

6 Şubat 2023 Depremi Üzerine X (Twitter) Verilerinin Konu İncelemesi: LDA ve BERTopic Yaklaşımlarının Karşılaştırması

Topic Analysis of X (Twitter) Data on the February 6, 2023 Earthquake: A Comparison of LDA and BERTopic Approaches

Dilara CANDAN KARADAŞ
Atatürk Üniversitesi
Bilgisayar Mühendisliği
Erzurum – Türkiye
dilarac@atauni.edu.tr
ORCID: 0009-0005-6465-9642

Tolga AYDIN
Atatürk Üniversitesi
Bilgisayar Mühendisliği
Erzurum – Türkiye
atolga@atauni.edu.tr
ORCID: 0000-0002-8971-3255

Gülşah TÜMÜKLÜ ÖZYER
Atatürk Üniversitesi
Bilgisayar Mühendisliği
Erzurum – Türkiye
gulsah.ozyer@atauni.edu.tr
ORCID: 0000-0002-0596-0065

Öz

6 Şubat 2023'te Türkiye ve Suriye'yi etkileyen, insanların hayatını kaybetmesine ve yıkıma yol açan depremler, bölgesel ve küresel çapta insani yardım ve kriz yönetimi süreçlerini başlatmıştır. Çalışmada, bu depremle ilgili X (Twitter) platformundan toplanan 29.185 veri incelenmiştir. Zaman içerisindeki konuların değişimini ve en çok konuşulan başlıkları belirlemek için Latent Dirichlet Allocation (LDA) ve BERTopic yöntemleri kullanılmıştır. LDA, afet bölgesindeki yardım çağrıları, bağışlar ve enkaz kurtarma gibi geniş kapsamlı konuları öne çıkarırken; BERTopic, konu detaylandırma ve ayrıştırmada daha başarılı sonuçlar vermiştir. Analizler, konuşulan konuların özellikle ilk iki günde yoğun olduğunu, sonraki günlerde ise dikkatlerin farklı noktalara kaydığını göstermiştir. Çalışmanın sonuçları, sosyal medya verilerinin afet yönetiminde kritik bir bilgi kaynağı olduğunu vurgulamakta ve yayınlara hem yöntemler arasındaki fark hem de kriz analitiği açısından anlamlı bir katkı sunmaktadır.

Anahtar Sözcükler: 6 Şubat 2023 depremi, LDA, BERTopic

Abstract

The earthquakes that affected Türkiye and Syria on February 6, 2023, resulted in the loss of lives and widespread destruction, triggering regional and global humanitarian aid and crisis management processes. In this study, 29,185 data points collected from the X (Twitter) platform related to this earthquake were analyzed. Latent Dirichlet Allocation (LDA) and BERTopic methods were employed to identify the evolution of topics over time and the most frequently

discussed issues. While LDA highlighted broad themes such as calls for aid, donations, and search and rescue efforts in the affected regions, BERTopic yielded more successful results in detailing and differentiating topics. The analyses revealed that discussions were particularly intense during the first two days, with attention shifting to different issues in the subsequent days. The findings of this study emphasize that social media data serve as a critical source of information in disaster management, offering a meaningful contribution to the literature both in terms of methodological differences and crisis analytics.

Keywords: February 6 2023 Earthquake, LDA, BERTopic

1. Giriş

Depremler, tarih boyunca milyonlarca insanın yaşamını etkileyerek toplumsal, ekonomik ve çevresel açıdan ciddi krizlere neden olmuştur. 6 Şubat 2023'te Türkiye ve Suriye'yi etkileyen Kahramanmaraş merkezli depremler, 50 binden fazla insanın hayatını kaybetmesine, milyonlarca kişinin yerinden edilmesine ve geniş çaplı bir insani krize yol açmıştır. Bu tür afetlerde, sosyal medya platformları bilgi paylaşımı, yardım çağrıları ve toplumun duygusal durumunun ifade edilmesinde önemli bir rol oynamaktadır. Kaynaklarda, sosyal medya analizlerinin afet yönetimindeki potansiyelini ele alan birçok çalışma bulunmaktadır. Örneğin yapılan bir çalışma, 6 Şubat 2023 Türkiye depremlerinden sonraki 15 günlük dönemde halkın sosyal medya kullanımını incelemiştir. Türkiye'de yaşayan 1698 kişiden anketle toplanan veriler, sosyal medyanın bilgi paylaşımı için güçlü bir kaynak olarak kullanıldığını göstermiştir. Afetler bağlamında sosyal medyanın önemi son yıllarda giderek artmakta ve kaçınılmaz

bir rol üstlendiği görülmektedir [1]. Özellikle kriz durumlarında, sosyal medyanın haber alma kaynağı olarak giderek daha fazla tercih edildiği gözlemlenmektedir [2]. Ayrıca sosyal medyanın, kriz anlarında müdahale süreçlerine katkısı ve toplulukların iyileşme sürecindeki etkisi, araştırmacıların dikkatini çeken önemli bir konu haline gelmiştir [3]. Depremlerin zararlarını ve yardımları içeren paylaşımlar bazı katılımcılarda olumsuz duygular yaratırken, sosyal medya fenomenleri ve STK'ların (Sivil Toplum Kuruluşlarının) paylaşımları sosyal medyanın faydalı olduğunu düşündürmüştür. Çalışma, kriz yönetiminde sosyal medya kullanımının etkinliğini artırmak için kullanıcı geri bildirimlerinin dikkate alınması gerektiğini vurgulamaktadır [4].

