

TÜRKÇE METİNLER İÇİN DUYGU ANALİZİ YAKLAŞIMI İLE İLETİŞİMDE BAĞLAM DAN BAĞIMSIZ MODELLERİN GELİŞTİRİLMESİ ÜZERİNE BİR ARAŞTIRMA: KARMA VERİ UYGULAMASI ÖNERİSİ

Çiğdem AYTEKİN
Marmara Üniversitesi, Türkiye
cigdem.aytekin@marmara.edu.tr
<https://orcid.org/0000-0002-1385-9864>

Mehmet Ali BAYRAM
Marmara Üniversitesi, Türkiye
malibayram91@gmail.com
<https://orcid.org/0000-0003-1298-4521>

ÖZ

Günümüzde kullanıcılar ürün ve hizmetlerle ilgili olumlu veya olumsuz duygularını, direkt olarak satış için sergilendikleri mecra veya diğer mecralarda yazdıkları yorumlar vasıtası ile herkese açık bir biçimde paylaşmaktadırlar. Müşteriler tarafından birinci ağızdan aktarılan bu yorumlar, şirketler açısından büyük önem taşımaktadır. Duygu Analizi, bu yorumların olumlu olumsuz gibi duygulara ayrılması gerekliliği ile ortaya çıkmıştır. İletişimde bağlam farklı biçimlerde ele alınmaktadır. Burada bağlam, onun fiziksel boyutu olan ortam kavramsallaştırması ile değerlendirilmiş ve mecra tabanlı bir araştırma gerçekleştirilmiştir. Bu doğrultuda, farklı bağlamları temsil ettiği düşünülen ürün yorumları, film yorumları ve kitap yorumları olmak üzere üç mecra seçilmiş ve bu mecralarda yer alan kullanıcı yorumları veri olarak kullanılmıştır. İnternet ortamında yorum yazılan her mecranın kendine özgü birtakım özellikleri bulunmaktadır. Dolayısıyla mecralara ait Duygu Analizi modellerinin de farklı olması gerekir. Ancak bu pratik bir yöntem değildir. Eğer tüm mecralarda Duygu Analizi yapabilmeye elverişli bir model bulanabilirse, tüm analizler bu model ile kolaylıkla yapılabilir. Bu problem den hareketle çalışmada, çok sayıda mecra için uygulanabilecek bir Karma Veri Uygulaması geliştirilmiştir.

Anahtar Kelimeler: Duygu Analizi, Bağlam, Mecra, Karma Veri Uygulaması

A RESEARCH ON DEVELOPING CONTEXT FREE MODELS ON COMMUNICATION WITH SENTIMENT APPROACH FOR TURKISH TEXTS: A MIXED DATA MODEL PROPOSAL

ABSTRACT

Nowadays users share their positive or negative opinions on products or services publically via the comments they leave on products sale site or any other social mediums. these comments directly shared by the end user carry significant importance to corporations. Sentiment analysis has sprouted from the need to categorize those said comments such as positive or negative. In communication, context is reviewed in different terms. here context was evaluated. in this paper, context was evaluated with its conceptual medium and a medium based review was conducted. in this regard, data was acquired from product, movie and book comments given the fact that those comments represent different contexts. On the internet, every single medium that bears comments with its own unique characteristics. thus, different mediums require different sentiment analysis models; though this approach is not feasible. if a model were to be developed which could be suitable to conduct sentimental analysis on all different media, that model would be utilized to accelerate the sentimental analysis processes. Moving with this objective, a mixed data model was developed to adress this problem.

Keywords: Sentiment Analysis, Context, Medium, Mixed Data Model

GİRİŞ

Günümüzde kullanıcılar sıklıkla satın aldıkları/tükettikleri ürün ve hizmetler konusunda ilgili mecralara olumlu veya olumsuz yorumlar yazmaktadırlar. Bu yorumlar e-ticaret şirketi açısından online itibar yönetimi vb. durumlar açısından değerlendirilebilirken, diğer kullanıcılar tarafında bir takım faydalar da üretmektedir. Bu kullanıcılar bir alım esnasında ilgili yorumlara ulaşarak görüşleri değerlendirmekte ve onları referans olarak kullanmaktadır. Yorumların hacmi giderek büyümektedir. Dolayısıyla, değer elde edebilmek adına otomatik olarak analiz edilmeleri gerekir. Duygu Analizi günümüzde bu amaçla sıklıkla kullanılan bir yöntem olarak karşımıza çıkmaktadır.

Literatürde Duygu Analizine ilişkin çok farklı çalışmalar bulunmaktadır. Bunların en güncellerinden birisi Kumar ve Geetanjali'ye aittir. Onlar yaptıkları çalışmada sistematik bir literatür taraması gerçekleştirmişlerdir. Bu yolla bağlama dayalı Duygu Analizi üzerine mevcut çalışmaları araştırarak analiz etmişler ve söz konusu araştırma alanındaki boşlukları ve gelecekteki yönelimleri belirlemişlerdir. Kumar ve Geetanjali “Bağlam Tabanlı Duygu Analizi” üzerine yapılan çalışmaları şu şekilde özetlemektedir (Kumar & Geetanjali, 2020:15357):

1. Duygu Analizinde kullanılan “bağlam” teriminin farklı tanımlarını öneren çalışmalar
2. Bağlam Tabanlı Duygu Analizi üzerine anketler ve literatür çalışmaları
3. Bağlam Tabanlı Duygu Analizi üzerine karşılaştırmalı çalışmalar:
 - a. Bağlam Tabanlı Duygu Analizinde kullanılan teknikleri karşılaştırma
 - b. Bağlam Tabanlı Duygu Analizi ile Bağlam Tabanlı Olmayan Duygu Analizinin karşılaştırılması
4. Duygu Analizine ayrılmış özel bir dizi çalışma
5. Sadece dergide yayınlanan yukarıda belirtilen bilgileri içeren çalışmalar.

Bu çalışmaya, birinci maddede belirtilen “bağlamın farklı tanımlarını öneren çalışmalar” konu edilmiş ve bağlam onun fiziksel boyutu olan “mecra” tanımlamasıyla kullanılmıştır. Öngörümüze göre, her bir mecra Duygu Analizi bakımından kendine özgü birtakım özelliklere sahiptir. Farklı mecralara yazılan yorumlarda kullanılan duygu ifade eden kelimelerin farklı olabilmesi, duygu ifade eden kelimelerin bir mecrada kullanılırken bir diğer mecrada çok az kullanılması veya hiç kullanılmaması, duygu ifade eden kelimelerin farklı mecralarda farklı anlamlara gelebilmesi vb. bu özelliklere örnek olarak verilebilir.

Bu noktadan hareketle ürün yorumları, film yorumları ve kitap yorumları mecrası olmak üzere üç mecra araştırmaya konu edilmiş ve bu mecralardaki yorumlara ilişkin duygu durumlarını tahmin edebilecek makine öğrenmesi modelleri oluşturulmuştur. Daha sonra her bir model, kendi verileri ve diğer mecra verileri ile test edilmiş ve olumlu/olumsuz yorumlar için başarı oranları karşılaştırılmıştır. Sonuçta görülmüştür ki, her bir mecra kendi verileri ile test edildiğinde daha yüksek, ancak diğer mecra verileri test edildiğinde daha düşük bir başarı göstermektedir. Bu durumda analizlerde her bir mecra için kendi modelinin kullanılması gerekir. Ancak, bu pratik bir yöntem değildir. Eğer başarılı sonuçlar veren tek bir model oluşturulabilirse, tüm mecralarda aynı model kullanılarak Duygu Analizi yapılabilir. Bu nedenle çalışmada mecra değişiminden en az etkilenen -bağlamdan bağımsız- Karma Veri Uygulaması önerisinde bulunulmuştur. Önerilen model, Kumar’ın üçüncü maddede belirttiği “karşılaştırmalı çalışmalar” ile örtüşmektedir. Zira öncelikle modellerin kendi verileri ile diğer mecra verileri arasında bir karşılaştırma yapılmış ve ardından bağlamdan bağımsız bir model ortaya atılmıştır.

