Işık (2012)'deki, New York bölgesini dikkate alan talep tahmin modelinde, 2011 yılının ilk 32 haftalık dönemi boyunca müşteriler tarafından çok çekici bulunan iki benzerlik kümesindeki nispi karışım oranları Tablo 2'de verilmiştir. Nispi karışım oranı, belirli bir ürün özelliğinin bu benzerlik kümesine ait

olan ve söz konusu zaman aralığında mevcut olan ürünlerin yüzde kaçında var olduğunu göstermektedir. Örneğin, 12 numaralı benzerlik kümesindeki araçların hepsi deri kaplamalı koltuklara sahiptir ancak hiçbirinde senkronizasyon sistemi yoktur. Önden çekişli araç oranına bakıldığında, iki benzerlik kümesindeki araçların neredeyse tamamının dört çeker olduğu anlaşılır. Buna göre, New York bölgesindeki tüketicilerin göz önünde bulundurulmuş çalışma döneminde en çekici buldukları benzerlik kümelerinde, en çok dört çeker ve sınırlı sürüm araçlara rastlanmaktadır. *Senkronizasyon sistemi*, *navigasyon merkezi*, *uydu radyosu*, *arka görüş kamerası*, *kör nokta bilgi sistemi* ve *ortam paketi* gibi teknolojik ürün özelliklerine; *ön farlara*, *açılır tavana*, *sürücü paketine* ve *deri koltuklara* değer verilmektedir. Tüketici yorumlarının metin analizi de *açılır tavan*, *arka görüş kamerası*, *ortam paketi* ve *deri koltukların* öne çıkan ürün özellikleri olduğunu göstermişti.

Tablo 2

Benzerlik kümeleri 12 ve 19 için nispi karışım oranları (Işıklı, 2012)

| Opsiyonlar | Benzerlik Kümesi 12 | Benzerlik Kümesi 19 |
|--|---------------------|---------------------|
| Önden Çekişli (FWD) | %9 | %7 |
| Standart Sürüm (SE) | %0 | %0 |
| Sınırlı Standart Sürüm (SEL) | %0 | %0 |
| Sınırlı Sürüm (LIMITED) | %100 | %99 |
| Spor Sürüm (SPORT) | %0 | %1 |
| Kömür Karası İç Döşeme (Charcoal Black Trim) | %73 | %53 |
| Senkronizasyon Sistemi (SYNC/Touch) | %0 | %0 |
| Paspas (Floor mats) | %36 | %59 |
| Ön Farlar (Headlamps) | %100 | %96 |
| Açılır Tavan (Moonroof) | %100 | %98 |
| Arka Görüş Kamerası (Rearview camera) | %100 | %100 |
| Portbagaj (Roof rack) | %0 | %1 |
| Uydu Radyosu (Satelliteradio) | %100 | %100 |
| Hız Kontrolü (Speed Control) | %0 | %69 |
| Römork Çekme Paketi (Trailertow Pkg) | %0 | %15 |
| Deri Koltuklar (Leather) | %100 | %100 |
| Yüksek Kaliteli Tekerlekler (Premium Wheels) | %0 | %1 |
| Ortam Paketi (Ambient Pkg) | %100 | %100 |
| Kör Nokta Bilgi Sistemi (BLIS) | %0 | %100 |
| Komfor Grubu (Comfort Grp) | %0 | %0 |
| Sürücü Paketi (Drivers Pkg) | %100 | %99 |
| Navigasyon Merkezi (Navigation Center) | %100 | %99 |
| Jant Tipi 1 | %0 | %6 |
| Jant Tipi 2 | %0 | %0 |
| Jant Tipi 3 | %100 | %94 |
| Jant Tipi 4 | %0 | %0 |
| Lastik Tipi 1 | %0 | %0 |
| Lastik Tipi 2 | %100 | %93 |
| Lastik Tipi 3 | %0 | %7 |
| Lastik Tipi 4 | %0 | %0 |