Yaşanan bir krizin ardından olayın ciddiyetine ve büyüklüğüne bağlı olarak toplumun bu olayı nasıl değerlendirdiği sosyal medya verilerinin analizi ile incelenebilmektedir. Bir çalışmada HAARP (Yüksek Frekanslı Etkin Aurora Araştırma Programı) komplo teorisini analiz ederek, bu teorisinin söyleminin zamanla nasıl değiştiğini ve dikkat artışını tetikleyen faktörleri araştırmaktadır. 2022 Ocak ile 2023 Mart arasında X API'si (Uygulama Programlama Arayüzü) ile HAARP hakkında bir milyondan fazla tweet toplanmış ve 6 Şubat 2023'teki Türkiye-Suriye depremi öncesi, sırası ve sonrasındaki tweetler üzerinde duygu analizi yapılmıştır. Çalışma, büyük depremlerin ardından HAARP tartışmalarının arttığını ve olumlu tweet duyguları ile tweet sayısının pozitif ilişkili olduğunu bulmuştur. Bu durum, sosyal medyanın krizlerde komplo teorilerine olan inancı pekiştirdiğini göstermektedir [5].

Sadece deprem özelinde kalmayıp farklı kriz durumları üzerinde de sosyal medya verileri üzerinden analizler yapılarak bazı çıkarımlar elde edilmiştir. Bu duruma örnek olarak yapılan bir çalışmada Sina Weibo verilerini kullanarak Pekin-Tianjin-Hebei bölgesindeki "23-7" aşırı yağış olayını analiz etmiş ve BERTopic, BERT, NER (Named Entity Recognition), duygu analizi ve afet zarar sözlükleri gibi yöntemleri entegre etmiştir. Çalışmada, Weibo'daki "yardım çağrıları" ve "afet zararları" gibi konular aracılığıyla potansiyel afet bölgelerinin belirlenmesi ve zararların değerlendirilmesi incelenmiştir. Duygu analizi ve zarar sözlükleri kullanılarak afetlerin psikolojik ve fiziksel etkileri nicel olarak değerlendirilmiştir. Çalışma, aşırı yağış koşullarına duyarlılığı göstererek belediye düzeyinde zararları analiz etmiş ve afet zararlarının hızlı bir şekilde değerlendirilmesi için etkili bir yöntem sunmuştur. Geleneksel afet verilerinin yerini tamamen almasa da Sina Weibo verilerinin afet olaylarını hızlı tanımlama ve raporlama açısından benzersiz bir değer taşıdığı vurgulanmıştır [6].

Sosyal medya, yalnızca doğal afetler değil, toplumsal olarak önemli görülen her konuda bilgi paylaşımının yanı sıra, halkın duygusal tepkilerini anlamak ve toplumda konuşulan konuların değişimini analiz etmek için de etkili bir araçtır. Örneğin yapılan bir çalışmada Birleşik Krallık'taki ana akım medyadan elde edilen sosyal medya verileri kullanılarak, plastikler hakkındaki kamuoyu algıları analiz edilmiştir. LDA tabanlı gözetimsiz makine öğrenimi modeli ve bir sözlük yöntemi, yorumların duygusunu değerlendirmek için

uygulanmıştır. Çalışma, plastiklerle ilgili 6 birinci düzey ve 13 ikinci düzey konu kategorisi belirlemiştir. Sonuçlar, halkın plastik ürünlere yönelik tutumunun genel olarak istikrarlı olduğunu ve duyguların zamansal ve mekânsal dağılımının, konuların sıklığıyla yüksek oranda ilişkili olduğunu göstermektedir [7].

Yapılan çalışmalar, kullanılan yöntemlerin sadece tek tür kriz odaklı ve tek bir amaca hizmet eden çalışmalar olmadığını; aynı zamanda çok dilli veri kümelerinde de başarılı olduğunu göstermektedir. Bu çalışmalar, kriz dönemlerinde çok dilli konu dinamiklerini ve iletişim trendlerini çözümlmek için yenilikçi bir yöntem sunmaktadır. Tunus sosyal ağlarındaki pandemi, spor ve siyasetle ilgili diyaloglar analiz edilmiş, çok dilli bir veri kümesi toplanarak işlenmiştir. Dil farklılıklarını ele almak için No-English-to-English Makine Çevirisi yöntemi kullanılmıştır. LDA ve HDP (Hierarchical Dirichlet Process) modelleriyle önemli konular çıkarılmış, ARIMA (AutoRegressive Integrated Moving Average) zaman serisi analiziyle konu trendleri incelenmiştir. Bu yöntem, kamuoyu duygularını yansıtan anahtar konuları başarıyla belirlemiş ve uyum puanı gibi metriklerle standart yaklaşımlardan daha iyi performans göstermiştir. Ayrıca, tartışmalardaki konusal değişimler ve trendler doğru bir şekilde tespit edilmiştir [8].