TÜRKÇE METİNLER İÇİN DUYGU ANALİZİ MECRALARI

Doğal Dil İşleme yaklaşımları üç grupta inceleyebilir: *Sembolik* yaklaşım, insan tarafından geliştirilen düzenlemelere ve sözlüklere dayanmaktadır. Başka bir deyişle, bu yaklaşımın temelini uzmanlar tarafından somutlaştırılan ve kaydedilen belirli bir dilde genel olarak onaylanmış konuşma düzenlemeleri oluşturur. *İstatistiksel* yaklaşım, semantik olguların gözlemlenebilir ve kalıcı örneklerine dayanır. İstatistiklere dayalı modeller, kalıcı temaları büyük metnin matematiksel yorumu yoluyla tanımlar. Bilgisayar sistemi, büyük metin örneklerindeki eğilimleri tanıyarak, gelecekteki girdi değişkenlerini ve dil çıktısının gelişimini yorumlamak için kullanacağı kendi anlamsal kurallarını geliştirebilir. *Birleştirici* yaklaşım ise, sembolik ve istatistiksel yaklaşımların bir karışımıdır. Bu yaklaşım, genel olarak onaylanmış dil kurallarıyla başlar ve bunları istatistiksel çıkarımdan elde edilen

girdilerden belirli uygulamalara dönüştürür (URL-1). İstatistiksel yaklaşıma örnek olarak Chen ve arkadaşları, Doğal Dil İşleme araştırma alanının son durumunu ortaya koymak amacıyla, alanla ilgili yayınlarda bibliyometri ve sosyal ağ analizini kullanarak veriye dayalı bir istatistiksel yöntem sunmuşlardır. 2007-2016 yılları arasında Web of Science indeksinde yer alan 3222 makale araştırmaya konu edilmiştir. İstatistiksel yöntemler kullanılarak bu makalelerin literatür dağılım özellikleri, problem edilen konuları ve yazarlar arasındaki işbirliği ilişkileri araştırılmıştır. Böylelikle, araştırmacılara doğal dil işleme alanının son durumu ile ilgili bilgiler sağlanabilmiştir (Chen, Chen, Zhang, & Hao, 2017:507). Günümüzde yapılan Doğal Dil İşleme çalışmalarının çok büyük bir kısmı istatistiksel alandadır.

Duygu Analizi bir istatistiksel Doğal Dil İşleme görevidir. Kavram ilk olarak Nasukawa ve Yi'nin 2003 yılındaki KCAP Konferansı'nda sundukları "Sentiment analysis: Capturing favorability using natural language processing" (Nasukawa & Yi, 2003) isimli çalışmalarında ortaya atılmıştır. Duygu Analizi, farklı mecralarda yer alan metin formatındaki verilerden, onlarda gizli olan olumlu, olumsuz gibi duygu durumlarının otomatik olarak keşfedilmesini hedefler.

Duygu Analizi ile ilgili Türkçe çalışmalar çok farklı biçimlerde kategorize edilebilir. Burada çalışmanın gereği doğrultusunda, analizlerde kullanılan mecralar düzeyinde kategorizasyona ilişkin bir literatür taraması yapılmıştır.

Twitter Mecrasında Türkçe Duygu Analizi Çalışmalarından Örnekler

Twitter'dan elde edilen verilere ilişkin Duygu Analizi markalar için online itibar yönetimi, müşteri memnuniyetini artırma; kamuoyunu ilgilendiren önemli bir konu hakkında kullanıcı duygularını öğrenme; finansal konularda kullanıcı eğilimlerini belirleme vb. gibi farklı amaçlara hizmet edebilir. Twitter'da Duygu Analizi ile elde edilen veriler hükümetler ve bireysel karar vericiler için de temel bir kaynak olabilir. Böylesine çalışmalarda tweet'ler olumlu, olumsuz ya da nötr olarak sınıflandırılabilirdiği gibi daha geniş bir skalaya da yayılabilir (1-5 veya 1-10 gibi).

Çoban ve arkadaşları yaptıkları çalışmada Twitter API'den toplanan Türkçe tweet'ler üzerinden bir Duygu Analizi gerçekleştirmişlerdir. SVM, Naive Bayes, Multinomial Naive Bayes ve KNN gibi makine öğrenme algoritmaları üzerinden sınıflandırmalar yapılmış ve sonuçlar karşılaştırmalı olarak incelenmiştir (Çoban, Özyer, & Özyer Tümüklü, 2015:1).

Akgül ve arkadaşları "Duygusal Twitter" adını verdikleri çalışmalarında belli bir anahtar kelime ile Twitter'dan veri çekmişlerdir. Yapılandırdıkları sistem, elde edilen tweet'leri hem sözlük hem de n-gram modeli ile olumlu, olumsuz ve nötr olarak otomatik biçimde etiketlemektedir. Ayrıca sistem tasarımı esnek, kişi ve kurumlara kendileri için özel sözlükler oluşturmalarına izin vermektedir. Kullanılan sözlük ve karakter tabanlı n-gram yöntemlerinde yaklaşık olarak sırasıyla %70 ve %69 başarı elde edilmiştir (Akgül, Ertano, & Diri, 2016:110).

Eliaçık ve Erdoğan çalışmalarında, finans ile alakalı Twitter kullanıcı verilerini göz önünde bulundurarak, o kullanıcının finans ile alakalı ilgi düzeyini ve finansal topluluk içerisindeki inandırıcılığını kullanan yeni bir topluluk düzeyi Duygu Analizi yöntemi önermişlerdir. Yapılan analizler sonucu önerilen yöntemin önceki yöntemlere göre finansal sosyal toplulukların duygu polaritesi ile borsa fiyatları arasındaki bağıdaklık oranını daha hassas hesapladığı görülmüştür (Eliaçık & Erdogan, 2015:792).

Onan ve Bayar'a göre, Duygu Analizi bir metin sınıflandırma problemi olarak modellenebilir. Bununla birlikte, metin sınıflandırmasının önemli problemlerinden biri de yüksek boyutluluk'tur. Bu doğrultuda çalışmada, Türk Twitter mesajlarını verimli bir şekilde temsil etmek için Latent Dirichlet tahsis yöntemi kullanılmış ve makine öğrenme sınıflandırıcılarının tahmin performansı değerlendirilmiştir. Ampirik analizde karşılaştırılan yöntemler arasında en yüksek kestirimci performans, Naive Bayes algoritması ile elde edilmiştir (Onan & Bayar, 2017).

Akın ve Şimşek çalışmalarında, bir kanalda sekiz aylık periyotta yayımlanan programlar hakkında yazılan tweet'leri veri seti olarak kullanmışlardır. Duygu Analizi ile bu veri setinde yer alan kanal ve program hakkındaki iletilerin olumlu, olumsuz ya da nötr olarak sınıflandırılan duygulardan hangisini içerdiği ortaya konmuştur. Bu bilgi ile kanal yöneticileri, program sorumluları için öngörülerde bulunabilmiş ve bu doğrultuda ilgili stratejileri geliştirebilecekleri belirtilmiştir (Karaöz Akın & Gürsoy Şimşek, 2018:249-250).