Tablo 2'den görülebileceği gibi, ele alınan ürün 27 adet ikili opsiyonun kombinasyonu olarak yapılandırılabilir. Yani, bu ürün teorik olarak 134.217.728 farklı şekilde yapılandırılabilir. Ancak 2011 yılında sadece 15.062 farklı konfigürasyonu piyasa sunulmuş bu üründen sadece 109.180 adet satılmıştır. Başka bir deyişle, ürünün konfigürasyon uzayının yaklaşık on binde 1'lik bölümü tüketiciye sunulmuştur ve konfigürasyon başına satılan araç sayısı 7'den biraz fazladır. Yalnızca bir adet üretilen 509 konfigürasyon, 2054 adet üretilen ise 12 konfigürasyon bulunmaktadır. Metin madenciliği uygulamasının vurguladığı dört ürün özelliği göz önünde bulundurularak oluşturulan konfigürasyon

uzayı sadece 16 eleman içerecektir. Bu durumda konfigürasyon başına düşecek araç sayısı 6.800'den fazladır. Dolayısıyla 13 satış bölgesine ayrılmış ABD pazarında, aylık satış tahmini yapmayı sağlayacak güvenilir bir istatistiksel model kurulma imkanı doğacaktır.

5. Sonuç ve Tartışma

Pazarlamacılar, perakendeciler ve imalatçılar, bir yandan depolama ve üretim maliyetlerini makul bir seviyede tutarken, diğer yandan tüketici ihtiyaçlarını karşılamak için ne düzeyde bir ürün çeşitliliğinin yeterli olduğunu bilmek isterler.

Müşteri memnuniyetsizliği maliyeti, müşterilerin rakiplere kaybedilmemesi için yeterli sayıda ürün çeşidi sunmayı önermektedir. Ancak ölçek ekonomisi ve üretim maliyetleri, imalatçıları az sayıda ürün çeşidi sunmaya zorlar. Ürün çeşitliliği yüksekken üretim maliyetleri de yüksektir ve ürün karmaşıklığı, temin sürelerini ve ürün kalitesini olumsuz yönde etkileyebilir.

Sunulacak ürün çeşitlerinin belirlenmesinde rekabetçi hedeflere sahip olmak, hemen hemen her imalatçı için mümkündür. Ancak çok geniş bir konfigürasyon uzayıyla sonuçlanan, birçok seçenek ve özelliğe sahip karmaşık bir ürün olan yapılandırılabilir ürün imalatçıları için bu durum özellikle zahmetlidir. Bu gibi ürünler, genellikle ikili kodlanmış (var-yok) ürün özelliklerinin kombinasyonları olarak görülebilir. Dikkate alınması gereken farklı yapılandırmaların bolluğu, stok yönetimi veya talep tahmini gibi işlemsel modellerin verimliliği üzerinde olumsuz bir etkiye sahip olabilir. Ürün özellikleri arasındaki etkileşim ve ürün özelliklerinin ikili kodlanmış olması da istatistiksel modellerin kullanımını kısıtlayabilir. Dahası, bir dayanıklı tüketim malı da olsa yapılandırılabilir bir ürünün çeşit sayısı arttığında, tüketicilerin satın alma öncesinde veya sırasında tüm ürün özelliklerini her zaman dikkate alması mümkün olmayacaktır. Bu nedenle, uygulayıcılar bu olası engellerin önüne geçmek için mevcut verileri önceden işlemek ve yeniden yapılandırmak için alternatif yöntemler benimsemelidir. Bütün bunların yanı sıra, bu gibi işlemsel modellere doğru bilgi girmek için pazarın veya pazar bölümünün dinamiklerinin iyi anlaşılması gerektiği unutulmamalıdır.

Bu çalışma, metin yığınlarından yeni ve alakalı bilgiler çıkaran keşfedici bir analizdir. Talep planlama sürecinin daha kolaylıkla yürütülebilmesi için tüketicilerin satın alma sonrası davranış ve algıları hakkında bilgi edinmenin değerini tartışmaktadır. Bu amaçla yapılandırılabilir bir ürün hakkındaki çevrimiçi tüketici değerlendirmelerini inceleyen bir metin madenciliği uygulaması sunulmaktadır.

