

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

YAPAY ZEKA DESTEKLİ MEDYA ANALİZİ:
TÜRKİYE'DE YAYINLANAN AFRİKA HABERLERİNİN
KONU MODELLEME TEKNİĞİ İLE İNCELENMESİ

Ali Onur TEPECİKLİOĞLU

Danışman
Dr. Öğr. Üyesi Kutan KORUYAN

İZMİR – 2025

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

YAPAY ZEKA DESTEKLİ MEDYA ANALİZİ:
TÜRKİYE'DE YAYINLANAN AFRİKA HABERLERİNİN
KONU MODELLEME TEKNİĞİ İLE İNCELENMESİ

Ali Onur TEPECİKLİOĞLU

Danışman
Dr. Öğr. Üyesi Kutan KORUYAN

İZMİR – 2025

YÜKSEK LİSANS
TEZ ONAY SAYFASI

Üniversite : Dokuz Eylül Üniversitesi
Enstitü : Sosyal Bilimler Enstitüsü
Adı ve Soyadı : ALİ ONUR TEPECİKLİOĞLU
Öğrenci No : 2022800668
Tez Başlığı : Yapay Zeka Destekli Medya Analizi : Türkiye' de Yayınlanan Afrika Haberlerinin Konu Modelleme Tekniği ile İncelenmesi

Savunma Tarihi : 17/07/2025
Danışmanı : Dr. Öğr. Üyesi Kutan KORUYAN

JÜRİ ÜYELERİ

<u>Ünvanı, Adı, Soyadı</u>	<u>Üniversitesi</u>	<u>İmza</u>
Dr. Öğr. Üyesi Kutan KORUYAN	-Dokuz Eylül Üniversitesi	
Prof.Dr.Vahap TECİM	-Dokuz Eylül Üniversitesi	
Prof.Dr. Mehmet Gökay ÖZERİM	-Yaşar Üniversitesi	

ALİ ONUR TEPECİKLİOĞLU tarafından hazırlanmış ve sunulmuş olan bu tez savunmada başarılı bulunarak oy birliği() / oy çokluğu() ile kabul edilmiştir.

Prof. Dr. Erhan DEMİRELİ
Müdür

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Yapay Zeka Destekli Medya Analizi: Türkiye’de Yayınlanan Afrika Haberlerinin Konu Modelleme Tekniğı İle İncelenmesi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

17/07/2025

Ali Onur Tepeciklioğlu



ÖZET

Yüksek Lisans Tezi

Yapay Zeka Destekli Medya Analizi: Türkiye’de Yayınlanan Afrika Haberlerinin

Konu Modelleme Tekniği ile İncelenmesi

Ali Onur TEPECİKLİOĞLU

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Yönetim Bilişim Sistemleri Anabilim Dalı

Yönetim Bilişim Sistemleri Programı

Bu çalışma, yönetim bilişim sistemleri ve uluslararası ilişkiler disiplinleri arasında bir diyalog kurarak, interdisipliner bir metodoloji ile Türkiye’de yayınlanan Afrika haberlerini incelemeyi amaçlamaktadır. Afrika’nın Türk dış politikasındaki artan önemine dair sınırlı sayıda nicel yöntemle dayalı çalışmanın bulunması, bu araştırmayı hem güncel hem de özgün kılmaktadır. Söz konusu haberlerin incelenmesinde yaklaşık 20.000 haber metnini içeren bir veri seti oluşturulmuş doğal dil işleme alanında geliştirilen BERTopic algoritması vasıtası ile Afrika ve Afrika ülkelerinin Türk medyasından nasıl çerçevelendiğini tespit etmek amacı ile konu modelleme tekniği uygulanmıştır. Elde edilen sonuçlar, Afrika haberlerinde baskın temaların güvenlik, darbeler, doğal afetler, salgın hastalıklar, ticaret, enerji ve kültürel ilişkiler etrafında şekillendiğini göstermiştir. Çalışmanın bulguları Türk medyasında Afrika, çatışma ve şiddet ile anılan ancak ticari fırsatları da içeren uzak ve sıcak bir coğrafya olarak çerçevelendiğine işaret etmektedir. Çalışma kapsamında tespit edilen temalarda zaman, ülke ve medya bazlı farklılıklar da incelenmiş ve Türkiye-Afrika ilişkilerinin seyrinin bu alt kırılımlara yansıdığı ortaya konmuştur. Bu analiz, Türk medyasının Afrika’ya dair söylemini nesnel biçimde ortaya koyarak, dış politika ve kamuoyu algısı bağlamında çıkarımlar sunmaktadır.

Anahtar Kelimeler: BERTopic, Doğal Dil İşleme, Çerçeve Analizi, Konu Modelleme, Türkiye – Afrika İlişkileri

ABSTRACT

Master's Thesis

AI-Supported Media Analysis: Examining News on Africa Published in Turkey

Using Topic Modelling Technique

ALİ ONUR TEPECİKLİOĞLU

Dokuz Eylül University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

This study aims to examine Africa-related news published in Türkiye through an interdisciplinary methodology that establishes a dialogue between the fields of Management Information Systems and International Relations. The limited number of quantitative methodology based studies addressing the growing significance of Africa in Turkish foreign policy renders this research both timely and original. A dataset comprising approximately 20.000 news articles was compiled, and topic modeling was employed using the BERTopic algorithm, a method developed in the field of Natural Language Processing, to identify how Africa and African countries are framed in Turkish media. The findings indicate that dominant themes in Africa-related news revolve around security, military coups, natural disasters, epidemics, trade, energy, and cultural relations. The results suggest that Africa is predominantly framed in Turkish media as a distant and hot geography often associated with conflict and violence, but also encompassing commercial opportunities. The study also investigates temporal, country-based, and media-specific variations in framing and demonstrates that the trajectory of Türkiye-Africa relations is reflected in these subdimensions. This analysis provides an objective account of the discursive construction of Africa in Turkish media and offers insights into foreign policy and public perception.

Keywords: BERTopic, Natural Language Processing, Framing Analysis, Topic Modeling, Türkiye–Africa Relations

YAPAY ZEKA DESTEKLİ MEDYA ANALİZİ: TÜRKİYE’DE YAYINLANAN AFRİKA HABERLERİNİN KONU MODELLEME TEKNİĞİ İLE İNCELENMESİ

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR	ix
TABLolar LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
EKLER LİSTESİ	xiii
GİRİŞ	1

BİRİNCİ BÖLÜM

MAKİNE ÖĞRENİMİ ve DOĞAL DİL İŞLEMENİN BİR ALT DALI OLARAK KONU MODELLEME

1.1. MAKİNE ÖĞRENİMİ NEDİR?	7
1.2. DERİN ÖĞRENME	8
1.3. DOĞAL DİL İŞLEME	9
1.4. KONU MODELLEME	11
1.4.1. Vektörleştirme	11
1.4.1.1. Vektörleştirme Teknikleri	12
1.4.1.1.1. Kelime Torbası (Bag of Words)	12
1.4.1.1.2. TF-IDF	13
1.4.2. Konu Modelleme İş Akışı	15
1.4.3. Konu Modelleme Teknikleri	17
1.4.3.1. Olasılıksal Modelleme Teknikleri	18
1.4.3.1.1. Latent Dirichlet Allocation (LDA)	18
1.4.3.1.2. Hierarchical Dirichlet Process (HDP)	19
1.4.3.1.3. Dynamic Topic Model (DTM)	19
1.4.2.2. Cebirsel Modeller	19

1.4.2.2.1 Latent Semantic Analysis (LSA)	20
1.4.2.2.2 Non-negative Matrix Factorization (NMF)	20
1.4.2.3. Melez Modelleme Teknikleri	21
1.4.2.3.1 Gömülü Temsillere (Embeddings) Dayalı Konu Modelleme Teknikleri	21
1.4.2.3.2. Sinir Ağlarına (Neural Networks) Dayalı Konu Modelleme Teknikleri	22
1.4.2.3.3. Kümeleme (Clustering) Tabanlı Melez Teknikler	23

İKİNCİ BÖLÜM

ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE YAPAY ZEKA

2.1. ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE YAPAY ZEKA KULLANIM ALANLARI	24
2.2. ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE KONU MODELLEME TEKNİKLERİNİN KULLANIMI	26
2.3. KONU MODELLEME TEKNİKLERİ İLE ÇERÇEVE ANALİZİ	29

ÜÇÜNCÜ BÖLÜM

KULLANILAN ARAÇLAR

3.1. PANDAS	36
3.2. MATPLOTLIB	38
3.3. MEDIACLOUD	40
3.4. BEAUTIFUL SOUP	44
3.5. BERTOPIC	45
3.5.1. Sentence-BERT: Gömme/Vektörleştirme	47
3.5.2. UMAP: Boyut İndirgeme	48
3.5.3. HDBSCAN: Kümeleme	48
3.5.4. CountVectorizer: Terim Frekanslarının Çıkarılması (BoW)	49
3.5.5. C-TF-IDF: Ağırlıklandırma	49

DÖRDÜNCÜ BÖLÜM
BERTOPIC İLE TÜRKİYE'DE YAYINLANAN AFRİKA HABERLERİNİN
İNCELENMESİ

4.1. İŞ AKIŞI	51
4.2. VERİ ELDESİ	52
4.2.1. URL Eldesi	52
4.2.2. URL'lerde Veri Ön Temizliği	52
4.2.3. Haber Metinlerinin Web Kazıma Yöntemi ile Eldesi	58
4.3. VERİ TEMİZLİĞİ	59
4.3.1. Metin Ön İşleme	59
4.3.2. AA Kaynaklı Haberlerin Temizlenmesi	60
4.3.3. Mükerrer Haber Metinlerinin Temizliği	62
4.3.4 Çok Kısa ve Çok Uzun Haberlerin Temizlenmesi	64
4.4. MODELİN KURULMASI VE ÇALIŞTIRILMASI	68
4.4.1. Haberlerin Kategorilere Ayrılması	68
4.4.2. Cümle Gömme Algoritmasının Çalıştırılması	72
4.4.3. Derleme Özel Duraklama Kelimeleri Listesi Oluşturulması	73
4.4.4. Model Parametrelerinin Seçilmesi	74
4.4.5. Modelin Çalıştırılması ve Sonuçlar	77
SONUÇ	87
KAYNAKÇA	110
EKLER	

KISALTMALAR

AA	Anadolu Ajansı
AB	Avrupa Birliği
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
BM	Birleşmiş Milletler
BoW	Bag of Words (Kelime Torbası)
c-TF-IDF	Class-based Term Frequency–Inverse Document Frequency
DDİ	Doğal Dil İşleme
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HDP	Hierarchical Dirichlet Process
KDE	Kernel Density Estimation
LSA	Latent Semantic Analysis
ML	Makine Öğrenmesi (Machine Learning)
NMF	Non-negative Matrix Factorization
regex	Regular Expressions
SBERT	Sentence BERT
STM	Structured Topic Modelling
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TF-IDF	Term Frequency-Inverse Document Frequency (Terim Frekansı-Ters Belge Frekansı)
Üİ	Uluslararası İlişkiler (Disiplin)
UMAP	Uniform Manifold Approximation and Projection
URL	Uniform Resource Locator
VAE	Variational Autoencoders
YZ	Yapay Zeka

TABLÖLER LİSTESİ

Tablo 1: DDİ Uygulamaları	10
Tablo 2: Kelime Doküman Matrisi Örneği	13
Tablo 3: TF-IDF Özellik Matrisi	14
Tablo 4: Media Cloud Turkey National Kategorisi Haber Sayıları	41
Tablo 5: Bert Mimarisi	46
Tablo 6: BERTopic Mimarisi	47
Tablo 7: BERTopic Konu Modelleme Süreci	50
Tablo 8: Eş Sesli Kelimeler İçeren Haber Başlıkları	53
Tablo 9: Haber Meta Veri Veri Seti Genel Görünümü	55
Tablo 10: Medya Kuruluşuna Göre Haber Meta Verisi Dağılımı	56
Tablo 11: Ülkelere Göre Haber Meta Verisi Dağılımı	57
Tablo 12: Veri Kelime Sayısı Temel İstatistikleri	65
Tablo 13: Veri Temizlik Aşamaları ve Kalan Satır Sayıları	66
Tablo 14: Medya Kuruluşuna Göre Haber Metni Verisi Dağılımı	66
Tablo 15: Ülke Başına Haber Metni Verisi Dağılımı	67
Tablo 16: Kategoriler ve Haber Türleri	69
Tablo 17: BERT Haber Kategorizasyonu Değerlendirme Metrikleri	70
Tablo 18: Veri Setinin Haber Kategorilerine Göre Dağılımı	72
Tablo 19: Kategori Bazında Model Parametreleri	76
Tablo 20: Siyaset Kategorisi Konu Modelleme Sonuçları	80
Tablo 21: Ekonomi Kategorisi Konu Modelleme Sonuçları	82
Tablo 22: Sağlık Kategorisi Konu Modelleme Sonuçları	84
Tablo 23: Kültür Kategorisi Konu Modelleme Sonuçları	85
Tablo 24: Siyaset Kategorisi İlk 5 Konudan Örnekler	88
Tablo 25: Ekonomi Kategorisi İlk 5 Konudan Örnekler	94
Tablo 26: Sağlık Kategorisi İlk 5 Konudan Örnekler	99
Tablo 27: Kültür Kategorisi İlk 5 Konudan Örnekler	104

ŞEKİLLER LİSTESİ

Şekil 1: BoW Gösterimi	12
Şekil 2: TF-IDF Gösterimi	14
Şekil 3: Konu Modelleme İş Akışı	15
Şekil 4: Pandas Kullanımı	37
Şekil 5: Pandas Kodu Çıktısı	37
Şekil 6: Matplotlib Kullanımı	39
Şekil 7: Matplotlib Kodu Çıktısı	39
Şekil 8: Media Cloud API Çağrısı	42
Şekil 9: Media Cloud API Çağrısı Kod Çıktısı	43
Şekil 10: Beautiful Soup Kütüphanesi Kullanımı	45
Şekil 11: İş Akış Şeması	51
Şekil 12: Medya Kuruluşuna Göre Haber Sayıları Frekans Dağılımı	54
Şekil 13: AA Web Sitesi Kazıma Fonksiyonu	58
Şekil 14: AA Benzerlik Skorları Frekans Dağılımı	61
Şekil 15: AA Kaynaklı Haberlerin Medya Kuruluşlarına Göre Oranı	61
Şekil 16: Kosinüs Benzerliği Yoğunluk Dağılımı - Tüm Veri Seti	63
Şekil 17: Metin Uzunlukları Yoğunluk Dağılımı	65
Şekil 18: BERT Sınıflandırma Modeli Karışıklık Matrisi ve Eşik Değer/Kapsama Oranı İlişkisi	71
Şekil 19: Cümle Gömmelerinin Gerçekleştirilmesi	73
Şekil 20: Modelin Çalıştırılması	77
Şekil 21: Siyaset Kategorisi Konu-Benzerlik Matrisi	81
Şekil 22: Ekonomi Kategorisi Konu-Benzerlik Matrisi	83
Şekil 23: Sağlık Kategorisi Konu-Benzerlik Matrisi	84
Şekil 24: Kültür Kategorisi Konu-Benzerlik Matrisi	86
Şekil 25: Siyaset Kategorisi Ülke/Konu Dağılımı Isı Haritası	90
Şekil 26: Siyaset Kategorisi Zaman/Konu Dağılımı Isı Haritası	91
Şekil 27: Siyaset Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası	93
Şekil 28: Ekonomi Kategorisi Ülke/Konu Dağılımı Isı Haritası	96
Şekil 29: Ekonomi Kategorisi Zaman/Konu Dağılımı Isı Haritası	97
Şekil 30: Ekonomi Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası	98
Şekil 31: Sağlık Kategorisi Ülke/Konu Dağılımı Isı Haritası	100
Şekil 32: Sağlık Kategorisi Zaman/Konu Dağılımı Isı Haritası	101

Şekil 33: Sağlık Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası	103
Şekil 34: Kültür Kategorisi Ülke/Konu Dağılımı Isı Haritası	105
Şekil 35: Kültür Kategorisi Zaman/Konu Dağılımı Isı Haritası	106
Şekil 36: Kültür Medya Kuruluşu/Konu Dağılımı Isı Haritası	107



EKLER LİSTESİ

EK 1: Afrika Ülke Adları ve Kısaltmaları

EK 2: Kullanılan Kütüphaneler ve Versiyonları



GİRİŞ

Yapay Zeka (YZ) alanının bir alt dalı olan doğal dil işleme (DDİ) çok sayıda dokümanın hızlıca ve isabetli bir biçimde incelenerek, sınıflandırılmasını, özetlenmesini, duygu analizi yapılmasını, konularının belirlenmesini, metinde geçen ad ve varlıkların tanımlanmasını, ilişki çıkarımı yapılmasını, metinden yeni metin üretilmesini ve daha bir çok işlemin yapılmasını sağlamaktadır (Arslan ve diğerleri., 2023: 2836). DDİ'nin temelde kullandığı veri metindir (*text as data*). Basitçe DDİ, metin verisini sayısallaştırarak çeşitli yapay zeka algoritmalarının bu metinler üzerinde kullanımı vasıtası ile yukarıda sayılan işlemleri gerçekleştirmektedir. DDİ'nin belki de en büyük avantajı, insan emeği ile analizi mümkün olmayan ya da çok uzun zaman alacak sayıda dokümanın incelenmesine olanak vermesidir. Bir diğer avantajı ise yapılan analizin insana ait olan önyargılardan (*human bias*) bağımsız olmasına olanak sağlamasıdır.

Yönetim bilişim sistemleri alanında gün geçtikçe önemini arttıran DDİ hem işletmeleri hem de karar destek sistemlerini dönüştürmektedir (Stoykova ve Shakev, 2023: 2). Çünkü DDİ günümüzde karar destek sistemlerinin hem objektif hem de subjektif birçok faktörden etkilenecek gittikçe karmaşıklaşan problemleri çözmesinde önemli kolaylıklar sağlamaktadır (Kacprzyk ve Zadrozny, 2010: 380). Buna ilaveten karar destek sistemlerine veri akışı her geçen gün artmaktadır ve verilerin niteliği çeşitlenmektedir. DDİ özellikle metni veri olarak işleyebilme ve karar alıcılara anlamlı bilgi sunabilme kapasitesini dikkate değer biçimde artırması bağlamında büyük önem kazanmaktadır (Métais, 2002: 249).

Dış politika yapım sürecinde en önemli unsurlardan biri, alınan kararlardır. Nihayetinde, bireylerin veya insan gruplarının verdikleri kararlar, uluslararası politikada meydana gelen gelişmeleri belirlemektedir. Bu bağlamda, karar destek sistemlerinin uluslararası ilişkilerdeki rolü artmaktadır. Uluslararası İlişkiler (Uİ) disiplini metnin veri kaynağı olarak en yoğun bulunduğu alanlardan biridir. Basın açıklamaları, uluslararası anlaşmaların metinleri, haberler ve akademik makale ile kitapların hepsi metin şeklinde veri üretmektedir. Dolayısıyla dil ve dil kullanımı uluslararası politikanın temel yapı taşlarındandır. Dili bir sosyal yapı olarak ele aldığımızda, kullanılan dil, tema, kavramlar ve kelimeler, siyasetteki materyal gerçekliklere asıl anlamını kazandırmaktadır. Wendt'in (1995: 73) yaklaşımından hareketle, materyal gerçekliğin ancak sosyal süreçler ile inşa edilen sosyal gerçeklik sayesinde anlam kazandığı söylenebilir. Örneğin, nükleer silahlar bir materyal

gerçeklik olarak ele alındığında, bu silahların tek başlarına bir anlam taşımadığı ifade edilebilir. Ancak belirli bir bağlama oturtulduklarında anlam kazandıkları görülmektedir. Başka bir ifade ile, Kuzey Kore'nin nükleer silahlarının Amerika Birleşik Devletleri için yarattığı tehdit ile İngiltere'nin elinde bulunan nükleer silahların yarattığı tehdit aynı düzeyde olmamaktadır. Bunun nedeni, Amerika Birleşik Devletleri ile İngiltere arasındaki paylaşılan değerler, kurumlar ve bilgi birikiminin, A.B.D'nin İngiltere'nin nükleer silahlarını bir korunma aracı ve müttefik gücünün göstergesi olarak anlamlandırması, buna karşın Kuzey Kore'nin nükleer silahlarını bir tehdit, uluslararası güvenlik sorunu ve saldırgan bir araç olarak tanımlamasıdır.

Bu süreçlerin tamamı ve devletler arası iletişim olarak adlandırılan diplomasi, doğal dilin bir parçası olarak değerlendirilmektedir. Dolayısıyla, Uİ dilin ve onun bir türevi olan metinlerin veri kaynağı olarak kullanımına yabancı değildir. Ancak, geleneksel olarak metin analizi, duygu durum analizi, bilgi çıkarımı ve söylem analizi gibi yöntemler Uİ'de genellikle nitel yöntemler ile gerçekleştirilmektedir (Ünver, 2019: 161). Nicel yöntemlerin bu alanda kullanımı ise henüz yeni olmakla birlikte, uluslararası politikada her an büyük miktarda metin verisi üretilmekte olup, bu veriler karar destek sistemlerinde kullanılmaya başlanmaktadır.

Uİ'nin metin odaklı bir disiplin olması, DDİ tekniklerinin bu alanda uygulanabilirliğini ve potansiyel faydalarını artırmaktadır.. Bu metinlerden bilgi çıkarımı, duygu analizi ve söylem analizi yapılması, Uİ araştırmalarına yeni bir boyut kazandırmaktadır. DDİ, nitel analiz yöntemlerinin yanı sıra nicel yöntemlerin de Uİ çalışmalarına entegre edilmesine olanak tanımakta ve büyük ölçekli veri analiziyle daha kapsamlı sonuçlar elde edilmesini sağlamaktadır.

Bu bağlamda, DDİ'nin en önemli katkılarından biri, karar destek sistemlerinin geliştirilmesidir. Alınan kararların sonuçları, uluslararası politikayı şekillendirdiğinden ve çoğu zaman doğrudan insan hayatını etkilediğinden, doğru ve hızlı bilgiye ulaşmak kritik öneme sahiptir. DDİ, hem hızlı hem de önyargısız analiz yapabilme kapasitesiyle bu süreci desteklemektedir. Örneğin, DDİ teknikleri kullanılarak bir ülkenin dış politika söylemi analiz edilebilir ve bu söylemlerin diğer devletlerle ilişkiler üzerindeki etkisi daha nesnel bir şekilde değerlendirilebilir. Ayrıca, DDİ, çok katmanlı ve boyutlu yapıdaki uluslararası politikada karmaşık konuları anlamaya yönelik avantajlar da sunmaktadır. Doğası gereği uluslararası politika sosyal, ekonomik, politik, askeri, kültürel vb. boyutların iç içe geçmesinden oluşur. Dahası, bireysel, yerel, sistemik katmanda olan biten hemen herşey uluslararası siyasette sonuçlar doğurmaktadır. Dolayısıyla uluslararası politikada gerçekleşen herhangi bir olayı anlamak bu

karmaşık ve iç içe geçmiş yapıları detaylı bir biçimde analiz etmeyi gerektirir. Bu açıdan bakıldığında DDİ, sunduğu araç seti ile Uİ çalışmalarında yenilikçi bir boyut açmaya güçlü bir adaydır.

DDİ'nin Uİ çalışmalarında kullanılmasının yaratacağı avantajlardan hareket ile bu çalışmanın temel araştırma sorusu şu şekilde belirlenmiştir: "Türkiye'de yayınlanan Afrika haberleri en fazla hangi ana konuları ve temaları içermektedir?" Bu soru "Türkiye'de yayınlanan Afrika haberlerinin temaları ve ele alınış biçimleri zaman içerisinde değişmekte midir?", "Afrika haberlerinin ele alınış biçimlerinde ön plana çıkan tema ve ana konular medya organlarında nasıl bir dağılım göstermektedir?", "Afrika haberlerinin ele alınış biçimlerinde bölgesel ya da ülkesel farklılıklar bulunmakta mıdır?" gibi alt araştırma soruları ile desteklenecektir.

Afrika, uluslararası politikanın giderek daha fazla dikkat çeken bir bölgesi haline gelmiştir. Bu durum, Afrika'nın doğal kaynakları, genç nüfusu ve stratejik konumu gibi unsurlardan kaynaklanmaktadır. Afrika'nın dünya siyasetindeki yeri ve bu kıtaya yönelik uluslararası politikalar, küresel güç dengelerinin anlaşılması için önemli bir bağlam sunmaktadır. Türkiye de Afrika'nın uluslararası politikada artan önemine paralel olarak bu kıtaya yönelik 1990'ların sonunda bir açılım stratejisi planlamış ve 2000'lerin başından itibaren kıtada aktif bir oyuncu olmak için çeşitli politikalar geliştirmeye başlamıştır (Eyrice Tepeciklioğlu ve Tepeciklioğlu, 2021: 1-5).

Türkiye'nin Afrika stratejisinin gelişmesi ile birlikte özellikle Uİ alanında Türkiye-Afrika ilişkileri çalışmalarında da bir artış görülmektedir (Eyrice Tepeciklioğlu, 2016). Ancak Türkiye Afrika ilişkilerini veriye dayalı olarak analiz eden çalışmalar sınırlı sayıdadır (Eyrice Tepeciklioğlu ve Tepeciklioğlu, 2024; Tepeciklioğlu ve diğerleri, 2024). Yorumlamanın ya da bilgi aktarmanın ötesine geçerek daha yüksek analiz kabiliyeti sunan veriye dayalı çalışmaların sayıca azlığı bu çalışmanın da çıkış noktalarından birini oluşturmaktadır.

Bu alana katkı yapma amacı ile seçilen konu Türkiye'de yayınlanan Afrika ile ilgili haberlerin analizidir. Medyanın yönlendirmede ve kamuoyu oluşturmadaki büyük etkisinden hareketle Afrika'nın Türk medyasında nasıl ele alındığını anlamak, Türkiye'nin dış politika hedeflerini ve Türk toplumunun Afrika'ya ve Afrika ülkelerine yönelik algısını değerlendirmek açısından önem taşımaktadır. Özellikle, Afrika'daki ülkelerin politik, ekonomik, sosyal ve kültürel boyutlarının nasıl çerçevelendiği, hem Türkiye-Afrika ilişkilerin uzun vadede nasıl şekilleneceğine dair ipuçları sunabilir hem de Türk toplumunun Afrika kıtasına ve Afrika ülkelerine bakış açısını değerlendirmeye olanak sağlayabilir.

Afrika kıtası Birleşmiş Milletlerde temsil edilen 54 farklı ülkeden oluşmaktadır (bkz. Ek 1). Bu ülkelerin her biri, tıpkı Avrupa ya da Asya ülkeleri gibi farklı sosyal, ekonomik, politik ve kültürel özelliklere sahiptir. Bu çalışma hem bu farklılıkların Türkiye’de nasıl anlaşıldığını hem de genel olarak Afrika kıtasının nasıl algılandığını analiz etmeyi hedeflemektedir. Bu bağlamda haberlerde yer alan temaların analiz edilmesi, Türkiye’nin Afrika ülkelerini ayrı ayrı ve Afrika kıtasını bir bütün olarak nasıl tanımladığına ve bu coğrafyayı uluslararası politikada nereye oturttuğuna dair derinlemesine bir perspektif kazandıracaktır.

Çalışmanın temel araştırma sorusu, Türk medyasında yayınlanan Afrika haberlerinde kullanılan ele alınış biçimlerini, yani ana konu ve temaları “çerçeve” (*frame*), bu çerçevelerin kullanılması yolu ile haberlerin aktarılmasını “çerçeveleme” (*framing*) olarak tanımlamaktadır. Çerçeveleme kavramı, haberlerin ve metinlerin analiz edilmesinde önemli bir teorik çerçeve sunmaktadır. Entman’ın (1993: 52) tanımına göre çerçeveleme, “bir gerçeklik algısının belirli yönlerini seçip bu yönleri iletişimde daha belirgin hale getirerek, belirli bir sorun tanımını, nedensel yorumu, ahlaki değerlendirmeyi ve/veya bir çözüm önerisini teşvik etme sürecidir.” Bu tanım, çerçevelemenin bir metni sadece tarif etmekle kalmadığını, aynı zamanda o metne bir anlam ve yön kazandırdığını vurgulamaktadır.

Çerçeveleme, okuyucuların belirli bir konuyu nasıl algıladığı üzerinde doğrudan etkili olmaktadır. Örneğin, bir doğal afet haberi, mağduriyet çerçevesiyle sunulduğunda, okuyucuların olaya insani yardım bağlamında yaklaşması teşvik edilebilir. Aynı haber, hükümetin başarısızlığı çerçevesiyle sunulduğunda ise, okuyucuların hükümete karşı eleştirel bir tutum geliştirmesi sağlanabilir. Dolayısıyla, çerçeveleme, hem bireylerin hem de toplumların algılarını ve tutumlarını şekillendiren güçlü bir araçtır.