Nizam ve Akın çalışmalarında, denetimli öğrenme yaklaşımını kullanarak Twitter mecrasında Duygu Analizi yapmışlardır. Veri seti, bazı gıda firmalarının çeşitli ürünlerine yapılan yorumlardan oluşmaktadır. Tweet’ler makine öğrenmesi yöntemlerinden unigram özelliğine göre analiz edilmiş, farklı sınıflandırma algoritmalarının gösterdikleri başarı oranları incelenmiştir. Elde edilen deneysel sonuçlarda sınıflar arası uygun dağılım gösteren dengeli veri setinin dengesiz veri setine göre daha iyi performans gösterdiği izlenmiştir. En iyi performans gösteren sınıflandırma algoritması ise %72,33 ortalama doğruluk başarı oranıyla Sequential Minimal Optimization Algoritması olmuştur (Nizam & Akın, 2014:1).

Onan yaptığı çalışmada, Türkçe tweet’ler üzerinde makine öğrenmesi sınıflandırıcılarını kullanarak Duygu Analizi gerçekleştirmiştir. Metin sınıflandırıcılarının başarımlarında veri temsil modeli büyük önem taşımaktadır. Çalışma kapsamında tweet’lerin sınıflandırılmasında, üç temel makine öğrenmesi sınıflandırıcısı (naive bayes algoritması, destek vektör makineleri ve lojistik regresyon) ve üç temel temsil modeli (1-gram, 2-gram ve 3-gram) ile bu temsil modellerinin farklı bileşenleri değerlendirilmiştir. Sonuçlar, en yüksek başarımın Naive Bayes algoritması ile ve veri seti 1-gram ve 2-gram öznitelik setlerinin birleştirilmesi ile oluşturulan öznitelik seti ile elde edildiğini göstermektedir (Onan, Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi, 2017:12).

Diğer Mecralarda Türkçe Duygu Analizi Çalışmalarından Örnekler

Çetin ve Eryiğit çalışmalarında Türkçe hedef tabanlı Duygu Analizini esas alarak restoran yorumlarından bir veri kümesi oluşturmuşlardır. Onlar, hedef terim ve hedef kategori bulma görevlerinde, kelime vektörlerinin pozitif etkileri ve Türkçe’nin serbest dizimli yapısı sebebiyle ortaya çıkan problemleri aşmak amacı ile tüm alt görevlerde bağlılık ayrıştırıcı kullanmanın faydalarını ortaya çıkarmışlardır (Çetin & Eryiğit, 2018:55).

Demir ve arkadaşları yaptıkları çalışmada sinema filmlerinin yorumlandığı bir web sitesinden elde edilen yorumlar, bir e-ticaret sitesinde yer alan kitapları değerlendirmeye yönelik okuyucu yorumları ve diğer yorumlar olmak üzere üç veri seti kullanmışlardır. Buradaki yaklaşım sözlük tabanlı olup Afinn sözlüğü, Bing sözlüğü, NRC sözlüğü ve SentiTurkNet’ten yararlanılmıştır. Farklı kombinasyonlarla yapılan analiz sonuçları en yüksek %82,85 ve en düşük %72,78 olarak gerçekleşmiştir. Hatalı olduğu tespit edilen sonuçlara ilişkin gözlemlenen durumlar şu şekilde sıralanabilir: Hatalı yazılan kelimeler, kinayeli yorumlar, art niyetle yazılmış yorumlar, birden fazla anlamlı bazı deyim ve kelimeler, karşılaştırma içeren cümlelerin farklı taraflar için farklı duygular taşınması gibi durumlar (Demir, Baban Chawai, & Doğan, 2019:62-64).

Kaynar ve Yıldız çalışmalarında, film yorumlarının içeriğine göre Naive Bayes, Merkez Tabanlı Sınıflayıcı, Çok Katmanlı Yapay Sinir Ağları (MLP) ve Destek Vektör Makineleri (SVM) gibi sınıflandırma algoritmalarını kullanarak Duygu Analizi yapmışlardır. Yapılan analizler sonucunda, gerek eğitim gerekse test verilerinde yapay sinir ağları ve destek vektör makinelerinin diğer yöntemlere oranla daha iyi sonuç verdiğini gözlemlemişlerdir (Kaynar & Yıldız, 2016).

Atan ve Çınar yaptıkları çalışmada, Borsa İstanbul’da işlem gören BIST30 şirketlerine ilişkin olarak 2014 yılında farklı haber kaynaklarında yayınlanmış 14.108 haber metnini veri seti olarak kullanmışlardır. Haber içeriklerinde yer alan bu ifadeler, Türkçeye çevrilmiş bir duygu sözlüğü yardımıyla sayısal değerlere dönüştürülmüştür. Daha sonra, bu sayısal skorlar ile aynı dönemde piyasada oluşan şirket değerleri arasındaki ilişkiler analiz edilmiştir. Ortaya çıkan temel sonuç, finansal piyasalarla yayınlanan haberler ve bunların duygu tonları ile finansal değerler arasında anlamlı ilişkilerin var olduğudur (Atan & Çınar, 2019:1).

Gözükara ve Özel yaptıkları çalışmada, veri seti olarak İngilizce ve Türkçe dillerinde yazılmış film ve ürün yorumlarını ve algoritma olarak da Destek Vektör Makineleri (DVM) sınıflayıcısını kullanmışlardır. Ayrıca doküman vektörü hesaplama yöntemlerinin Türkçe ve İngilizce veri kümeleri üzerindeki detaylı karşılaştırmaları da çalışmada yer almaktadır (Gözükara & Özel, 2016:467).

Sütcü ve Aytekin yaptıkları çalışmada, beyazperde.com sitesinde yer alan filmlere ilişkin olarak paylaşılan kullanıcı yorumlarının örneklem boyutta Duygu Analizini yapmışlardır. Filmlere ilişkin geri bildirimlerin bu şekilde yapay zekâ yöntemleri ile otomatik olarak yapılandırılması, başta filmi izleme niyetinde olanlar olmak üzere, filmin oyuncularını, yapımcılarını ve reklamverenlerini açısından paha

biçilmez değerdedir. Böylesine bir uygulama, medya işletmelerine etkin stratejiler oluşturma konusunda yeni ve geliştirilebilir olanaklar sunar (Sütçü & Aytakin, 2019:61).

Yukarıda örnekleri verilen Duygu Analizi çalışmaları Twitter ve diğer mecralar olmak üzere başlıca kategorilere ayrılmıştır. Hazır veri setlerinin kullanıldığı çalışmalar da bulunmaktadır, ancak burada çalışmanın amacı gereği dahil edilmemişlerdir. Twitter mecrası, yapılan bölümlenmede analizlerin yoğun olarak yapıldığı bir ortam olarak karşımıza çıkmaktadır. Zira bu ortam her geçen gün artan kullanıcı sayısı ile çok fazla sayıda ve değişik konularda tweet barındırmaktadır ve hesapların önemli bir bölümü korumalı değildir. Diğer mecralar ise haber siteleri, film yorumları, e-ticaret yorumları, restoran yorumları, kitap yorumları siteleri gibi daha alt kategorilere ayrılabilir. Duygu Analizi çalışmalarında amaç yeni bir yöntem geliştirme, algoritma performansı ölçme, hatalı durumları tespit etme, bireysel veya kurumsal kullanıcılar için karar vermeye dayalı sonuçlar üretme vb. olabilir. Burada önemli olan, yapılacak olan çalışmanın hedefi doğrultusunda bir değer yaratmaktır.