Tüketici değerlendirmeleri, bireysel müşterilerinin özel tercihlerini fark etmelerini sağlayarak perakendecilere bire bir pazarlama stratejileri geliştirmek açısından kolaylık sağlayabilmektedir. Hatta bu sayede bir ürünle ilgili tüketicilerin dilinde dolaşacak sözcüklerin yönetilmesi de mümkün olmaktadır. Müşteri görüşleri günlük yaşamda çok önemli bir rol oynamaya başlamıştır. Tüketicilerin

bir karar verirken başkalarının görüşlerine başvurması yeni bir şey değildir; ancak önceleri daha ziyade tanıdıklarının görüşlerini dikkate alan günümüz tüketicileri, artık hiç tanımadıkları insanların yorumlarından da etkilenebilmektedir. Günümüzde pek çok İnternet kullanıcısı birçok ürün için fikirlerini bloglar, inceleme siteleri ve sosyal paylaşım siteleri aracılığıyla paylaşmaktadır. Ticari ve kurumsal organizasyonlar da ürünleri, destekleri ve hizmetleriyle ilgili tüketici görüşleri bulmak için fazlasıyla isteklidirler.

Bu çalışmadaki metin madenciliği uygulamasında, veri kümesi bir ön işleme tabi tutulduktan sonra ürün özellikleri belirlenmiştir. Bazı sözcükler kısaltma şeklinde, bazılarıysa hatalı yazılmış olabileceğinden, eş anlamlı sözcükler aynı biçimde yazılmıştır. Yorum sayısı az olduğu söz konusu işlem yazar tarafından ayrıca kontrol edilmiştir. Sonra yorumların taraflı mı tarafsız mı olduğu belirlenerek ürün özelliklerinin hangi gruba dahil olduğu bulunmuştur. Klasik bir duygu analizi de yapılmıştır. Çeşitli sınıflandırma algoritmalarından yararlanarak ürünü tavsiye eden yorumlarda öne çıkan ürün özellikleri belirlenmiştir.

Burada sunulan yaklaşım, dikkate alınan tüketici yorumu sayısı görece az olmasına rağmen nicel yöntemlerle paralel sonuçlar vermiştir ve metin madenciliği tekniklerinin ana veya yardımcı modül olarak benzer çerçevelere dahil edilmesinin yararlı olacağını göstermiştir. Kullanılan sınıflandırma algoritmaları öngörü amaçlı değil, sınıflandırma amaçlı kullanılmıştır; zira tüketici davranışı dinamiktir ve tarafsız olmayan cümlelerin içinde sıklıkla geçen bir sözcük kısa-orta vadede bile sınıf değiştirebilir. Benzer şekilde, müşterinin ürünü tavsiye etmemesine neden olan bir özellik, zamanla çok aranan bir hale gelebilir. Bu yüzden yapılan işlemler bir defaya mahsus değildir; gerektiğinde tekrarlanmalıdır. Önerilen yaklaşım halihazırda tüketici yorumlarındaki bölgesel farklılıkları göz önünde bulundurma kabiliyetine sahip değildir. Tüketici yorumlarının bölgesel bazda ayırt edilmesi mümkün olduğunda, bu bilgi önerilen çerçeveye kolayca dahil edilebilir.

Dikkate alınan yorumlar, genel derecelendirmenin yanı sıra, ürünün performans, konfor, yakıt tasarrufu, sürüş keyfi, iç tasarım, dış tasarım, üretim kalitesi ve güvenilirlik açılarından da derecelendirmelerini içermektedir. Gelecek çalışmalarda, daha fazla sayıda müşteri yorumunun bu bilgileri kullanarak incelenmesi düşünülebilir. Daha büyük bir veri setiyle, müşteri

derecelendirmeleri kullanılarak sınıflandırma yapılabilir. Alternatif başka bir modelde, ifadelerin skorlanmasıyla kurulabilir. Bunun için bir sıfatla başlayan sözcük gruplarına pozitif bir skor, olumsuzluk içeren sözcüklere negatif bir skor verilerek sınıflandırma yapılabilir. Bu çalışmanın metodolojisi, basit bir duygu analizi ve basit özellik sayıları içermektedir. Ancak daha büyük bir veri seti kullanılarak, yorumların kutupluluğunu ve nesnellliğini ölçmek için duygu analizi uygulanabilir. Kutupluluk, yorumun olumlu, olumsuz veya nötr olduğunu gösteren -1 ile 1 arasında bir reel sayı olarak analize katılabilir. Yorumun nesnelligiyse, 0 ile 1 arasında bir reel sayı olarak bir yorumun öznellikten ne kadar uzak olduğunu gösterecektir.