Çerçeveleme kavramını daha somut bir şekilde açıklamak için iki farklı örnek verilebilir. İlk olarak gündelik hayatta da karşılaşılabilecek, herhangi bir şehirde yaşanan bir trafik kazası haberini ele alalım. Birinci çerçeve olarak, kazanın ardından olay yerine müdahale eden ambulansın gecikmesi şu şekilde sunulabilir: “Hatalı park edilmiş araçlar nedeniyle ambulans olay yerine 15 dakika içinde ulaşabildi.” Bu ifade, hatalı park eden araçları sorunun kaynağı olarak vurgular ve kamuoyunun dikkatini bireylerin park etme alışkanlıklarının değiştirilmesine yöneltebilir. Aynı haber, farklı bir çerçeve ile şöyle aktarılabilir: “Trafikteki yoğunluk nedeniyle ambulans olay yerine 15 dakika sonra ulaşabildi.” Bu ifade, gecikmenin kaynağı olarak genel trafik sorununu öne çıkarır ve kamusal altyapı veya trafik düzenlemelerine odaklanılmasını teşvik

eder. Her iki çerçeve de aynı olayı anlatır, ancak seçilen kelimeler ve vurgu yapılan unsurlar, okuyucunun olaya ilişkin algısını, çıkarımlarını ve çözüm önerilerini önemli ölçüde değiştirebilir. Çerçeveleme bilgileri eğip bükme ve çarpıtma değil, belli kısımlarını ön plana çıkarma bağlamında önem kazanır.

İkinci bir örnek olarak, uluslararası politika ile ilgili yayınlanmış bir haber, çerçeve kullanımı ile bir gerçeğin belirli yönlerini seçip bu yönü iletişimde vurgulamanın önemini oldukça iyi yansıtmaktadır. CNN'in uluslararası edisyonunun 11 Ekim 2024 tarihli, Gazze'de yaşanmakta olan açlık riski hakkında yaptığı haberin (Lockwood, 2024) başlığı İngilizce'den Türkçe'ye çevirisi ile "BM, Ekim ayının başından bu yana Kuzey Gazze'ye yiyecek girmedeğini ve bunun 1 milyon insanı açlık riskiyle karşı karşıya bıraktığını söyledi." şeklindedir (*UN says no food has entered northern Gaza since start of October, putting 1 million people at risk of starvation*). Bu başlıktan görülebileceği üzere, haberin edilgen bir dil kullanılarak ve açlık riskinin nedeninin sadece bu bölgeye yiyecek girmedeği ile açıklanarak verilmesi, gerçeğin vurgulanan yönünü değiştirmektedir. Bu bölgeye yiyecek girmemesinin açık nedeni İsrail ordusunun ablukası iken bu bilginin başlıkta seçici bir biçimde yer almaması, çerçevelerin gerçeğin belli kısımlarını ön plana çıkararak vurgulaması ve çerçevelemenin iletişimdeki gücü ile metinlerin farklı bağlamlarda nasıl anlam kazandığını göstermesi açısından iyi bir örnek teşkil etmektedir.

Uluslararası politikada olayların medya tarafından nasıl çerçevelendiğini, hangi özelliklerinin ön plana çıkarıldığını derinlemesine anlamak ve analiz etmek için DDI'de konu modelleme (*topic modelling*) tekniği oldukça güçlü bir araç sunmaktadır. Çerçeveleme, bir metindeki temaların ve konuların analiz edilmesiyle anlam kazanır. Konu modelleme, büyük metin yığınlarını analiz ederek gizli temaları ve konuları otomatik olarak ortaya çıkarmayı amaçlayan bir DDI yöntemidir. Bu yöntem, metinlerdeki kelime örüntülerini istatistiksel olarak analiz ederek bir dizi konu belirler ve her belgeyi bu konuların bir veya birkaçına göre sınıflandırır. Böylece, metinlerde öne çıkan temalar, hangi çerçevelerin kullanıldığını daha sistematik bir şekilde ortaya koyabilir. Özellikle metinlerin anlam yapısını keşfetmek, içerik analizi yapmak veya veri kümelerini daha yönetilebilir hale getirmek amacıyla kullanılır. Konu modelleme tekniği, metin veri setlerinden anlamlı bilgiler çıkararak araştırma sorularına yanıt bulmak veya yeni bakış açıları geliştirmek için etkili bir araç sunar.

Bu çalışmanın araştırma sorusu yanıtlanırken, Türkiye'de yayımlanan Afrika ile ilgili haberlerinin çerçevelerini anlamak amacıyla yenilikçi bir konu modelleme tekniği olan BERTopic kullanılacaktır. Bu yöntem için öncelikle Türkiye'de

yayınlanmış Afrika ülkeleri ve Afrika'nın geneline dair haberlerden bir veri seti oluşturulacak, daha sonra bu veri seti, kategorilere ayrılacak, sonrasında ise her bir kategoriye ayrı ayrı konu modelleme uygulanacaktır. BERTopic büyük ölçekli metin verilerinden anlamlı bilgiler çıkararak, UI bağlamında çerçeveleri tespit etme çalışmalarına yeni bir boyut kazandırmaya adaydır. Özellikle, uluslararası politikada metin verisi yoğunluğunun her geçen gün artması, bu tür nicel analiz yöntemlerinin önemini artırmaktadır. Konu modelleme ile elde edilen temalar, Afrika'ya yönelik algının nasıl şekillendiğini ve hangi tema, sorun ve çözüm önerilerinin ön plana çıktığını ortaya koyacaktır.



BİRİNCİ BÖLÜM

MAKİNE ÖĞRENİMİ ve DOĞAL DİL İŞLEMENİN BİR ALT DALI OLARAK KONU MODELLEME

1.1. MAKİNE ÖĞRENİMİ NEDİR?

Makine öğrenimi (ML), bilgisayarların kendisine sunulan verilerden öğrenerek ve deneyimlerden faydalanarak görevlerini otomatik olarak iyileştirme yeteneği olarak tanımlanabilir. Geleneksel programlamadan farklı olarak makine öğrenimi algoritmaları açıkça kodlanmış kurallara dayanmaktan ziyade, verilerden çıkarımlar yapar ve bu çıkarımları modele dönüştürür (El Naqa & Murphy, 2015: 6). Bu yaklaşım, insan zekasını taklit ederek kararlar almayı ve tahminler yapmayı hedefleyen YZ'nin bir alt dalıdır.

Makine öğrenimi temelde dört ana öğrenme türünü içerir: denetimli öğrenme, denetimsiz öğrenme, yarı denetimli öğrenme ve pekiştirmeli öğrenme. Denetimli öğrenme (*supervised learning*), etiketlenmiş veri setlerinden öğrenirken, denetimsiz öğrenme (*unsupervised learning*), verilerdeki gizli yapıları keşfetmek için etiketlenmemiş verileri kullanır. Yarı denetimli öğrenme (*Semi-Supervised Learning*), hem etiketlenmiş hem de etiketlenmemiş verilerle çalışan bir makine öğrenimi paradigmasıdır. Bu yöntem, geleneksel denetimli öğrenme (yalnızca etiketlenmiş veri) ve denetimsiz öğrenme (yalnızca etiketlenmemiş veri) arasında bir köprü kurar. Amacı, az sayıda etiketlenmiş veriyi ve çok sayıda etiketlenmemiş veriyi bir arada kullanarak daha etkili modeller oluşturmaktır. Pekiştirmeli öğrenme (*reinforcement learning*) ise, bir sistemin ödül ve ceza mekanizmalarıyla öğrenmesini sağlar (Ayodele, 2010: 19)

Makine öğrenimi, geçmiş verilerden öğrenip akıllı kararlar alabilme yeteneğiyle dördüncü sanayi devrimi çağında birçok alanda popüler hale gelmiştir. Bu teknoloji, tahmin analitiği ve akıllı karar verme, siber güvenlik, nesnelerin interneti ve akıllı şehirler, trafik tahmini ve ulaşım, sağlık hizmetleri, e-ticaret, DDİ, görüntü ve desen tanıma, sürdürülebilir tarım gibi çeşitli uygulama alanlarında kullanılmaktadır. Örneğin, sağlıkta hastalık teşhisi ve tedavisi, e-ticarette müşteri davranışlarına dayalı ürün önerileri, tarımda mahsul verimi tahmini ve nesnelerin interneti ile enerji tüketimi gibi pek çok çözüm sağlanabilmektedir. Makine öğrenimi, verimliliği artırmak ve daha bilinçli kararlar alınmasını sağlamak için birçok sektörde yenilikçi çözümler sunmaktadır (Sarker, 2021: 15-17).

1.2. DERİN ÖĞRENME

Derin öğrenme (*deep learning*), ML'nin bir alt dalı olarak, çok katmanlı yapay sinir ağları kullanarak karmaşık örüntüleri tanıma, öğrenme ve genelleme yeteneklerini geliştiren bir algoritmalar bütünüdür. Bu yöntem, özellikle hiyerarşik özellik öğrenimi kabiliyeti sayesinde, ham verilerden soyut temsiller oluşturabilmektedir. Makine öğreniminin gelişiminde önemli bir dönüm noktası olan derin öğrenme, biyolojik sinir ağlarının işleyişinden esinlenmiş modelleri daha fazla katmanla derinleştirerek öğrenme kapasitesini artırmayı hedefler. Her bir katmanın, veriden giderek daha soyut özellikler çıkarması sayesinde model çok katmanlı bir kavrayış geliştirir. Bu yapı vasıtasıyla model örüntüleri insan müdahalesine ihtiyaç duymaksızın, verinin kendi yapısına dayanarak öğrenme potansiyeline sahip olmaktadır (Amaratunga, 2021: 5-6).

Bu yapı, yalnızca teknik bir gelişim değil, aynı zamanda insan benzeri öğrenme biçimlerine daha yakın bir yaklaşımı da beraberinde getirmektedir. Derin sinir ağları, insan beynindeki nöronların işleyişine benzer biçimde, örüntüleri doğrudan deneyim yoluyla öğrenir. Yapay sinir ağları üzerindeki bu modelleme anlayışı, veriden öğrenme sürecinde araştırmacı müdahalesini en aza indirerek, modeli kendi deneyimiyle bilgi edinmeye yönlendirir (Saxe ve diğerleri, 2021: 55).

Derin öğrenmenin en dikkat çeken özelliklerinden biri, geleneksel makine öğrenmesi yaklaşımlarında sıkça başvurulman manuel özellik mühendisliğini büyük ölçüde ortadan kaldırmasıdır. Derin öğrenme modelleri, veriyi ham haliyle alıp, katmanlar aracılığıyla kendi temsil biçimlerini otomatik olarak oluşturabilir. Bu durum, doğal dil işleme gibi karmaşık ve bağlama duyarlı alanlarda büyük avantaj sağlar. Özellikle metinlerin anlamını daha iyi yakalamak için geliştirilen dil temsilleri, bu çok katmanlı yapılar sayesinde bağlamsal bilgiyle zenginleşir (Fan ve diğerleri, 2021: 264-266).

Bu yaklaşım, günümüzde dil modelleme gibi alanlarda aktif olarak kullanılmaktadır. Özellikle cümle gömme gibi yöntemlerde, metnin anlamını temsil eden sayısal vektörlerin oluşturulmasında derin öğrenme temelli yöntemler tercih edilmektedir. Bu temsil biçimleri, dilin anlamını katmanlar boyunca öğrenen bir modelin ürünüdür ve gömme temelli doğal dil işleme modellerinin temelini oluşturmaktadır. Dolayısıyla derin öğrenme doğal dil işleme alanında bağlamsal anlamın yakalanmasını mümkün kılan güçlü bir yaklaşım olarak değerlendirilebilir. Veriye dayalı olarak öğrenen bu modeller, çok katmanlı yapıları sayesinde dilin

karmaşıklığını daha iyi temsil etmekte ve bu sayede konu modelleme gibi görevlerde yüksek performans sağlamaktadırlar (Amaratunga, 2021: 6).

1.3. DOĞAL DİL İŞLEME

Makine öğreniminin bir alt dalı olan DDİ, insan dilinin bilgisayarlar tarafından anlaşılıp işlenmesini sağlar. DDİ, dilbilim, bilgisayar bilimi ve istatistik gibi alanların kesiştiği noktada bulunmaktadır ve gerçek insanlar tarafından üretilen metin veya konuşma verilerini anlamlandırma, analiz etme ve bu verilere dayalı eylemler gerçekleştirme amacı gütmektedir. İnsan dilindeki belirsizlik, karmaşıklık ve çeşitlilik göz önüne alındığında, DDİ'nin temel amacı, makinelerin doğal dili insanlar gibi anlamasına olanak tanımadır (Liddy, 2001: 1-3).

DDİ'nin temelleri, 1950'lerde Noam Chomsky'nin (1955) dilbilim çalışmaları ve Alan Turing'in (1951) "makinelere düşünebilir mi?" sorusuyla ortaya attığı Turing Testi'ne kadar geri götürülebilir. 1960'larda daha çok bir dilin diğerine makineler aracılığı ile çevirisine odaklanan bu alandaki çalışmalar, zamanla büyük miktarda veri ve artan hesaplama gücünün etkisiyle istatistiksel ve makine öğrenimi yöntemlerine yönelerek çeşitlenmiştir (Johri ve diğerleri, 2021: 368-370). Özellikle derin öğrenme tekniklerinin 2010'lardan itibaren DDİ'ye uygulanması, metin ve konuşma işleme kapasitesinde önemli sıçramalara neden olmuştur.

DDİ'nin gelişimi, hem dilbilimsel yapılar üzerine yapılan çalışmalara hem de veri bilimindeki yeniliklere paralel olarak ilerlemiştir. İlk dönemlerde gramer kurallarını izleyen dil işleme sistemleri, günümüzde doğal dilin bağlamsal anlamını yakalamak için sinir ağları ve dönüştürücü tabanlı modeller (*transformer*) modelleri gibi tekniklere dayanmaktadır (Ruder ve diğerleri, 2019: 569-570). GPT (*Generative Pre-training Transformer*) ve BERT (*Bidirectional Encoder Representations from Transformers*) gibi modeller, bağlam odaklı dil temsilleri sayesinde dilin daha derin bir şekilde anlaşılmasını sağlamıştır. Bu modeller, DDİ'nin birçok alanda uygulanmasına temel oluşturarak hem akademik hem de endüstriyel çalışmalarda kullanılmaktadır.

DDİ, farklı problemleri çözmek için çeşitli tekniklere sahiptir. Bunlardan bazıları Tablo 1'de verilmiştir:

Tablo 1: DDİ Uygulamaları

Teknik İsmi	Teknik Açıklaması
Ad Varlık Tanıma (<i>Named Entity Recognition, NER</i>):	Metin içindeki özel isimleri (kişi adları, yer isimleri, organizasyonlar vb.) tanıyan bir tekniktir.
Duygu Analizi (<i>Sentiment Analysis</i>):	Metindeki duygusal tonu (olumlu, olumsuz veya nötr) belirleme amacıyla metinlerin işlenip istatistiksel yöntemler ile incelenmesini temel alan bir yaklaşımdır.
Dil Modelleme (<i>Language Modeling</i>):	Bir metindeki kelime sıralarını tahmin ederek dilin istatistiksel yapısını anlamaya yardımcı olan bir tekniktir.
Makine Çevirisi (<i>Machine Translation</i>):	Bir dildeki metni başka bir dile çevirmek için kullanılır.
Konuşma-Metin Dönüştürme (<i>Speech to Text</i>):	Sesli konuşmayı metne dönüştüren bir DDİ alanıdır
Metin-Konuşma Dönüştürme (<i>Text to Speech</i>)	Metni konuşmaya dönüştüren bir DDİ alanıdır.
Doğal Dil Üretimi (<i>Natural Language Generation, NLG</i>)	Yapılandırılmış veriyi anlamlı doğal dil ifadelerine dönüştürür. Büyük dil modelleri (<i>Large Language Models, LLM</i>) olarak adlandırılan ChatGPT ve Gemini gibi modeller bu alanda yer almaktadır.
Konu Modelleme (<i>Topic Modeling</i>)	Metinlerdeki gizli konu yapısını belirlemek ve metinleri bu konulara göre kategorilere ayırmak için kullanılan bir yaklaşımdır.

Kaynak: Yazar tarafından derlenmiştir.

Bu teknikler, farklı DDİ uygulama alanlarına yönelik özelleştirilmiş çözümler geliştirmek için temel yaklaşımlardır ve günümüzde birçok sektör tarafından odaklanılan problemin niteliğine uygun olarak kullanılmaktadır. Bu çalışmanın amacı da büyük ölçekli haber verisinden belirli bir konunun, bu tez özelinde Afrika'nın, nasıl ele alındığını incelemek olduğundan, bu problemin çözümüne en uygun DDİ tekniği konu modelleme olarak belirlenmiştir. Ancak konu modelleme, genel bir teknik adıdır ve bu çalışmanın da kullanacağı BERTopic dahil olmak üzere, bir çok alt tekniği

altında barındırmaktadır. Çalışmanın temel aldığı teknik olması sebebiyle DDI'nin bir alt dalı olarak konu modelleme daha ayrıntılı olarak incelenecektir.

1.4. KONU MODELLEME

Veri madenciliği, düzensiz ve yapılandırılmamış metin verisinden yapılandırılmış ve düzenli veri etme süreci olarak tanımlanabilir (Kaya ve Gülbandır, 2022: 46). Veri madenciliğinin bir türü olan konu modelleme oldukça çeşitli teknikler kullansa da temelde bir kelimenin belirli bir metinde ne sıklıkta ve hangi örüntülerde geçtiğine odaklanmaktadır. Konu modelleme algoritmaları tek bir metinden ziyade, *corpus* adı verilen bir dokümanlar koleksiyonu, Türkçe adı ile derlem üzerinde eğitilmektedirler. Konu modelleme algoritmaları öncelikle girdi olarak ele aldıkları derlemdeki temaları keşfetmeye çalışır, sonrasında ise derlemde bulunan her bir dokümanın keşfedilen bu temaları, bir diğer deyişle konuları ne kadar yansıttığını belirlerler (Gillings ve Hardie, 2023).

1.4.1. Vektörleştirme

Konu modelleme tekniklerinin temel çalışma prensibi farklı dokümanlarda bulunan kelime kullanım örüntülerini tespit etmek ve bu örüntülerin dağılım olasılıklarını hesaplamaktır (Alghamdi ve Alfalqi, 2015:147). DDI süreçlerinde bilgisayarlar, kelimeleri insanlar gibi semantik bir düzen içerisinde algılamazlar. Her ne kadar bilgisayarların insanlar tarafından kullanılan dili anlaması *doğal dil işleme* olarak adlandırılrsa da, bilgisayarlar 1 ve 0 temsili ile çalışmaya devam etmektedir. Bir başka ifade ile dilin bilgisayar tarafından anlaşılması için kelimelerin vektörleştirilmesi gerekmektedir. Vektörleştirme, basitçe, kelimelerin sayısal olarak temsil edilmesi olarak adlandırılabilir. Kelimeler, dokümanlar veya cümleler belirli matematiksel gösterimlerle (vektörler) ifade edilerek bilgisayarların bunları analiz etmesine olanak tanınmaktadır. Bu vektörler sayesinde bilgisayarlar, kelimeler arasındaki benzerlikleri öğrenebilir, kelimeleri anlamlı gruplara ayırabilir ve dilin yapısını kavrayabilir. Vektör temsilleri sayesinde yapay zeka modelleri, metinleri sınıflandırabilir, özetleyebilir, çeviri yapabilir ve insan dilini daha doğal bir şekilde işleyebilir.

1.4.1.1. Vektörleştirme Teknikleri

DDİ’de konu modelleme için kullanılan vektörleştirme teknikleri çeşitli olmakla birlikte en yoğun olarak kullanılan iki temel teknik bulunmaktadır. Bunlar “kelime torbası” (*Bag of Words, BoW*) ve “terim frekansı-ters belge frekansı” (*Term Frequency-Inverse Document Frequency, TF-IDF*) olarak adlandırılmaktadır.

1.4.1.1.1. Kelime Torbası (Bag of Words)

BoW, metindeki kelimeleri dikkate alarak bir kelime sıklığı matrisi oluşturur. BoW bir belgedeki kelimelerin tam sıralamasını göz ardı ederken, her terimin frekansını önemli bir faktör olarak ele alır. Başka bir ifade ile BoW yöntemi, her kelimenin belgede kaç kez geçtiğini sayarak, metni sayısal bir forma dönüştürür. Bu sayede metinler, bilgisayarlar tarafından işlenebilir hale gelir (Joachims, 1998: 138).

BoW algoritmasının çalışma yöntemi şu şekildedir;

- Öncelikle bir kelime sözlüğü (*vocabulary*) oluşturulur (tüm metinlerde geçen benzersiz kelimelerin listesi).
- Her belge, bu sözlükteki kelimelerin kaç kez geçtiğini gösteren bir vektörle temsil edilir.
- Sadece kelime frekansı (*term frequency*) dikkate alınır; kelimenin metindeki konumu veya bağlamı göz önünde bulundurulmaz.

Örneğin, şu iki cümleyi ele alalım: “Kedi mama yiyor ve kedi uyuyor”, “Köpek mama yiyor ve kedi koşuyor”. Bu cümlelerin birer python listesi olarak vocabulary ve BoW gösterimleri aşağıdaki şekil ve tablodaki gibi olacaktır.

Şekil 1: BoW Gösterimi

```
sentence_1 = "Kedi mama yiyor ve kedi uyuyor"
sentence_2 = "Köpek mama yiyor ve kedi koşuyor"
vocabulary = ["kedi", "mama", "iyor", "ve", "uyuyor", "köpek",
               "koşuyor"]
bow_cumle1 = [2, 1, 1, 1, 1, 0, 0]
bow_cumle2 = [1, 1, 1, 1, 0, 1, 1]
```

Kaynak: Yazar tarafından derlenmiştir.

BoW yöntemi, tek bir belge için kelime sıklığı vektörü oluşturarak metni sayısal hale getirir. Ancak, birden fazla belge üzerinde çalışıldığında, her belgenin BoW vektörleri bir araya getirilerek kelime-doküman matrisi (*document-term matrix*, *DTM*) oluşturulur. Bu matris, her satırın bir kelimeyi, her sütunun ise bir belgeyi temsil ettiği büyük bir tablo gibi düşünülebilir. Böylece, belgeler arasındaki kelime dağılımlarını karşılaştırmak ve metin madenciliği uygulamalarında kullanmak mümkün hale gelir.

Tablo 2: Kelime Doküman Matrisi Örneği

Kelime	Cümle 1 (Kedi mama yiyor ve kedi uyuyor.)	Cümle 2 (Köpek mama yiyor ve kedi koşuyor.)
kedi	2	1
mama	1	1
iyor	1	1
ve	1	1
uyuyor	1	0
köpek	0	1
koşuyor	0	1

Kaynak: Yazar tarafından derlenmiştir.

1.4.1.1.2. TF-IDF

TF-IDF, kelimelerin önemini belirlemek için BoW modeline ek olarak bir ağırlıklandırma mekanizması kullanmaktadır. Yani bu yöntemde, ilk önce kelimelerin doküman ve derlem içerisindeki frekansları tespit edilir, daha sonra ise bu frekanslar ağırlıklandırılır. Bu sayede ortaya her bir kelime için bir TF-IDF skoru ortaya çıkar (Ramos, 2003: 2).

TF-IDF algoritmasının çalışma prensibi şu şekildedir;

- **TF (Term Frequency - Terim Frekansı):** Kelimenin belirli bir belgede kaç kez geçtiğini sayılarak, bu sayı o dokümandaki toplam kelime sayısına bölünür.

$$TF = \text{Kelimenin Dokümandaki Frekansı} / \text{Dokümandaki Toplam Kelime Sayısı}$$

- IDF (*Inverse Document Frequency - Ters Döküman Frekansı*): Kelimenin nadirliğini ölçme amacı taşır. Eğer bir kelime çok fazla belgede geçiyorsa, o kelimenin önemi azalır.

$$IDF = \log(\text{Toplam Doküman Sayısı} / \text{Kelimenin Geçtiği Doküman Sayısı})$$

- TF-IDF Skoru: TF ve IDF değerlerinin çarpımıdır.

$$TF-IDF = TF * IDF$$

Örneğin şu üç cümleyi ele alalım; “Kedi mama yiyor ve kedi uyuyor”, “Köpek mama yiyor ve kedi koşuyor”, “Kedi ve köpek birlikte oynuyor”. Bu cümlelerin TF-IDF skorlarının scikit-learn kütüphanesi kullanılarak hesaplanmış hali ve matris gösterimi aşağıdaki şekil ve tablodaki gibi olacaktır.

Şekil 2: TF-IDF Gösterimi

```
from sklearn.feature_extraction.text import TfidfVectorizer

sentence_1 = "Kedi mama yiyor ve kedi uyuyor"
sentence_2 = "Köpek mama yiyor ve kedi koşuyor"
sentence_3 = "Kedi ve köpek birlikte oynuyor"

vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform([sentence_1, sentence_2,
sentence_3])
```

Kaynak: Yazar tarafından derlenmiştir.

Tablo 3: TF-IDF Özellik Matrisi

birlikte	kedi	koşuyor	köpek	mama	oynuyor	uyuyor	ve	iyor
0	0.598068	0	0	0.385061	0	0.506309	0.299034	0.385061
0	0.31877	0.539725	0.410475	0.410475	0	0	0.31877	0.410475
0.55249	0.32631	0	0.420183	0	0.55249	0	0.32631	0

Kaynak: Yazar tarafından derlenmiştir.

Bu yöntem, sık geçen ama anlamsal açıdan pek katkısı olmayan kelimelerin etkisini azaltırken, az geçen ancak daha anlamlı olan kelimelerin önemini artırmayı amaçlamaktadır.

1.4.2. Konu Modelleme İş Akışı

Hemen tüm konu modelleme teknikleri benzer prensipler ile ve iş akışı ile çalışmaktadır. Aşağıda konu modelleme tekniği uygulanırken takip edilmesi gereken iş akışı verilmiştir:

Şekil 3: Konu Modelleme İş Akışı

Veri eldesi → Veri ön işleme → Model seçimi → Modelin çalıştırılması → Sonuçların değerlendirilmesi

Kaynak: Yazar tarafından derlenmiştir.

Konu modellemenin ilk aşaması veri eldesidir. Burada dikkat edilmesi gereken prensip, elde edilecek metinlerinin türlerinin benzer olmasıdır. Bu veriler makaleler, blog yazıları, sosyal medya gönderileri, forum yorumları, haber metinleri, akademik makaleler, romanlar, kitaplar, raporlar ve benzerleri olabilir. Farklı metin türleri kullanılarak konu modelleme yapılabilir ancak bu şekilde yapılan konu modellemelerde isabet oranları düşük olacaktır. Dolayısıyla veri türlerinin benzer özelliklere sahip olmaları oldukça önemlidir. Verilerin tutarlı bir formatta olması, örneğin, tüm metinlerin UTF-8 gibi aynı kodlama standardında olması, verideki gürültü (*noise*), yani anlamsız veya alakasız bilgilerin minimize edilmesi, eksik veya bozuk verilerin düzeltilmesi ya da çıkarılması ve belge uzunluklarının dengeli olması alınacak sonucun isabetli olması üzerinde etkilidir. Veri eldesinde bir başka önemli nokta ise veri sayısıdır. Konu modelleme algoritmaları, genellikle büyük veri kümeleri üzerinde daha iyi sonuçlar verir. Küçük veri kümelerinde model, temaları doğru bir şekilde öğrenemeyebileceğinden az sayıda veri ile çalışmak modelin isabet oranını düşürecektir.

İkinci aşama olan veri temizleme, konu modelleme sürecinin en kritik adımlarından biridir. Bu aşamada, bir önceki aşamada elde edilen ham veri üzerinde yapılan düzenlemeler ve ön işleme adımları, modelin performansını doğrudan etkiler. Veri temizleme sürecinde ilk olarak metinlerdeki konuların keşfedilmesi bağlamında gereksiz ve anlamsız bilgilerin ayıklanması gerekmektedir. Bu ayıklama süreci, özellikle noktalama işaretleri, sayılar, özel karakterler ve URL'ler gibi metnin anlamsal yapısına katkıda bulunmayan unsurların kaldırılmasını içerir. Ayrıca, metinlerde sıkça geçen ancak anlam taşımayan “ve”, “ama”, “ise” gibi duraklama kelimeleri (*stop words*) de çıkarılmalıdır. Bu tür kelimeler, metinlerin tematik yapısını bozabileceğinden, modelin doğruluğunu olumsuz etkileyebilir.

Bir diğ er  nemli adım, kelimelerin k klerine indirgenmesidir. Bu, lematizasyon (*lemmatization*) veya k klendirme (*stemming*) teknikleri kullanılarak ger ekleřtirilir.  rneğ in, “kořuyor”, “kořtu” ve “kořar” gibi farklı  ekimleri olan kelimeler, “koř” k k ne indirgenerek modelin daha tutarlı bir řekilde  alıřması saėlanır. Bu iřlem,  zellikle T rk e gibi sondan eklemeli dillerde b y k  nem tařır. Ayrıca, metinlerdeki b y k/k   k harf farklılıklarının giderilmesi de  nemlidir. T m metinlerin k   k harfe d n řt r lmesi, kelimelerin tutarlı bir řekilde iřlenmesini saėlar.    nk  bilgisayarlar, harfleri belirli bir kod řemasına g re tanımaktadırlar. B y k harf ve k   k harfler bu kod řemasında farklı temsil edildikleri i in bilgisayarlar, “araba” ve “Araba” řeklindeki aynı iki kelimeyi farklı kelimelermiř gibi algırlarlar.

Veri temizleme s recinde dikkat edilmesi gereken    nc  bir nokta, metinlerin uzunluklarının dengeli hale getirilmesidir.    nk  derlem i erisinde yer alan ařır  kısa dok manlar modelin  ğrenme s recine katkı saėlamazken, ařır  uzun dok manlar ise belirli konuların diğ erlerine bask n  ıkmasına neden olabilir. Bu nedenle, metinlerin belirli bir uzunluk aralıėında tutulması, modelin daha verimli  alıřmasını saėlayacaktır.