Bu çalışmada “diğer mecralar”, değer yaratılmak istenen hedef doğrultusunda çalışmaya konu edilmiştir. Film yorumları, ürün yorumları, kitap yorumları sitelerinden elde edilen metinler veri seti olarak kullanılmış ve bu yorumlara ilişkin “bağlam” durumu üzerine bir araştırma gerçekleştirilmiştir.

İLETİŞİMDE BAĞLAM

Amerikalı siyaset bilimci Lasswell, 1948 yılında yazdığı bir makaleye şu cümleyle başlar (Mcquail & Windahl, 2010:27):

“Bir iletişim eylemi en kolay şekilde şu sorular yanıtlanarak açıklanabilir:

Kim? Ne söyler? Hangi kanal ile? Kime? Ne gibi bir etki ile?”

1956 yılında ortaya konan Gerbner Genel İletişim Model’inde ise, aşamalar Laswell’e benzer biçimde şu şekilde sunulur (Mcquail & Windahl, 2010:38):

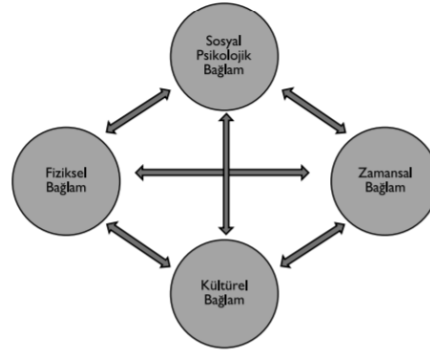
1. Bir kişi
2. Bir olayı algılayıp
3. Tepkide bulunduğu
4. Belli bir ortamda
5. Bazı araçlar kanalıyla
6. Kullanılabilir bir malzeme hazırlar
7. Bunun bir biçimi
8. Ve **bağlamı** vardır
9. İçeriği aktararak içerik iletir
10. Ve bazı sonuçlara yol açar.

Gerbner’in modeli iletişimde doğrusal modeller içinde yer alır. Oysa Jakobson iletişime gösterebilimsel açıdan yaklaşmış ve iletişimin oluşturu etmenlerini Şekil 1’deki gibi modellemiştir (Fiske, 2003:56). Bu modelde, doğrusal modellerde yer alan iletinin **bağlam**, temas ve kod gibi öğelerle genişletildiği görülmektedir.

Gönderen	Bağlam	Alıcı
	İleti Temas Kod	

Şekil 1. İletişimin Oluşturucu Etmenleri

Bir öge olarak iletişim modelleri içerisinde önemli bir yere sahip olan bağlam, iletişimin gerçekleştiği çevre olarak da ele alınmaktadır. Bu paralellikte Türk Dil Kurumu Sözlüğü bağlamı dilbilimsel açıdan “bir dil birimini çevreleyen, ondan önce veya sonra gelen, birçok durumda söz konusu birimi etkileyen, onun anlamını, değerini belirleyen birim veya birimler bütünü” olarak tanımlamıştır (URL-2). Küçüğe göre de, insanlar arasında gerçekleşen bütün iletişimler bir bağlam içinde meydana gelir ve bu bağlamlar birbirleriyle etkileşim içerisinde. Bağlam, iletişim sürecinde içinde bulunan birbirleriyle ilişkili koşullardır. *Başka bir deyişle bağlam, bir söz ya da davranışın içinde geliştiği ve ona anlam kazandıran çevre olarak da tanımlanabilir.* Buradan bağlamın çok boyutlu ve kapsayıcı bir terim olduğu anlaşılmaktadır. İletişimin içinde geliştiği bağlam Şekil 2’de görüldüğü gibi fiziksel bağlam, sosyal-psikolojik bağlam, zamansal bağlam ve kültürel bağlam olmak üzere dört boyutla açıklanabilir (Küçük, 2012:14).



Şekil 2. İletişim Bağlamı

Fiziksel bağlam, iletişimin meydana geldiği gözle görülür somut çevredir. İçinde bulunduğunuz oda, koridor, park ya da toplantı salonu bu bağlama örnek olarak verilebilir. Fiziksel bağlamın mesajınızın içeriği (ne söylediğiniz) ve biçimi (nasıl söylediğiniz) üzerinde bazı etkileri olur. *Sosyal-psikolojik bağlam*, katılımcılar arasındaki statü ilişkileri, roller ve toplumun kültürel kuralları gibi öğeleri içerir. Ayrıca, belli bir durum içinde resmi olma ya da olmama gibi zıt durumları da içerir. *Zamansal bağlam*, iletişimin gerçekleştiği zamanı kapsar. Bu zaman bir gün olabileceği gibi, çok daha uzun tarihsel bir süreç de olabilir. Bu çerçevede mesajın etkisinin ve uygunluğunun zamana bağlı olduğu söylenebilir. Örneğin, bir değerle ilgili tutum ve değerler tarihsel süreç içinde değişebilir. *Kültürel bağlam*, kültür ile ilgili her şeyden beslenir. Kültür, bir toplumun tarihsel süreç içinde ürettiği ve gelecek kuşaklara aktardığı tüm özelliklerin toplamıdır. Kültürel faktörler iletişimde neyi nasıl söylediğimizi etkiler. Dolayısıyla bir toplumdaki deyiş biçimi, farklı bir toplumunki ile örtüşmeyebilir (Küçük, 2012:14-15).

Ayrıca, bağlama ilişkin yüksek bağlamlı/düşük bağlamlı biçiminde yaklaşımlar da bulunmaktadır. Edward Hall, “Beyond Culture” adlı kitabında bağlamı bir duyu olarak tanımlar ve iletişim bilimleri açısından yüksek bağlam/düşük bağlam kültürlerinden bahseder. Kültürün dünya görüşlerini nasıl etkilediğine dair kanıtlar sunar (Hall, 1989). Hofstede’ye göre, yüksek bağlamlı iletişim daha çok geleneksel kültürlerde, düşük bağlamlı iletişim ise modern kültürlerde bulunur. Bu nedenle yüksek bağlamlı/düşük bağlamlı ayrımı, kısmen geleneksel/modern ayrımıyla örtüşür (Hofstede, 2001:30). Bu yaklaşımların yukarıda sözü edilen kültürel bağlam içinde değerlendirileceği söylenebilir. Bu çalışmada ise bağlam fiziksel boyutuyla ele alınacaktır.

YÖNTEM VE BULGULAR

Amaç ve Önem

Bu çalışmada amaç, Duygu Analizi ve makine öğrenmesi yöntemleri ile iletişimde bağlamın etkisini ortaya koymak ve bu etkiyi ortadan kaldıracak bağlamdan bağımsız modellerin geliştirilmesi için bir Karma Veri Uygulaması önerisinde bulunmaktadır. Böylelikle, mecranlarda Duygu Analizi yapılırken her bir bağlama ilişkin modeller yerine, karma veri kümelerinden oluşturulacak bir model kullanılabilir ve çok sayıda mecranın Duygu Analizi bu model ile pratik bir şekilde yapılabilir. Gerçekleştirilen

araştırma, bağlamdan bağımsız modellerin geliştirilmesi için önerdiği yöntemin ilk olması bakımından önemlidir.