Bu çalışma, konu belirlemeye yönelik araştırmalara bir katkı olarak, X verileri üzerinden felaket etkilerini hızlı bir şekilde tespit etmeye yönelik bir çerçeve sunmaktadır. Anahtar sözcük ve konum filtreleme gibi yöntemlerle X verileri işlenmiş ve BERT tekniği kullanılarak felaket bölgelerinden gelen tweetler kategorize edilmiştir. Analiz, toplam 376 konu belirlemiş, bunların 196'sının felaket etkileriyle doğrudan ilişkili olduğunu göstermiştir. Bu etkiler arasında elektrik kesintileri, yol kapanmaları, su basan yollar ve altyapı hasarları gibi unsurlar yer almaktadır. Çalışma, acil durum yönetimi için değerli bilgiler sağlayacak bir yöntem sunmaktadır [9].

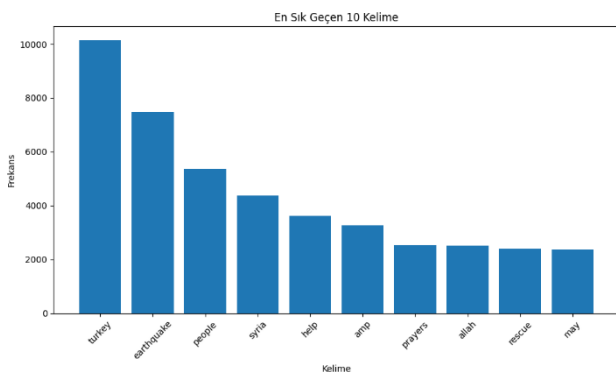
Sonuç olarak, yayınlarda sosyal medya verilerinin afet yönetimi, bilgi paylaşımı ve toplumun duygusal tepkilerini anlamada önemli bir rol oynadığı açıkça görülmektedir. Hem geleneksel yöntemler olan LDA gibi modellerin hem de BERTopic gibi modern tekniklerin farklı kriz durumlarında etkili sonuçlar verdiği kanıtlanmıştır. Bu bağlamda, 6 Şubat 2023 depremleri üzerine yapılan analizler, sosyal medya platformlarının kriz yönetiminde nasıl kullanılabileceğini anlamak için kritik bir zemin oluşturmaktadır. Bu çalışmanın amacı, 6 Şubat 2023 tarihinde meydana gelen depremin ardından sosyal medyada tartışılan konuların, sonraki beş günlük süreçteki değişimini analiz etmektir. Analizlerde, konu modelleme yöntemlerinden BERTopic ve LDA kullanılarak, kriz dönemlerinde toplumsal ilgi ve odaklanmanın konusal olarak nasıl değiştiği incelenmiştir. Çalışmanın bir diğer önemli hedefi, bu iki yöntemin benzer amaçlarla kullanılmalarına rağmen, farklı algoritmik yaklaşımlar ve modelleme süreçlerinden kaynaklanan sonuç farklılıklarını ortaya koymaktır. Görselleştirmeler aracılığıyla, her iki yöntemin ürettiği konu dağılımlarındaki farklılıklar karşılaştırılmış ve bu farkların kriz konulu veri analizi bağlamındaki etkileri tartışılmıştır. Bu kapsamda, çalışmanın hem kriz dönemlerinde iletişim dinamiklerini anlamaya hem

Bu çalışmanın 2. Bölümü'nde kullanılan veriye ve ön işleme ait bilgiye, 3. bölümde kullanılan yöntemle, 4. ve 5. Bölümlerde sırasıyla yapılan deneylere ve performans sonuçlarına yer verilmiştir.

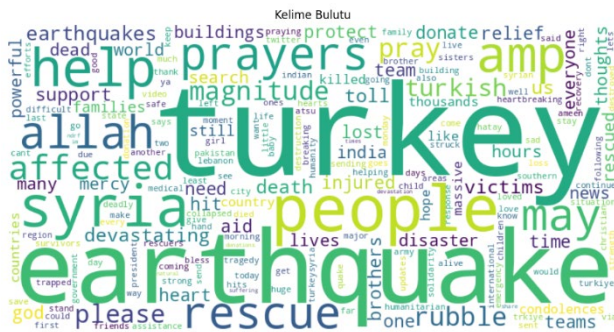
2. Veri Kümesi ve Ön İşleme

2.1. Veri Kümesi

Bu çalışmada, Türkiye’de 6 Şubat 2023 tarihinde meydana gelen depremle ilgili sosyal medya verilerini içeren hazır bir veri kümesi kullanılmıştır [10]. Veri kümesi, deprem sonrasındaki beş gün boyunca toplanan ve her biri bir güne ait verilerden oluşan toplam beş CSV (Comma-Separated Values – Virgülle Ayrılmış Değerler) dosyasını içermektedir. Çalışmada bu beş dosya birleştirilerek toplamda 29.185 İngilizce metinden oluşan bir veri kümesi elde edilmiştir. Veri kümesinde, metin uzunlukları incelenmiş ve en uzun metnin 56 sözcükten oluştuğu, ortalama sözcük sayısının ise 19,47 olduğu belirlenmiştir. Ayrıca, deprem sonrası sosyal medya etkileşimlerini yansıtan bu metinlerde en sık geçen sözcükler analiz edilmiştir. Bu analizde, "turkey" (10.145 kez), "earthquake" (7.491 kez) ve "people" (5.360 kez) gibi ifadelerin öne çıktığı görülmüştür. Şekil-1’de en sık geçen 10 sözcüğün sıklığı (frekansı) görselleştirilirken, Şekil 2’de ise bu sözcüklere ilişkin bir sözcük bulutu sunulmuştur.