Veri eldesi ve veri  n iřlemeden sonra gelen    nc  ařama, model se imi ařamasıdır. Bu ařama, konu modelleme s recinin en teknik ve kritik adımlarından biridir. Model se imi yapılırken, kullanılacak algoritmanın veri setinin yapısına ve  alıřmanın amacına uygun olması b y k  nem tařır.  rneğ in, Latent Dirichlet Allocation (LDA), konu modelleme alanında en yaygın kullanılan algoritmalarından biridir ve  zellikle b y k metin k llyatlarında tematik yapıları ortaya  ıkarmak i in etkilidir. Bunun yanı sıra, Non-Negative Matrix Factorization (NMF) gibi alternatif y ntemler de k   k veya orta  l ekli veri setlerinde bařarılı sonu lar verebilir. Son yıllarda, derin  ğrenmeden yararlanan modeller( rneğ in, BERTopic) de konu modelleme alanında pop ler hale gelmiřtir. Bu modeller,  zellikle karmařık ve b y k veri setlerinde daha y ksek doėruluk oranları sunabilir. Dolayısıyla bu g rece geniř algoritma se enekleri arasından veri tipine ve arařtırmanın amacına uygun olan modelin se ilmesi, konu modelleme ile alınacak sonucu doėrudan etkileyecektir.

Model se iminin ardından, model parametrelerinin ayarlanması gerekmektedir. Bu parametreler, modelin performansını doėrudan etkileyen hiperparametrelerdir.  rneğ in, LDA algoritmasında konu sayısı (k), alpha ve beta parametreleri gibi hiperparametrelerin doėru bir řekilde belirlenmesi, modelin daha tutarlı ve anlamlı sonu lar  retmesini saėlar. Konu sayısı, genellikle veri setinin b y kl ğ ne ve i eriğ ine g re belirlenir.  ok az konu se ilmesi, temaların genel

kalmasına neden olurken, çok fazla konu seçilmesi ise temaların aşırı özelleşmesine ve anlamsız hale gelmesine yol açabilir. Bu nedenle, konu sayısının belirlenmesinde coherence score gibi metriklerden yararlanılabilir. Ayrıca, alpha ve beta parametreleri, konuların belgeler ve kelimeler arasındaki dağılımını kontrol eder. Bu parametrelerin optimize edilmesi, modelin daha dengeli ve anlamlı sonuçlar üretmesine yardımcı olur.

Model parametrelerinin ayarlanması sürecinde, çapraz doğrulama (*cross-validation*) ya da farklı parametreler ile deneme yanılma gibi yöntemler kullanılarak modelin genelleme yeteneği test edilmelidir. Bu, modelin farklı veri alt kümeleri üzerinde tutarlı bir performans sergileyip sergilemediğini anlamak açısından önemlidir.

1.4.3. Konu Modelleme Teknikleri

Konu modelleme teknikleri literatürde genellikle olasılıksal (*probabilistic*), cebirsel (*algebraic*) ve melez (*hybrid*) modeller olmak üzere üç ana kategoriye ayrılmaktadır (Churchill ve Singh, 2022). Bu üç kategorideki hemen her model denetimli, denetimsiz, ya da yarı denetimli öğrenme yöntemleri ile eğitilebilirler.

Denetimsiz öğrenme yöntemleri, veri kümesinde önceden belirlenmiş etiketler olmadan konu yapısını keşfetmek için kullanılmaktadır. Bu yaklaşım, özellikle büyük, yapılandırılmamış ve karmaşık veri kümelerindeki gizli örüntüleri, temaları veya kümeleri belirleme açısından avantaj sağlamaktadır. Konu modelleme bağlamında, denetimsiz yöntemler genellikle belgelerdeki kelime dağılımlarını veya sözcükler arasındaki ilişkileri analiz ederek temel konuları ortaya çıkarmayı amaçlayan yöntemlerdir. Bu öğrenme yöntemi belirli anlamların belirli bağlamlarda ve benzer/aynı kelimeler kullanılarak yakalanabileceği varsayımına dayanır. Bu tür algoritmaların temelinde kümeleme, boyut azaltma ve olasılıksal modelleme gibi teknikler bulunmaktadır. Böylece denetimsiz olarak eğitilen modeller, verinin içsel yapısını inceleyerek benzer içerikleri gruplandırabilmekte ve ana temaları otomatik olarak belirleyebilmektedir.

Denetimli öğrenme yöntemleri, modelin önceden etiketlenmiş verilerle eğitildiği ve yeni veriler için tahminler yapabildiği bir yaklaşımdır. Konu modelleme bağlamında denetimli yöntemler, önceden insan tarafından kategorisi/konusu etiketlenmiş metinlerden öğrenerek, yeni belgelerin hangi konuya ait olduğunu belirleme sürecinde kullanılmaktadırlar. Bu yöntem, özellikle verilerin önceden belirli

kategorilere ayrıldığı durumlarda yüksek doğruluk ve yorumlanabilirlik sağlaması nedeniyle tercih edilmektedir. Denetimli konu modelleme algoritmalarında genellikle sınıflandırma ve regresyon gibi teknikler kullanılarak, modelin eğitim sürecinde öğrendiği desenleri yeni verilere uygulaması hedeflenmektedir.

Yarı denetimli öğrenme yöntemleri, hem etiketli hem de etiketlenmemiş verileri birlikte kullanarak modelin öğrenme ve tahmin performansını arttırmayı amaçlamaktadır. Konu modelleme bağlamında, sınırlı sayıda etiketli belgeyle eğitilen bir model, büyük ölçüde etiketlenmemiş verilerden faydalanarak daha geniş kapsamlı tahminler yapabilmektedir. Bu yaklaşım, özellikle etiketleme sürecinin maliyetli veya zaman alıcı olduğu durumlarda avantaj sağlamaktadır. Yarı denetimli öğrenme algoritmaları, denetimli ve denetimsiz yaklaşımların birleşimi olarak düşünülebilir ve genellikle kümeleme ile sınıflandırma tekniklerini bir arada kullanarak, veri içindeki gizli yapıları daha etkili bir şekilde keşfetmeyi hedefler. (R. S. T. Lee, 2024, s. 125).

Her üç öğrenme yönteminin de verinin yapısı, kullanılacak model, zaman ve maliyet bağlamında avantajları ve dezavantajları bulunmaktadır. Dolayısıyla konu modellemede model eğitilirken bu kriterler göz önüne alınarak öğrenme yöntemi seçilmelidir.

1.4.3.1. Olasılıksal Modelleme Teknikleri

Olasılıksal modelleme teknikleri, metinlerde altta yatan temaları keşfetmek için olasılık teorisine dayanmaktadır. Bu teknikler, kelime dağılımları tespit edilerek bu dağılımlar ile temalar arasındaki ilişkileri modelleyerek çalışmaktadırlar. Bir derlem içerisindeki belgelerin ve kelimelerin olasılıksal ilişkilerini temel alan bu modeller denetimli, yarı denetimli ya da denetimsiz olarak eğitilebilirler.

1.4.3.1.1. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA), belgelerin belirli konuların karışımı olduğunu ve her konunun belirli bir kelime dağılımına sahip olduğunu varsayan bir modeldir. Model, belgelerdeki kelimelerin gözlemlenmesinden yola çıkarak, gizli (latent) temaları ve bu temaların belgelerdeki dağılımını tahmin eder. Birçok alanda yaygın olarak kullanılan bu modelde Dirichlet dağılımı, konu ve kelime dağılımlarını modellemek için kullanılır. Bu yöntem, özellikle büyük derlemlerde tematik analiz için tercih edilmektedir (Jelodar ve diğerleri, 2019).

1.4.3.1.2. Hierarchical Dirichlet Process (HDP)

Hierarchical Dirichlet Process (HDP), LDA modelinin genişletilmiş ve daha esnek bir versiyonu olarak nitelendirilebilir. LDA'nın aksine, HDP'de konu sayısı önceden belirlenmeden model çalıştırılmaktadır. Bu özellik, bilhassa konu sayısının önceden bilinmediği veya veri setinin dinamik olarak değişebileceği durumlarda modelin veriye uyum sağlama yeteneğini artıran bir etmen olarak öne çıkmaktadır. HDP, hiyerarşik bir Bayesian modelidir ve Dirichlet Süreçleri (Dirichlet Processes) temeline dayanmaktadır. HDP belgelerin konu dağılımlarını modellemek amacıyla iki seviyeli bir hiyerarşi kullanılmaktadır. Birinci yani üst seviyede, tüm belgeler arasında paylaşılan küresel bir konu kümesi bulunurken, alt seviyede her belge kendi konu dağılımına sahiptir. Bu ikili hiyerarşik yapı, belgelerin benzer temaları paylaşmasına olanak tanıırken, aynı zamanda her belgenin kendine özgü bir tematik yapıya sahip olabileceğini de kabul ederek esnek bir konu modellemesi sunmaktadır. Bu sayede, veri büyüdükçe veya çeşitlendikçe model yeni konuları keşfedebilmekte ve var olan konuların önem derecesini dinamik olarak ayarlayabilmektedir (Teh ve diğerleri, 2004).

1.4.3.1.3. Dynamic Topic Model (DTM)

Dynamic Topic Model (DTM), zaman içinde değişen temaları modellemek için geliştirilmiş bir olasılıksal konu modelleme yöntemidir. Bu model, temaların statik olmadığı, aksine zamanla evrimleşebileceği varsayımına dayanır. DTM de, Latent Dirichlet Allocation (LDA) modelinin genişletilmiş bir versiyonu olarak kabul edilir ve temaların hem içeriğinin (kelime dağılımlarının) hem de popülerliğinin (belgelerdeki dağılımlarının) zaman içinde nasıl değiştiğini analiz etmeyi amaçlar. Bu yaklaşım, özellikle tarihsel metinler, haber arşivleri veya sosyal medya verileri gibi zaman serisi verilerinde tematik değişimleri izlemek için oldukça kullanışlıdır (Wang ve diğerleri, 2015).

1.4.2.2. Cebirsel Modeller

Cebirsel modelleme teknikleri, metin verilerini sayısal bir uzaya dönüştürerek vektör temsilleri üzerinden kategorilere ayıran konu modelleme tekniklerinin genel adıdır. Genellikle doğrusal cebire dayanan bu modeller, metinler arasındaki benzerlikleri ve yapısal örüntüleri matrisler aracılığıyla analiz ederler. Bu yaklaşımda,

belge ve kelime ilişkileri genellikle belge-terim matrisi (*document-term matrix*) üzerinden temsil edilmekte, ardından bu matrisler çeşitli boyut indirgeme teknikleriyle sadeleştirilerek anlamlı konu kümeleri elde edilmektedir. Cebirsel modeller, verideki temel bileşenleri ortaya çıkarmaya yönelik matematiksel işlemlerle çalıştığı için, özellikle yüksek boyutlu veri ile yapılan konu modelleme çalışmalarında hız ve hesaplama açısından avantaj sunmaktadırlar.

1.4.2.2.1 Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA), metinler arasında gizli anlamsal ilişkileri ortaya çıkarmak amacıyla geliştirilen ilk cebirsel konu modelleme tekniklerinden biri olarak kabul edilebilir. LSA belge-terim matrisi üzerinde tekil değer ayrıştırması (*Singular Value Decomposition* - SVD) uygulayarak hem belgeleri hem de kelimeleri düşük boyutlu bir anlamsal uzaya indirgeyerek çalışmaktadır. Bu dönüşüm, belgeler ve terimler arasındaki benzerlikleri daha anlamlı şekilde modellemeyi ve metin koleksiyonları içerisinde örtük (*latent*) konuların keşfedilmesini mümkün kılmaktadır. LSA, sözcüklerin yalnızca frekanslarına değil aynı zamanda bağlamlarına da odaklanarak konu modelleme yapılabileceği mantığına dayanmaktadır. Bu bağlamlar, incelenen dokümanlarda sıkça birlikte geçen kelimelerin benzer anlamlara sahip olabileceği, bu kelimelerin tekil olarak geçtiği dokümanların ise bu anlamı içermediği varsayımı ile tespit edilmektedir. Bu sayede metinlerdeki anlamsal yakınlıklar, yalnızca kelime düzeyinde değil, daha soyut düzeydeki tematik örüntüler aracılığıyla da ortaya konulabilmektedir (Deerwester ve diğerleri, 1990).

LSA'nın temel avantajı, bilgi kaybını azaltarak büyük boyutlu metin verilerini daha yönetilebilir ve anlamlı hale hem basit hem de hızlı biçimde getirebilmesidir. Ancak, olasılıksal bir model olmaması nedeniyle yorumlanabilirlik açısından LSA, sınırlılıklar taşımaktadır.

1.4.2.2.2 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF), belge-terim matrisinin yalnızca pozitif değerler (negatif olmayan değerler) kullanılarak ayrıştırılmasına dayanan bir cebirsel konu modelleme yöntemidir. Bu yaklaşım, her belgeyi belirli oranlarda konuların birleşimi olarak, her konuyu ise bazı kelimelerin ağırlıklı birleşimi olarak temsil eder. NMF'in en önemli avantajı, elde edilen faktör matrislerinde yalnızca pozitif

değerlerin bulunması nedeniyle, kelime ve belge temsillerinin daha anlamlı ve yorumlanabilir hale gelmesidir. Bu durum, özellikle konuların daha “ayrık” ve açık olarak ortaya çıkmasını sağlar. LSA’nın aksine, negatif değerlerin olmaması, NMF ile elde edilen konuların içerik bakımından daha doğrudan tanımlanmasına olanak vermektedir.

NMF, parçaların öğrenilmesi yolu ile bütünün tanınabileceği prensibine dayanan bir algoritmadır. Genellikle düşük boyutlu temsillerin elde edilmesinde, belge kümelerinin görselleştirilmesinde ve yorumlanabilir konuların keşfinde tercih edilmektedir. Özellikle metin madenciliği ve doğal dil işleme alanlarında, hızlı çalışması ve anlamsal olarak tutarlı konu kümeleri üretmesi nedeniyle oldukça yaygın biçimde kullanılmaktadır. (Lee ve Seung, 1999).

1.4.2.3. Melez Modelleme Teknikleri

Melez konu modelleme teknikleri, olasılıksal ve cebirsel modelleme tekniklerinin güçlü yönlerini birleştirmeyi amaçlayan tekniklerdir. Bu çalışmanın da kullanacağı tekniği içeren melez modelleme teknikleri, genellikle birden fazla teknikten faydalanarak daha tutarlı, esnek ve yorumlanabilir konu temsilleri üretir. Melez teknikler, olasılıksal modelleme tekniklerinin tematik derinlik sağlayan yapısı ile cebirsel modelleme tekniklerinin hesaplama açısından verimli yapısını bir araya getirerek, büyük ve karmaşık metin verileri üzerinde daha isabetli sonuçlar elde etmeyi hedeflerler. Son yıllarda, özellikle gömülü temsiller (*embeddings*) ile çalışan ve sinir ağları vasıtasıyla entegre edilen yeni nesil modellerin çoğu bu melez yapıda tasarlanmaktadır. Bu tür modeller, hem kelime-seviye hem de belge-seviye ilişkileri daha iyi öğrenebilmekte ve çeşitli uygulamalarda (örneğin haber analizi, sosyal medya izleme, metin sınıflandırma) oldukça etkili sonuçlar vermektedirler (Alsmadi ve diğerleri, 2024).

1.4.2.3.1 Gömülü Temsillere (Embeddings) Dayalı Konu Modelleme Teknikleri

Gömülü temsillere dayalı konu modelleme teknikleri, kelimeleri ve belgeleri bağlamsal vektörler aracılığıyla temsil eden ve bu temsiller üzerinden konuları tespit etmeyi amaçlayan modellerdir. Bu yaklaşımlar, klasik istatistiksel yöntemlerin sınırlı kaldığı anlamsal benzerlikleri daha doğru biçimde yakalayabilmek amacıyla

geliştirilmiştir. Özellikle önceden eğitilmiş dil modelleri (*pre-trained language models*) ile üretilen kelime ve cümle gömmeleri (*word/sentence embeddings*), metinlerin bağlamsal olarak temsil edilmesine olanak sağlar. Bu temsiller vasıtasıyla, sadece kelime sıklığına değil, kelimelerin kullanım bağlamına dayalı semantik örüntüler de analiz edilebilir. Bu tarz tekniklerin en bilinen örneklerinden biri olan Embedded Topic Model (ETM), klasik LDA'nın kelime gömme ile birleşiminden oluşmakta ve konuların anlamlı biçimde modellenmesini sağlamaktadır (Dieng ve diğerleri, 2020).

BERTopic tekniği de bir yönü ile gömme tabanlı temsiller üzerinden çalıştığı için bu gruba dahil edilebilir. Benzer şekilde Top2Vec modeli de LDA ile LSA'nın güçlü yönlerini birleştiren bir konu modelleme tekniğidir. Bu teknik, belge temsillerini gömme vektörlerine dönüştürdükten sonra kümeleme yöntemi ile konu oluşturma yoluna gitmekte ve herhangi bir ön işleme gerek duymadan bağlamsal benzerlikleri esas almaktadır (Angelov, 2020).

1.4.2.3.2. Sinir Ağlarına (Neural Networks) Dayalı Konu Modelleme Teknikleri

Sinir ağlarıyla geliştirilmiş olasılıksal modeller, klasik konu modelleme yaklaşımlarının varyasyonel çıkarım (*variational inference*) süreçlerini derin öğrenme mimarileri ile birleştirmeyi amaçlayan modern yöntemlerdir. Bu modellerde temel amaç, belgelere ait konu dağılımlarını ve konulara ait kelime dağılımlarını öğrenirken sinir ağı yapılarını kullanarak daha esnek ve güçlü temsiller üretmektir. Özellikle Variational Autoencoders (VAE) temelli modeller bu çerçevede öne çıkmaktadır. Örneğin ProdLDA modeli, LDA'nın varyasyonel formunu sinir ağları üzerinden yeniden tanımlayarak daha keskin ve ayırık konular elde edilmesine imkan tanımaktadır (Srivastava ve Sutton, 2017). Benzer şekilde NVITM (*Neural Variational Inference for Text*) gibi modeller, belgelerden elde edilen sürekli vektör temsilleri üzerinden konu dağılımlarını hesaplamakta ve metinlerin anlamsal yapısını derinlemesine modellemektedir. Bu modeller, klasik yöntemlerin sınırlı kaldığı büyük ölçekli ve karmaşık veri kümelerinde daha istikrarlı ve anlamlı sonuçlar üretme potansiyeline sahiptirler.

1.4.2.3.3. Kümeleme (Clustering) Tabanlı Melez Teknikler

Kümeleme tabanlı melez konu modelleme yöntemleri, metinleri önce bağlamsal olarak sayısal temsillere dönüştüren, ardından bu temsiller üzerinde boyut indirgeme ve kümeleme işlemleri gerçekleştiren bir yaklaşım sunmaktadır. Bu tekniklerde ilk aşamada her belge, önceden eğitilmiş dil modelleri aracılığıyla 1.3.2.3.1. numaralı başlıkta ayrıntıları açıklanan gömülü temsillere dönüştürülür. Özellikle BERT ve türevleri (örn. Sentence-BERT), metinleri bağlam bilgisiyle zenginleştirilmiş vektörler olarak ifade ederek her belgenin anlamsal bir konum kazanmasını sağlar. Bu vektör temsiller, yüksek boyutlu uzaylarda yer aldığından, çoğu zaman UMAP (*Uniform Manifold Approximation and Projection*) gibi boyut indirgeme algoritmaları kullanılarak daha sıkıştırılmış ve görselleştirilebilir bir yapıya dönüştürülür. Bu sayede HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) gibi yoğunluk tabanlı kümeleme algoritmaları ile benzer bağlamsal özellikler taşıyan belgeler bir araya getirilerek konular oluşturulur.

Bu yaklaşımın en bilinen ve kapsamlı örneklerinden biri olan BERTopic, gömülü temsilleri, boyut indirgeme tekniklerini ve yoğunluk temelli kümeleme yöntemlerini entegre biçimde kullanarak, konu modelleme sürecine esneklik ve bağlamsal duyarlılık kazandırmaktadır. BERTopic modelinde, her konu, bu konuyu temsil etme olasılığı algoritma tarafından yüksek olarak belirlenen anahtar kelimelerle özetlenerek yorumlanabilir hale getirilir. Bu yapı, klasik olasılıksal modellerin istatistiksel temsillerinin ötesine geçerek, metinler arası semantik ilişkileri daha güçlü biçimde ortaya koymaktadır (Grootendorst, 2022).

İKİNCİ BÖLÜM

ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE YAPAY ZEKA

2.1. ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE YAPAY ZEKA KULLANIM ALANLARI

Uluslararası ilişkiler bir disiplin olarak kuruluşundan itibaren genellikle nitel analiz yöntemlerine dayalı olarak gelişmiştir. Özellikle yorumlayıcı ve tarihsel bağlam çerçevesinde çözümler yoluyla disiplin içerisinde politik süreçler anlamlandırılmaya ve açıklanmaya çalışılmıştır. İstatistiki yöntemler disipline 1960'larda girmeye başlamıştır (Kaplan, 1966) . Sistemik modellerin ve matematiksel yöntemlerin Uİ çalışmalarında kullanımının siyasetin doğası gereği içerdiği moral temelleri göz ardı edeceği ve bağlamın belirleyiciliğini yakalayamayacağı gerekçesiyle yoğun eleştirilere tabii tutulmuştur (Bull, 1966). Dolayısıyla bu yöntemlerin nitel yöntemlerle diyalogları sınırlı kalmış ve özellikle Avrupa kökenli Uİ çalışmalarında kendilerine yeterince yer bulamamışlardır.

Ancak son yıllarda YZ teknolojilerinde yaşanan gelişmeler uluslararası politikanın çeşitli yönlerinin analizinde yeni olanaklar sunmaya başlamıştır. Bu teknolojiler, daha önce manuel yollarla ve doğrudan araştırmacıların eliyle yapılan analizlerin daha hızlı, sistematik ve önyargıdan arındırılmış biçimde yapılmasını sağlamıştır. Bu durum, uluslararası siyasette yer alan aktörlerin söylemlerinin izlenmesi, dış politika stratejilerinin anlamlandırılması ve kamuoyunun yönlendirilme biçimlerinin nesnel biçimde incelenmesine imkan tanımaktadır. Dahası YZ ve alt dalları, stratejik öngörü, kriz erken uyarı sistemleri ve politik eğilim tespiti gibi karar destek sistemlerinin geliştirilmesi noktasında da önemli rol oynamaktadır.

YZ teknolojileri, aynı zamanda Uİ'de daha derinlikli ve sistematik analizlere olanak tanımaktadır. Uİ'deki neredeyse tüm başat teoriler uluslararası politikada aktör merkezli analiz çerçevelerinde söz konusu aktörlerin rasyonel kararlar alacağı varsayımı ile hareket etmektedirler. YZ destekli teknolojiler Uİ'ye bu varsayımın ötesine geçme olanağı sunarak kararların arka planındaki bilgi işleme süreçlerine odaklanmayı mümkün kılmaktadır. Bu durum, disiplinin geleneksel aktör merkezli analizlerinden farklı olarak, karar alma davranışlarını biçimlendiren zihinsel çerçeveleri ve bilgi temelli süreçleri modellemeye imkan sunmaktadır. YZ bu bağlamda yalnızca büyük veri kümeleri üzerinde analiz yapılmasına değil, aynı

zamanda teori inşasına katkı sağlayabilecek araçlar sunarak metodolojik çeşitliliği artırmaktadır (Schrodt, 2019).

Bu çerçevede YZ'nin dış politika davranışlarının simülasyonu, kriz senaryolarının modellenmesi ve politik belirsizlik ortamlarında alternatif sonuçların tahmini gibi alanlarda kullanılabilir olması, disiplindeki mevcut paradigma sınırlarını genişletme potansiyeline sahiptir. YZ'nin özellikle karar süreçleri için öngörüler üretme kapasitesi, Ul'yi sadece analitik değil, aynı zamanda doğrudan politika yapıcılara yönelik bir açılım aracı haline getirmektedir. Bu bağlamda YZ'nin, hem nitel hem de nicel yöntemleri destekleyen hibrit araştırma yaklaşımlarına olanak tanıyarak, disiplinin araştırma sınırlarını genişlettiği ifade edilebilir (Meleouni ve Efthymiou, 2023).

YZ'nin Ul'de kullanım alanları artarken, büyük ölçüde metin temelli bilgi kaynaklarıyla çalışan bir alan olması sebebi ile disiplin, sayısallaştırılmış metinlerin sistematik analizine imkan tanıyan yöntemlerden giderek daha fazla yararlanmaktadır. Bu bağlamda, DDİ teknikleri, özellikle büyük ölçekli belge kümeleriyle çalışan araştırmacılar açısından önemli bir araç haline gelmektedir. Metinlerin yalnızca içeriksel değil, aynı zamanda yapısal özelliklerini de analiz edebilen DDİ uygulamaları, diplomatik söylemlerin, haber metinlerinin ya da uluslararası aktörlerin tutum ve davranışlarının belirli kalıplar içinde nasıl inşa edildiğini nesnel ölçütlere dayalı olarak inceleme imkanı sunmaktadır. Bu yönüyle DDİ, yalnızca veri hacmiyle başa çıkmayı değil, aynı zamanda metinlerdeki örüntülerin daha önce olmadığı kadar derinlemesine çözümlenmesini de mümkün kılmaktadır (Marwala, 2023).

Disiplinin metodolojik çeşitliliği dikkate alındığında, DDİ'nin hem nitel hem de nicel araştırma geleneklerine katkı sağlayabileceği görülmektedir. Özellikle vaka analizi, söylem analizi, çerçeveleme ve anlatı yapılarının çözümlenmesi gibi yöntemlerin daha sistematik, şeffaf ve tekrarlanabilir hale gelmesi, DDİ tabanlı araçlarla mümkün olabilmektedir. DDİ teknikleri ile, metinlerdeki özne-eylem-nesne (*subject-verb-object* - SVO) yapılarının çıkarılarak aktör-çıkar-eylem tespitinin yapılabilmesi, aktör davranışının izlenmesi, söylemlerin zaman ve mekan ilişkilerinin çözümlenmesi gibi işlemler, Ul'de geleneksel yaklaşımların sınırlarını aşarak daha ayrıntılı analizlerin yapılabilmesini sağlamaktadır (Franzosi ve diğerleri, 2022).

2.2. ULUSLARARASI İLİŞKİLER DİSİPLİNİNDE KONU MODELLEME TEKNİKLERİNİN KULLANIMI

Yukarıda da ifade edildiği üzere, uluslararası ilişkiler disiplininde DDİ yöntemlerinin kullanımı, büyük veri kümelerinin sistematik ve derinlemesine analiz edilmesi bakımından önemli bir dönüşüm potansiyeli sunmaktadır. Özellikle söylem analizi, çerçeveleme ve anlatı yapılarını incelemeye yönelik DDİ tabanlı yöntemlerin disipline entegre edilmesi, metinlerin içeriksel ve yapısal unsurlarını daha önce mümkün olmayan bir ölçekte incelemeye olanak vermektedir. Konu modelleme teknikleri de DDİ'nin sağladığı bu analitik olanaklardan birini oluşturmakta ve Uİ çalışmalarında giderek artan oranda tercih edilmektedir. Bu teknikler sayesinde, geniş metin veri kümeleri içerisindeki tematik yapılar otomatik olarak keşfedilmekte, böylece uluslararası siyasetteki söylem, algı ve politika üretim süreçlerinin arka planındaki temalar ve söylemsel kalıplar nesnel bir biçimde ortaya çıkartılabilmektedir.

Konu modelleme tekniklerinin gelişmesi ile birlikte bu yöntemlerin analitik potansiyelinin Uİ'ye yaptığı katkılar birçok araştırmacı tarafından değerlendirilmiştir. Bunun bir sonucu olarak farklı konu modelleme teknikleri kullanılarak Avrupa Birliği'nde (AB) pestisit kullanımı hakkındaki tartışmaları AB kurumlarının nasıl yönlendirdiğinden (Lock, 2020) ulusal siber güvenlik stratejilerinin tespitine (Kolini ve Janczewski, 2017) birçok çalışma yapılmış ve yapılmaya devam etmektedir.

Örneğin Kirill ve diğerleri (2020) medya içeriklerindeki propaganda unsurlarının konu modelleme teknikleri ile nasıl tespit edilebileceği incelenmektedir. Fazla sayıda haber metni verisi içerisinde propaganda içeren haberleri, manuel etiketleme süreçlerini en aza indirerek hızlı ve etkili şekilde belirlemeyi amaçlayan çalışmada yazarlar, propaganda içerikleri üreten ve tarafsız kabul edilen haber kaynaklarından derlenen iki ayrı derlem üzerinde, konu modelleme yöntemiyle hangi temaların propagandacı haber kaynakları ile daha ilişkili olduğunu ortaya koymuştur. Çalışmanın sonuçları, konu modelleme tekniğinin propaganda içeriğini tespit etmekte etkili olduğunu ve geniş çaplı metin analizlerinde kullanılabileceğini göstermektedir.

Konu modellemenin Uİ'de sıklıkla başvuru alan bir yöntem olan söylem analizi için kullanımı da yaygın kabul görmüş araştırma konularından biridir. Örneğin Jacobs ve Tschötschel (2019) ile Brokes ve McKeany (2019) konu modellemenin söylem analiziyle ilişkisini farklı boyutlarıyla ele almaktadırlar.