Verilerin Toplanması ve Örneklem

Her bir mecranın, bağlamın *fiziksel* boyutu olan ortamı temsil ettiği varsayılarak, ürün yorumları mecrası, film yorumları mecrası ve kitap yorumları mecrası olmak üzere üç mecra seçilmiştir. Dolayısıyla her bir mecra, Duygu Analizi bakımından makine öğrenmesine kılavuzluk edecek kendine özgü birtakım özelliklere sahiptir:

- Film yorumu yazan bir kullanıcının kullandığı duygu ifade eden kelimeler, ürün yorumları yazan bir kullanıcının kullandığı kelimelerden farklıdır. Örneğin, film yorumlarında “sıkıcı”, “eğlenceli”, “sürükleyici”, “heyecan verici” gibi duygu ifade eden kelimeler kullanılır, ancak bu kelimeler ürün yorumları için çoğunlukla geçerli değildir. Zira filmler kültür ürünleridir. Aynı durum kitaplar için de geçerlidir.
- Kullanıcıların yorum yazmada kullandıkları duygu ifade eden kelimeler bir mecrada kullanılırken, bir diğer mecrada kullanılmayabilir. Örneğin, “...film izlerken ağladım...” yorumundaki “ağlamak” kelimesinin sıklığı, kitap yorumları mecrasında daha düşük değerde olabilir. Zira “kitap okurken ağlamak” filme göre çok daha az gerçekleşen bir durumdur.
- Bir duygu durum bildiren kelime, farklı mecralardaki yorumlarda farklı anlamlarda kullanılıyor olabilir. Örneğin, “filmin finali de çok vasat ve çok bilindik” yorumundaki “bilindik” kelimesi olumsuz bir anlam ifade ederken, bir ürün yorumunda “bilindik bir marka” biçiminde olumlu anlamda kullanılmış olabilir.

Bu doğrultuda üç mecraya ilişkin yorumlar ve onlara ilişkin kullanıcı puanlamaları (1/2 yıldız, yıldız ya da 1-5 aralığında rakamsal puanlar) bir veritabanına kaydedilmiştir. Verinin büyük miktarda olması manuel olarak toplanmasını zor hale getirdiğinden, otomatik veri toplama işlemi için python 3.7.5 yazılım dili kullanılmıştır. Bu işlemin mümkün olan en kısa sürede ve sorunsuz tamamlanabilmesi için de urllib3 kütüphanesi seçilmiştir.

Ürün yorumları mecrası olarak yorumbudur.com (28.01.2020-30.01.2020) ve hepsiburada.com (30.01.2020-03.02.2020) siteleri, film yorumları mecrası için beyazperde.com (08.03.2020-09.03.2020) sitesi ve kitap yorumları mecrası için de kitapyurdu.com (05.05.2020-09.05.2020) sitesi kullanılmıştır.

yorumbudur.com sitesindeki yorumları elde edebilmek amacıyla, öncelikle her bir ürüne ait bağlantı adresini alabilmek üzere gerekli fonksiyonlar oluşturulmuştur. Bu yolla ana sayfadan /urunler/ şeklinde başlayan 120 bağlantıya istek gönderilerek 2.841 ürün bağlantı adresine ulaşılmış ve sonra tümüne yapılan istek sonucunda 122.366 adet farklı ürüne ait bağlantı adresi elde edilip urunlinkleri.txt isimli dosya ile algoritmaya kaydedilmiştir. Daha sonra, ürünlere ait yorumların bulunduğu sayfaların bağlantı adreslerine ulaşabilmek için de istek gönderilmiş, /yorumlar/ şeklinde başlayan bağlantılar elde edilmiş ve listeye eklenmiştir. Son olarak listenin tamamı yorumbaglantilari.txt dosyasına kaydedilmiştir. Bu işlem klasik bir programlama ile normal bir bilgisayarda çok uzun süreceği için paralelleştirme yapılmıştır. Google Cloud sunucu bilgisayarlarında 10 adet Python dosyası aynı anda çalıştırılmak suretiyle işlem başlatılmış ve ortak ürünlerin elenmesiyle birlikte 25.292 adet bağlantı elde edilmiştir. Bu doğrultuda elde edilen bağlantılara tekrar istekler gönderilerek ürünlere ait yorumların başlığı, içeriği ve puanı elde edilmiş, daha sonra gereksiz boşluk ve noktalama işaretlerinden temizlenerek Pandas kütüphanesi aracılığı ile csv dosyalarına yazılmıştır. Böylelikle 254.100x4 adet veriden oluşan bir veritabanı elde edilmiştir. Benzer biçimde, hepsiburada.com sitesinden 2.657.072x4 adet, beyazperde.com sitesinden 192.073x3 adet ve kitapyurdu.com sitesinden 819.408x3 adet veriden oluşan veritabanları elde edilmiştir. Veritabanında ürün ile ilgili sütun başlıkları sıra-puan-başlık-yorum (4 adet) biçiminde olurken, film ve kitap ile ilgili sütun başlıkları mecranın tasarımı gereği sıra-puan-yorum (3 adet) biçiminde olmaktadır. Veri toplama işlemi sonrasında yorumbudur klasöründe 254.101 yorum, hepsiburada klasöründe 2.657.073 yorum, beyazperde klasöründe 192.074 yorum ve kitapyurdu klasöründe ise 819.409 yorum bulunmaktadır.

İstatistiksel yaklaşıma uygun olarak doğal dil işleme yapabilmek için öncelikle benzersiz her bir kelimeyi belirli bir sayı ile temsil etmek gerekir. Bu nedenle de öncelikle metindeki tüm kelimelerin birbirinden ayrıştırılması, sonra da bir sayıya atanmaları gerekir. Jetonlaştırma adı verilen bu işlem sayesinde metin verilerin tümü sayısal verilere dönüştürülmüş olur. Bu işlem sırasında veri içinde yer alan kelimeler çok kullanılanlardan az kullanılanlara doğru sıralanır ve en çok kullanılan belli sayıdaki kelime seçilir. Bu yaygın bir yaklaşımdır, ancak bu yaklaşımda model hazırlandıktan sonra kullanım aşamasına hazır olan kelimeler, model eğitilirken kullanılan verinin içinde bulunmaması durumunda değerlendirilmeye alınmayacaktır. Bu araştırmada durumun alternatifi olarak üç yöntem önerilmektedir:

1. Toplanan tüm veriler birleştirilir, elde edilen toplam veri içinde en çok kullanılan ilk 20.000 kelimenin bulunabilmesi için Python Keras kütüphanesi kullanılabilir ve böylelikle elde edilen kelimeler jetonlaştırma işleminde kullanılabilir.
2. Sak ve arkadaşları (Sak, Güngör, & Saraçlar, 2008) 2008 yılında Türkçedeki tüm kelime ve onlara ilişkin kullanım sıklıkları üzerine bir araştırma yapmışlardır ve bu dosya Bilgin (URL-3) tarafından web sitesinde csv formatında yayınlanmıştır. Dolayısıyla bu 1.337.898 kelimele derlemedeki kullanım sıklıklarına bakılabilir ve çok kullanılanlardan az kullanılanlara doğru sıralama yapılabilir. Daha sonra en çok kullanılan 20.000 kelime alınarak jetonlaştırma işleminde kullanılabilir.
3. Birinci ve ikinci madde birlikte uygulanarak kelime listeleri birleştirilebilir ve kullanım sıklığına göre tekrar sıralanarak ilk 20.000 kelime alınabilir.

Burada üçüncü maddede önerilen yöntem kullanılmıştır. Listelerde tekrar eden kelimeler döngüler yardımıyla silinmiş ve çok kullanılanlardan az kullanılanlara doğru sıralanarak 28.592 kelime elde edilmiştir. 20.000 kelime yaklaşımına uygun olarak sondaki 8.592 kelime silinmiş ve tüm platformlarda kullanılabilir biçimde json formatında kaydedilmiştir.