Şekil-1. Veri setinden sık geçen 10 sözcük ve sıklıkları

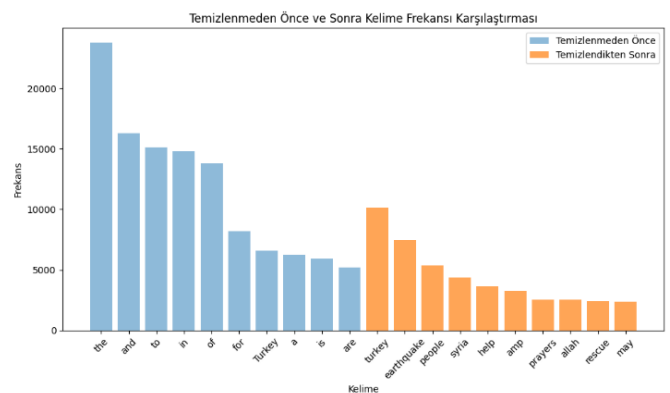


Sekil-2. Veri setine ait sözcük bulutu

2.2. Ön işleme

Çalışmada kullanılan veri kümesi üzerinde, metinlerin modelleme ve analiz için uygun hale getirilmesi amacıyla bir dizi ön işleme adımı gerçekleştirilmiştir. İlk olarak,

metinlerdeki URL'ler ve web adresleri temizlenmiş; ardından, metinlerde bulunan *mentionlar* (@kullanıcıadı) ve *hashtag* işaretleri (#) kaldırılmıştır. Daha sonra, yalnızca harflerden oluşmayan karakterler, noktalama işaretleri ve sayılar metinlerden çıkarılmıştır. Bu işlemlerin ardından, metinlerin tamamı küçük harfe çevrilmiş ve doğal dil işleme sürecinde gereksiz olan durdurma sözcükleri (örneğin "and", "the", "is") metinlerden arındırılmıştır. Ayrıca, metinlerde anlam taşımayan tek harfli sözcükler temizlenmiş ve fazladan oluşabilecek boşluklar düzeltilmiştir. Bu ön işlemler sonucunda, metinler hem istatistiksel analiz hem de konu modelleme yöntemleri için daha verimli ve temiz bir formata dönüştürülmüştür. Bu adımlar, veri kümesindeki gürültüyü azaltarak model sonuçlarının doğruluğunu ve güvenilirliğini artırmayı hedeflemiştir. Şekil-3'te, metinler temizlenmeden önce ve sonra en sık geçen sözcüklerin sıklıklarındaki değişim karşılaştırmalı olarak sunulmuştur.



Şekil-3. Ön işlem öncesi ve sonrası sözcüklerin sıklık dağılımı

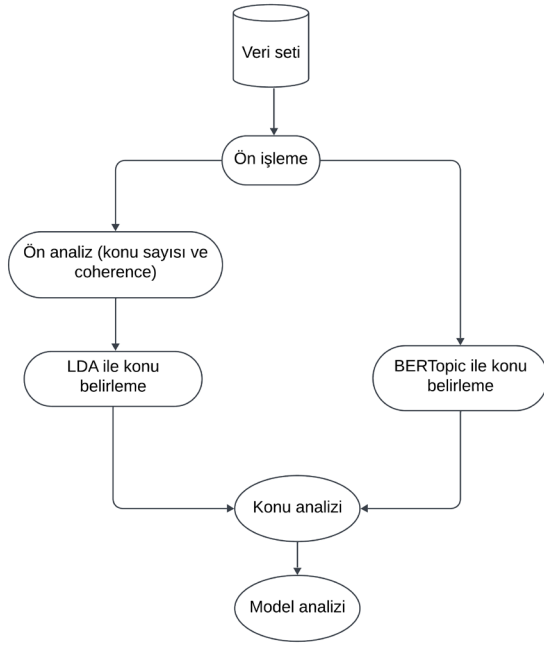
Seçilen ön işleme adımlarının uygulanmasının ardından veri kümesindeki değişimler Şekil-4'te görsel olarak; Çizelge-1'de ise sayısal olarak ifade edilmiştir.

Çizelge-1: Ön İşleme Sürecinin Verilere Etkisi

Özellik	Ön İşlem Öncesi	Ön İşlem Sonrası
Toplam Farklı Sözcük Sayısı	50.382	23.197
En Uzun Metin Uzunluğu	56	35
Ortalama Sözcük Sayısı	19,47	10,77
Toplam Veri Sayısı	29.185	29.185
Toplam Sözcük Sayısı	568.245	314.191

3. Yöntem

Çalışmanın bu kısmında seçilen ve ön işlem süreçleri ile deneysel işlemlere hazırlana veri kümesi ile hangi yöntemlerin izlendiği ve bu yöntemlerin neden uygulandığı ile ilgili ayrıntılı bilgiler verilmektedir. Şekil 4’te, konu modelleme sürecinde kullanılan adımlar ve LDA ile BERTopic yöntemlerinin nasıl karşılaştırıldığı görselleştirilmiştir.