Jacobs ve Tschötschel, konu modellemenin söylem analizi çalışmalarındaki metodolojik sınırlılıkları aşma potansiyeline vurgu yapmaktadırlar. Yazarlara göre

konu modelleme, geniş metin kümelerinin sistematik analizinde sadece sınıflandırma yapmanın ötesinde, söylemlerin yapısını ve kelimelerin bağlamsal kullanımını ortaya koymak için yöntemsel avantajlar sunmaktadır. Ayrıca konu modelleme, geleneksel yöntemlerin oldukça zorlandığı ölçeklendirme, tekrar edilebilirlik ve sistematizasyon sorunlarını çözmede oldukça etkili olarak değerlendirilmiştir.

Öte yandan Brookes ve McEnery (2019) ise çalışmalarında, İngiltere Ulusal Sağlık Sistemi hakkında yazılmış 228.000'den fazla çevrimiçi hasta yorumunu inceleyerek LDA tekniğinin söylem analizinde ne ölçüde işe yaradığını değerlendirmektedirler. Çalışmada, aynı veri kümesi hem konu modelleme hem de geleneksel söylem analizi yöntemleriyle analiz edilerek iki yaklaşım karşılaştırılmıştır. Yazarlara göre sonuçlar, konu modelleme ile elde edilen temaların çoğunlukla yüzeyde tutarlı görünse de metinlerin içeriğiyle tam anlamıyla örtüşmediğini ortaya koymuştur. Yazarlar, konu modellemenin büyük veri kümelerinde hızlı tematik ayrıştırma yapabilme ve metinlerin genel "ne hakkında olduğu" bilgisini sunabilme açısından avantaj sunduğunu, ancak bu çıktıların yalnızca nitel analizle desteklendiğinde anlamlı hale gelebileceğini ifade etmektedirler.

Son yıllarda konu modelleme, uluslararası örgütlerde üretilen toplantı tutanakları gibi metin verilerinde aktör davranışını, aktörlerin ilişkili olduğu konuları ve aktör önceliklerini belirleme açısından da Uİ'de yaygın olarak kullanılmaya başlanmıştır. Örneğin Wojciechowska ve diğerleri (2023) tarafından kaleme alınan çalışmada UNESCO Dünya Mirası ve Somut Olmayan Kültürel Miras Komiteleri toplantı tutanaklarını analizi vasıtasıyla devletler arası söylemlerde örtük biçimde yer alan gerilimlerin otomatik olarak tespiti amaçlanmıştır. 1973-2023 yılları arasında 113 toplantı tutanağından ve yaklaşık 6,3 milyon kelimedenden oluşan bir derlem içerisinden seçilen 1270 paragraflık örneklem yazarlar tarafından manuel olarak etiketlenmiştir. Gerilim içeren söylemleri tespit etmek için RoBERTa tabanlı bir sınıflandırma modeli geliştirilmiş ve %72 oranında doğruluk elde edilmiştir. Buna ek olarak, BERTopic yöntemiyle yürütülen konu modelleme süreci sayesinde toplantılarda öne çıkan temalar belirlenmiş ve devletler arasında doğrudan ifade edilmeyen ancak var olan gerilimleri otomatik olarak tespit edebilecek bir model geliştirilmiştir. Temaların ve bulguların doğruluğu ile anlamlılığı insan değerlendirmeleriyle de doğrulanmıştır. Çalışma konu modellemenin diplomatik metinlerin yapısal çözümlemesinde etkin biçimde kullanılabileceğini göstermesi bakımından önemli bir örnek teşkil etmektedir.

Uluslararası örgütler içerisindeki aktör davranışını tespit etmeye yönelik bir başka çalışma ise BM Güvenlik Konseyi içerisinde yapılan tartışmalarda BM Genel

Sekreterliği'nin rolüne odaklanmaktadır. Afganistan temalı tartışmalarda Sekreteryası'nın nasıl bir söylemsel rol üstlendiğini analiz eden çalışma 1996-2017 yılları arasında gerçekleşen 155 oturumdan elde edilen 1600 konuşmayı veri olarak ele almaktadır. Yapısal konu modelleme (*structured topic modelling - STM*) yöntemi kullanılarak analiz edilen veride 13 farklı konu belirlenmiş ve her yıl için konuşmacı-konu ağları oluşturulmuştur. Çalışma, BM Genel Sekreterliği'nin varoluş amacına uygun olarak çoğunlukla tarafsız bir dil benimsediğini, ancak zaman zaman "güvenlik ve reform", "seçimler" ve "ekonomik kalkınma" gibi temalar aracılığıyla kendi gündemini oluşturabildiğini göstermektedir. Bu temaların diğer aktörler tarafından da sahiplenilmesi, BM bürokrasisinin yalnızca arka planda kalan teknik bir aktör değil, aynı zamanda söylemsel olarak etkili bir uluslararası aktör olabileceğini ortaya koymaktadır (Eckhard ve diğerleri, 2023).

Aktör önceliklerinin belirlenmesine odaklanan bir başka çalışma, Pakistan'ın 1970–2018 yılları arasında Birleşmiş Milletler Genel Kurulu'nun yıllık genel tartışmalarında yaptığı konuşmaları konu modelleme yöntemiyle analiz ederek ülkenin dış politika önceliklerini ortaya koymayı hedeflemektedir. Çalışmada, toplam 49 konuşmadan oluşan derlem üzerinde LDA algoritması uygulanmış ve dokuz ana tema tespit edilmiştir. Bu temalar arasında terörle mücadele, terörizmin refaha oluşturduğu tehdit, barış ve güvenlik, ekonomik kalkınma, nükleer tehdit, Afgan barış süreci ve uluslararası topluma çağrı gibi başlıklar öne çıkmaktadır. Araştırma, hem nicel hem nitel analiz yöntemlerini birleştirerek yürütülmüş ve makine ile tespit edilen temaların anlamlandırılmasında yorumlayıcı yaklaşımlardan yararlanılmıştır. Sonuçlar, Pakistan'ın küresel söylemde güvenlik, istikrar ve kalkınma temaları etrafında şekillenen tutarlı bir dış politika söylemi izlediğini ortaya koymaktadır. Hem izlediği yöntem hem de vardığı sonuçlar açısından çalışma, konu modelleme yönteminin ulusal söylemlerin dış politikadaki yansımalarını ortaya koymak için etkin biçimde kullanılabileceğini göstermektedir (Khan ve diğerleri, 2022).

Aktörler arasındaki ilişkilerde zaman içerisinde öne çıkan rekabet alanlarının tespit amacı ile de konu modelleme tekniklerinden faydalanılmaktadır. Tang (2024) tarafından gerçekleştirilen çalışmada, 2013-2023 arasındaki Çin-ABD ilişkilerinin tematik evrimini analiz etmek amacıyla LDA yöntemiyle konu modelleme uygulanmıştır. Akademik makaleler, haber içerikleri, hükümet raporları ve sosyal medya gönderilerinden oluşan çok kaynaklı ve büyük bir veri seti üzerinde yürütülen analizde, ekonomi, teknoloji, askeri strateji, diplomatik hamleler, ideoloji ve finansal hegemonya olmak üzere altı ana temanın temel rekabet alanları olduğu belirlenmiştir.

Zaman serisi analizinde özellikle 2018 sonrası dönemde ekonomik ve teknolojik rekabetin ön plana çıktığı, finansal hegemonya kavramının ise ABD'nin Çin'in yükselişini sınırlandırmaya yönelik stratejilerinde merkezi bir rol oynadığı tespit edilmiştir. Ayrıca birlikte geçme ağı (*co-occurrence network*) analizleriyle bazı anahtar kavramların birlikte var olma örüntüleri görselleştirilmiş, böylece konuların yalnızca içeriksel değil, bağlamsal ilişkileri de ortaya konulmaya çalışılmıştır. Çalışma, konu modelleme yönteminin yalnızca tematik ayrıştırma değil, aynı zamanda söylemsel eğilimlerin zaman içerisindeki dönüşümünü anlamada da etkili bir araç olduğunu göstermektedir.

Uİ'de konu modelleme, mülteci krizi gibi genel temaların nasıl ele alındığının tespiti için de kullanılmaktadır. Grant ve diğerlerinin (2021) yürüttüğü araştırma 1970'lerde yaşanan üç büyük mülteci krizine (Uganda, Şili ve Vietnam) ilişkin BM, ABD ve Birleşik Krallık arşivlerinde yer alan belgeler üzerinden konu modelleme yöntemini kullanarak bu krizlere dair söylemin hangi temalar etrafında şekillendiğini belirlemeyi amaçlamaktadır. Ağustos 1972 ile Aralık 1979 arasındaki dönemi kapsayan ve 55000 sayfayı aşkın belgelerden oluşan derlem üzerinde LDA yöntemiyle 16 konu belirlenmiştir. Analiz sonucunda sezgisel olarak bilinen aktör davranışları ve ulusal mülteci politikalarının yanı sıra, özellikle "insan hakları söylemi" ve "hareket özgürlüğü" gibi daha önceden açıkça farkedilmeyen temaların da ön plana çıktığı ortaya konmuştur. Çalışma bu yönü ile konu modelleme tekniklerinin yalnızca belgeleri sınıflandırmakla kalmayıp, bu belgelerdeki örtük söylemsel örüntüleri açığa çıkarmada Uİ için etkili bir araç olabileceğini göstermektedir.

2.3. KONU MODELLEME TEKNİKLERİ İLE ÇERÇEVE ANALİZİ

Konu modelleme tekniklerinin çerçeve analizi amacıyla kullanımı, hem Uİ hem de iletişim çalışmaları başta olmak üzere sosyal bilimlerin pek çok alanında metin temelli araştırmalara yeni bir ivme kazandırmıştır. Konu modelleme çerçeve analizinde, özellikle büyük ölçekli veri setlerinde temaların sistematik biçimde belirlenmesi ve bu temaların hangi anlatı yapıları veya bakış açılarıyla sunulduğunun nesnel olarak tespit edilmesi bakımından dikkate değer kolaylıklar sağlamıştır.

Çerçeveleme, belirli gerçeklik unsurlarının ön plana çıkarılarak ele alınan gerçekliğin nasıl algılandığını yönlendiren bir süreç olarak, gerek haber metinlerinde gerekse uluslararası politik söylemlerde sıkça başvurulan bir stratejidir. Bu bağlamda konu modelleme teknikleri, metinlerde baskın biçimde yer verilen temaların, hangi

bağlamlar içerisinde sunulduğunu anlamak, örtük anlam yapılarını ortaya koymak ve medya organları ya da siyasi aktörler tarafından kullanılan çerçeveleri nesnelleştirmek amacıyla çerçeve analizinde bir araç olarak kullanılmaktadır. İnsan yorumuna dayalı manuel kodlama gereksinimini azaltması ve özellikle denetimsiz tekniklerin analiz edilecek temaları önceden tanımlamaya gerek duymaması, verinin kendi içindeki örüntülere (yani kelime kullanım sıklıkları, bağlam ilişkileri vb.) dayanarak çerçevelerin otomatik ve nesnel olarak ortaya çıkarılabilmesi, konu modellemenin çerçeve analizinde sunduğu önemli avantajlardır. Bir başka deyişle konu modelleme ile çerçeveler önceden araştırmacının belirlediği kategorilere uydurulmaya çalışılmaz, veride zaten baskın olarak bulunan tematik yapıları istatistiksel yöntemlerle açığa çıkarır. Dolayısıyla, konu modelleme çerçeve analizine nesnellik, ölçeklenebilirlik ve tekrarlanabilirlik gibi önemli katkılar sunmaktadır.

Konu modellemenin çerçeve analizinde nasıl kullanılabileceğini değerlendiren çalışmalarında Eisele ve diğerleri (2023), farklı konu modelleme tekniklerini farklı düzeylerde denetim içeren yaklaşımlarla karşılaştırmaktadır. 2009-2019 yılları arasında Avusturya'da yayın yapan 12 gazetenin AB ile ilgili yayınladıkları haberlerden oluşturulan derlem yazarlarca çerçeve kategorileriyle etiketlenmiştir. LDA, seeded LDA, naive bayes, lojistik regresyon, destek vektör makineleri (*support vector machines* - SVM) gibi algoritmalar bu veri üzerinde karşılaştırmalı olarak test edilmiştir. Çalışma, hem denetimsiz hem de yarı denetimli yöntemlerin tematik yapıları ortaya çıkarma potansiyelini değerlendirmesi açısından önemli bir örnek teşkil etmektedir.

Çalışma sonucunda, LDA gibi gözetimsiz yöntemlerin temaları genel hatlarıyla saptamada etkili olmakla birlikte, önceden belirlenmiş çerçevelerin açık biçimde ayrıştırılmasında yetersiz kaldığı, buna karşılık, etiketli veri kullanımı mümkün olduğunda denetimli yöntemlerin daha yüksek doğruluk sunduğu ortaya konmuştur. Seeded LDA gibi yarı denetimli yaklaşımların ise her iki yöntem arasında ara bir model sunduğu ve orta düzeyde başarı sağladığı belirtilmiştir. Bu yönüyle çalışma, çerçeve analizinde konu modelleme kullanımının yöntemsel sınırlarını ve potansiyelini ortaya koymakta; özellikle etiketli verinin sınırlı olduğu durumlarda gözetimsiz tekniklerin sağlayabileceği katkıya dikkat çekmektedir.

Avrupa'da 2015–2016 yılları arasında yaşanan mülteci krizinin medya söylemine nasıl yansıdığını inceleyen çalışmalarında Heidenreich ve diğerleri (2019) beş farklı Avrupa ülkesinde yayımlanan 130.000'i aşkın haber içeriği üzerinde LDA temelli konu modelleme yöntemi kullanarak karşılaştırmalı bir çerçeve analizi

gerçekleştirmektedir. Almanya, Birleşik Krallık, İsveç, İspanya ve Macaristan'a ait önde gelen haber kaynaklarından elde edilen metinlerde, her ülke için ayrı ayrı uygulanan model vasıtasıyla her biri 10 temadan oluşan konu kümeleri çıkarılmıştır. Yazarlar tespit edilen bu temalar üzerinden, her ülkenin medya organlarında mülteci krizinin nasıl çerçevelendiği sistematik biçimde analiz etmeyi amaçlamışlardır.

Çalışma, kriz dönemlerinde medya çerçevelerinin hem ortaklaşan temalar barındırdığını (örneğin güvenlik, insani yardım, ulusal siyaset) hem de ülkeye özgü farklılıklar gösterdiğini ortaya koymaktadır. Örneğin Macaristan'da sınır güvenliği ve AB politikaları çerçevesi baskınken, Almanya'da barınma ve ulusal düzeydeki siyasi tartışmalar ön plana çıkmıştır. İsveç'te insani yaklaşım odaklı söylem yaygınlaşırken, Birleşik Krallık'ta Brexit ve göçmen kampları konuları belirgin hale gelmiştir. Çalışma, konu modelleme yönteminin büyük ölçekli medya verisi üzerinde zaman ve ülke bağlamındaki söylemsel örüntüleri ortaya koyma kapasitesini göstermesi bakımından çerçeve analizine önemli bir katkı sunmaktadır.

Mülteci krizine ilişkin bir diğer çalışma ise Finlandiya'da konu hakkındaki kamusal tartışmalara odaklanmaktadır. 2015–2016 yılları için hem geleneksel medya hem de sosyal medya içeriklerinden oluşturulan veri seti üzerinde karşılaştırmalı olarak çerçeveleri tespit etmek amacı ile LDA yöntemi uygulanmış, ardından sosyal medyada paylaşılan bağlantılar üzerinden hiperlink ağı analizi gerçekleştirilmiştir. Toplamda 27.504 haber içeriği ile 1 milyonu aşkın sosyal medya ve forum mesajı üzerinde gerçekleştirilen analiz, geleneksel medyada daha kurumsal ve idari süreçlere odaklı çerçevelerin öne çıktığını, buna karşılık sosyal medyada göçmenlerin tehdit unsuru olarak çerçevelendiği söylemlerin daha baskın olduğunu tespit etmiştir. Etiketlenmemiş büyük veri ile yapılan çalışma, konu modellemenin yalnızca kategorize etme değil, aynı zamanda söylemsel örüntülerin ve çerçeve çeşitliliğinin nesnel biçimde izlenebileceği bir analiz aracı olarak işlev görebileceğini göstermesi açısından önem arz etmektedir (Pöyhkäri ve diğerleri, 2021).

Bir başka karşılaştırmalı çalışma ise Almanya, İspanya, Birleşik Krallık ve İsviçre medyasının 2014–2022 yılları arasında Suriyeli ve Ukraynalı mültecilere ilişkin söylemlerini incelemektedir. Yazarlar öncelikle, 84.623 haber metni üzerinde LDA yöntemi ile konu modelleme yapmış, daha sonra elde edilen konuları, kavramsal kategorilere göre manuel olarak sınıflandırmıştır. Bu iki aşamalı analiz sonucunda, mülteci söylemlerinin on bir farklı çerçevede sunulduğu tespit edilmiş ve bu çerçeveler “kader”, “tehdit”, “değer” ve “bağlam” olmak üzere dört ana başlık altında gruplandırılmıştır. Çalışmanın bulgularına göre, Suriyeli mülteciler sıklıkla güvenlik,

siyasal tartışma konusu ve sisteme getirdikleri yük bağlamında ele alınırken, Ukraynalı mülteciler insani yardım, mağduriyet ve değer temelli çerçevelerle sunulmuştur. Konu modelleme yönteminin farklı çerçeveleri tespit etmedeki gücünü ortaya koyan çalışmada çerçeveler arasındaki farklılığın kültürel ve dinsel önyargılarla ilişkili olabileceği vurgulanmaktadır (Hoffmann ve Hameleers, 2024).

Konu modelleme teknikleri bir ülke medyasının bir konuya bakış açısını zaman içerisinde izlemek için de kullanılabilir. Örneğin 2012 yılında Hindistan'ın Delhi şehrinde gerçekleşen bir toplu tecavüz vakasını konu alan Shah'ın çalışması (2019), cinsel şiddet haberlerinin Hindistan'da nasıl çerçevelendiğini ve zaman içinde çerçeveleme biçimlerinin nasıl değiştiğini konu modelleme yöntemiyle analiz etmektedir. Times of India, Hindustan Times ve Indian Express gazetelerinden derlenen 7.250 haber metni zaman içerisindeki değişimleri de yakalayabilmek için olay tarihinden önceki ve sonraki altı ay ile olay tarihinden üç yıl sonraki altı ay olmak üzere üç dönem halinde sınıflandırılmıştır. Her dönem için LDA uygulanarak tespit edilen temalar içeriksel olarak “episodik”, “tematik” ya da her ikisi olarak sınıflandırılmıştır.

Çalışma, olayın hemen ardından haber söyleminin büyük ölçüde bireysel olaylara ve fail-mağdur anlatılarına dayalı “episodik” çerçevelerle şekillendiği, olayın üzerinden geçen 3 yıl sonunda ise toplumsal nedenlere ve yapısal eşitsizliklere odaklanan “tematik” çerçevelerin ön plana çıkmaya başladığını tespit etmiştir. Bu açıdan çalışma konu modelleme tekniklerinin yalnızca anlık temaları değil, söylemsel eğilimlerin uzun vadeli dönüşümünü de izleyebilecek bir araç olarak kullanılabileceğini göstermesi bakımından dikkat çekicidir.

Çerçeve analizinde LDA ile BERTopic tekniklerini birlikte kullandığı çalışmada Verbytska (2024), 2022–2023 yılları arasında Rusya-Ukrayna Savaşı'nın Batı ve Doğu medyasında nasıl çerçevelendiğini incelemektedir. Bu kapsamda yazar, LDA ile metinlerin istatistiksel yapısına dayalı olasılıksal konu kümeleri oluşturmuş, bu konu kümeleri üzerinde BERTopic modeli ile de bağlamsal ve anlamsal ilişkilere duyarlı bir çerçeveleme analizi gerçekleştirmiştir. Çalışma, medya organlarının çerçeve tercihlerinde anlamlı farklılıklar bulunduğunu ortaya koymaktadır. Batı medyasını temsilen örnek olarak seçilen Euronews haberlerinde enerji krizi, küresel ittifaklar ve mülteci sorunu gibi daha sistemsel ve uluslararası boyutlara odaklanan çerçeveler (örneğin enerji kaynakları, dış politik düzenlemeler ve insani yardım) öne çıkarken, Doğu medyasını temsil eden örnek olarak seçilen Kyiv Post haberlerinde doğrudan savaş deneyimi, askeri destek ve ulusal direniş anlatıları

etrafında şekillenen çerçevelerin (örneğin güvenlik ve savunma, uluslararası silah yardımı ve gazetecilik pratikleri) baskın olduğu gözlemlenmiştir. Ayrıca yazara göre Kyiv Post, medya ve sanat yoluyla oluşturulan kamuoyu desteğini doğrudan direnişin bir parçası olarak işlerken, Euronews bu temaları daha çok sembolik dayanışma ve kültürel temsil çerçevesinde sunmaktadır. Bu karşılaştırmalı çerçeveleme analizi, medya kuruluşlarının coğrafi ve ideolojik konumlarının söylem üretimine nasıl yansıdığını ortaya koyması bakımından dikkat çekicidir.

BERT tabanlı sınıflandırma modelleri tek başlarına da çerçeve tespiti amacıyla kullanılmaktadırlar. Jumle ve diğerleri (2024), İsviçre’de gerçekleştirilen iki halk oylamasına ilişkin haberleri temel alan çalışmalarında, çeşitli arama motorlarından elde ettikleri 372 web sayfası ve 2.736 paragraftan oluşan bir veri kümesi kullanmıştır. Bu veriler, beş genel haber çerçevesine —sorumluluk atfı, insan ilgisi, çatışma, moral ve ekonomik çerçeve— dayalı olarak manuel biçimde etiketlenmiştir. Etiketlenen veri setiyle, *HuggingFace* Transformers kütüphanesi kullanılarak *bert-base-uncased* modeli eğitilmiştir. Özellikle insan ilgisi, ekonomik çerçeve ve sorumluluk atfı gibi çerçevelerde yüksek doğruluk oranları elde edilirken, örnek sayısının azlığı nedeniyle moral çerçeve kategorisinde performans sınırlı kalmıştır. Çalışma, transformer tabanlı modellerin yalnızca önceden tanımlanmış çerçeveleri tanımada değil, aynı zamanda farklı bağlam ve metin türlerinde genellenebilir biçimde kullanılabileceğini göstermektedir.

Asnawi ve diğerleri (2024), Endonezya haber içerikleri üzerinde yaptıkları çalışmada, BERTopic ve LDA modellerini karşılaştırarak her iki yöntemin kısa ve yapılandırılmamış metinlerdeki performansını değerlendirmektedir. Özellikle çevrim içi haber verisi gibi tematik olarak dağınık ve dinamik yapıya sahip metinlerde, LDA’nın konu bütünlüğü ve tutarlılık açısından sınırlı kaldığı, buna karşın BERTopic’in gömme ve kümeleme algoritmalarını birleştirerek daha tutarlı ve yorumlanabilir temalar sunduğu belirtilmektedir. Yazarlara göre BERTopic’in sunduğu bu avantajlar, konu modellemenin yalnızca kelime frekansına değil, aynı zamanda kavramsal benzerliklere dayalı olarak gerçekleştirilmesini mümkün kılmakta, bu da özellikle haber söyleminde öne çıkan çerçeve yapılarının daha isabetli biçimde ortaya çıkarılmasına katkı sağlamaktadır. Çalışma konu modelleme yönteminin veri setinin yapısal özelliklerine göre dikkatle seçilmesi gerektiğini vurgulamaktadır.

BERTopic yöntemi İngilizce dışındaki dillerde de iyi sonuçlar vermesi açısından tercih edilmektedir. Örneğin Nagy’nin (2024) çalışması, 2014 Rusya-Ukrayna krizine dair Macar çevrimiçi haber medyasındaki çerçeve yapılarını analiz

etmek amacıyla konu modeli ağı analizi (*analysis of topic model networks*) yöntemini BERTopic algoritmasıyla birleştirerek uygulamaktadır. Macarca metinlerde bağlam duyarlılığını artırmak amacıyla oluşturulan gömülü temsiller UMAP ile boyutları azaltıldıktan sonra HDBSCAN ile kümelenecek anlamlı çerçeve toplulukları ortaya çıkarılmıştır. Böylece çalışmada insani ilgi, uluslararası diplomasi ve ekonomik sonuçlar gibi temalara dayanan çerçeve yapıları otomatik biçimde tespit edilmiştir.

Çalışma ilaveten, MH17 uçak kazası ve Ukrayna'daki genel çatışma haberlerinden oluşan iki alt derlemi ayrı ayrı analiz etmiş, insani kayıplar, uluslararası suçlama, ekonomik yaptırımlar ve çatışma unsurlarının öne çıktığı farklı çerçeve kümeleri ortaya koymuştur. Bununla birlikte, analiz sonuçlarının medya kuruluşlarının ideolojik yönelimine (sol, sağ, merkez, aşırı sağ) veya hükümete yakınlık derecesine göre belirgin bir çerçeve farkı göstermediğini belirten yazar bunun sebebi olarak haberlerin büyük ölçüde tek bir kaynak olan Macar Haber Ajansı'ndan alınarak yeniden üretilmesinden kaynaklandığını öne sürmektedir. Bu yönüyle çalışma, konu modelleme tekniklerinin özellikle homojen ve ajans temelli medya ortamlarında çerçeve analizinde sunduğu olanakları ve sınırlarını göstermesi açısından dikkat çekicidir.

Konu modelleme ile çerçeve analizi bu sunduğu bu avantajlar bakımından son yıllarda literatürde sıklıkla başvurulmuş bir teknik haline gelmiştir. Konu modelleme teknikleri, çerçeve analizine nicel bir boyut kazandırarak tematik yapıların nesnel, tekrarlanabilir ve ölçeklenebilir biçimde ortaya konulmasına olanak tanımaktadır. Özellikle LDA gibi olasılıksal yaklaşımlar, geniş veri setlerinde temel temaların tespiti açısından etkili olurken, bağlamsal duyarlılığı sınırlı kalmakta, kısa ve yapısal olarak dağınık metinlerde yorumlanabilirlik açısından zorluklar yaratabilmektedir. Buna karşın, BERTopic modeli bağlamsal gömme, boyut indirgeme ve yoğunluk tabanlı kümeleme gibi modern teknikleri bir araya getirerek, metinler arasındaki anlamsal ilişkileri daha isabetli biçimde tespit edebilmekte, bu yönüyle özellikle haber söylemi gibi içeriksel çeşitliliğin yüksek olduğu veri setlerinde çerçeve yapılarını daha tutarlı ve ayrık biçimde ortaya koyabilmektedir.

Bu çalışmada ele alınan araştırma soruları doğrultusunda, Türkiye'de Afrika'ya dair haber söyleminin hangi temalar etrafında çerçevelendiğini incelemek hedeflenmektedir. Haber metinlerinin yapısal çeşitliliği, metin uzunluklarındaki dengesizlik, ve kavramsal yakınlıkların bağlamsal biçimde değerlendirilmesinin gerekliliği, bu araştırma açısından BERTopic yöntemini LDA'ya kıyasla daha uygun bir araç haline getirmektedir. BERTopic'in Türkçe dilinde de etkin biçimde

alışabilmesi, erevelerin yalnızca kelime sıklıklarına deėil, semantik bütönlüklerine göre de tespit edilebilmesine imkan sunarak, bu alışmanın amaçlarıyla yapısal bir uyum göstermektedir. Bu doėrultuda, ereve analizinde BERTopic kullanımının yöntemsel katkıları kadar, ampirik olarak da anlamlı bulgular üretmesi beklenmektedir.



ÜÇÜNCÜ BÖLÜM

KULLANILAN ARAÇLAR

3.1. PANDAS

Pandas, Python programlama dili için geliştirilmiş, güçlü ve esnek veri yapıları sağlayan açık kaynaklı bir veri analiz kütüphanesidir. Pandas kütüphanesi temel olarak etiketli (*label-based*) eksenlere sahip veri yapıları sunarak, özellikle zaman serileri ve tabular (iki boyutlu) veriler üzerinde yüksek seviyede veri manipülasyonu ve analiz imkanı sağlar. Veri bilimi, istatistik, sosyal bilimler ve finans gibi farklı disiplinlerde yaygın olarak kullanılan bu kütüphane, Python'un bilimsel ekosistemini tamamlayan temel yapı taşlarından biri haline gelmiştir. Pandas ismi, çok boyutlu veri kümelerini ifade eden “panel data” kavramından türetilmiştir.

Kütüphanenin sunduğu temel veri yapıları olan Series (tek boyutlu veri) ve DataFrame (iki boyutlu veri), hem R programlama dilindeki “data.frame” sınıfına benzer şekilde çalışmakta hem de onu birçok yönüyle aşmaktadır. Özellikle eksik verilerin işlenmesi, otomatik veri hizalama (*alignment*), veri seçimi ve etiketleme konularında sağladığı programlama arayüzü (*application programming interface – API*) sayesinde, Pandas kullanıcı dostu bir analiz ortamı sunmaktadır. Pandas, farklı veri türlerini aynı yapı içinde barındırabilen sütun temelli bir yapıda çalıştığı için, heterojen veri kümeleriyle esnek bir biçimde çalışmaya imkan tanır. Ayrıca zaman serileri üzerinde frekans dönüşümü, kaydırma (*shifting*) ve ileri/geri doldurma (*forward/backward filling*) gibi işlemler kolaylıkla gerçekleştirilebilir. Pandas, CSV, Excel, Text ve daha birçok formatta veriyi okumakta ve yazmakta kullanılabilen bir kütüphanedir (McKinney, 2011).