Verilerin Model Eğitimi için Hazırlanması ve Yöntem

yorumbudur.com ve hepsiburada.com web sitelerinden alınan ve sıra-puan-başlık-yorum isimli sütunlardan oluşan veriler öncelikle başlık ve yorum sütunları birleştirilerek sıra-puan-yorum olmak üzere üç sütunlu veriye dönüştürülmüştür. Puan satırında 1 yıldızdan başlayıp birer puan artacak şekilde 5 yıldız kadar ilerleyen 5 farklı puanın 1 puan olanları 0.0 ve 5 puan olanları 1.0 olarak yeniden düzenlenmiştir. Yorum kısmı puanı veren tarafından boş bırakılmış, sadece puanlaması yapılmış olan yorumlar ve nötr puan verilen yorumlar veriden çıkarılmıştır. Elde edilen 133.494 olumsuz ve 1.935.356 olumlu yorum, beyazperde.com web sitesinden alınıp süzülen veri sayısına eşit olabilmesi için 52.610'a düşürülmüş ve böylelikle 105.220 yorum elde edilmiştir.

Eldeki veriler içinde en az sayıya sahip olan beyazperde.com web sitesinden alınan ve sıra-puan-yorum olmak üzere iki sütundan oluşan verilerin, öncelikle puan satırında bulunan 0.5 yıldızdan başlayıp yarım puan artacak şekilde 5.0'a kadar ilerleyen 10 farklı puanın 3'ten düşük olanları 0.0 ve 4'ten büyük olanları 1.0 olacak şekilde yeniden düzenlenmiştir. Yorum kısmı puanı veren tarafından boş bırakılarak sadece puanlama kısmı bulunan yorumlar ve nötr puan verilen yorumlar veriden çıkarılarak 124.673 yoruma düşürülmüştür. Bu durumda olumsuz yorum sayısının 52.610 olduğu tespit edildiğinden, olumlu yorumlardan da sadece 52.610 yorum alınmış ve böylelikle toplam yorum sayısı 105.220 olarak kaydedilmiştir.

kitapyurdu.com web sitesinden alınan ve sıra-puan-yorum isimli üç sütundan oluşan verinin puan satırında bulunan 1 yıldızdan başlayıp birer puan artacak şekilde 5 yıldız kadar ilerleyen 5 farklı puanın 3'ten düşük olanları 0.0 ve 4'ten büyük olanları 1.0 olarak yeniden düzenlenmiştir. Yorum kısmı, puanı veren tarafından boş bırakılarak sadece puanlama kısmına sahip yorumlar ve nötr puan verilen yorumlar veriden çıkarılmıştır. Elde edilen 68.841 olumsuz ve 613.338 olumlu yorum, beyazperde.com web sitesinden alınıp süzülen veri sayısına eşit olabilmesi için 52.610'ar veriye indirilmiş ve böylelikle 105.220 adet veri elde edilmiştir.

Daha sonra, açıklanan biçimde hazırlanan toplam 315.660 verinin her bir grubundan 1/3'ünün alınması suretiyle, eşit sayıda olumlu ve olumsuz örnek barındıran 105.220 adet karma veri kümesi oluşturulmuştur. Örneklem sayısı ise, daha önce oluşturulan 4 veri grubunun her birinden rastgele 500

olumlu, 500 nötr ve 500 olumsuz yorum alınarak elde edilmiştir. Böylece toplam 6.000 veri sunucuya kaydedilmiştir.

Diğer yandan, araştırmanın verileri ikincil veriler olup yukarıda belirtilen belli bir dönemi kapsamaktadır. İz sürme yöntemi ile dört web sitesi tespit edilmiş ve yorumlardan oluşan veri tabanları elde edilmiştir. Araştırma amacı bakımından ise keşfedici gruptadır.

Model

Farklı mecralardan elde edilen verilerle eğitilmiş makine öğrenmesi modelleri oluşturabilmek için, yaygın makine öğrenmesi kütüphanelerinden Tensorflow Kütüphanesi seçilmiş ve pip kurulum aracı vasıtasıyla kurulumu gerçekleştirilmiştir.

Her bir model için elde edilen veriler Sklearn Kütüphanesinden yararlanılarak rastgele karıştırılmış ve verinin %90'ı modelin eğitimi, %10'u da modelin kendi kendini değerlendirmesi için ayrılmıştır. Modeli eğitebilmek için verinin yorum kısmı puan kısmından ayrılmış, yorum kısmındaki metinsel ifadeler daha önce hazırlanan jetonlaştırma işlemi ile sayısal ifadelere dönüştürülmüştür. Jetonlaştırma işlemi için her bir yorumdaki kelime, jetonlaştırma listesinde bulunması halinde, ona karşılık gelen sayısal ifade uzunluğu en fazla 50 elemanı geçmeyecek şekilde ayarlanmış ve listeye eklenmiştir. İşlem sonunda oluşan listenin eleman sayısının 50'den küçük olması durumunda başına sıfırlar eklenmiş ve her bir yoruma karşılık 50 sabit uzunluğunda jetonlaştırılmış yorum listesi oluşturulmuştur (Şekil 3). Bu sayede aralarındaki ilişkinin doğru bir şekilde ortaya çıkabilmesi mümkün olmuştur.

```
1. def tokenlestir(yorumListesi):
2.     y_yorumlar = []
3.     for yorum in yorumListesi:
4.         y_yorum = []
5.         for kelime in str(yorum).lower().split():
6.             if len(y_yorum) < 50 and kelime in json_tokenizer:
7.                 y_yorum.append(json_tokenizer[kelime])
8.
9.         if len(y_yorum) < 50:
10.            sifirlar = list(np.zeros(50 - len(y_yorum), dtype=int))
11.            y_yorum = sifirlar + y_yorum
12.
13.        y_yorumlar.append(y_yorum)
14.    return np.array(y_yorumlar, dtype=np.dtype(np.int32))
```

Şekil 3. Jetonlaştırma İşlem Kodları

Tensorflow kütüphanesinde bulunan Keras Kütüphanesi içerisinde yer alan derin öğrenme modellerinden sıralı model seçilmiştir. Daha sonra dışardan gelecek veriyi doğru şekilde kabul edecek ilk katman olarak, hazırlanan verinin vektör boyutuna uygun uzunluğu 50 olan ve en büyük değeri 20.001'den küçük olan listeleri giriş vektörü olarak kabul eden gömülü giriş katmanı eklenmiştir. Çıkış katmanı için ise, Keras Kütüphanesinin tek bir değer üreten ve aktivasyon fonksiyonu sigmoid olan derinlik katmanı eklenmiştir. Eğitim boyunca doğruluk oranının artırabilmesi için başarı ölçütü olarak binary_crossentropy ve optimizasyon algoritması olarak da Adam Algoritması kullanılmıştır.

Modelin her bir devirde 512 veriyi aynı anda işleyeceği ve toplam 25 devir boyunca tekrar edecek olan eğitim işlemi başlatılmış, bu işlem boyunca gelişim aşaması grafik olarak kayıt altına alınmıştır. Son olarak, modeli test etmek için ayrılan %10'luk verinin kendini değerlendirmesi sağlanmıştır. Mecralara ilişkin modellerin genel doğruluk ve kayıp oranları **Hata! Başvuru kaynağı bulunamadı.**'de görülmektedir.