Şekil-4. Çalışmanın akış diyagramı

3.1. Coherence ile En İyi Konu Sayısı Seçimi

Bu aşamada, LDA modeli için en uygun konu sayısını belirlemek amacıyla coherence sonuçları hesaplanmıştır. Coherence sonucu, modelin oluşturduğu konuların tutarlılığını ve anlamlılığını ölçen bir ölçüdür. Bu ölçü, bir modelin gerçek dünyadaki konusal yapıyı ne kadar iyi yansıttığını anlamak için kritik bir rol oynamaktadır. Coherence sonuçlarının yüksek olması, modelin ürettiği konuların daha anlamlı ve ayrıştırılabilir olduğunu gösterir.

Öncelikle, Gensim kütüphanesi ile veri kümesindeki metinler lokmalarına (token) ayrılmıştır ve bir sözlük nesnesi ile derlem oluşturulmuştur. Bu işlem sırasında, analizde gürültü yaratabilecek çok nadir ve çok sık geçen sözcükler filtrelenmiş ve yalnızca anlamlı sözcükler analize dahil edilmiştir. Ardından, farklı konu sayıları (2’den 20’ye kadar) için LDA modelleri eğitilmiş ve her bir modelin coherence sonucu hesaplanmıştır. Optimal konu sayısını belirlemek için, her bir modelin coherence sonucu bir grafik üzerinde görselleştirilmiş ve en yüksek sonuç veren konu sayısı optimal olarak seçilmiştir.

3.2. LDA ile Konu Modelleme

LDA modeli, ön işleme sonrası elde edilen temiz metinler üzerinde çalıştırılmış ve belirlenen optimal konu sayısına göre eğitilmiştir. Gensim kütüphanesi ile eğitilen LDA modeli, her bir metni belirlenen konular arasında dağıtacak şekilde tasarlanmıştır. Model, metinlerdeki sözcük dağılımına dayanarak her bir metin için en olası konuyu belirlemiş ve genel konu dağılımlarını çıkarmıştır. Model çıktıları arasında, her bir konunun temsil ettiği en sık kullanılan sözcükler de bulunmaktadır. Bu sözcükler, her konunun konusal içeriğini anlamak ve sonuçları yorumlamak için kullanılmıştır. Ayrıca, günlük bazda konu dağılımları çıkarılarak, deprem sonrası sosyal medya etkileşimindeki konusal değişimler analiz

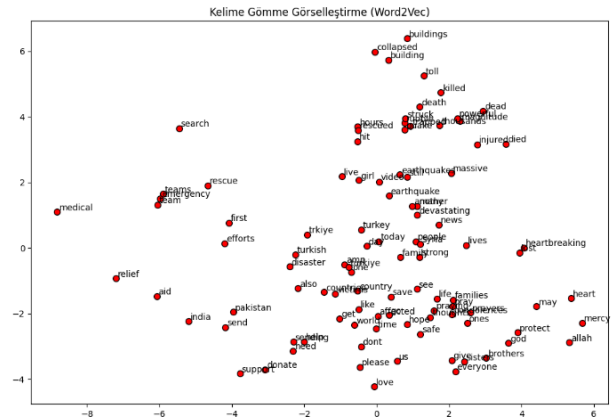
edilmiştir. Sonuçlar, bir çizgi grafiği ile görselleştirilmiş ve her bir günün hangi konulara odaklandığı ortaya konmuştur.

3.3. BERTopic ile Konu Modelleme

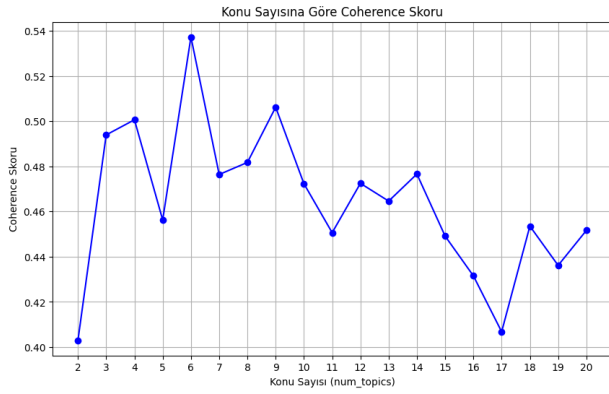
Bu aşamada, konuları belirlemek için BERTopic modeli kullanılmıştır. BERTopic, LDA modelinden farklı olarak, metinlerin anlamını ve bağlamını yakalamak için derin öğrenmeye dayalı sözcük yerleştirme (word embedding) tekniklerini kullanır. Model, metinleri daha derin anlam ilişkilerine göre gruplandırarak konuları belirlemiştir. BERTopic modeli eğitildikten sonra, her bir metin belirlenen konular arasında sınıflandırılmış ve konu başlıkları, model tarafından otomatik olarak atanmıştır. Ayrıca, günlük bazda konu dağılımları çıkarılmış ve görselleştirilmiştir. BERTopic’in sunduğu konu dinamikleri ve anahtar sözcükler bir çizgi grafiği üzerinde gösterilmiş, her bir gün için öne çıkan konular detaylı bir şekilde analiz edilmiştir.