Bu çalışmada, Pandas kütüphanesi oluşturulan veri setinin ön işleme sürecinde yoğun biçimde kullanılmıştır. Elde edilen haber metinlerin veri çerçeveleri (DataFrame) haline getirilmesi, tarih, başlık ve içerik sütunlarının temizlenmesi, dil filtrelemesi, gereksiz karakterlerin kaldırılması ve eksik verilerin ayıklanması gibi işlemler Pandas'ın sunduğu fonksiyonlar aracılığıyla gerçekleştirilmiştir. Ayrıca “groupby”, “dropna”, “fillna”, “merge”, “pivot_table” gibi işlevler veri kümesi üzerinde analiz yapılmasını mümkün kılmıştır. Şekil 4'te basit bir örnek ile Pandas kütüphanesinin çalışma prensibi gösterilmiştir.

Şekil 4: Pandas Kullanımı

```
import pandas as pd
# Örnek veri sözlüğü
veri = {
    "öğrenci": ["Ali", "Ayşe", "Mehmet", "Zeynep", "Ahmet", "Osman",
               "Emre", "Fatma", "Meryem", "Ela"],
    "ders": ["Matematik", "Matematik", "Fizik", "Fizik", "Matematik",
            "Matematik", "Matematik", "Fizik", "Fizik", "Matematik"],
    "puan": [85, 90, 78, 92, 88, 35, 44, 60, 69, 20]
}
# Matematik dersi alan öğrencilerin filtrelenmesi
df = pd.DataFrame(veri)

# Matematik dersi alan öğrencilerin filtrelenmesi
matematik_df = df[df["ders"] == "Matematik"]

# Derse göre ortalama puanların hesaplanması
ortalama_puanlar = df.groupby("ders")["puan"].mean()

# Tüm verinin ekrana yazdırılması
print(df, end="\n\n")
# Matematik dersi alan öğrencilerin ekrana yazdırılması
print(matematik_df, end="\n\n")
# Ortalama puanların ekrana yazdırılması
print(ortalama_puanlar, end="\n\n")
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 4'teki kod bloğunun çıktısı Şekil 5'te gösterilmiştir.

Şekil 5: Pandas Kodu Çıktısı

```
# Tüm veri
öğrenci      ders  puan
0      Ali  Matematik   85
1     Ayşe  Matematik   90
2   Mehmet    Fizik    78
3   Zeynep    Fizik    92
4    Ahmet  Matematik   88
5    Osman  Matematik   35
6     Emre  Matematik   44
7    Fatma    Fizik    60
8   Meryem    Fizik    69
9      Ela  Matematik   20
# Matematik dersi alan öğrenciler
öğrenci      ders  puan
0      Ali  Matematik   85
1     Ayşe  Matematik   90
4    Ahmet  Matematik   88
```

5	Osman	Matematik	35
6	Emre	Matematik	44
9	Ela	Matematik	20

```
# Ortalama Puanlar
ders
Fizik          74.750000
Matematik      60.333333
Name: puan, dtype: float64
```

Kaynak: Yazar tarafından derlenmiştir.

3.2. MATPLOTLIB

Matplotlib, bilimsel, mühendislik ve finansal verilerin görselleştirilmesine olanak tanıyan, açık kaynaklı bir Python kütüphanesidir. Kütüphane, yüksek kaliteli iki boyutlu grafiklerin yalnızca birkaç satır kod ile üretilebilmesine olanak sağlayacak şekilde tasarlanmıştır. Kolay kullanım imkanı sağlayan arayüzü sayesinde kütüphane ile Python komut satırından etkileşimli biçimde grafikler oluşturulabilir. Matplotlib aynı zamanda grafikleri grafik kullanıcı arayüzlerine gömülü olarak da üretilebilmektedir. Matplotlib'in esnek yapısı, yayın kalitesinde çıktılar alınmasına olanak tanıyan vektörel ve raster formatları desteklemesiyle öne çıkmaktadır. JPEG, PNG, PDF, SVG gibi farklı formatlara çıktı alma özelliği kenar yumuşatma (*antialiasing*), şeffaflık (*alpha blending*), yazı tipi yönetimi, matematiksel notasyon desteği ve eksen etiketleme gibi özelliklerle bütünsel bir yapı sunmaktadır (Barrett ve diğerleri, 2005).

Matplotlib, farklı veri türlerine yönelik geniş bir görselleştirme yelpazesi sunmaktadır. Kütüphane ile zaman serisi grafikleri, çubuk grafikler, çizgi grafikler, histogramlar, pasta grafikler, kutu grafikleri, saçılım grafikleri ve alan grafikleri gibi temel görselleştirme türlerinin yanı sıra; karmaşık çoklu eksen yapıları, finansal veri grafikleri, hata çubukları, yoğunluk dağılımları gibi gelişmiş görselleştirmeler de oluşturulabilmektedir. Ayrıca, eksen etiketleme, başlık ekleme, yazı tipi özelleştirme, renk ve çizgi stillerini değiştirme gibi seçeneklerle grafiklerin görsel ve işlevsel yönleri kolaylıkla geliştirilebilmektedir. Matplotlib, özellikle pandas, NumPy ve SciPy gibi bilimsel Python kütüphaneleriyle uyum içinde çalışmaktadır.

Bu çalışmada matplotlib, pandas ile işlenen verilerin görselleştirilmesinde kullanılmıştır. Görselleştirme, yapılan analizlerin anlaşılmasını kolaylaştırmakta ve çıkarımların desteklenmesine katkı sunmaktadır. Şekil 6'da, Şekil 4'te pandas ile

oluşturulan basit veri çerçevesinden derslere göre ortalama puanları çubuk grafik şeklinde görselleştiren bir kod parçası örneği bulunmaktadır:

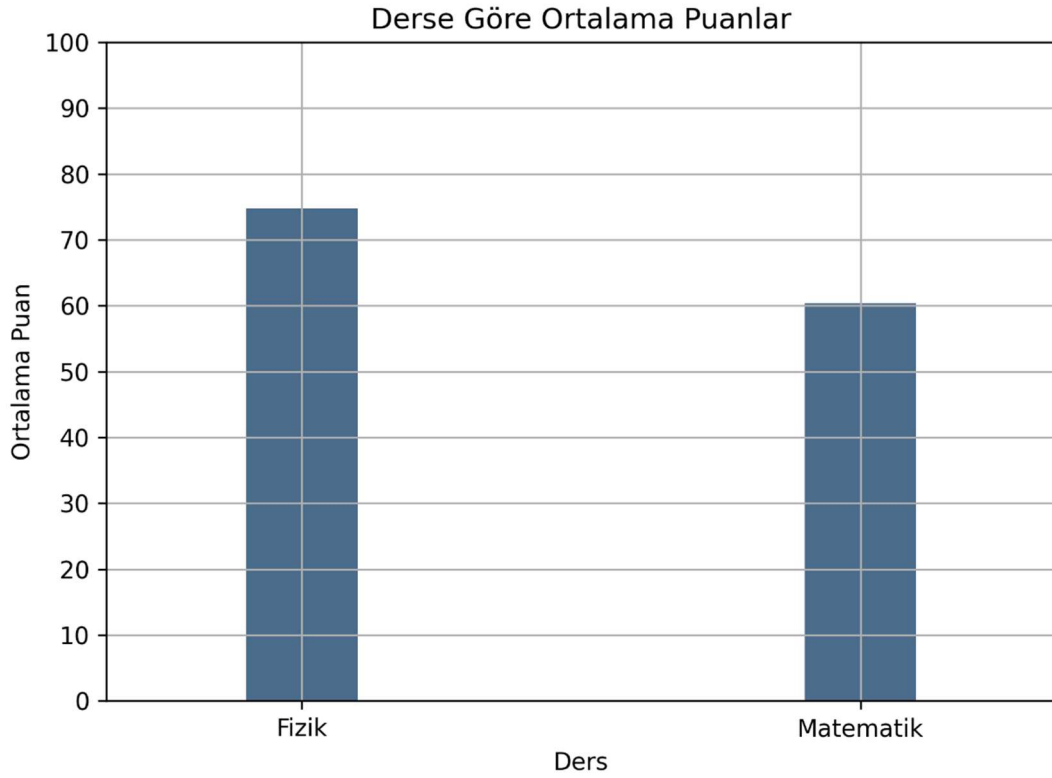
Şekil 6: Matplotlib Kullanımı

```
import matplotlib.pyplot as plt
ortalama_puanlar.plot(kind="bar", title="Derse Göre Ortalama Puanlar")
plt.xlabel("Ders")
plt.ylabel("Ortalama Puan")
plt.tight_layout()
plt.show()
```

Kaynak: Yazar tarafından derlenmiştir.

Şekil 6'daki kod bloğunun çıktısı Şekil 7'de gösterilmiştir.

Şekil 7: Matplotlib Kodu Çıktısı



Kaynak: Yazar tarafından derlenmiştir.

3.3. MEDIACLOUD

Media Cloud, küresel medyayı internette bulunan açık kaynaklar üzerinden analiz etmeye yönelik geliştirilmiş, açık kaynaklı bir haber toplama ve analiz platformudur. 2007 yılında Harvard Üniversitesi bünyesindeki Berkman Klein Center tarafından başlatılan bu girişim, haber üretimi, yayılımı, temsil ve etki gibi alanlarda araştırmalar yürütmeyi hedefleyen araştırmacılar için kapsamlı bir veri setine erişim sunmaktadır. Media Cloud telif hakları nedeni ile haber metinlerinin kendilerine değil, meta veri olarak adlandırılabilen, URL, yazar, medya organı adı, tarih gibi bilgilerine erişim imkanı sağlamaktadır. Platform, 2008'den bu yana toplanmış bir milyardan üzerinde haber içeriğinin meta verisini barındırmaktadır. Platformun veri toplama süreci iki temel bileşenden oluşmaktadır; düzenli olarak haber içeriği çeken RSS benzeri araçlar ve belirli bir konuyu derinlemesine incelemek amacıyla kullanılan ve konu haritalayıcı (*topic mapper*) olarak adlandırılan konu temelli bağlantı tarayıcısı. İlk bileşen yaklaşık 60 bin medya kaynağından günlük ortalama bir milyon haber toplamaktadır. İkinci bileşen olan ise belirli bir anahtar kelime setiyle başlatılan haberleri kategorize etmektedir. Bu yapı sayesinde Media Cloud kullanıcılara, konu temelli ve tarihsel temelli arama yapma olanağı sunmaktadır (Roberts ve diğerleri, 2021).

Media Cloud veri tabanına kullanıcı arayüzü ile web tabanlı erişimin yanı sıra REST tabanlı API üzerinden de ulaşılabilmektedir. Platforma ücretsiz bir biçimde üye olunduktan sonra alınacak API anahtarı ile örneğin python requests benzeri bir kütüphane kullanılarak API çağrıları yapılabilmektedir. Daha kolay veri çekebilmek amacı ile Github'da bulunan python istemcisi de kullanılabilir. API veri çekme ve basit filtreleme gibi temel işlevlerin yanı sıra, medya kimliği, yani hangi medya organına ya da koleksiyona ulaşmak istendiği, yayın tarihi, etiket, dil gibi çeşitli parametrelerle ile filtreleme yapmaya uygun olarak dizayn edilmiştir. Bunlara ilaveten istenirse indekslenen haberlerin sadece başlıklarında da arama yapmaya izin veren API, platform tarafından derlenmiş coğrafi çevrimiçi haber koleksiyonları barındırmaktadır. Bu haber koleksiyonları, hemen her ülke için ulusal ve yerel düzeyde yayın yapan medya organlarından oluşmaktadır. Örneğin bu çalışma için de kullanılan "Turkey - National" isimli koleksiyonda 113 haber kaynağı bulunmaktadır. Bu koleksiyonda taranan haber sayısı haftalık ortalama 639'dur. Tablo 4'te "Turkey - National" coğrafi kategorisinde Nisan 2025 itibarı ile platformda en çok haber içeriği bulunan ilk yirmi haber kaynağının isimleri ve taranan haftalık haber sayısı bulunmaktadır.

Tablo 4: Media Cloud Turkey National Kategorisi Haber Sayıları

Haber Kaynağı	Haftalık Haber	Haber Kaynağı URL'si
haberturk	4206	http://www.haberturk.com/
Haber Port	3236	http://www.haberport.com/
yenicaggazetesi.com.tr	2781	http://yenicaggazetesi.com.tr/
aa	2297	http://www.aa.com.tr/
gazetebirlik.com	2200	http://gazetebirlik.com/
cumhuriyet.com.tr	2192	http://cumhuriyet.com.tr/
aksam.com.tr	1912	http://aksam.com.tr/
ntv.com.tr	1881	http://ntv.com.tr/
gunboyugazetesi.com.tr	1788	http://gunboyugazetesi.com.tr/
takvim.com.tr	1713	http://takvim.com.tr/
stargazete	1595	http://www.stargazete.com
milliyet	1557	http://www.milliyet.com.tr
sabah	1525	http://www.sabah.com.tr/
hurriyet.com.tr	1400	http://hurriyet.com.tr/
aydinlik.com.tr	1391	http://aydinlik.com.tr/
sozcu.com.tr	1218	http://sozcu.com.tr/
evrensel	936	http://www.evrensel.net/
Ozgur Kocaeli	894	http://www.ozgurkocaeli.com.tr/
milligazete.com.tr	851	http://milligazete.com.tr/
yenisafak	824	http://www.yenisafak.com.tr/

Kaynak: Yazar tarafından <https://search.mediacloud.org/collections/34412131> adresinden derlenmiştir.

Bu çalışmada Media Cloud API, Türkiye'deki medya organlarının Afrika'ya ilişkin haber çerçevelerini incelemek amacıyla veri toplama aracı olarak python istemcisi aracılığı ile kullanılmıştır. Arama, Media Cloud'un sunduğu "Turkey - National" isimli coğrafi koleksiyon kapsamında gerçekleştirilmiş, böylece yalnızca Türkiye merkezli ulusal medya kuruluşlarının haber içeriklerine erişim sağlanmıştır. Verinin toplandığı sırada mevcut en geniş tarih aralığı olan Haziran 2020 – Nisan 2025 kullanılarak, platformda ilgili koleksiyonda yer alan tüm veriler taranmıştır. Arama sonucunda elde edilecek haberlerin doğrudan Afrika ve Afrika ülkeleri ile ilgili olma olasılığını maksimum hale getirmek için, aramalar yalnızca haber başlıklarında gerçekleştirilmiştir. Bu kapsamda "Afrika" kelimesine ek olarak Afrika kıtasını oluşturan 54 ülkenin her birinin adı ayrı birer anahtar kelime (*keyword*) olarak sorgu dizisine dahil edilmiştir. Aşağıda yer alan Şekil 8 ve Şekil 9'da örnek bir Media Cloud

API çağrısı kodu ve bu çağrının döndürdüğü python sözlüğü formatında sonuç gösterilmektedir.

Şekil 8: Media Cloud API Çağrısı

```
import os, mediacloud.api
from importlib.metadata import version
import datetime as dt
from IPython.display import JSON
MC_API_KEY = os.getenv("MC_API_KEY")
search_api = mediacloud.api.SearchApi(MC_API_KEY)
f'Using Media Cloud python client v{version("mediacloud")}'

dates = {
    'start_date' : dt.date(2021, 1, 1),
    'end_date' : dt.date(2021, 1, 5)
}

turkey_national_code = "34412131"

query_keyword = "Afrika"
my_query = f"article_title:{query_keyword}"

all_stories = []
more_stories = True
pagination_token = None
while more_stories:
    page, pagination_token = search_api.story_list(
my_query,
dates['start_date'],
dates['end_date'],
collection_ids=[turkey_national_code],
pagination_token=pagination_token
    )
    all_stories += page
    more_stories = pagination_token is not None

print(all_stories[:3])
```

Kaynak: Yazar tarafından derlenmiştir.

Yukarıdaki kod bloğunun çıktısı Şekil 9'da verilmiştir.

Şekil 9: Media Cloud API Çağrısı Kod Çıktısı

```
[{'id':  
'd8e4b1739cbc595c63a9ae5ad038bf312aa5c8817be3c0702ca025baadf8e633',  
'indexed_date': datetime.datetime(2024, 8, 17, 16, 52, 38, 769774,  
tzo=datetime.timezone.utc),  
'language': 'tr',  
'media_name': 'trthaber.com',  
'media_url': 'trthaber.com',  
'publish_date': datetime.date(2021, 1, 1),  
'title': "Afrika 2020'de terör ve iç çatışma sarmalından çıkamadı",  
'url': 'http://www.trthaber.com/haber/dunya/afrika-2020de-teror-ve-ic-  
catisma-sarmalindan-cikamadi-543110.html'},  
{'id':  
'ab7b17c2f3a2b92b445c4712c75873041912b79e7444dc622937610a2ae035e2',  
'indexed_date': datetime.datetime(2024, 8, 17, 16, 51, 46, 84291,  
tzo=datetime.timezone.utc),  
'language': 'tr',  
'media_name': 'aa.com.tr',  
'media_url': 'aa.com.tr',  
'publish_date': datetime.date(2021, 1, 1),  
'title': "Güney Afrika Cumhuriyeti'nde Kovid-19 vaka sayısı 1 milyon  
57 bini geçti",  
'url': 'https://www.aa.com.tr/tr/guncel/guney-afrika-cumhuriyetinde-  
kovid-19-vaka-sayisi-1-milyon-57-bini-gecti/2095511'},  
{'id':  
'fa18f142683bb1ebdeaf7df173aa5763e7f8f70d6705a43e7fa29a85ca016d40',  
'indexed_date': datetime.datetime(2024, 8, 17, 15, 26, 55, 585917,  
tzo=datetime.timezone.utc),  
'language': 'tr',  
'media_name': 'aa.com.tr',  
'media_url': 'aa.com.tr',  
'publish_date': datetime.date(2021, 1, 2),  
'title': 'Afrika kıtasında Kovid-19 vakası sayısı 2 milyon 800 bini  
geçti',  
'url': 'https://www.aa.com.tr/tr/guncel/afrika-kitasinda-kovid-19-  
vakasi-sayisi-2-milyon-800-bini-gecti/2096372'}
```

Kaynak: Yazar tarafından derlenmiştir.

3.4. BEAUTIFUL SOUP

Beautiful Soup, HTML ve XML belgeleri üzerinde çalışmak üzere tasarlanmış, Python tabanlı bir bilgi çıkarım (*information extraction*) kütüphanesidir. Kütüphane, yarı düzenli ya da düzensiz yapıdaki web sayfalarından metin tabanlı bilgi çekmek amacıyla geliştirilmiştir. Temel olarak kütüphane, bir web sayfasının HTML yapısını bir DOM (*Document Object Model*) ağacına dönüştürmekte ve bu yapı üzerinde gezinme, obje seçme ve metin ayıklama işlemleri yapılmasına imkan tanımaktadır. Kütüphane kullanıcının belirli HTML etiketlerine ya da sınıflara göre içerik aramasına ve filtrelemesine olanak sağlamaktadır.

Beautiful Soup kütüphanesinin en önemli avantajı, hedeflenen veri noktalarına doğrudan ulaşılmasına olanak tanıyan yapısıdır. Bu yönüyle, özellikle yapısal HTML şablonlarının tekrar eden alanlarında, örneğin başlık, tarih, kaynak, paragraf, tablo gibi alanların hızlı ve hassas biçimde çekilmesinde oldukça etkilidir. Kütüphane DOM ağacında gezinirken derinlik veya genişlik öncelikli arama yöntemlerini benimsemeye, HTML elementlerine ait öznitelikler üzerinden filtreleme yapmaya ve içerik üzerinde ön işleme adımlarını entegre biçimde gerçekleştirmeye olanak sağlamaktadır (Zheng ve diğerleri, 2015).

Bu çalışmada Beautiful Soup, Media Cloud platformundan elde edilen haber bağlantılarının HTML içeriklerinden, haber metinlerini çıkarmak amacıyla kullanılmıştır. Media Cloud yalnızca haberin bağlantı bilgisini sağladığından, tam metin erişimi Beautiful Soup kullanılarak haber sayfalarından manuel biçimde sağlanmıştır. Bu işlem sırasında haber sayfalarının HTML yapıları incelenmiş, başlık, tarih ve içerik gibi öğeleri temsil eden HTML etiketleri belirlenmiş ve bu alanlara özgü veri çıkarımı gerçekleştirilmiştir. Özellikle sayfa yapılarının birbirinden farklı olması nedeniyle, her medya kaynağı için özel kurallar tanımlanmış, böylece otomatikleştirilebilir ama esnek bir veri çekme modülü oluşturulmuştur. Bu süreç, sonrasında uygulanacak konu modelleme ve çerçeve analizi işlemleri için gerekli olan temiz ve tutarlı bir derlem elde edilmesini mümkün kılmıştır. Şekil 10'da Beautiful Soup kütüphanesinin çalışma mantığı kod bloğu ile gösterilmiştir.

Şekil 10: BeautifulSoup Kütüphanesi Kullanımı

```
from bs4 import BeautifulSoup

html = """
<html>
  <body>
    <div class="haber">
      <h2 class="baslik">Türkiye'den Afrika'ya Yeni Yardım Paketi</h2>
      <span class="tarih">2023-06-21</span>
      <p class="icerik">TİKA tarafından Somali'de gerçekleştirilen
yardım operasyonu genişletiliyor.</p>
    </div>
  </body>
</html>
"""

soup = BeautifulSoup(html, "html.parser")

baslik = soup.find("h2", class_="baslik").text
tarih = soup.find("span", class_="tarih").text
icerik = soup.find("p", class_="icerik").text

print("Başlık:", baslik)
print("Tarih:", tarih)
print("İçerik:", icerik)