Mecra	Modelin Genel Doğruluk Oranı	Modelin Genel Kayıp Oranı
yorumbudur.com hepsiburada.com	%98	%0,3
beyazperde.com	%97	%1,2
kitapyurdu.com	%97	%1,2
karma veri uygulaması	%97	%1

Böylelikle, doğal dil işleminin istatistiksel yaklaşımına uygun olarak dili işleyen ve bu yolla yorumlardaki duygu durumunu tahmin edebilecek makine öğrenmesi modelleri oluşturulmuştur.

Modellerin Testi ve Bulgular

Oluşturulan modeller her bir mecra için iki farklı şekilde kullanılmış ve aralarında bir karşılaştırma yapılmıştır (Tablo 1, Tablo 2, Tablo 3):

- Modelin kendisi ile aynı mecradaki verilerle test edilmesi sonucunda elde edilen bulgular
- Modelin diğer mecralardan alınan verilerle test edilmesi sonucunda elde edilen bulgular

Tablo 1. Ürün Yorumları Mecrası Modelinin Kendi Verileri ve Diğer Mecra Verileri ile Test Edilmesi Sonucunda Elde Edilen Bulgular

Kendi Verileri			
	Olumlu (500 Adet)	Nötr (500 Adet)	Olumsuz (500 Adet)
Başarı	444	17	489
Hata	56	483	11
Başarı Oranı	%88,8	%3,4	%97,8
Diğer Mecra Verileri			
	Olumlu (1500 Adet)	Nötr (1500 Adet)	Olumsuz (1500 Adet)
Başarı	701	74	1351
Hata	799	1426	149
Başarı Oranı	%46,7	%4,9	%90

Tablo 2. Film Yorumları Mecrası Modelinin Kendi Verileri ve Diğer Mecra Verileri ile Test Edilmesi Sonucunda Elde Edilen Bulgular

Kendi Verileri			
	Olumlu (500 Adet)	Nötr (500 Adet)	Olumsuz (500 Adet)
Başarı	456	36	452
Hata	44	464	48
Başarı Oranı	%91,2	%7,2	%97,8
Diğer Mecra Verileri			
	Olumlu (1500 Adet)	Nötr (1500 Adet)	Olumsuz (1500 Adet)
Başarı	903	176	1061
Hata	597	1324	439
Başarı Oranı	%60,2	%11,7	%70,7

Tablo 3. Kitap Yorumları Mecrası Modelinin Kendi Verileri ve Diğer Mecra Verileri ile Test Edilmesi Sonucunda Elde Edilen Bulgular

Kendi Verileri			
	Olumlu (500 Adet)	Nötr (500 Adet)	Olumsuz (500 Adet)
Başarı	370	95	425
Hata	130	405	75
Başarı Oranı	%74	%19	%85
Diğer Mecra Verileri			
	Olumlu (1500 Adet)	Nötr (1500 Adet)	Olumsuz (1500 Adet)
Başarı	1085	125	1126
Hata	415	1375	374
Başarı Oranı	%72,3	%8,3	%75

Tablolar incelendiğinde görülmektedir ki, her bir model olumlu ve olumsuz yorumlar için kendi alanındaki verilerle test edildiğinde daha yüksek, ancak diğer mecralardan alınan verilerle test edildiğinde daha düşük bir başarı göstermektedir (modellerin nötr yorumlar karşısında gösterdiği düşük başarı için web ve mobil ortamlarda bir uygulama gerçekleştirilmiş ancak bu çalışmaya dahil edilmemiştir). Mecra bağlamın fiziksel boyutunu oluşturmaktadır. Buna göre, Duygu Analizi çalışmaları yapılırken farklı mecralar için farklı modellerin kullanılması gerekir. Ancak, her mecra için Duygu Analizine ilişkin farklı modellerin kullanılması pratik bir yöntem değildir. Eğer başarılı sonuçlar veren tek bir model oluşturulabilirse, tüm mecralarda aynı model kullanılarak Duygu Analizi yapılabilir. Bu nedenle mecra değişiminden en az etkilenen -bağlamdan bağımsız- **Karma Veri Uygulaması** önerisinde bulunulmuştur. Bu model karma veriler ile eğitilmiştir. Karma veriler, yukarıda açıklandığı şekliyle birden fazla ve birbirinden farklı mecralardan elde edilmiştir.

Tablo 4’de Karma Veri Uygulaması ile elde edilen sonuçlar görülmektedir. Modelin olumlu yorumlar için başarısı %79,8 ve %77,8; olumsuz yorumlar için başarısı %85,8 ve %85,4 olarak gerçekleşmiştir. Bu sonuçların birbirine çok yakın olması Karma Veri Uygulamasının başarısını ortaya koymaktadır.

Tablo 4. Karma Veri Uygulamasının Kendi Verileri ve Diğer Mecra Verileri ile Test Edilmesi Sonucunda Elde Edilen Bulgular

Kendi Verileri			
	Olumlu (500 Adet)	Nötr (500 Adet)	Olumsuz (500 Adet)
Başarı	399	45	429
Hata	101	455	71
Başarı Oranı	%79,8	%9	%85,8
Diğer Mecra Verileri			
	Olumlu (1500 Adet)	Nötr (1500 Adet)	Olumsuz (1500 Adet)
Başarı	1168	123	1281
Hata	332	1377	219
Başarı Oranı	%77,8	%8,2	%85,4

SONUÇ

Günümüzde kullanıcılar ürün ve hizmetlerle ilgili olumlu veya olumsuz duygularını, direkt olarak satış için sergilendikleri mecra veya diğer mecralarda yazdıkları yorumlar vasıtası ile herkese açık bir biçimde paylaşmaktadırlar. Müşteriler tarafından birinci ağızdan aktarılan bu yorumlar, şirketler açısından büyük önem taşımaktadır. Duygu Analizi, bu yorumların olumlu olumsuz gibi duygulara ayrılması gerekliliği ile ortaya çıkmıştır.

Duygu Analizi bir doğal dil işleme görevidir. Güçlü bilgisayarların devreye girmesi ile istatistiksel yaklaşımlarla dilin belirli düzeylerde modellenebileceği düşünülmüş ve çalışmalar bu yöne kaydırılarak makine öğrenmesi modelleri ile başarılı sayılabilecek Duygu Analizi uygulamaları gerçekleştirilmiştir.

Duygu Analizi farklı disiplinlerdeki araştırmalara konu olmuştur. Bunların büyük bir çoğunluğu bilgisayar mühendisliği alanına aittir. Diğer alanlar işletme, matematik, ekonomi, endüstri mühendisliği, bilim ve teknoloji, ekonometri, jeodezi ve fotogrametri, mekatronik mühendisliği, iletişim bilimleri olarak sıralanabilir. Bu çalışmadaki yönelim ise, iletişim çatısı altında bilişim yaklaşımıdır.

İletişimde bağlam farklı biçimlerde ele alınmaktadır. Burada bağlam, onun fiziksel boyutu olan ortam kavramsallaştırması ile değerlendirilmiş ve mecra tabanlı bir araştırma gerçekleştirilmiştir. Bu doğrultuda, farklı bağlamları temsil ettiği düşünülen üç mecra seçilmiş ve bu mecralarda yer alan kullanıcı yorumları veri olarak kullanılmıştır.