4. Bulgular

Bu çalışmada, LDA modeli için en uygun konu sayısını belirlemek amacıyla farklı konu sayılarında (2’den 20’ye kadar) coherence sonuçları hesaplanmıştır. Yapılan analizler sonucunda, en yüksek coherence sonucuna sahip olan model, konu sayısı 6 olduğunda elde edilmiştir (Coherence Sonucu: 0,5373). Bu sonuç, 6 konunun veri kümesindeki konusal dağılımı en iyi şekilde temsil ettiğini göstermektedir. Word2Vec modeli, sözcükler arasındaki bağlamsal ilişkileri modelleyerek LDA’nın ürettiği konuların tutarlılığını artırmada önemli bir rol oynamıştır. Word2Vec ile elde edilen sözcük yerleştirmeleri, benzer anlamdaki sözcükleri aynı vektör alanında gruplandırarak LDA’nın daha anlamlı ve yorumlanabilir konular üretmesine katkı sağlamıştır. Bu yaklaşım sayesinde, LDA modelinin coherence sonucu artırılmış ve elde edilen konular, kriz sonrası sosyal medya etkileşimlerini daha doğru bir şekilde analiz etmek için uygun bir yapı sunmuştur. Şekil 5’te, Word2Vec modeli ile oluşturulan sözcük yerleştirmelerinin iki boyutlu uzaydaki dağılımı görselleştirilmiş; Şekil 6’da ise farklı konu sayıları için hesaplanan coherence sonuçları sunularak en uygun konu sayısının belirlenme süreci gösterilmiştir.



Şekil-5. Sözcük Yerleştirme Görselleştirmesi- Word2Vec



Şekil-6. Konu Sayısına Göre Coherence Sonucu

Şekil 7’de, 6-10 Şubat 2023 tarihleri arasındaki sosyal medya paylaşımlarında LDA ile belirlenen konuların günlük dağılımını göstermektedir. Görselde, her bir konu, ilgili günlerdeki sıklıklarına göre çizilmiş ve bu sıklıklar üzerinden kriz sürecinde hangi konuların öne çıktığı analiz edilmiştir. Konu 3 ("turkey", "prayers", "allah", "syria", "people", "affected") kriz sonrası yoğun bir şekilde dikkat çeken ve dua, dayanışma gibi konular etrafında şekillenen mesajlarla en yüksek sıklığa ulaşmıştır. Ancak, bu konu, 7 Şubat’tan itibaren belirgin bir şekilde azalmış ve diğer konularla paralel bir düşüş göstermiştir. Benzer şekilde, Konu 4 ("earthquake", "turkey", "magnitude", "syria", "people") depremle doğrudan ilişkili konusal bir yapıya sahiptir ve başlangıçta oldukça yüksek bir sıklık sergilemiştir. Ancak, bu konunun günlük sıklıklarında da gözle görülür bir düşüş yaşanmış, sosyal medyadaki ilginin zamanla azaldığı gözlemlenmiştir. Konu 0 ve Konu 1 gibi konular ("rescue", "rubble", "aid", "disaster") ise yardım çağrıları ve kurtarma çalışmalarıyla ilişkili anahtar sözcükler barındırmaktadır. Bu konuların sıklıklarının, deprem sonrası ikinci ve üçüncü günlerde arttığı, ancak ilerleyen günlerde diğer konularla birlikte azaldığı görülmektedir. Bu durum, yardım çalışmalarının kriz anında daha fazla vurgulandığını ancak zaman geçtikçe sosyal medya etkileşimlerinin genel olarak azaldığını göstermektedir. Genel olarak, görseldeki dağılım, kriz sonrası sosyal medya etkileşimlerinin zamanla azaldığını ve başlangıçta yoğun olan duygusal tepkilerin yerini

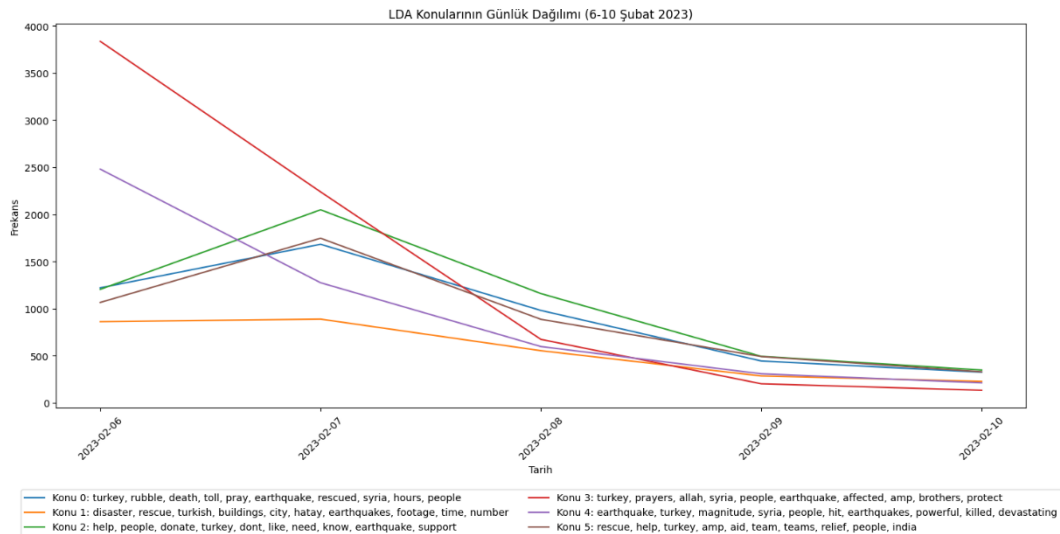
daha rasyonel ve yardım odaklı tartışmalara bıraktığını ortaya koymaktadır. Bu analiz, kriz ile ilgili konuların zamanla nasıl şekillendiğini anlamak için önemli bir içgörü sunmaktadır.