Başlık: Türkiye'den Afrika'ya Yeni Yardım Paketi
Tarih: 2023-06-21
İçerik: TİKA tarafından Somali'de gerçekleştirilen yardım operasyonu
genişletiliyor.
```

Kaynak: Yazar tarafından derlenmiştir.

3.5. BERTOPIC

BERTopic, metin verileri üzerinde konu modelleme gerçekleştirmek amacıyla geliştirilen modern bir tekniktir. BERTopic derin öğrenme tabanlı bir dil modeli olan BERT (*Bidirectional Encoder Representations from Transformers - Dönüştürücü Tabanlı Çift Yönlü Kodlayıcı Temsilleri*) üzerine inşa edilmiştir. BERT DDİ görevlerinde yüksek doğruluk sağlayan bir modeldir. Bu mimaride “çift yönlü – *bidirectional*” ifadesi, modelin hem sol hem de sağ bağlamı aynı anda dikkate alarak her kelimeyi çift yönlü olarak işlemesini ifade eder. Bu sayede kelimeler yalnızca önceki kelimelere değil, sonraki kelimelere göre de anlamlandırılır. “Kodlayıcı temsiller - *encoder representations*” ifadesi ise, dönüştürücü mimarisinin yalnızca

kodlayıcı kısmının kullanıldığını belirtmektedir, bu kısım girişteki tüm kelimeler arasındaki bağlam ilişkilerini çok katmanlı dikkat mekanizmaları yoluyla işler ve her bir kelimeye ait bağlamsal temsil vektörlerini üretir. “dönüştürücü” ise, saf dikkat mekanizması kullanan yenilikçi mimarının adıdır. BERT, bu yapıyı kullanarak maskelenmiş dil modeli ve sonraki cümleyi tahmin etme görevleriyle ön eğitimden geçirilmekte ve bu sayede çeşitli dil görevleri için güçlü temsiller öğrenmektedir (Devlin ve diğerleri, 2019).

Tablo 5: Bert Mimarisi

Bidirectional Encoder Representations from Transformers	Dönüştürücü Tabanlı Çift Yönlü Kodlayıcı Temsilleri
Bidirectional	Çift yönlü (Metindeki her kelimeyi hem önceki hem de sonraki bağlama göre işler.)
Encoder	Kodlayıcı (Dönüştürücü mimarisinin girdileri işleyen, yani “anamlandıran” kısmı)
Representations	Temsiller (Her kelimenin bağlamsal sayısal gösterimleri / vektör temsilleri.)
from Transformers	Dönüştürücülerden (Temsiller dönüştürücülerden elde edilir)

Kaynak: Yazar tarafından derlenmiştir.

Bu mimari kullanılarak geliştirilen BERTopic (*Bidirectional Encoder Representations from Transformers for Topic Modeling*), BERT’in sunduğu bağlamsal anlam yakalama gücünü konu modelleme alanına entegre eden bir yaklaşımdır. Klasik LDA gibi yaklaşımların kelime sıklığına dayalı basitleştirici varsayımlarının ötesine geçe BERTopic, metinleri sadece kelime sıklıklarına göre değil, derin anlamsal yapısına göre gruplandırarak konu başlıklarını belirlemektedir. Bu teknik, sosyal medya içerikleri (Nanayakkara ve diğerleri, 2024), haber metinleri (Al-Zaman ve Rashid, 2025), akademik yayınlar (Madrid-García ve diğerleri, 2024) ve müşteri geri bildirimleri (Koruyan, 2022) gibi çok çeşitli metin veri kümeleri üzerinde konu tespiti ve tematik analizinde güçlü bir araç olarak öne çıkmaktadır. BERTopic’in esnek ve modüler yapısı sayesinde, farklı gömme, boyut indirgeme veya kümeleme teknikleri vasıtasıyla özelleştirme yapılarak üzerinde çalışılan veri setlerine uygun modeller geliştirilebilmektedir.

Bu yöntemde belgeler önce SBERT (*Sentence BERT*) gibi bir model ile sayısal temsillere dönüştürülür; ardından bu temsiller UMAP gibi boyut indirgeme algoritmaları ile düşük boyutlara indirgenir ve HDBSCAN gibi yoğunluk tabanlı bir kümeleme algoritması ile konu kümeleri elde edilir. Son olarak her küme için temsil edici anahtar kelimeler, CountVectorizer ve c-TF-IDF (*Class-based Term Frequency–Inverse Document Frequency* - Sınıf Tabanlı Terim Frekansı – Ters Belge Frekansı) yöntemleri kullanılarak seçilir. Bu çok aşamalı yapı sayesinde BERTopic, özellikle kısa ve bağlamsal açıdan zengin metinlerde, geleneksel yaklaşımlara kıyasla daha anlamlı ve yorumlanabilir konu grupları üretmektedir (Uğuz ve Tülü, 2024).

Tablo 6: BERTopic Mimarisi

Dokümanlar		
SBERT (Sentence BERT)	↓	Gömme (Vektörleştirme)
UMAP	↓	Boyut Azaltma
HDBSCAN	↓	Kümeleme
CountVectorizer	↓	Terim frekanslarının çıkarılması (BoW)
c-TF-IDF	↓	Ağırlıklandırma
Konu Temsilleri		

Kaynak: <https://maartengr.github.io/BERTopic/algorithm/algorithm.html#visual-overview> adresinden yazar tarafından derlenmiştir.

3.5.1. Sentence-BERT: Gömme/Vektörleştirme

BERTopic yönteminin ilk adımını, her bir metnin sayısal bir vektör temsiline dönüştürülmesi oluşturmaktadır. Bu amaçla kullanılan SBERT cümleleri bağlamlarına göre anlamca temsil eden bir tekniktir. SBERT, BERT temel alınarak geliştirilmiştir ve her cümleyi sabit uzunlukta, anlamı yansıtan, yani cümle düzeyinde semantik ilişkileri koruyarak bir vektörle dönüştürür. Bu vektörler, cümlelerin birbirine ne kadar benzediğini ölçmek için doğrudan kosinüs benzerliği gibi yöntemlerle karşılaştırılmaya elverişlidir (Mandelbaum ve Shalev, 2016; Reimers ve Gurevych, 2019). Yapılan araştırmalarda, SBERT'in, önceden eğitilmiş BERT ağına göre cümle benzerliği ve

metin kümelenmesi gibi görevlerde hem doğruluk hem de işlem süresi açısından üstün performans gösterdiği ortaya konmuştur (Choi ve diğerleri, 2021)

Bu adımda SBERT ile oluşturulan gömmeler derlem içerisindeki her bir dokümanı çok boyutlu bir vektör uzayında semantik olarak temsil eder. Bu temsiller, sonraki aşamalarda kullanılacak olan boyut indirgeme, kümeleme ve konu çıkarımı süreçlerinin yapı taşı olarak kullanılmaktadır.

3.5.2. UMAP: Boyut İndirgeme

SBERT tarafından oluşturulan vektör temsilleri çok boyutlu olduğu için, bu vektörlerin hem görselleştirilebilir hem de kümelenebilir bir yapıya dönüştürülmesi amacı doğrultusunda çok boyutlu olan vektörlerin boyutlarının azaltılması gerekmektedir. Boyut azaltma için BERTopic modelinde UMAP kullanılmaktadır. UMAP, yüksek boyutlu verileri, içerdikleri yapısal ilişkileri büyük ölçüde koruyarak daha düşük boyutlara indirgemeye yarayan bir algoritmadır. Özellikle SBERT'in ürettiği bağlamsal vektörlerin iki veya üç boyuta indirgenmesiyle, belgeler arasındaki anlamsal benzerlikler daha görünür hale gelir ve kümeleme süreci daha verimli, kararlı ve hızlı bir şekilde gerçekleştirilebilir (McInnes ve diğerleri, 2020).

3.5.3. HDBSCAN: Kümeleme

BERTopic konu modelleme tekniğinde boyut indirgeme işlemi sonrasında elde edilen vektör temsilleri, belgelerin içerik açısından benzerliklerine göre gruplandırılmak üzere HDBSCAN algoritmasına tabi tutulmaktadır. HDBSCAN yoğunluk tabanlı klasik kümeleme yöntemlerinden farklı olarak, küme sayısının önceden belirlenmesine ihtiyaç duymadan farklı yoğunluk düzeylerine sahip kümeleri ayırt edebilir. Bu özelliği sayesinde, algoritma farklı uzunlukta ve içerikteki metinlerin oluşturduğu dengesiz veri kümelerinde yüksek performans göstermektedir. HDBSCAN, her bir belgeyi çevresindeki yoğunluğa göre değerlendirir ve yoğunluğu belirli bir eşik altında olan belgeleri “gürültü” (*noise*) olarak tanımlar, böylece anlamlı kümelerden sapmalar sistematik biçimde ayırt edilebilir.

Bu algoritmada kümeleme işlemi veri noktalarının oluşturduğu yoğunluk yapısına göre hiyerarşik bir ağaç yapısında gerçekleştirilir. Bu yapı içerisinde, yalnızca anlamlı ve kararlı kümelerin çıkarılması için küme kararlılığı adlı bir ölçüt kullanılır. Bu ölçüt, bir kümenin farklı yoğunluk seviyelerinde ne kadar “hayatta

kaldığını” göstererek, istikrarlı kümelerin öne çıkarılmasını sağlar BERTopic bu algoritma yardımı ile sayede, anlamsal olarak benzer haber başlıklarını bir araya getirirken, içeriği yeterince yoğun olmayan veya benzersiz olan başlıkları otomatik olarak aykırı olarak tanımlar. Bu da daha temiz ve yorumlanabilir bir konu kümeleme yapısı oluşturulmasına katkı sağlamaktadır (Campello ve diğerleri, 2013).

3.5.4. CountVectorizer: Terim Frekanslarının Çıkarılması (BoW)

Kümeleme işlemi sonucunda her bir konu kümesine ait belgeler belirlendikten sonra, bu belgelerde geçen kelimelerin istatistiksel özelliklerinin çıkarılması gerekmektedir. Bu amaçla BERTopic yönteminde CountVectorizer kullanılmaktadır. CountVectorizer, her belgeyi kelime frekanslarına dayalı olarak sayısal bir vektöre dönüştüren klasik bir BoW yöntemidir. Bu yöntemde, her bir belge içinde geçen kelimeler sayılır, ardından elde edilen bu sayılar, belgelerin içerdiği kelime yoğunluğunu gösteren bir matris oluşturur. Oluşturulan matrisin satırları belgeleri, sütunları ise kelimeleri temsil eder ve her hücrede o kelimenin ilgili belgede kaç kez geçtiği yer alır.

BERTopic bağlamında CountVectorizer, her bir konuya ait belge kümesinden o konuya özgü kelime dağılımını çıkarmak için kullanılmaktadır. Bu işlem, daha sonra uygulanacak olan c-TF-IDF yöntemi için gerekli olan ham veriyi sağlar. Bu aşamada, duraklama kelimelerinin filtrelenmesi, minimum belge sıklığına göre kelime seçimleri gibi ön işlemler de genellikle CountVectorizer aracılığıyla gerçekleştirilmektedir.

3.5.5. C-TF-IDF: Ağırlıklandırma

BERTopic yönteminin önemli aşamalarından biri olan konu temsili oluşturma süreci, her bir küme için ayırt edici nitelikteki terimlerin seçilmesi vasıtasıyla gerçekleştirilir. Bir başka ifade ile kümelenen her bir konu için bu konuyu en iyi temsil eden kelimeler algoritmik olarak belirlenmektedir. Bu amaçla, klasik TF-IDF yöntemine dayanan ancak onun belge düzeyindeki uygulamasını küme düzeyine genişleten bir teknik olan c-TF-IDF kullanılır. Bu yöntem, her bir konu kümesini tek bir “sınıf” olarak değerlendirir ve o sınıfa özgü kelimeleri belirlemek amacıyla özel bir TF-IDF hesaplaması yapar. Yani, her küme içindeki tüm belgeler birleştirilerek büyük bir metin haline getirilir ve ardından, bu metindeki kelimelerin diğer kümelerdeki kullanımlarıyla karşılaştırılarak her kelimeye bir ayırt edicilik puanı atanır.

c-TF-IDF yöntemi, klasik TF-IDF'e kıyasla daha konu merkezli bir değerlendirme yapmaktadır. Terim frekansı, belirli bir konudaki kullanım sıklığını temsil ederken, ters belge sıklığı, terimin diğer tüm konulara kıyasla ne kadar özel olduğunu gösterir. Bu sayede, yalnızca sık geçen değil, aynı zamanda bir kümeye özgü olan kelimeler öne çıkarılabilmektedir. BERTopic, bu kelimeleri her bir konu kümesini tanımlayan anahtar kelimeler olarak sunar ve sonuç olarak daha tutarlı, anlamlı ve birbirinden daha net ayrıışan konu temsilleri elde edilir. Konu temsilleri, konu modelleme tekniğinde elde edilen sonuçların açıklanabilirliğini ve yorumlanabilirliğini büyük ölçüde artırmaktadır (Grootendorst, 2022)

Tablo 7: BERTopic Konu Modelleme Süreci

Örnek Doküman: "Türkiye, son yıllarda Afrika kıtasına yönelik dış politikasında insani yardım odaklı bir yaklaşım benimsemiştir. TİKA ve Kızılay gibi kurumlar aracılığıyla Afrika'da birçok sağlık ve eğitim projeleri yürütülmektedir. Bununla birlikte ticaret hacmi de önemli ölçüde artış göstermiştir."		
Aşama	Açıklama	Örnek Sonuç
SBERT	Doküman 768 boyutlu vektöre dönüştürülmesi	[0.13, -0.41, 0.85, ..., -0.67]
UMAP	Vektörün boyutunun azaltılması (2 Boyut)	[1.22, -0.45]
HDBSCAN	Metnin kümeye atanması - Gürültü olup olmadığının tespiti	Küme 4
CountVectorizer	Metindeki kelimelerden frekans vektörü oluşturulması	{"türkiye": 1, "afrika": 2, "yardım": 1, "ticaret": 1, "proje": 1}
C-TF-IDF	Küme için ayırt edici terimlerin hesaplanması	afrika (2.6), insani (2.1), ticaret (1.7)
KeyBERT	Metin için en temsili kelimelerin seçimi	afrika, yardım, ticaret

Kaynak: Yazar tarafından derlenmiştir.

DÖRDÜNCÜ BÖLÜM

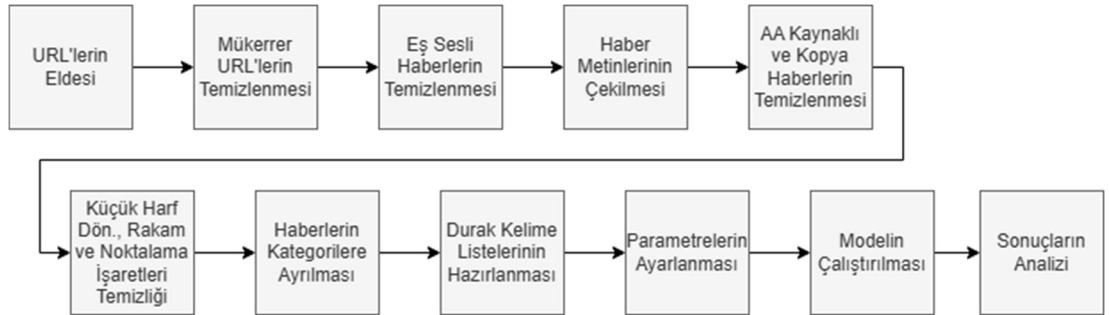
BERTOPIC İLE TÜRKİYE'DE YAYINLANAN AFRİKA HABERLERİNİN İNCELENMESİ

4.1. İŞ AKIŞI

Bu çalışma kapsamında araştırma sorusunun yanıtlanması için Türkiye’de yayınlanan Afrika ile ilgili haberlere BERTopic ile denetimsiz olarak konu modelleme uygulanmıştır. Bu uygulama python yazılım dilinin 3.10.7 versiyonu ile gerçekleştirilmiştir. Bir sanal ortam kurularak tüm kütüphane ve modüllerin ana python sürümünden izole edilmesi sağlanmıştır. Bu sanal ortam jupyter notebook versiyon 7.3.3 arayüzü ile kullanılmıştır. Kullanılan temel kütüphaneler, pandas versiyon 2.2.3, matplotlib versiyon 3.10.0, mediacloud versiyon 4.3.0, scikit-learn versiyon 1.5.2, beautifulsoup versiyon 4.12.3 ve bertopic versiyon 0.16.4’tür. Kullanılan tüm kütüphanelerin listesi Ek-2’de verilmiştir. Kullanılan tüm kütüphaneler “pip install ‘kütüphane ismi’” komutu ile yüklenebilir.

Analiz aşamasında temel iş akışı veri eldesi, veri ön işleme, haberlerin kategorilere ayrılması, modelin kurulması ve sonuçların değerlendirilmesi olarak özetlenebilir. İş akış şeması Şekil 11’de verilmiştir.

Şekil 11: İş Akış Şeması



Kaynak: Yazar tarafından derlenmiştir.

4.2. VERİ ELDESİ

4.2.1. URL Eldesi

Bu çalışmada Türkiye’de yayınlanan Afrika haberlerinin konu modelleme yöntemiyle analiz edilebilmesi amacıyla büyük ölçekli bir haber verisi derlenmiştir. Veri toplama sürecinde Afrika kıtasına ya da 54 Afrika ülkesinden birine yahut birkaçına doğrudan haber başlığında atıf yapan Türkçe haber içerikleri hedeflenmiştir. Haber başlıklarında arama yapılmasının sebebi, elde edilecek verinin doğrudan Afrika ile ilgili olma olasılığını maksimize etmektir. Bu kapsamda, Media Cloud API aracılığıyla, 54 Afrika ülkesinin ismine ek olarak Afrika anahtar kelimesi ile haber başlıklarında arama yapılarak 64 farklı medya kuruluşundan toplam 54.626 habere ait URL, tarih, başlık gibi meta veriye erişim sağlanmıştır. Bunun yanı sıra, Media Cloud tarafından taranmayan ancak dikkate değer sayıda Afrika ile ilgili haber yayınlayan haber kaynaklarından biri olan T24 de kendi API’si yardımıyla aynı anahtar kelimeler ve aynı tarih aralığı ile taranıp veri setine dahil edilmiştir. T24’ten elde edilen metaveri 4092 adettir. Toplanan toplam 58.718 ham meta veri 01.01.2020 ile 17.04.2025 arasını kapsamaktadır.

4.2.2. URL’lerde Veri Ön Temizliği

Haber meta verilerinin toplanması aşamasında kullanılan Media Cloud API’nin sözdizimi (*syntax*) kuralları sondan eklemeli ve fazla sayıda eş sesli kelime barındıran Türkçe ile beklenen uyumlulukta çalışmamaktadır. Bu nedenle toplanan ham verinin, mükerrer URL’lerden, eş sesli kelimelerden kaynaklı haberlerden ve İngilizce yazımı Türkçe ile aynı olan ülke isimleri nedeni ile arama sonucunda gelen İngilizce haberlerden temizlenmesi gerekmiştir. Bu temizlik, aynı URL’ye sahip satırlar için pandas kütüphanesinin `drop_duplicates()` fonksiyonu ile “url” sütununda gerçekleştirilmiştir. Eş sesli kelimeler için haber başlıklarını içeren “title” sütununda basit regex (*regular expressions*) aramaları ile veri temizliği yapılmıştır. Türkçe dışındaki dillerde gelen haber meta verileri için temizlik Media Cloud’dan dönen verinin “language” sütunu kullanılarak gerçekleştirilmiştir. Eş sesli kelimeler için yapılan temizliğin isabeti, her bir ülke ile ilgili haberlerden minimum %10 örneklem alınarak manuel olarak kontrol edilmiştir. Karışıklıklar çok spesifik olduğu için kontrol edilen veride ortalama 0,98 gibi yüksek bir oranla isabet sağlanmış, kontrol

aşamasında fark edilen satırlar da veriden düşürülmüştür. Eş sesli kelimeler içeren haber başlıklarına kimi örnekler Tablo 8’de verilmiştir.

Tablo 8: Eş Sesli Kelimeler İçeren Haber Başlıkları

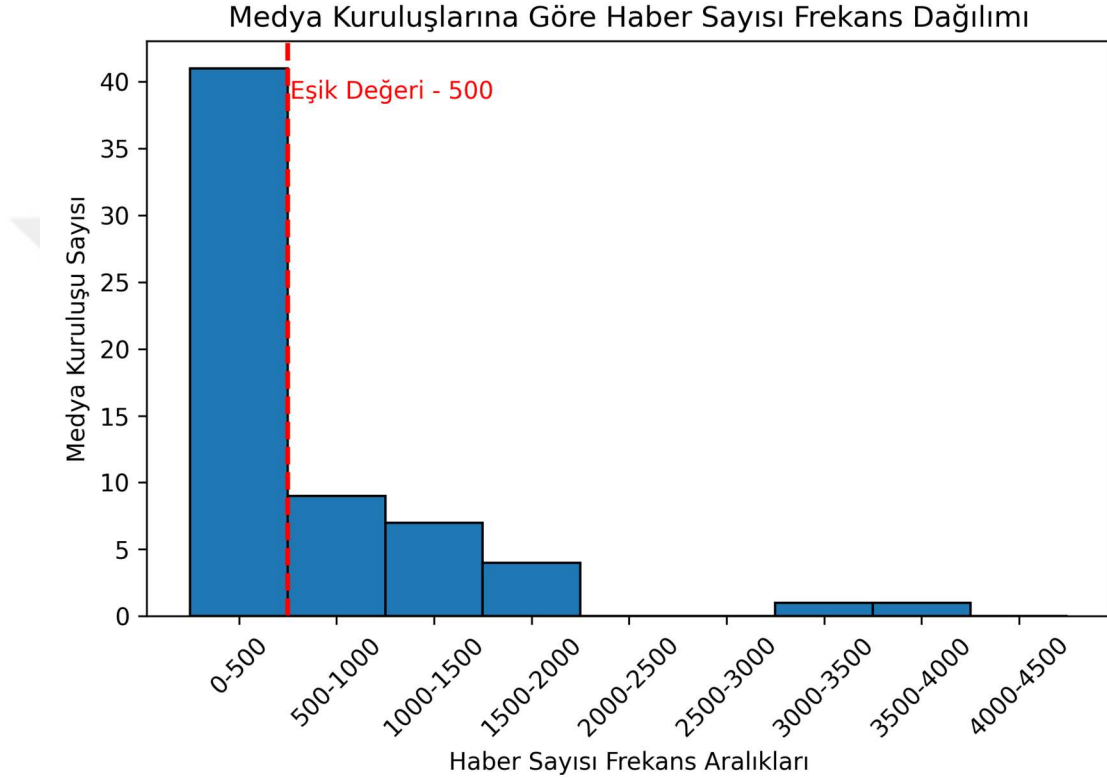
Kelime	Haber Başlığı
Sudan	“Deniz suyu sıcaklığı hava sıcaklığını 5 derece geçince, vatandaşlar <u>sudan</u> çıkmak istemedi”
	“Antalya ve Alanya’da <u>sudan</u> ucuza arsa satışı başladı: Kentsel Dönüşüm Başkanlığı’ndan büyük fırsat”
	“ <u>Sudan</u> sebeple kardeşi ve yengesini öldürdü”
Mali	“ <u>Mali</u> Müşavirler Gelir Vergisi Beyan Süresinin Uzatılmasını Talep Ediyor”
	“Yasada sona gelindi... Kulüp yöneticilerine <u>mali</u> sorumluluk”
	“Ford’dan Almanya’daki birimine dev <u>mali</u> destek!”
Mısır	“ <u>Mısır</u> Çarşısı’nda Haşereyle Mücadele: Fatih Belediyesi İlaçlama Yaptı”
	“Diyette <u>mısır</u> ekmeği yenir mi? Diyetle mısır ekmeği yemek kilo aldırır mı?”
	“Kontrolden çıkıp <u>mısır</u> tarlasına uçtu: 5 yaralı”
Togo	“Ankara Büyükşehir Belediyesi’nden <u>TOGO</u> Kuleleri için yıkım kararı”
Gine	“Papua Yeni <u>Gine</u> Petrol Bakanı Maladina’nın Aile İçerisinde Şiddet Suçlamaları Düşürüldü”
Cezayir	“Ben Bu Cihana Sığmazam 67. bölüm 2. fragmanı yayınlandı! <u>Cezayir</u> , Habip ve Pırıl’ın intikamını alacak!”

Kaynak: Yazar tarafından derlenmiştir.

Media Cloud veri tabanı, yukarıda da sözü edildiği üzere haber metinlerinin tam içeriğine erişim sağlamamaktadır. Bu durum, çalışma kapsamında haberlerin tam metinlerine erişilebilmesi amacıyla web kazıma (*scraping*) tekniklerinin kullanılmasını zorunlu kılmıştır. Elde edilen meta veri setinde toplam 64 farklı haber kaynağı bulunmakta olup, her bir haber kaynağının internet sitesine ait yapının birbirinden farklı olması, kazıma sürecinin teknik karmaşıklığını artırmakta ve her bir haber kaynağı için özel olarak yapılandırılmış ayrı kod blokları gerektirmektedir. Bununla

birlikte, bazı kaynakların çok sınırlı sayıda haber içermesi, bu kaynaklar için web kazıma işlemi yürütmenin emek, zaman ve çalışmaya katkısı açısından rasyonel bir tercih olmadığı düşünülmüştür. Bu bağlamda, kazıma işleminin kapsamını belirlemek üzere erişilen haber meta verilerinin medya kaynaklarına göre frekans dağılımı görselleştirilmiş (bkz. Şekil 12) ve analiz edilmiştir.

Şekil 12: Medya Kuruluşuna Göre Haber Sayıları Frekans Dağılımı



Kaynak: Yazar tarafından derlenmiştir.

Yapılan fayda-maliyet analizi neticesinde, yalnızca 500 ve üzeri haber içeriğine sahip olan medya organlarının kazıma işlemine dahil edilmesinin uygun olacağı değerlendirilmiştir. Bu doğrultuda web kazıma işleminin ilk aşamada elde edilen meta verinin mükerrer bağlantıya sahip, eş sesli, Türkçe olmayan ve 500 den daha az sayıda haber meta verisi sağlayan haber kaynaklarına ait satırlardan temizlenmesi sonucunda kalan satır sayısı 33.856'dır. Haberlerin meta verisinden oluşturulan veri setinin genel görünümü, en az 500 habere sahip medya kuruluşlarına ve ülkelere göre dağılımları sırasıyla Tablo 9, 10 ve 11'de verilmiştir.

Tablo 9: Haber Meta Veri Veri Seti Genel Görünümü

date	media_name	url	title	country
2021-10-22	ntv.com.tr	https://www.ntv.com.tr/turkiye/...	Cumhurbaşkanı Erdoğan: Hedefimiz Afrika ülkele...	afrika
2022-04-26	cumhuriyet.com.tr	https://www.cumhuriyet.com.tr/spor/....	Kevin N'Koudou'nun milli takım tercihi Kamerun	kamerun
2021-10-19	milliyet.com.tr	https://www.milliyet.com.tr/siyaset/....	Cumhurbaşkanı Erdoğan, Togo'da üç Afrika ülkes...	afrika
2020-01-21	turkiyegazetesi.com.tr	https://www.turkiyegazetesi.com.tr/gundem/....	Cumhurbaşkanı Erdoğan Afrika turuna çıkıyor	afrika
2024-12-19	aa.com.tr	https://www.aa.com.tr/tr/dunya/libya-b...	Libya Başbakanı, Suriye'den Rus kuvvetlerinin ...	libya
2022-09-03	dunya.com	https://www.dunya.com/dunya/misir-...	Mısır ve ABD'den Akdeniz'de ortak askeri tatbikat	misir
2023-07-22	gazetevatan.com	https://www.gazetevatan.com/gundem/...	Cumhurbaşkanı Erdoğan, Cezayir Cumhurbaşkanı T...	cezayir
2023-11-15	trthaber.com	https://www.trthaber.com/haber/dunya/ingilter...	İngiltere'de Yüksek Mahkeme, hükümetin "Ruanda...	ruanda

Kaynak: Yazar tarafından derlenmiştir.

Tablo 10: Medya Kuruluşuna Göre Haber Meta Verisi Dağılımı

Haber Kaynağı	URL Sayısı	Haber Kaynağı	URL Sayısı
aa.com.tr	4871	takvim.com.tr	1158
trthaber.com	3906	ntv.com.tr	1106
t24.com.tr	3078	yenicaggazetesi.com.tr	899
sabah.com.tr	1892	hurriyet.com.tr	879
milliyet.com.tr	1748	aydinlik.com.tr	847
gazetebirlik.com	1551	milligazete.com.tr	824
aksam.com.tr	1513	dunya.com	694
haberturk.com	1446	dirilispostasi.com	637
cumhuriyet.com.tr	1417	sozcu.com.tr	561
haberport.com	1339	evrensel.net	531
gunes.com	1266	yeniasya.com.tr	518
turkiyegazetesi.com.tr	1175		

Kaynak: Yazar tarafından derlenmiştir.

Tablo 11: Ülkelere Göre Haber Meta Verisi Dağılımı

Ülke Adı	URL Sayısı	Ülke Adı	URL Sayısı	Ülke Adı	URL Sayısı
Libya	6629	Orta Afrika Cumhuriyeti	351	Liberya	72
Mısır	4187	Güney Sudan	293	Namibya	62
Afrika	3931	Gana	242	Malavi	59
Güney Afrika Cumhuriyeti	2517	Çad	229	Burundi	55
Cezayir	1701	Kamerun	209	Togo	54
Sudan	1640	Uganda	198	Zambiya	53
Tunus	1388	Fildişi Sahili	190	Moritanya	51
Fas	1359	Tanzanya	187	Eritre	45
Somali	1011	Gine	185	Benin	39
Nijerya	919	Angola	160	Lesotho	32
Burkina Faso	846	Gine-Bissau	160	Botsvana	31
Etiyopya	815	Sierra Leone	142	Mauritius	25
Senegal	502	Gabon	129	Esvatini	24
Kongo Cumhuriyeti	482	Madagaskar	98	Komor Adaları	21
Demokratik Kongo cumhuriyeti	443	Ekvator Ginesi	93	Sao Tome ve Principe	20
Nijer	432	Zimbabve	88	Şeyseller	20
Ruanda	420	Gambiya	86	Yeşil Burun Adaları	15
Kenya	398	Cibuti	83		
Mali	360	Mozambik	75		

Kaynak: Yazar tarafından derlenmiştir.

4.2.3. Haber Metinlerinin Web Kazıma Yöntemi ile Eldesi

Veri eldesindeki bir sonraki aşama Tablo 10'da listelenen haber kaynaklarından haber metinlerinin web kazıma yolu ile çekilmesi olmuştur. Bu amaç doğrultusunda python programlama dilinin beautiful soup kütüphanesinden faydalanılmıştır. 23 haber kaynağının her biri için ayrı ayrı hazırlanan kod blokları vasıtası ile kazıma işlemi gerçekleştirilmiş ve elde edilen veriler 'text' isimli sütuna kaydedilmiştir. Bu aşamada hazırlanan fonksiyondaki dikkat unsurları Türkçe karakter normalizasyonu, haber kaynağı sunucularına aralıklı ve manueli taklit eden istekler göndererek DDoS önlemlerini bertaraf etmek, doğru html elemanları tespiti ile mümkün olduğu kadar temiz ve doğru veriye ulaşmak olarak sıralanabilir. Bir örnek teşkil etmesi amacı ile, Anadolu Ajansı (AA) web sitesini kazıyarak haber metinlerini elde eden kod bloğu aşağıda verilmiştir (Şekil 13). Her haber kaynağı için aynı özelliklere sahip ancak web sitesinin yapısına göre farklı html elemanlarını kazıyarak veri elde edecek şekilde fonksiyon uyarlanmıştır.

Şekil 13: AA Web Sitesi Kazıma Fonksiyonu

```
import requests
from bs4 import BeautifulSoup
import numpy as np
import time
import random
import unicodedata

headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.3'
}

# Türkçe karakter normalizasyonu fonksiyonu
def normalize_unicode_turkish(text):
    return unicodedata.normalize("NFC", text)

# Haber metinlerini çeken fonksiyon
def scrape_text_only(url):
    try:
        # Sunucuya yüklenmek için rastgele gecikme
        time.sleep(random.uniform(0.5, 1.0))

        response = requests.get(url, timeout=10, headers=headers)
        response.raise_for_status()
        response.encoding = 'utf-8' # Türkçe karakterler için encoding

        soup = BeautifulSoup(response.text, 'html.parser')
        article_content = soup.find(class_='detay-icerik')
        if article_content:
            # <strong>, <em> vb. inline taglerden kaynaklı boşluk kayıplarını önleme
            paragraphs = article_content.find_all('p')
            raw_text = " ".join(p.get_text(" ", strip=True) for p in paragraphs)

            normalized_text = normalize_unicode_turkish(raw_text)
            return normalized_text if normalized_text else np.nan

        return np.nan

    except Exception as e:
        print(f"Error scraping {url}: {e}")
```

```

return np.nan

# Daha hızlı veri eldesi için threading
from concurrent.futures import ThreadPoolExecutor, as_completed

def parallel_scrape(urls, max_workers=5):
    results = [None] * len(urls)
    with ThreadPoolExecutor(max_workers=max_workers) as executor:
        future_to_idx = {executor.submit(scrape_text_only, url): idx for idx, url in enumerate(urls)}
        for future in as_completed(future_to_idx):
            idx = future_to_idx[future]
            try:
                results[idx] = future.result()
            except Exception:
                results[idx] = np.nan
    return results