Kaynaklar

- Abrahams, A.S., Fan, W., Jiao, J., Wang, G.A. ve Zhang, Z.J. (2014). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, 24(6), 975-990. Doi: <https://doi.org/10.1111/poms.1230>.
- Berezina, K., Bilgihan, A., Cobanoglu, C. ve Fevzi, O. (2015). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *Journal of Hospitality Marketing and Management*, 25(1), 1-24. Doi: <https://doi.org/10.1080/19368623.2015.983631>.
- Berger, I.E., Cunningham, P. ve Drumwright, M.E. (2007). Mainstreaming corporate social responsibility: Developing markets for virtue. *California Management Review*, 49, 132-157. Doi: <https://doi.org/10.2307/41166409>.
- Boone, T., Ganeshan, R., Hicks, R.L. ve Sanders, N.R. (2018). Can Google trends improve your sales forecast? *Production and Operations Management*, 27, 1770-1774. Doi: <https://doi.org/10.1111/poms.12839>.
- Cady, F. (2017). *The Data Science Handbook*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Chan, S.W.K. ve Chong, M.W.C. (2017). Sentiment analysis in financial texts. *Decision Support Systems*, 94, 53-64. Doi: <https://doi.org/10.1016/j.dss.2016.10.006>.
- Chong, A.Y.L., Li, B., Ngai, E.W.T., Ch'ng, E. ve Lee, F. (2016). Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. *International Journal of Operations and Production Management*, 36(4), 358-383. Doi: <https://doi.org/10.1108/ijopm-03-2015-0151>.
- Cui, R., Gallino, S., Moreno, A. ve Zhang, D.J. (2018). The operational value of social media information. *Production and Operations Management*, 27, 1749-1769. Doi: <https://doi.org/10.2139/ssrn.2702151>.
- Eminagaoglu, M. (2020). A new similarity measure for vector space models in text classification and information retrieval. *Journal of Information Science*, 1-14. Doi: <https://doi.org/10.1177/0165551520968055>.
- Fareri, S., Fantoni, G., Chiarello, F., Coli, E. ve Binda, A. (2020). Estimating industry 4.0 impact on job profiles and skills using text mining. *Computers in Industry*, 118, 103222. Doi: <https://doi.org/10.1016/j.compind.2020.103222>.
- Ferguson, S. ve Donndelinger, F. (2010, September). *A Multidisciplinary Approach to Market Segmentation and Product Family Definition*. Paper presented at the 13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, Fort Worth, Texas, USA. Doi: <https://doi.org/10.2514/6.2010-9228>.
- Fisher, M.L. (1997, March-April). What is the right supply chain for your product? *Harvard Business Review*, 75(2), 105-116.
- Hong, J.W. ve Park, S.B. (2019). The identification of marketing performance using text mining of airline review data. *Mobile Information Systems*, 2019. Doi: <https://doi.org/10.1155/2019/1790429>.
- Hotho, A., Nürnberger, A. ve Paaß, G. (2005). A brief survey of text mining. *Ldv Forum*, 20, 19-62.
- Işıklı, E. (2012). *Decision Support Models for the External Variety Of Configurable Products* (Doctoral dissertation, Wayne State University, USA). Erişim adresi : https://digitalcommons.wayne.edu/oa_dissertations/600.
- Işıklı, E. ve Ketenci, M. (2020). Using Consumer Reviews for Demand Planning: Case of Configurable Products. In C. Kahraman, S. Cebi, S. Cevik Onar, B. Oztaysi, A. Tolga, I. Sari (Eds.) *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making: Proceedings of the INFUS 2019 Conference* (pp. 354-361). Cham:

- Springer Nature. Doi: https://doi.org/10.1007/978-3-030-23756-1_44.
- Jo, T. (2019). *Text Mining: Concepts, Implementation, and Big Data Challenge. Studies in Big Data*. Cham: Springer International.
- King, T. (2017). IDC: Data Creation to Reach 163 Zettabytes by 2025. Erişim adresi : <https://solutionsreview.com/data-management/idc-data-creation-to-reach-163-zettabytes-by-2025/>
- Lee, C.C. ve Hu, C. (2004). Analyzing hotel customers' ecomplaints from an internet complaint forum. *Journal of Travel and Tourism Marketing*, 17(2-3), 167-181. Doi: https://doi.org/10.1300/J073v17n02_13.
- Li, X.Q., Gao, F., Xiang, W.G., Shi, T.Y. ve Li, P. (2018). Application of text mining techniques in railway safety supervision system. *IOP Conference Series: Earth and Environmental Science*, 189(6), 1-7. Doi: <https://doi.org/10.1088/1755-1315/189/6/062009>.
- Linoff, G.S. ve Berry, M.J. (2010). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 3rd edition, VitalSource Bookshelf version. USA: John Wiley & Sons, Inc.
- Lowenstein, M.W. (2014). *Customers Inside, Customers Outside: Designing and Succeeding With Enterprise Customer-Centricity Concepts, Practices, and Applications*. USA: Business Expert Press, LLC.
- Manning, C.D., Raghavan, P. ve Schütze, H. (2009). *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press.
- Maurer, C. ve Schaich, S. (2011). Online Customer Reviews Used as Complaint Management Tool. In R. Law, M. Fuchs, F. Ricci (Eds.), *Information and Communication Technologies in Tourism 2011* (pp. 499-511). Springer, Vienna. https://doi.org/10.1007/978-3-7091-0503-0_40.
- Murphy, R. (2019). Local Consumer Review Survey. Retrieved from <https://www.brightlocal.com/research/local-consumer-review-survey/>.
- Podium. (2017). *The Complete Guide to Online Reviews*. Erişim adresi : <https://www.podium.com/resources/podium-state-of-online-reviews/>.
- Ramdas, K. (2003). Managing product variety: An integrative review and research directions. *Production and Operations Management*, 12, 79-101. Doi: <https://doi.org/10.1111/j.1937-5956.2003.tb00199.x>.
- Shah, K., Patel, H., Sanghvi, D. ve Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. *Augmented Human Research*, 5(1), 1-16. Doi: <https://doi.org/10.1007/s41133-020-00032-0>.
- Shilakes, C.C. ve Tylman, J. (1998). *Enterprise Information Portals*, Merrill Lynch, Inc., New York, NY, USA.
- Song, B., Yoon, B., Lee, C. ve Park, Y. (2017). Development of a service evolution map for service design through application of text mining to service documents. *Research in Engineering Design*, 28(2), 251-273. Doi: <https://doi.org/10.1007/s00163-016-0240-5>.
- Tabachnick, B. G. ve Fidell, L. S. (2019). *Using Multivariate Statistics*, 7th edition New York, USA: Pearson.
- Trattner, A., Hvam, L., Forza, C. ve Herbert-Hansen, Z. N. L. (2019). Product complexity and operational performance: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 25, 69-83. Doi: <https://doi.org/10.1016/j.cirpj.2019.02.001>.
- Turgutlu, K. ve Işıklı, E. (2016). Daily Touristic Plan Recommendation Using Text Mining. In M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.) *Position Papers of the 2016 Federated Conference on Computer Science and Information Systems* (pp. 41-48). Warsaw, Poland: Polskie Towarzystwo Informatyczne. Doi: <https://doi.org/10.15439/2016f594>.
- Wong, C.U.I. ve Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - The case of Macau. *Tourism Management Perspectives*, 23, 19-29. Doi: <https://doi.org/10.1016/j.tmp.2017.03.009>.
- Xia, L. ve Bechwati, N.N. (2008). Word of mouse: The role of cognitive personalization in online consumer reviews. *Journal of Interactive Advertising*, 9(1), 3-13. Doi: <https://doi.org/10.1080/15252019.2008.10722143>.
- Yan, J., Zhang, Z., Lin, K., Yang, F. ve Luo, X. (2020). A hybrid scheme-based one-vs-all decision trees for multi-class classification tasks. *Knowledge-Based Systems*, 198, 105922. Doi: <https://doi.org/10.1016/j.knsys.2020.105922>.

Ye, Q., Law, R. ve Gu, B. (2008). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28, 180-182. Doi: <https://doi.org/10.1016/j.ijhm.2008.06.011>.

Zhang, Y. ve Jatowt, A. (2020). Estimating a one-class naive Bayes text classifier. *Intelligent Data Analysis*, 24(3), 567-579. Doi: <https://doi.org/10.3233/ida-194669>.

Žižka, J., Dařena, F. ve Svoboda, A. (2019). *Text Mining with Machine Learning Principles and Techniques*. Boca Raton: CRC Press. Doi: <https://doi.org/10.1201/9780429469275>.