Şekil8’de ise, BERTopic yöntemi kullanılarak belirlenen konuların 6-10 Şubat 2023 tarihleri arasındaki günlük dağılımını göstermektedir. Görselde, her bir konu, ilgili günlerdeki sıklıklarına göre çizilmiş ve kriz sürecinde hangi konuların öne çıktığı analiz edilmiştir.

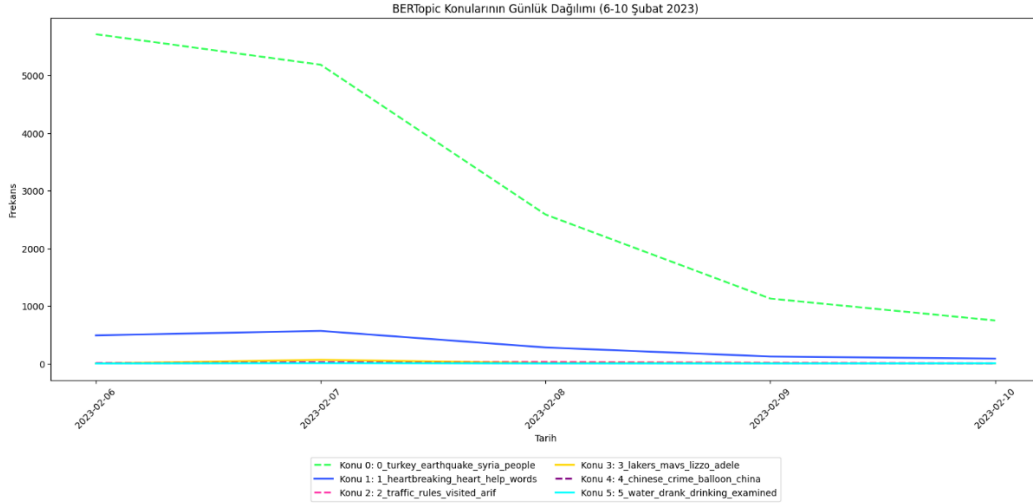
Konu 0 ("turkey", "earthquake", "syria", "people") açık ara en yüksek sıklığa sahip konudur ve deprem sonrası sosyal medyadaki genel konuyu temsil etmektedir. Bu konunun sıklığı, depremle ilgili kriz ortamının en yoğun olduğu ilk günlerde zirve yapmış, ancak sonraki günlerde düzenli bir azalma göstermiştir. Bu durum, depremle ilgili sosyal medya etkileşimlerinin zamanla azaldığını ve kriz sonrası ilginin düşüşe geçtiğini işaret etmektedir.

Konu 1 ("heartbreaking", "heart", "help", "words"), duygusal ve yardım odaklı mesajları içeren bir konudur. İlk günlerde görece düşük bir sıklığa sahip olmasına rağmen, sonraki günlerde sabit bir artış göstermiş ve sosyal medya paylaşımlarında istikrarlı bir şekilde yer bulmuştur. Bu durum, deprem sonrası dayanışma ve yardım çağrılarının sosyal medyada sürekliliğini koruduğunu göstermektedir. Diğer konular ise (örneğin, Konu 3, "lakers", "mavs", "lizzo", "adele") deprem bağlamının dışında kalan, sosyal medyada arada sırada bahsi geçen ve depremle ilişkili olmayan konuları temsil etmektedir. Bu tür konuların sıklığının düşük olması, sosyal medya etkileşimlerinin büyük ölçüde depremle ilgili olduğunu doğrulamaktadır.

Sonuç olarak, BERTopic yöntemiyle elde edilen konuların dağılımı, sosyal medya kullanıcılarının deprem sonrası ilk günlerde genel krizle ilgili içeriklere yoğunlaştığını, ancak bu ilginin zamanla azaldığını ortaya koymaktadır. Ayrıca, duygusal ve yardım odaklı konuların sosyal medyada daha uzun süre gündemde kaldığı gözlemlenmiştir.



Şekil-7. LDA Konularının Günlük Dağılımı



Şekil-8. BERTopic Konularının Günlük Dağılımı

5. Sonuçlar

LDA ve BERTopic yöntemleriyle elde edilen günlük konu dağılımlarının analizine bakıldığında, her iki modelin sonuçları arasında önemli farklılıklar görülmektedir. LDA modeli, sözcük sıklıklarına ve dağılımlarına dayalı istatistiksel bir yaklaşım sunar. Bu nedenle, elde edilen konular daha dengeli ve konusal olarak ayrıştırılmıştır. Örneğin, LDA grafiğinde, belirli günlerde farklı konuların öne çıktığı açıkça görülmektedir. Konu 3 gibi dua ve destek odaklı konular, kriz sonrası ilk günlerde yüksek bir sıklık gösterirken, kurtarma ve yardım çağrılarıyla ilişkilendirilen Konu 0 ve Konu 1 daha uzun bir süre boyunca gündemde kalmıştır. Bu durum, LDA'nın sosyal medya etkileşimlerini farklı konusal yönlerden analiz etmek için daha ayrıntılı bir yapı sunduğunu göstermektedir.