# Fonksiyonun çağırılması ve Dataframe'e kayıt edilmesi
mask = urls_df['media_name'] == "aa.com.tr"
urls_to_scrape = urls_df.loc[mask, 'url'].tolist()
scraped_texts = parallel_scrape(urls_to_scrape, max_workers=8)

urls_df.loc[mask, 'text'] = scraped_texts

```

Kaynak: Yazar tarafından derlenmiştir.

Web kazıma işlemleri sonucunda taranan 33.856 web adresinin 32.721'ine erişim sağlanmış ve bu adreslerden, bağlantısı çalışır durumda olan ve hedeflenen html etiketi boş veri (NaN) içermeyen haber metinleri elde edilmiştir.

4.3. VERİ TEMİZLİĞİ

Hemen hemen tüm yapay zeka uygulamalarında olduğu gibi DDİ'de de, eğitilen algoritmanın sağlıklı sonuçlar üretebilmesi için veri temizliği kritik öneme sahiptir. Bu nedenle elde edilen 32.721 adet ham haber metni verisi detaylı bir temizlik aşamasına tabii tutulmuştur.

4.3.1. Metin Ön İşleme

Derlenen 32.721 satırlık ham veri üzerinde öncelikle normalizasyon işlemleri uygulanmıştır. Bu kapsamda, metinlerde yer alan tüm karakterler küçük harfe dönüştürülmüş, sayılar ve noktalama işaretleri temizlenmiştir. Bu işlem sonucunda elde temizlenen haber metinleri 'clean_text' sütununa kaydedilmiştir. Bu adımların temel amacı, metin içerisindeki biçimsel çeşitliliğin azaltılarak dilin yapısal birliğinin sağlanmasıdır. Büyük-küçük harf farklılıkları, noktalama işaretleri ve sayılar modelin öğrenme sürecinde anlamsal bir katkı sağlamamakta, aksine kelime örüntülerinin yapay olarak çoğalmasına ve modelin gereksiz yere karmaşıklaşmasına yol açmaktadır. Bu durum, modelin temaları ve konuları doğru bir şekilde tespit etme yeteneğini olumsuz etkilemektedir. Dolayısıyla, normalizasyon işlemleriyle metinler

daha tutarlı ve analiz edilebilir bir yapıya kavuşturulmuş, konu modelleme algoritmasının performansının artırılması hedeflenmiştir.

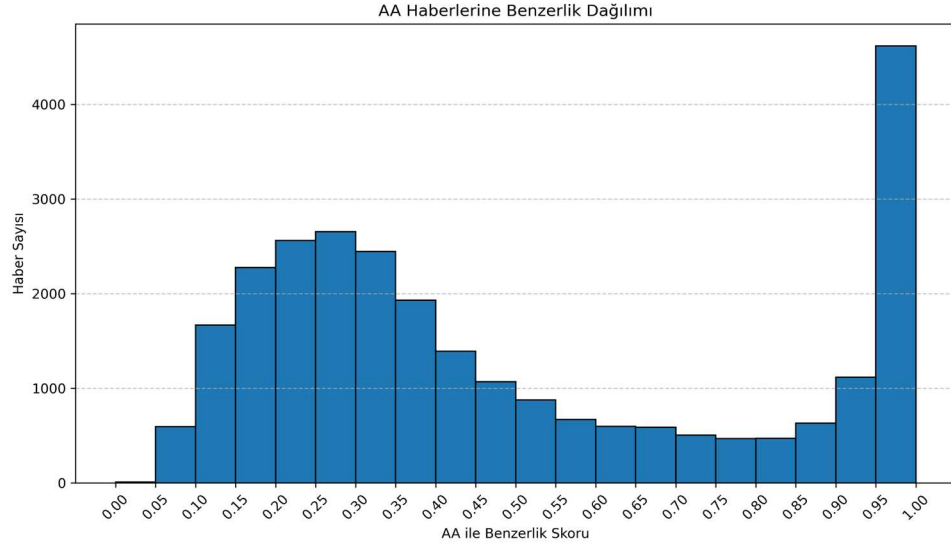
4.3.2. AA Kaynaklı Haberlerin Temizlenmesi

Veri temizlik işleminin ikinci aşaması ise Türkiye'deki medya ekosistemindeki haberlerin dikkate değer bir kısmının AA kaynaklı olmasından ileri gelmiştir. AA Türkiye'deki en eski, en geniş ağı sahip ve en büyük haber sağlayıcı konumundadır. Derlenen veri setinde AA'nın haberleri Media Cloud'dan elde edilen bağlantılar vasıtasıyla zaten taranmıştır. Bu durum, AA'nın servis ettiği haberlerin taranan 23 haber kaynağının hemen hepsinde hiçbir değişikliğe uğramadan tekrar yayınlanması neticesinde, aynı haberlerin veri setinde tekrar eden birçok kopyasının bulunması sonucuna yol açmaktadır. Konu modelleme uygulamalarında tekrar eden veriler bu kümelerle verilen ağırlığı arttıracığından algoritmanın işleyişini olumsuz yönde etkileyecek ve ortaya çıkan sonucun sağlıklı olmamasına yol açacaktır. Bu sebeple AA'nın servis ettiği haberlerin diğer haber kaynaklarından temizlenmesi ihtiyacı oluşmuştur.

Bu temizlik aşamasında kosinüs benzerliği (*cosine similarity*) kullanılmıştır. Kosinüs benzerliği, iki metin vektörü arasındaki açıyı ölçerek benzerliği hesaplayan bir yöntemdir. 1'e yakın değerler yüksek benzerliği, 0'a yakın değerler ise düşük benzerliği ifade eder. Bu yöntem metin madenciliği ve belge karşılaştırmalarında yaygın olarak kullanılmaktadır (Pradhan ve diğerleri, 2015)

İlgili sütunda yer alan her haber metni TF-IDF tekniği ile vektörize edilerek AA'dan elde edilen haber metinleri ile karşılaştırılmış ve kosinüs benzerliği hesaplanarak bir benzerlik skoru elde edilmiştir. Bu skor 'aa_similarity_score' sütununa kaydedilmiş ve ilgili haberin en çok benzediği, yani benzerlik skorunun en yüksek olduğu AA haberinin indeksi de daha sonra manuel kontrol amacı ile 'aa_similarity_index' sütununa kaydedilmiştir. Hesaplanan benzerlik skorlarının frekans dağılımı Şekil 14'te verilmiştir. Şekil incelendiğinde 0,85 ve üzeri benzerlik skoruna sahip haberlerin veri setine dahil edilmemesi normal dağılım açısından sağlıklı olduğu değerlendirilmiştir. Benzerlik skoru 1 olan, yani bire bir aynı olan haberlerin sayısı 699'dur. Bire bir aynı haberler dahil 0,85 ve üzeri benzerlik skoruna sahip haber sayısı ise 7.155'tir.

Şekil 14: AA Benzerlik Skorları Frekans Dağılımı

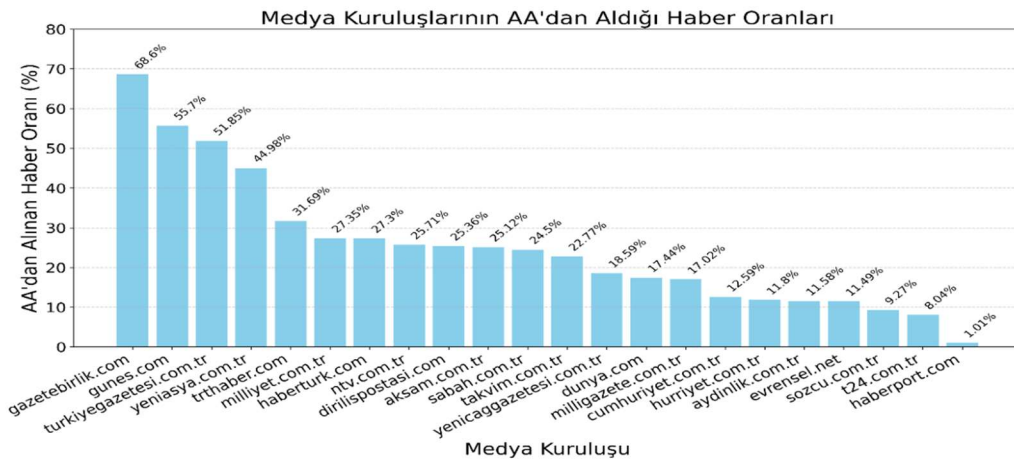


Kaynak: Yazar tarafından derlenmiştir.

Benzerlik skorlarının frekans dağılımı incelenerek eşik değeri belirleme yöntemi Singh ve Singh (2021) tarafından da önerilmiştir. Söz konusu çalışmada haber metinleri TF-IDF vektörleri ve kosinüs benzerliği yöntemiyle karşılaştırılmış, insan kontrolleri ve skor dağılımı kullanılarak 0,788 eşik değeri tanımlanmıştır. Bu yaklaşım, tarafımızca uygulanan histogram analizi ve 0,85 eşik değeri seçimiyle yöntemsel olarak paralellik göstermektedir.

Şekil 15'te taranan ve haber metni elde edilen medya kuruluşlarının AA kaynaklı haber sayılarının bu medya kuruluşlarından elde edilen tüm haberlere oranı verilmiştir.

Şekil 15: AA Kaynaklı Haberlerin Medya Kuruluşlarına Göre Oranı



Kaynak: Yazar tarafından derlenmiştir.

Şekilden de anlaşılabileceği üzere kimi haber kaynaklarından elde edilen haberlerin yarısından fazlası, kimi haber kaynaklarının ise dörtte biri AA kaynaklıdır.

4.3.3. Mükerrer Haber Metinlerinin Temizliği

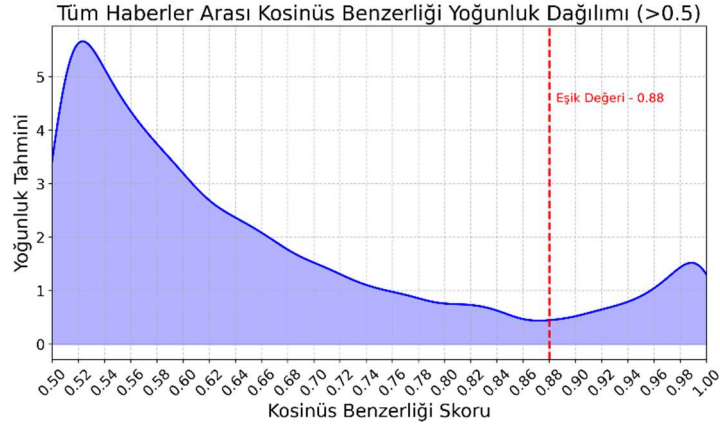
Türk medya ekosisteminde bir başka sıkça karşılaşılan durum ise, bir medya kaynağı tarafından daha önce yayınlanmış haberin ya çok küçük farklılıklar ile (örneğin bir kaç paragrafın yahut kelimenin yerinin değiştirilmesi) ya da hiç değiştirilmeden farklı bir başlık ve farklı bir URL ile yine aynı medya kuruluşu tarafından yayınlanmasıdır. Bu nedenle her haber kuruluşu için ayrı ayrı kendi iç benzerlik skorları hesaplanmış ve 0,85 eşik değeri üzerinde benzerliğe sahip haberler veri setinden çıkarılmıştır. Bu aşamada düşürülen haber sayısı 690'dır.

Medya kuruluşlarının aynı haberleri tekrar tekrar dolaşıma sokmalarının yanı sıra birbirlerinin haberlerini neredeyse hiçbir değişiklik yapmadan kopyalamaları da Türk medya ekosisteminde yaygın görülen bir durumdur. Dahası, AA dışında Türkiye'de haber servis eden diğer iki büyük ajans olan İhlas ve Demirören haber ajanslarından kaynaklanan haberler birçok medya kuruluşu tarafından aynı biçimde yayınlanmaktadır. Media Cloud'dan elde edilen meta verilerde İhlas ve Demirören haber ajansları bulunmamaktadır. Bu sebeple AA'dan doğrudan alınan haberler ve farklı bağlantılarda tekrar yayınlanan haberler temizlendikten sonra dahi konu modelleme algoritmasının sağlıklı çalışmasını engelleyecek sayıda tekrar eden verinin temizlenmesi zorunluluğu doğmuştur.

Bu noktada AA'ya 0,85 ve üzeri skora sahip benzerlik taşıyan satırlar ve farklı bağlantı adresi ile tekrar yayınlanan haberler düşürüldükten sonra geri kalan 24.876 satırlık AA dahil tüm veri seti için tekrar çiftler halinde kosinüs benzerliği skorları hesaplanmış ve bir benzerlik matrisi oluşturulmuştur.

Bu benzerlik matrisi üzerinden her haber için diğer haberler ile olan benzerlikleri değerlendirmek amacıyla kosinüs benzerliği eşik değeri olarak 0,88 belirlenmiştir. Eşik değeri, Şekil 16'de görülebilecek tüm haber çiftleri arasındaki benzerlik skorlarının yoğunluk tahmini grafiği (*kernel density estimation - KDE*) ile görselleştirilerek belirlenmiştir. Bu eşik, yalnızca çok yüksek benzerlik taşıyan haberlerin kümelenmesini ve gereksiz haber kaybının önlenmesini amaçlamaktadır.

Şekil 16: Kosinüs Benzerliği Yoğunluk Dağılımı - Tüm Veri Seti



Kaynak: Yazar tarafından derlenmiştir.

Belirlenen eşik doğrultusunda 0,88 ve üzeri benzerlik skoruna sahip tüm haber çiftleri tespit edilmiş, bu çiftler temel alınarak bağlantılı bileşenler yaklaşımıyla kümeler oluşturulmuştur. Benzer olduğu değerlendirilen haberlerden oluşan her kümede yalnızca bir temsilci haberin kalması, diğerlerinin veri setinden çıkarılması hedeflenmiştir.

Benzer haberlerden oluşan kümedeki haberlerin hangi haber tarafından veri setinde temsil edileceğinin tespiti için bir karar ağacı takip edilmiştir. Öncelikle, kopya haberlerin kaynak haberini tespit amacı ile her kümedeki en eski tarihli haber seçilmiştir. Eğer birden fazla haber aynı tarihte yayınlanmış ise, çalışmanın önceki adımlarında hesaplanan medya kuruluşlarının özgün haberlerini koruma oranı (bkz. Şekil 15) en yüksek olan medya kuruluşuna ait haber seçilmiştir. Bir başka ifade ile, AA'dan haber alma oranı en düşük olan medya kuruluşuna temsilci haber seçerken öncelik verilmiştir. Eğer bu kriterlerde de eşitlik bozulmamışsa, küme içerisindeki satır index numarası en küçük olan haber temsilci olarak atanmıştır. Bu karar mekanizması, haberlerin hem tarihsel orijinalliğini hem de medya kuruluşlarının özgün haber üretme kapasitesini dikkate alacak şekilde tasarlanmıştır.

Bu işlemler neticesinde veri setine üç yeni sütun eklenmiştir:

- 'duplicate_group_id': Her küme için bir benzersiz küme kimliği atanmıştır. Küme dışı kalan yani benzer haber taşımayan satırlara "-1" değeri verilmiştir.
- 'is_representative': Kümedeki temsilci haberler için "1", temsilci olmayanlar için "0", küme dışı satırlar için "-1" değeri atanmıştır.

- 'cluster_size': Kümenin büyüklüğü belirtilmiştir. Küme dışı satırlar için bu değer "1" olarak atanmıştır.

Bu sütunlar sayesinde hangi haberlerin kümелendiği, hangi haberlerin temsilci olarak seçildiği ve küme büyüklüklerinin dağılımı analiz edilebilir hale getirilmiştir. Özgün ve temsilci haberler dışındaki mükerrer haber metinlerinin toplam sayısı 3.312'dir. Bu satırlar düşürüldükten sonra kalan satır sayısı 21.564 olmuştur.

4.3.4 Çok Kısa ve Çok Uzun Haberlerin Temizlenmesi

Doğal dil işleme (DDİ) ve konu modelleme uygulamalarında, belirli bir kelime sayısının altındaki veya üstündeki metinlerin ana veri setinden çıkarılması yaygın bir uygulamadır. Her ne kadar BERTopic gibi modeller kısa metinlerle de başarılı sonuçlar üretebilse de, bu çalışmada kullanılan haber metinleri bağlamında ve analiz düzlemi haberlerin bir olguyu nasıl çerçevelediği olduğunda, çok kısa metinlerin yeterince bilgi taşımayacağı açıktır. Kısa metinlerin konu modelleme performansını düşürdüğü, minimum kelime sınırının genellikle verinin kalitesine göre değerlendirilmesi gerektiği belirtilmektedir (Murshed ve diğerleri, 2023; Yan ve diğerleri, 2013). Uzun metinlerde ise bilgi eksikliği değil, tam tersine birden fazla konuyu barındırma problemi söz konusu olabilir. Bu, konu modelleme algoritmasının yalnızca bir konuya yoğunlaşmasına ve diğer bilgilerin kaybına neden olabilir. Literatürde uzun metinlerin bölünerek işlenmesi önerilmektedir (Sbalchiero ve Eder, 2020). Ancak bu çalışmada tarih/konu, medya kuruluşu/konu ve ülke/konu gibi boylamlar temelinde analiz yapılacağından, metin bölme işlemi uygun bulunmamıştır.

Bu nedenle çalışma kapsamında toplanan veri setinden hem kısa hem de uzun metinler veri setinden çıkarılmıştır. Çıkarma işlemi yapılırken öncelikle tüm verinin 'clean_text' sütunundaki kelime sayıları hesaplanmış ve Tablo 12'de sunulan genel istatistiksel bilgiler incelenmiştir.

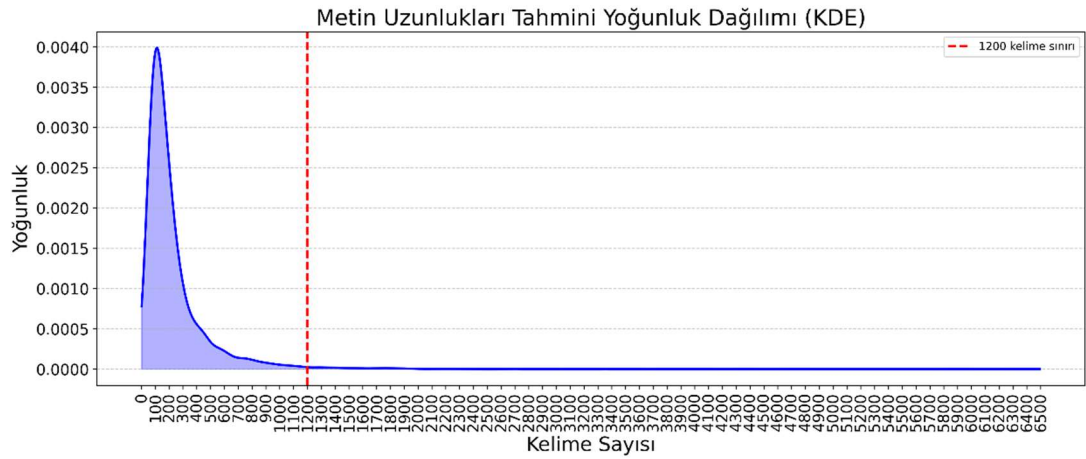
Tablo 12: Veri Kelime Sayısı Temel İstatistikleri

İstatistik	Değer
Minimum kelime sayısı	2
Maksimum kelime sayısı	6496
Ortalama kelime sayısı	236.16
Medyan	157
Q1	98
Q3	266

Kaynak: Yazar tarafından derlenmiştir.

Genel istatistikler açıkça verinin sağa çarpık dağılıma sahip olduğunu göstermektedir. Daha sonra uzun metinler için bir eşik belirlenebilmesi amacı ile metin uzunluklarının tahmini yoğunluk dağılımları çıkarılmıştır. Şekil 17’de de görülebileceği üzere 1200 kelime sonrası çalışmada derlenen veri seti için aykırı değer teşkil etmektedir.

Şekil 17: Metin Uzunlukları Yoğunluk Dağılımı



Kaynak: Yazar tarafından derlenmiştir.

Kalan veri kümesinde kısa haber metinleri incelendiğinde 15 kelime altındaki verilerin anlamlı bilgi içermediği tespit edilmiştir. Çok kısa ve çok uzun haber metinlerin çıkarılması sonucunda veriden 345 satır silinmiş ve kalan veri 21.219 satır olmuştur. Bu aşamada, bir sonraki başlık altında incelenecek nedenlerden ötürü 1.251 spor haberleri de veriden çıkarılmıştır. Dolayısıyla bu çalışmanın konu modelleme için

ele aldığı veri setinin temizlik aşamaları sonucu nihai gözlem sayısı 19.968'dir ($N = 19.968$).

Tablo 13: Veri Temizlik Aşamaları ve Kalan Satır Sayıları

Açıklama	Satır Sayısı	Düşürülen Satır Sayısı
Tüm URL'ler	58.718	-
Eş sesli, 500'den az haber ve dil temizliği	33.856	24.862
Web kazıma ile elde edilebilen haber metinleri	32.721	1.135
AA benzerlik temizliği	25.566	7.155
Medya kuruluşu içi benzerlik temizliği	24.876	690
Medya kuruluşları arası benzerlik temizliği	21.564	3.312
15 kelimenin altındaki metin temizliği	21.493	71
1.200 kelimenin üzerindeki metin temizliği	21.219	274
Spor haberlerinin çıkarılması	19.968	1.251
Nihai gözlem sayısı	19.968	-

Kaynak: Yazar tarafından derlenmiştir.

Tablo 14: Medya Kuruluşuna Göre Haber Metni Verisi Dağılımı

Haber Kaynağı	Haber Metni Sayısı	Haber Kaynağı	Haber Metni Sayısı
aa.com.tr	4463	milligazete.com.tr	526
t24.com.tr	2566	hurriyet.com.tr	524
trthaber.com	1701	yenicaggazetesi.com.tr	498
haberport.com	1258	sozcu.com.tr	455
sabah.com.tr	1012	gazetebirlik.com	382
cumhuriyet.com.tr	875	turkiyegazetesi.com.tr	381
milliyet.com.tr	808	evrensel.net	375
aksam.com.tr	770	dunya.com	287
haberturk.com	631	dirilispostasi.com	276
takvim.com.tr	611	yeniasya.com.tr	224
aydinlik.com.tr	573	gunes.com	207
ntv.com.tr	565		

Kaynak: Yazar tarafından derlenmiştir.

Tablo 15: Ülke Başına Haber Metni Verisi Dağılımı

Ülke Adı	Haber Sayısı	Ülke Adı	Haber Sayısı	Ülke Adı	Haber Sayısı
Libya	3704	Orta Afrika Cumhuriyeti	208	Zambiya	42
Mısır	2764	Güney Sudan	189	Liberya	42
Afrika	2107	Uganda	136	Moritanya	40
Güney Afrika Cumhuriyeti	1392	Çad	134	Ekvator Ginesi	39
Sudan	1059	Gana	127	Malavi	36
Cezayir	987	Tanzanya	123	Togo	35
Tunus	849	Fildişi Sahili	105	Burundi	34
Somali	675	Gine-Bissau	91	Eritre	30
Fas	653	Gine	88	Botsvana	25
Nijerya	586	Gabon	88	Benin	25
Etiyopya	508	Angola	86	Lesotho	23
Burkina Faso	480	Sierra Leone	82	Mauritius	22
Kongo Cumhuriyeti	315	Madagaskar	66	Seyşeller	15
Demokratik Kongo Cumhuriyeti	313	Zimbabve	65	São Tomé ve Príncipe	15
Ruanda	288	Cibuti	61	Komor Adaları	11
Nijer	285	Mozambik	56	Esvatini	11
Kenya	237	Gambiya	53	Yeşil Burun Adaları	5
Mali	236	Kamerun	49		
Senegal	227	Namibya	46		

Kaynak: Yazar tarafından derlenmiştir.

4.4. MODELİN KURULMASI VE ÇALIŞTIRILMASI

4.4.1. Haberlerin Kategorilere Ayrılması

Çalışma analiz yöntemi olarak BERTopic konu modelleme tekniğini kullanacaktır. Daha önce de ifade edildiği üzere BERTopic, önceden etiketlenmemiş veri üzerinde, yani denetimsiz olarak oldukça iyi sonuçlar veren bir algoritmadır. Özellikle cümle gömmeleri sayesinde semantik yapıyı iyi öğrenen ve kümeleme algoritmaları ile uyumlu çalışan bir yapıya sahiptir. Ancak bu çalışma kapsamında incelenecek verinin büyüklüğü ve karmaşıklığı, yani verinin çok fazla sayıda ve çok fazla konuda (haber kategorisinde) haber içermesi, BERTopic algoritmasının çerçeve yakalamada kimi nüansları kaçırmaya, yahut makro düzeyde iyi performans gösterirken, daha alt düzeylerde daha kötü bir performans göstermesine yol açabilir. Bir başka ifade algoritmanın kategoriler arası ortak temalara/çerçevelere fazlaca yoğunlaşıp kategoriye özgü gözden kaçırması olasılığı mevcuttur. Ayrıca, büyük ve dengesiz veri kümelerinde, bazı düşük sayıda haber içeren kategorilerde (örneğin spor haberleri yahut kültür haberleri gibi) temaların yeterince temsil edilmemesi riski bulunmaktadır. Bu sebeple çalışmada uygulanacak yöntemin hem evrensel hem de yerel farklılıkları mümkün olduğunca yakalayacak şekilde yapılandırılması önem taşımaktadır. Çok katmanlı yöntemlerin izlenmesi literatürde de yaygın biçimde rastlanılan bir tekniktir (Alfaro ve diğerleri, 2016; Krstovski ve diğerleri, 2016; Mao ve diğerleri, 2022; Sayed ve diğerleri, 2023; Voskergian ve diğerleri, 2024).

Bu noktadan hareket ile çalışma kapsamında yürütülen çerçeve analizinin sağlıklı bir biçimde gerçekleştirilebilmesi amacıyla, incelenen haber metinlerinin içeriklerine göre anlamlı kategorilere ayrılması uygun görülmüştür. Bu yaklaşım, özellikle konu modelleme teknikleri ile elde edilen sonuçların daha derinlemesine anlaşılması ve yorumlanması açısından önem arz etmektedir. Metinlerin kategorilere ayrılması, analiz edilen veri kümesinin karmaşıklığını azaltmakta ve her bir kategorinin özgün bağlamı içerisinde ele alınmasına imkan tanımaktadır.

Haberlerin kategorilere ayrılması ve her kategoriye ayrı ayrı konu modelleme yapılarak çeşitli avantajları bulunmaktadır. Öncelikle her kategori, kendi bağlamında derinlemesine analiz edilebilmekte ve bu sayede temalar daha özgün bir şekilde ortaya çıkarılabilmektedir. Ayrıca, veri kümesinde bulunan sınıflar arasındaki dengesizlik sorunu ortadan kalkmakta, her sınıf kendi içerisinde değerlendirildiği için daha homojen analiz sonuçları elde edilmektedir. Bunun yanı sıra, kategoriler arası

karşılaştırma yapma olanağı ortaya çıkmakta ve kesişen ya da ayrışan konular/çerçeveler daha net bir biçimde tanımlanabilmektedir. Öte yandan tüm derlem seviyesindeki bazı genel çerçeveler, kategori bazında konu modelleme yapıldığında algoritma tarafından gözden kaçırılabilir.

Tüm bu nedenlerle, çalışma kapsamında BERTopic ile konu modellemenin kategori bazında gerçekleştirilmesi uygun bir yöntem olarak değerlendirilmiştir. Bu amaç doğrultusunda öncelikle kategorizasyon için ince ayarlanmış bir BERT modeli çalıştırılarak haberler beş ayrı kategoriye ayrılmıştır. Bu kategoriler siyaset, ekonomi, kültür, sağlık ve spor olarak belirlenmiştir. Kategoriler ve açıklamaları aşağıdaki tabloda verilmiştir.

Tablo 16: Kategoriler ve Haber Türleri

Kategori	Haber Türleri
Siyaset	• Devletlerarası diplomatik ilişkiler ve görüşmeler. • Resmi ziyaretler, zirveler, anlaşmalar. • Ulusal ve uluslararası siyasi gelişmeler. • Siyasi liderlerin açıklamaları ve faaliyetleri. • Çatışmalar, askeri müdahaleler ve güvenlik politikaları. • Adli olaylar (özellikle siyasi yönü olan veya siyasi figürleri içeren).
Ekonomi	• Ticaret anlaşmaları ve ekonomik iş birlikleri. • Yatırım projeleri, finansal düzenlemeler. • Afrika ülkelerinin ekonomik göstergeleri (büyüme, işsizlik vs.). • Tarım, enerji ve doğal kaynaklar ekonomisi. • Türkiye-Afrika ekonomik ilişkileri. • Türkiye-Afrika ekonomisi karşılaştırmaları
Sağlık	• Salgın hastalıklar, halk sağlığına yönelik tehditler. • Doğal afetler ve iklim değişikliğine bağlı can kayıpları. • Kuraklık, sel, fırtına gibi hava olaylarıyla ilgili insan sağlığını etkileyen durumlar. • Gıda güvenliği ve su kıtlığı sorunları. • Uluslararası sağlık yardımları ve müdahaleleri. • Altyapı yetersizliği kaynaklı ölümler
Kültür	• Afrika kültürüyle ilgili genel tanıtım yazıları. • Sanat eserleri, sergiler, müzik ve diğer kültürel etkinlikler. • Afrika ülkelerinin turistik destinasyon olarak tanıtılması. • Ünlü kişilerin Afrika ülkelerinde yaptığı kültürel ziyaretler ve tatiller. • Hayvanlara ve doğal yaşama dair içerikler. • Geleneksel yaşam tarzları, festivaller ve yerel kültürel uygulamalar.
Spor	• Futbol başta olmak üzere tüm spor dallarında Afrika ülkelerinin faaliyetleri. • Karşılaşma skorları • Sporcuların performansları ve başarıları • Transfer haberleri

Kaynak: Yazar tarafından derlenmiştir.

Haber metinlerinin kategorilere ayrılması işlemi için, Türkçe dilinde önceden eğitilmiş bir dil modeli olan BERTurk (Schweter, 2020) kullanılarak yarı denetimli öğrenme yaklaşımı benimsenmiştir. Bu amaçla, haber metinlerinin içeriklerine göre

elle etiketlenmiş bir %8 lik bir örneklem veri kümesi oluşturulmuştur. Bu veri kümesinde; siyaset: 363, spor 357, sağlık 350, ekonomi 348, kültür 318 satır veri ile temsil edilmektedir.

Bu veri %80 eğitim ve %20 test verisi olarak ayrılmıştır. Kategorizasyon amacı ile BERT modeli üzerinde ince ayar (*fine-tuning*) gerçekleştirilmiştir. Model ayrılan eğitim verisi ile eğitilmiş ve eğitim sırasında, sınıflandırma katmanı güncellenmiş ve modelin beş kategori (siyaset, ekonomi, sağlık, kültür, sport) üzerinde tahmin yapabilmesi sağlanmıştır. Model eğitimi tamamlandıktan sonra, doğruluk %20'lik test verisi üzerinde test edilmiş ve F1 skoru ve hata matrisi gibi performans ölçütleri hesaplanmış ve sonuçların güvenilirliği değerlendirilmiştir.

Tablo 17: BERT Haber Kategorizasyonu Değerlendirme Metrikleri

	precision	recall	f1-score	support
kültür	0.940299	0.984375	0.961832	64
ekonomi	0.917808	0.957143	0.937063	70
sağlık	0.932432	0.985714	0.958333	70
siyaset	0.96875	0.849315	0.905109	73
spor	1	0.985915	0.992908	71
accuracy	0.951149	0.951149	0.951149	0.951149
macro avg	0.951858	0.952493	0.951049	348
weighted avg	0.952341	0.951149	0.950587	348

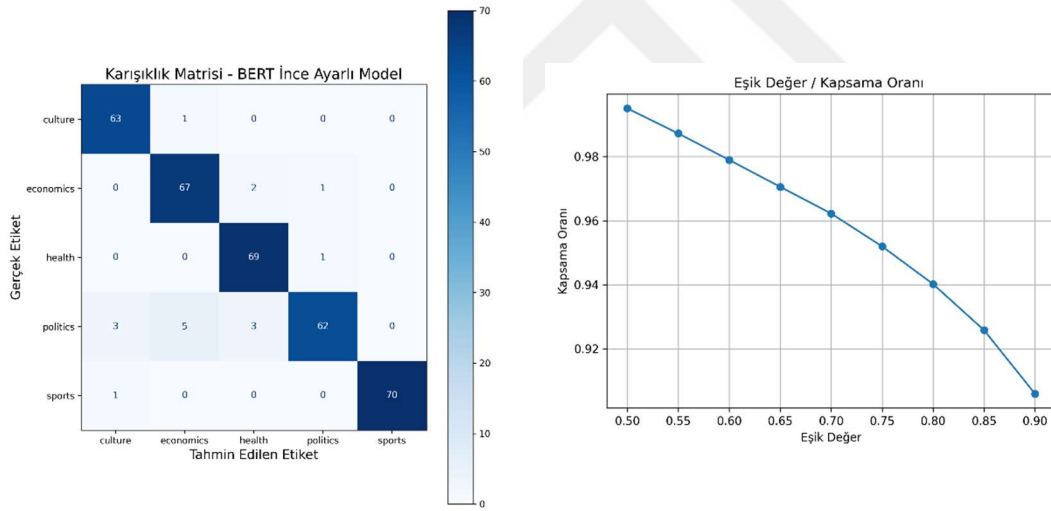
Kaynak: Yazar tarafından derlenmiştir.

Modelin performansına ilişkin hesaplanan değerlendirme metrikleri, sınıflandırma modelinin genel olarak yüksek bir doğruluk ve dengeli bir başarı sağladığını ortaya koymuştur. Beş kategoriye ait kesinlik (*precision*), duyarlılık (*recall*) ve F1 skorları incelendiğinde, tüm sınıflarda istikrarlı bir başarı düzeyine ulaşıldığı görülmektedir. Kültür kategorisinde F1 skoru 0,96, ekonomi kategorisinde 0,93, sağlık kategorisinde 0,96, siyaset kategorisinde 0,91 ve spor kategorisinde ise 0,99 olarak hesaplanmıştır. Özellikle spor kategorisinde elde edilen %99 F1 skoru, modelin bu sınıfı neredeyse hatasız olarak sınıflandırabildiğini göstermektedir. Bununla birlikte spor kategorisi için *precision* değerinin 1,00 çıkması, modelin yalnızca çok belirgin

örnekleri sınıflandırmış olabileceğini ve daha belirsiz örnekleri atlamış olabileceğini düşündürmektedir. Bu kadar yüksek bir doğruluk, veri sayısının görece az olduğu durumlarda modelin aşırı uyum (*overfitting*) göstermesiyle de açıklanabilir. İlaveten, siyaset kategorisinde duyarlılık değerinin diğer sınıflara kıyasla bir miktar düşük kalması, bu kategoride bazı gerçek örneklerin yanlış sınıflara atandığını düşündürmektedir. Bu durum, siyasete dair içeriklerin diğer kategorilerle, özellikle de ekonomi kategorisi ile örtüşen temalar barındırabilmesiyle ilişkilendirilebilir.

Modelin genel doğruluğu (*accuracy*) %95,1 olarak ölçülmüştür. Bu sonuçlar, eğitilen BERT tabanlı sınıflandırma modelinin hem dengeli hem de güvenilir bir performans sergilediğini ve haber metinlerinin kategorilere ayrılmasında etkin bir araç olarak kullanılabileceğini göstermektedir. Modele ait karışıklık matrisi ve eşik değer/kapsama oranı ilişkisi modelin oldukça sağlıklı tahmin yaptığına işaret etmektedir.

Şekil 18: BERT Sınıflandırma Modeli Karışıklık Matrisi ve Eşik Değer/Kapsama Oranı İlişkisi



Kaynak: Yazar tarafından derlenmiştir.

Eğitilen model ile tüm veri seti üzerinde kategori tahmini yapılmıştır. Bu sınıflandırmaya göre veri setindeki haberlerin kategorilere göre dağılımı Tablo 18'de verilmiştir.

Tablo 18: Veri Setinin Haber Kategorilerine Göre Dağılımı

Kategori	Satır Sayısı
Siyaset	13553
Ekonomi	2628
Sağlık	2402
Kültür	1385
Spor	1251

Kaynak: Yazar tarafından derlenmiştir.

Veri setindeki “spor” olarak etiketlenen haberler manuel olarak incelendiğinde birbirlerine oldukça benzedikleri tespit edilmiştir. Bunun temelde iki sebebi Türkiye’deki spor alanının fazlasıyla futbol tarafından işgal edilmesi ve Türk medya ekosistemindeki spor haberlerinin analiz/yorumdan ziyade bilgi (örneğin transfer yahut maç skoru) içermesi olarak öne sürülebilir. Yapılan ön çalışmada bu derece benzeşen haberlere çalışmanın tekniği olan konu modellemenin uygulanmasının herhangi anlamlı bir sonuç ortaya koymayacağından hareketle spor haberleri analizin dışarısında tutulmuştur.

4.4.2. Cümle Gömme Algoritmasının Çalıştırılması

BERTopic algoritmasının çalışmasının temel prensibi metin verisinde yer alan cümlelerin “gömme” (*embedding*) olarak adlandırılan bir çeşit vektörleştirme aşamasına tabii tutulmasıdır. Bu vektörleştirme aşamasında genellikle önceden eğitilmiş modeller kullanılmaktadır. Bu modeller, çok dilli olarak eğitilip dilden dile farklılaşan semantik yapıları ve anlam bütünlüğünü öğrenebilme kapasitesine sahip derin öğrenme yöntemleri neticesinde oluşturulmaktadır.

Bu çalışmada haber metinlerinin vektörleştirilmesi için xlm-r-bert-base-nli-stsb-mean-tokens (Reimers ve Gurevych, 2019) cümle gömme modeli tercih edilmiştir. Bu model, Türkçe’nin de yer aldığı çok dilli olarak eğitilmiştir ve bağlamdan anlam çıkarabilme yetenekleriyle öne çıkmaktadır. Modelin özellikle çok dilli ve yapısal çeşitlilik gösteren veri kümelerinde (örneğin düşük kaynaklı diller dahil) önceden eğitime tabi tutulmadan başarılı sonuçlar verdiği ampirik olarak da gösterilmiştir. (Kapočiūtė-Dzikiene ve diğerleri, 2021).

Ayrıca, modelin Türkçe içeriklerdeki başarısı da yapılan güncel araştırmalarla doğrulanmıştır. Saleem ve Mehrotra (2023), 2023 Maraş Depremleri sırasında Twitter’den toplanan çok dilli (ağırlıklı olarak Türkçe) tweet’leri analiz etmek için BERTopic algoritmasıyla farklı gömme modellerini karşılaştırmış ve en iyi sonuçların

“xlm-r-bert-base-nli-stsb-mean-tokens” modeliyle elde edildiğini bildirmiştir. Bu durum, modelin kısa, gürültülü ve anlam çeşitliliği yüksek metinlerde bile etkili performans sergilediğini göstermektedir. Dolayısıyla, söz konusu modelin seçimi, hem çok dilli hem de Türkçe haber metinleri üzerinde anlamlı ve güvenilir konu modellemesi gerçekleştirilmesi için uygun bir tercih olarak değerlendirilmiştir. Şekil 19’da cümle gömmelerinin gerçekleştirilme aşaması python kodu olarak gösterilmiştir.

Şekil 19: Cümle Gömmelerinin Gerçekleştirilmesi