İnternet ortamında yorum yazılan her mecranın kendine özgü birtakım özellikleri bulunmaktadır. Duygu ifade eden kelimelerin farklı mecralarda farklı anlamlarda kullanılışları, kelimelerin mecradan mecraya farklılık göstermesi, bir yorumda kullanılan kelimelerin bir diğer mecrada hiç kullanılmaması bu özelliklere örnek olarak verilebilir. Dolayısıyla mecralara ait Duygu Analizi modellerinin de farklı olması gerekir. Ancak bu pratik bir yöntem değildir. Eğer tüm mecralarda Duygu Analizi yapabilmeye elverişli bir model bulanabilirse, tüm analizler bu model ile kolaylıkla yapılabilir. Bu problemten hareketle çalışmada, çok sayıda mecra için uygulanabilecek bir Karma Veri Uygulaması geliştirilmiştir. Bu yolla öncelikle her bir mecradan toplanan verilerle ayrı ayrı makine öğrenmesi modelleri oluşturulmuş ve her bir model kendi verileri ve diğer mecra verileri ile test edilmiştir. Bu testlerde görülmüştür ki, modeller olumlu/olumsuz yorumlar için kendi verileri ile eğitildiklerinde, diğer mecra verileri ile eğitilen modellere göre daha başarılı sonuçlar üretmektedir. Bu durum bağlamın etkisini ortaya koyar. Bağlamdan bağımsız daha genel bir model oluşturulabilmek için ise, her bir mecradan rasgele belli bir sayıda alınan verilerle karma bir veri kümesi oluşturulmuş ve iki test yinelenmiştir. *Bu modelin iki test sonucundaki başarı oranları birbirine benzerdir.* Dolayısıyla Karma Veri Uygulaması çok sayıda mecranın Duygu Analizinde kullanılabilir.

Modellerin başarı oranlarında daha yüksek doğruluk değerlerine ulaşabilmek için daha çok veri gerekir. Ancak verinin işlenmeye hazırlanmasında birtakım zorluklar bulunmaktadır. Dil bilgisi ve yazım kurallarına uygunluk bunlardan birisidir. Oysa bu kurallara uyma sıradan bir kullanıcıdan beklenemez. Zira bu mecralar informel ortamlardır. Bu nedenle modellerin eğitimi için kullanılan verinin daha önce yazılmış normalleştirme algoritmalarından biri ile veya yeni bir normalleştirme algoritması kullanılarak eğitilmesi gerekir. Ayrıca incelenen mecra ne kadar fazla olursa, o kadar daha genel modeller oluşturulabilir.

İletişim alan yazınında yeni medya alanına yönelik çalışmalar gittikçe hız kazanmış durumdadır. Bu çalışmada önerilen model, toplu duygu analizlerine ilişkin bir ölçme yöntemidir. Bu ve benzeri yöntemler, yeni medya alanında yeni yöntemleri tartışmaya açması bakımından önemlidir.

KAYNAKLAR

- Akgül, E. S., Ertano, C., & Diri, B. (2016). Twitter verileri ile duygu analizi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22(2), s. 106-110. doi:10.5505/pajes.2015.37268
- Atan, S., & Çınar, Y. (2019). Borsa İstanbul'da Finansal Haberler İle Piyasa Değeri İlişkisinin Metin Madenciliği Ve Duygu (Sentiment) Analizi İle İncelenmesi. Ankara Üniversitesi SBF Dergisi, 74(1), s. 1-34.
- Chen, X., Chen, B., Zhang, C., & Hao, T. (2017). Discovering the Recent Research in Natural Language Processing Field Based on a Statistical Approach. International Symposium on Emerging Technologies for Education (s. 507-517). Cape Town: Springer.
- Çetin, F., & Eryiğit, G. (2018, 1). Türkçe Hedef Tabanlı Duygu Analizi İçin Alt Görevlerin İncelenmesi – Hedef Terim, Hedef Kategori Ve Duygu Sınıfı Belirleme. Bilişim Teknolojileri Dergisi, 11(1), s. 43-56.
- Çoban, Ö., Özyer, B., & Özyer Tümüklü, G. (2015). Sentiment analysis for Turkish Twitter feeds. 2015 23rd Signal Processing and Communications Applications Conference (SIU), (s. 1-5). Malatya.
- Demir, Ö., Baban Chawai, A., & Doğan, B. (2019). Türkçe Metinlerde Sözlük Tabanlı Yaklaşımla Duygu Analizi. International Periodical of Recent Technologies in Applied Engineering(2), s. 58-66. doi: 10.35333/porta.2019.98
- Eliaçık, A. B., & Erdogan, N. (2015). Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi. Proceedings of the 9th Turkish National Software Engineering Symposium, (s. 781-793). İzmir.
- Fiske, J. (2003). İletişim Çalışmalarına Giriş. (S. İrvan, Çev.) Ankara: Bilim ve Sanat.
- Gözükar, F., & Özel, S. (2016, 12). Türkçe ve İngilizce Yorumların Duygu Analizinde Doküman Vektörü Hesaplama Yöntemleri için Bir Deneysel İnceleme. Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 31(2), s. 467-481.
- Hall, E. T. (1989). Beyond Culture. Arizona: Anchor Books.
- Hofstede, G. (2001). Culture's Consequences:. London: Sage Publication.
- Karaöz Akın, B., & Gürsoy Şimşek, U. (2018, 7). Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi. Bilişim Teknolojileri Dergisi, 11(3), s. 245-253.
- Kaynar, O., & Yıldız, M. (2016). Makine Öğrenmesi Yöntemleri ile Duygu Analizi . International Artificial Intelligence and Data Processing Symposium (IDAP'16) , (s. 234-241). Malatya .
- Kumar, A., & Geetanjali, G. (2020). Systematic literature review on context-based sentiment analysis in social multimedia. Multimedia Tools and Applications(79), s. 15349–15380.
- Küçük, M. (2012). İletişim Kavramı ve İletişim Süreci. N. Orhon, & U. Eriş (Dü) içinde, İletişim Bilgisi (s. 2-19). Eskişehir: Anadolu Üniversitesi.
- Mcquail, D., & Windahl, S. (2010). İletişim Modelleri. (K. Yumlu, Çev.) Ankara: İmge Kitapevi.
- Nasukawa, T., & Yi, J. (2003). Sentiment Analysis Capturing favorability using Natural Language Processing. Proceedings of the 2nd International Conference on Knowledge Capture , (s. 1-10). Sanibel Island.
- Nizam, H., & Akın, S. (2014). Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması. XIX. Türkiye'de İnternet Konferansı, (s. 1-6). İzmir.
- Onan, A. (2017). Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi. Yönetim Bilişim Sistemleri Dergisi, 3(2), s. 1-14.
- Onan, A., & Bayar, C. (2017). Türkçe Twitter Mesajlarında Gizli Dirichlet Tahsisine Dayalı Duygu Analizi. Akademik Bilişim Konferansı. Aksaray.

Sak, H., Güngör, T., & Saraçlar, M. (2008). Turkish Language Resources: Morphological Parser, Morphological Disambiguator and Web Corpus. 6th International Conference Advances in Natural Language Processing , (s. 417-427). Gothenburg.

Sütcü, C., & Aytekin, Ç. (2019). An Example of Pragmatic Analysis in Natural Language Processing Sentimental Analysis of Movie Reviews. Communication and Technology Congress – CTC 2019, (s. 61-74). İstanbul.

ELEKTRONİK KAYNAKLAR

URL-1 <https://bezoarsoftware.com/approaches-in-natural-language-processing/> (Erişim Tarihi: 10.09.2020)

URL-2 <https://sozluk.gov.tr/> (Erişim Tarihi: 16.09.2020)

URL-3 <https://cogsci.boun.edu.tr/content/frequency-effects-processing-morphologically-complex-turkish-words> (Erişim Tarihi: 21.09.2020)