BERTopic modeli ise, sözcükler arasındaki bağlamsal ilişkileri anlamlandırmak için sözcük yerleştirme yöntemlerini kullanır. Bu yaklaşım, belirli bir konuyu diğerlerinden daha baskın hale getirerek daha genel bir analiz sağlar. Nitekim BERTopic grafiğinde, Konu 0 ("turkey", "earthquake", "syria", "people") depremle ilgili genel konuyu temsil eden en baskın konu olarak açık ara öne çıkmıştır. Diğer konuların sıklıkları ise görece düşük ve sabit kalmıştır. BERTopic'in bu bağlamsal odaklı yaklaşımı, sosyal medya paylaşımlarındaki genel eğilimleri ve büyük konuları yakalamada etkili olsa da konusal çeşitliliği ve ayrışmayı ikinci plana atabilir. İki yöntem arasındaki bu farklılıklar, temel olarak kullanılan modelleme yöntemlerinden kaynaklanmaktadır. LDA modeli, sözcüklerin dağılımına dayalı bir istatistiksel yapı sunarken, BERTopic bağlamı ve anlam ilişkilerini vurgulayan bir yöntemdir. Ayrıca, optimal konu sayısı seçimi de bu farklılıkları etkilemiştir. LDA için coherence sonucuna dayalı olarak optimal konu sayısı 6 olarak belirlenmişken, BERTopic'te kullanılan bağlamsal modelleme nedeniyle belirgin bir şekilde baskın bir konu öne çıkmıştır.

Sonuç olarak, LDA, zaman içindeki konusal değişimleri ve detaylı konuları daha ayrıntılı bir şekilde analiz edebilirken, BERTopic, genel konuların öne çıktığı daha bağlamsal bir analiz sunmaktadır. Bu iki yöntem, kriz sonrası sosyal medya analizlerinde birbirini tamamlayıcı nitelikte kullanılabilir ve her biri farklı boyutları anlamada önemli katkılar sağlayabilir.

Bununla birlikte, çalışmanın bazı sınırlılıkları da göz önünde bulundurulmalıdır. Kullanılan veri kümesi yalnızca Türkçe paylaşımları içermekte ve belirli bir zaman aralığını kapsamaktadır; bu da sonuçların genellenebilirliğini sınırlayabilir. Diğer taraftan veriler sadece deprem bölgesindeki insanların paylaşımları ile sınırlı değildir; bu durum özellikle deprem bölgesi için değerlendirme yapmak için uygun değildir. Ayrıca, yalnızca iki konu modelleme yöntemi kullanılmış olup, diğer yöntemlerle karşılaştırma yapılmamıştır. Gelecek çalışmalar, çok dilli ve konuma bağlı veri kümeleriyle benzer analizlerin yürütülmesi, farklı konu modelleme tekniklerinin karşılaştırılması ve zamana bağlı duygu değişimlerinin entegre edilmesi gibi yönlerde genişletilebilir. Bu tür çalışmalar, afet iletişimi ve sosyal medya analizleri konusunda daha kapsamlı içgörüler sunabilir.

Teşekkür

Hakemlere değerli görüşler ve önerilerinden dolayı teşekkür ederiz.

Kaynakça

- [1] Argın, Y. (2023). DOĞAL AFETLERDE SOSYAL MEDYA KULLANIMI: 2023 KAHRAMANMARAŞ DEPREMİ ÖZELİNDE TWITTER ÖRNEĞİ. İnsanlar Sanat Tasarım ve Mimarlık Araştırmaları Dergisi, 3(1), 140-165.
- [2] Alam, F., & Imran, M. (2020). CrisisDPS: Crisis Data Processing Services. In Proceedings of the 17th International Conference on Information Systems for Crisis Response and Management.

- [3] Zhang, Y., & Vos, M. (2019). Social media monitoring: aims, methods, and challenges for international companies. *Corporate Communications: An International Journal*.
- [4] Mızrak, S. (2024). Public's social media use during the Kahramanmaraş earthquakes on 6 February 2023. *International Journal of Disaster Risk Reduction*, 108, 104541.
- [5] Erokhin, D., & Komendantova, N. (2024). Earthquake conspiracy discussion on Twitter. *Humanities and Social Sciences Communications*, 11(1), 1-10.
- [6] Wang, C., Zhang, X., & Wu, J. (2024). Disaster information mining from a social perception perspective: A case study of the “23· 7” extreme rainfall event in the Beijing–Tianjin–Hebei region. *International Journal of Disaster Risk Reduction*, 115, 105056.
- [7] Xue, Y., Kambhampati, C., Cheng, Y., Mishra, N., Wulandhari, N., & Deutz, P. (2024). A LDA-Based Social Media Data Mining Framework for Plastic Circular Economy. *International Journal of Computational Intelligence Systems*, 17(1), 8.
- [8] Jaballi, S., Hazar, M. J., Zrigui, S., Mahjoubi, A., Nicolas, H., & Zrigui, M. (2024). Decoding Multilingual Topic Dynamics and Trend Identification through ARIMA Time Series Analysis on Social Networks: A Novel Data Translation Framework Enhanced by LDA/HDP Models. *Journal of Electrical and Computer Engineering*, 2024(1), 6669491.
- [9] Bhuvaneswari, A., & Kumudha, M. (2024, April). Topic Modeling Based Clustering of Disaster Tweets Using BERTopic. In *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSociCon)* (pp. 1-6). IEEE
- [10] <https://www.kaggle.com/datasets/esatakkasoglu/turkey-06-02-2023-eartquake-tweets/data?select=englishV2>