```
from sentence_transformers import SentenceTransformer
import numpy as np

# Temizlenmiş metinler listesinin hazırlanması
documents = df['clean_text'].tolist()

# Cümle gömme modelinin çağırılması
embedding_model = SentenceTransformer("xlm-r-bert-base-nli-stsb-mean-tokens")

# Belgeler için gömme vektörlerinin (embeddings) çıkarılması
embeddings = embedding_model.encode(
    documents,
    show_progress_bar=True,
    batch_size=32
)

# Elde edilen gömmelerin daha sonra kullanılmak üzere belirtilen dizine
# kaydedilmesi
np.save(
    r"[DOSYA_YOLU]/all_embeddings.npy",
    embeddings
)
```

Kaynak: Yazar tarafından derlenmiştir.

4.4.3. Derleme Özel Duraklama Kelimeleri Listesi Oluşturulması

Bu çalışmanın ana konusunu oluşturan haber metinlerinde yürütülen çerçeve analizlerinde temel hedef, olaylara ilişkin anlatıların ve temaların belirlenmesidir. Bu bağlamda, metinlerde yer alan olay yeri, aktörler, devlet başkanları veya uluslararası kuruluş isimleri gibi kişi, kurum ve lokasyon adları genellikle yalnızca haberin konusu veya coğrafi kapsamı hakkında bilgi vermekte, doğrudan çerçevenin içeriğini oluşturan fikir ve söylemleri taşımamaktadır. Çalışmada kullanılan haber derlemi, farklı medya kuruluşlarından derlenmiş geniş bir veri setine dayanmakta olduğundan,

metinlerde benzer kiři, kurum ve yer adlarının yoğun řekilde tekrarlandığı gözlemlenmiştir.

Literatürde, içerik taşımayan veya analiz hedefiyle doğrudan ilişkili olmayan kelimelerin çıkarılmasının, metin madenciliği süreçlerinin doğruluğunu ve verimliliğini artırdığı geniş şekilde kabul görmektedir. Örneğin Schofield ve diğerleri (2017) stopword çıkarımının, model eğitimi öncesinde veya sonrasında uygulanmasının modelin genel performansı üzerinde önemli bir fark yaratmadığını, önemli olanın konu dışı terimlerin model sonuçlarını gölgelememesini sağlamak olduğunu belirtmektedirler. Bu çerçevede, kiři, kurum ve yer adlarının mümkün olduğunca azaltılması, modelin içeriksel kavramlara daha iyi odaklanmasını destekleyici bir adım olarak değerlendirilmektedir.

Daha spesifik olarak, Baradad ve Mugabushaka (2015) tarafından yapılan çalışma, derleme özgü durak kelimeler listesi oluşturmanın standart durak kelime listelerine kıyasla metin analizi kalitesini artırdığını göstermektedir. Alan veya derlem özelindeki metinlerde bazı terimler, genel anlamda içerik taşıyor gibi görünse de, belirli bir bağlamda tekrarlayan ve ayırt edici bilgi sağlamayan yapılar haline gelebilmektedir. Bu nedenle, çalışmada yalnızca genel geçer durak kelime listelerine dayanılmamış, konu modelleme sürecinde içerik odaklı anlam kümelerinin oluşumunu sınırlayabilecek yoğun tekrar eden kelimelere de odaklanılmıştır. Derlemin yapısal özelliklerine özgü olarak kiři, kurum ve yer adları da içerik taşımayan unsurlar olarak tanımlanmış, toplam 1640 kelimeden ve edat, bağlaç vb. kelimelere ek olarak kiři, kurum ve yer isimlerinden oluşan geniş bir durak kelimeler listesi hazırlanarak vektörizasyon sürecinde bu kelimeler çıkarılmıştır. Bu yaklaşım, modelin çerçeve kavramlarını daha tutarlı ve doğru bir şekilde temsil etmesini sağlamak amacıyla tercih edilmiştir.

4.4.4. Model Parametrelerinin Seçilmesi

BERTopic ile konu modelleme gerçekleştirirken parametrelerin verinin tipine ve boyutuna göre ayarlanması büyük önem arz etmektedir. Özellikle boyut indirgeme (UMAP) ve kümeleme (HDBSCAN) algoritmalarının parametreleri elde edilecek sonuç üzerinde doğrudan etkilidir. Parametrelerin bu denli etkili olmasının temel nedeni, BERTopic'in modüler yapısının her bir bileşeninin (gömme modeli, boyut indirgeme, kümeleme ve terim ağırlıklandırma gibi) çıktıların kalitesini belirlemede kritik rol oynamasıdır. Modelin performansı, bu bileşenlerin birbirleriyle nasıl

etkileşime girdiğine ve veri özelliklerine nasıl adapte edildiğine bağlıdır. Borçin ve Jose (2024) bu parametrelerin metin ön işleme, gömme modeli kalitesi ve terim ağırlıklandırma stratejileriyle birlikte konu modellemenin tutarlılığı ve çeşitliliği üzerinde anlamlı etkiler yarattığını ampirik olarak göstermiştir.

BERTopic modelinde boyut indirgeme adımı, yüksek boyutlu gömme vektörlerinin daha düşük boyutlarda anlamlı yapıları koruyacak şekilde temsil edilmesini sağlamaktadır. Bu nedenle bu aşamada kullanılan UMAP algoritmasının parametreleri, özellikle “n_neighbors” ve “n_components”, elde edilecek konu kümelerinin kalitesini doğrudan etkilemektedir. Bu parametrelerden “n_neighbors”, her bir noktanın komşuluk yapısını tanımlar. Düşük “n_neighbors” değerleri (örneğin 5-10), yerel yapıyı ön plana çıkararak daha küçük ve ince kümeler oluşturur. Bu, özellikle veri satırı sayıca az alt kategorilerde, azınlıkta kalan ama önemli konuların yakalanmasına olanak sağlar. Yüksek “n_neighbors” değerleri (örneğin 50-100) ise daha geniş komşuluklar kurarak daha genel ve büyük kümeler üretir, bu da büyük haber kategorilerinde baskın konuları öne çıkarabilir.

UMAP algoritmasında bir diğer parametre olan “n_components” ise verinin indirgeneceği boyut sayısını belirler. Bir başka ifade ile vektörleştirilen verinin ne kadar bilgi taşıyacağı “n_components” parametresi ile belirlenmektedir. Düşük n_components değerleri düşük bilgiye yüksek değerler ise yüksek bilgiye işaret eder. Daha yüksek bilgi daha ayrıntılı kümelerin oluşmasını sağlayabilir ancak bu sefer de ince ayrımlar ve farklı sözdizimleri sanal olarak fazla sayıda kümeye yol açabilmektedir (Ordun ve diğerleri, 2020).

Boyutları indirgenmiş vektörlerin kümelenmesi, yani konuların oluşturulması aşamasında kullanılan HDBSCAN algoritması ise, yoğunluk temelli bir yöntemdir ve özellikle “min_cluster_size” ile “min_samples” parametreleri model çıktılarının kalitesini belirlemede kritik rol oynar. “min_cluster_size”, bir kümenin oluşabilmesi için gerekli minimum örnek sayısını tanımlar. Düşük değerler, daha küçük ve detaylı kümelerin oluşmasına izin verirken daha büyük değerler bir küme oluşabilmesi için daha fazla eleman gerektiğini belirlemektedir. Dolayısıyla veri setindeki satır sayısı arttıkça bu değerin çok küçük bırakılması fazla fark içermeyen birbirine benzer sayıca çok kümelerin oluşmasına yol açacaktır. Veri setindeki satır sayısı azaldıkça bu değerin yüksek belirlenmesi ise farkların kaybolarak çok büyük kümeler oluşmasına neden olacaktır. Bir başka ifade ile yüksek “min_cluster_size” değerleri daha büyük ve yoğun kümelerin oluşmasına izin verirken düşük değerler ayrıntılı ve seyrek kümelerin oluşmasına neden olur.

Bir diğer parametre olan “min_samples” ise bir noktanın çekirdek nokta olarak kabul edilmesi için gerekli minimum komşu sayısını belirler. Düşük “min_samples” değerleri kümelerin daha esnek ve küçük olmasına neden olurken, yüksek değerler yalnızca yoğun bölgelerdeki noktaların kümelere dahil edilmesini sağlar. Bu, düşük değerlerin daha fazla küçük kümeler ve potansiyel olarak daha fazla gürültü üretmesine, yüksek değerlerin ise daha katı ve stabil kümeler oluşturmaya neden olmaktadır (Stewart ve Al-Khassaweneh, 2022).

Bu parametrelerin çalışma prensipleri, BERTopic algoritmasının üreteceği sonuçlara etkileri, çalışmanın ele aldığı kategorilerin birbirlerinden hem içerik hem de sayısal olarak birbirlerinden önemli farklar barındırması da göz önünde bulundurularak, her bir kategori için ayrı parametreler belirlenip model kurulmuştur. Model parametrelerine ilişkin bilgiler Tablo 19’da verilmiştir.

Tablo 19: Kategori Bazında Model Parametreleri

Parametre Açıklamaları				
<u>random_state</u> → Kodun tekrarlanabilirliği için atanan parametre, <u>prediction_data</u> → Her bir satır için olasılık hesaplanmasını sağlayan parametre, <u>vectorizer_model</u> → Kelimeleri vektörize etmek için kullanılan model <u>ngram_range</u> → Alınacak ngram sayısı, Türkçe’deki tamlama çokluğu dikkate alınarak 2 olarak seçilmiştir. <u>nr_topics</u> → BERTopic in kaç konu üreteceği. Tam denetimsiz öğrenme için “auto” olarak seçilmiştir. <u>language</u> → Çalışılan dokümanların dili.				
Kategori	UMAP	HDBSCAN	Açıklama	BERTopic Modeli
Siyaset	umap_model = UMAP(n_neighbors=15, n_components=8, random_state=42)	hdbscan_model = HDBSCAN(min_cluster_size=50, min_samples=5, prediction_data=True)	Büyük veri sayısı, geniş kümeler, örtüşen konular fazlalıkta	topic_model = BERTopic(vectorizer_model=CountVectorizer(stop_words=all_stopwords, ngram_range=(1, 2) umap_model=umap_model, hdbscan_model
Ekonomi	umap_model = UMAP(n_neighbors=8, n_components=8, random_state=42)	hdbscan_model = HDBSCAN(min_cluster_size=30, min_samples=4, prediction_data=True)	Daha küçük veri sayısı, orta büyüklükte kümeler, konu farklılıkları fazla	

Sağlık	umap_model = UMAP(n_neighbors=10, n_components=8, random_state=42)	hdbscan_model = HDBSCAN(min_cluster_size=30, min_samples=4, rediction_data=True)	Daha küçük veri sayısı, orta büyüklükte kümeler, konu farklılıkları oldukça az düzeyde.	=hdbscan_model, calculate_probabilities=True, nr_topics="auto", language="turkish")
Kültür	umap_model = UMAP(n_neighbors=10, n_components=5, random_state=42)	hdbscan_model = HDBSCAN(min_cluster_size=15, min_samples=10, rediction_data=True)	En küçük veri sayısı, küçük kümeler, konu farklılıkları az düzeyde.	

Kaynak: Yazar Tarafından Derlenmiştir.

4.4.5. Modelin Çalıştırılması ve Sonuçlar

Aşağıda (bkz. Şekil 20) bir örneği verilen python kodu kullanılarak model siyaset, ekonomi, sağlık ve kültür etiketi taşıyan veriler için ayrı ayrı çalıştırılmıştır.

Şekil 20: Modelin Çalıştırılması

```
# Gerekli kütüphanelerin yüklenmesi
import pandas as pd
import numpy as np
from umap import UMAP
from hdbscan import HDBSCAN
from bertopic import BERTopic
from sklearn.feature_extraction.text import CountVectorizer
from bertopic.vectorizers import ClassTfidfTransformer

# Veri kümesinin yüklenmesi
df = pd.read_csv("../veri.csv", parse_dates=['publish_date'])

# Daha önce çıkarılmış olan gömmelerin yüklenmesi
embeddings = np.load(r"[DOSYA_YOLU]/all_embeddings.npy")

# Önceden oluşturulmuş stopwords listelerinin yüklenmesi
with open(r"[DOSYA_YOLU]/stopwords_generic.txt", encoding="utf-8") as f:
    generic = f.read().splitlines()
with open(r"[DOSYA_YOLU]/locations.txt", encoding="utf-8") as f:
    locations = f.read().splitlines()
with open(r"[DOSYA_YOLU]/organizations.txt", encoding="utf-8") as f:
    organizations = f.read().splitlines()
with open(r"[DOSYA_YOLU]/people.txt", encoding="utf-8") as f:
    people = f.read().splitlines()

# Tüm stopwords listelerinin birleştirilmesi
all_stopwords = list(set(generic + locations + organizations + people))

# Belirlenen kategoriye ait metinlerin ve gömmelerin filtrelenmesi
target_docs = df.loc[df['predicted_category'] == 'target_category', 'clean_text'].tolist()
target_embeddings = embeddings[df['predicted_category'] == 'target_category']

# CountVectorizer modelinin tanımlanması
```

```

vectorizer_model = CountVectorizer(
    stop_words=all_stopwords,
    ngram_range=(1, 2)
)

# UMAP boyut indirgeme modelinin tanımlanması
umap_model = UMAP(
    n_neighbors=10,
    n_components=8,
    random_state=42
)

# HDBSCAN kümeleme modelinin tanımlanması
hdbscan_model = HDBSCAN(
    min_cluster_size=30,
    min_samples=4,
    prediction_data=True
)

# BERTopic modelinin tanımlanması
topic_model = BERTopic(
    vectorizer_model=vectorizer_model,
    umap_model=umap_model,
    hdbscan_model=hdbscan_model,
    calculate_probabilities=True,
    verbose=True,
    nr_topics="auto",
    language="turkish",
)

# BERTopic modelinin eğitilmesi ve konu olasılıklarının çıkarılması
topics, probs = topic_model.fit_transform(target_docs, target_embeddings)

# Konu bilgileri tablosunun çıkarılması
topic_info = topic_model.get_topic_info()

```

Kaynak: Yazar tarafından derlenmiştir.

Kodun verideki her bir alt küme için ayrı ayrı çalıştırılması sonucunda BERTopic'in "get_topic_info()" fonksiyonu kullanılarak algoritmanın oluşturduğu konu modelleme tablosu görüntülenmiş, örtüşen kümeler tespit edilerek bu kümeler birleştirilmiştir. BERTopic modelleme fonksiyonu her çağrıldığında, "-1" numaralı, aykırı konuların (*outlier*) toplandığı bir küme üretmektedir. Bu kümedeki dokümanlar için bir olasılık skoru hesaplanmış ve BERTopic algoritmasının "reduce_outliers()" fonksiyonu, "*distributions*" ve "*c-tf-idf*" parametreleri ile zincirleme olarak çağrılarak bu dokümanlar ilgili konulara dağıtılmıştır.

Daha sonra elde edilen konu modelleme tablosundaki temsilci kelimeler (*Representation*) ve temsilci dokümanlar (*Representative_Docs*) incelenerek kategori bazında konulara isim verilmiştir. Çalışmada yer alan konu modelleme sonuçlarına, her bir kategori için yapılan deneylerde optimum sayıda anlamlı konular elde edilinceye kadar deneme yanılma yöntemi ile parametreler değiştirilerek varılmıştır (Optimum sonuçların alındığı parametreler için bkz. Tablo 19). Her bir kategori için modelin konuları ne derece ayırabildiğini değerlendirmek için konu-benzerlik matrisi

hazırlanmış, çeşitlilik (*diversity score*) ve tutarlılık (*coherence score*) skoru hesaplanmıştır (Sawant ve diğerleri, 2022).

Bu metriklerin ilki olan benzerlik matrisi konular arası ne derece örtüşme olduğunu göstermektedir. Tüm kategoriler için ayrı ayrı oluşturulan benzerlik matrisleri konu ayrımlarının model tarafından oldukça başarılı bir biçimde gerçekleştirdiğine işaret etmektedir. Modelin bu alandaki başarısı, ekonomi hariç her konunun temsilci kelimeleri ile hesaplanan yüksek ($>0,85$) çeşitlilik skorları ile de desteklenmektedir. Tutarlılık skoru ise her bir konunun temsilci kelimeleri arasında anlamsal bir bütünlük olup olmadığını ölçen temel metriklerden biridir. Tüm kategoriler için orta düzeyde ($\sim 0.50-0.55$) seyreden bu skorun haber dilinin betimleyici değil bilgi iletici olması ve aynı kavramların farklı biçimlerde ifade edilmesi, BERTopic algoritmasında kötü sonuç verdiği için kelime köküne indirgeme (*lemmatizing*) işlemi yapılmamış olması, Türkçe'nin sondan eklemeli bir dil olması ve kullanılan cümle gömme modelinin Türkçe'ye özel olmaması gibi sebeplerden kaynaklandığı değerlendirilmektedir. Bu bağlamda, ulaşılan değerler modelin anlamsal ayırt ediciliğini koruduğunu ve çalışmanın analitik amaçları için yeterli düzeyde bütünlük sağladığını göstermektedir.

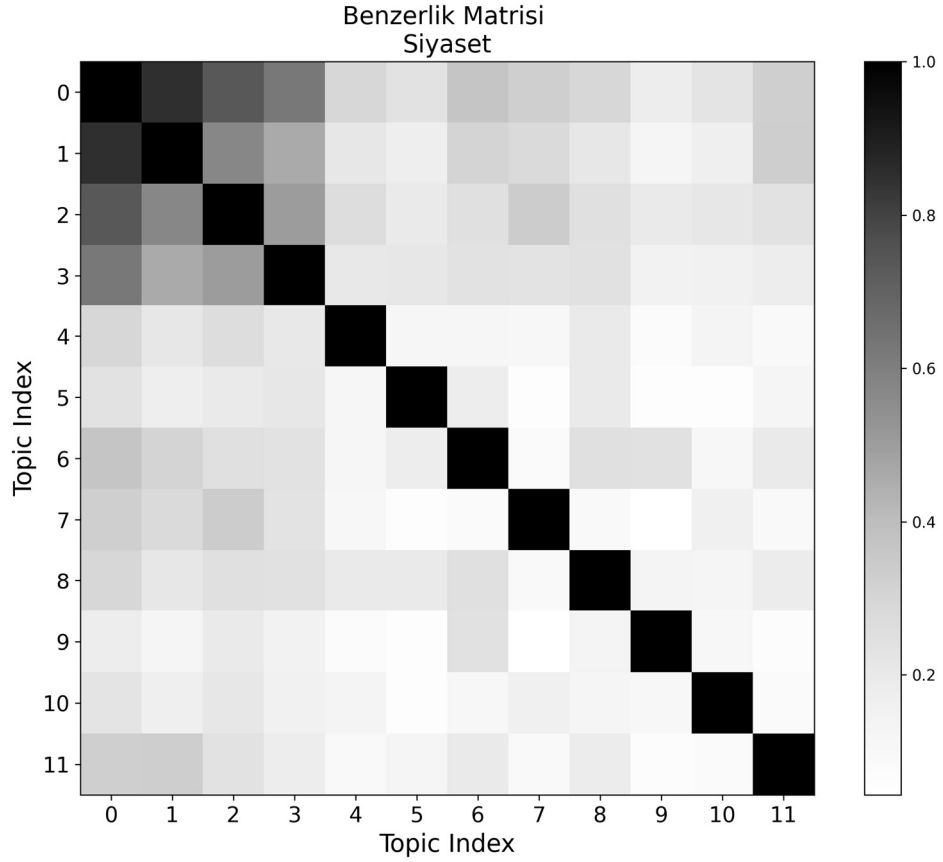
Birinci sırada doküman sayısına sahip “siyaset” kategorisi için yapılan konu modelleme sonucunda hiç bir kümeye ait olmayan ve algoritma tarafından “-1” numaralı konu olarak tanımlanan ve kategorideki dokümanların %14,95'ini içeren küme hariç 11 adet tema belirlenmiş ve bu temalar aşağıdaki tabloda görülebileceği gibi adlandırılmıştır. 13.553 dokümandan oluşan bu alt küme için konu modelleme sonuçları tablosu ve konu benzerlik matrisi aşağıda verilmiştir. Bu kategorinin çeşitlilik skoru 0,8750 olarak hesaplanmıştır. Çeşitlilik skorunun yüksek olması konuların birbirinden ayrışmasının anlamlı olduğunu göstermektedir. Tutarlılık skoru ise 0,5325 olarak hesaplanmıştır.

Tablo 20: Siyaset Kategorisi Konu Modelleme Sonuçları

No	Verilen İsim	Oran %	Kısa İsim	Temsilci Kelimeler	Sayı
0	Diplomatik Görüşmeler, Ortaklıklar ve Uluslararası Güvenlik Sorunları	40,67	Diplomasi ve Uls. Güvenlik	'ortak', 'güvenlik', 'ateşkes', 'savunma', 'mutabakat', 'bölgesel', 'barış', 'işbirliği', 'ziyaret', 'bağlı'	5512
1	Darbeler ve Askeri Müdahaleler	12,68	Darbe	'darbe', 'güvenlik', 'anayasa', 'terör', 'görevden', 'askıya', 'silahlı', 'saldırıda', 'asker', 'saldırı'	1719
2	İç Karışıklıklar, İsyan ve Şiddet	8,21	İç Karışıklıklar	'isyancı', 'çatışmalar', 'yerinden', 'soykırımı', 'soykırım', 'dışı', 'binlerce', 'silahlı', 'kuvvet', 'sınır'	1113
3	Doğal Afetler Karşısında Yardımlaşma	4,01	Doğal Afetler	'yardım', 'kurtarma', 'arama kurtarma', 'arama', 'sel', 'insani', 'deprem', 'yaralıları', 'tıbbi', 'tahliye'	544
4	Savunma Sanayii İşbirliği	3,96	Sav. San. İşbrl.	'hava', 'deniz', 'savunma', 'deniz görev', 'görev grubu', 'görev', 'siha', 'tatbikatın', 'helikopter', 'grubu'	537
5	Siyasi Yöne Sahip Adli Olaylar	3,76	Adli Olaylar	'kusurlu', 'motokurye', 'şüphelinin', 'adli', 'polis', 'soruşturma', 'serbest', 'gözaltına', 'şüpheli', 'ölümüne'	510
6	Akdeniz'de Düzensiz Göç ve Göçmen Krizleri	3,34	Düzensiz Göç	'düzensiz', 'göçmen', 'göçmenlerin', 'göçmenin', 'göç', 'sahil', 'teknenin', 'sahil güvenlik', 'taşıyan', 'ulaşmak'	453
7	Türkiye'den Terörü Kınama ve Taziye Mesajları	3,04	Terör, Taziye, Kınama	'üzüntüyle', 'terör', 'başsağlığı', 'taziyelerimizi', 'diliyoruz', 'halkına', 'kaybedenlerin', 'içten taziyelerimizi', 'içten', 'kardeş'	412
8	Afrika'da Seçimler ve Siyasi Dönüşümler	2,86	Seçimler	'oy', 'seçim', 'seçimlerde', 'aday', 'anayasa', 'adayı', 'sandalye', 'oyların', 'seçimde', 'ikinci'	388
9	Afrika'da Büyük Kazalar ve Felaketler	1,62	Kaza ve Felaketler	'uçacağın', 'sonucu', 'kaçak', 'iniş', 'taşıyan', 'kazada', 'yolcu', 'kurtarma', 'soruşturma', 'kazanın'	220
10	Afrikanlı Siyasi Liderlerin Ölümleri	0,88	Lider Ölümleri	'tedavi', 'hastanede', 'cenaze', 'tedavi gördüğü', 'toure', 'gördüğü', 'gördüğü hastanede', 'mehdi', 'kalp', 'çatışmada'	119

Kaynak: Yazar tarafından derlenmiştir.

Şekil 21: Siyaset Kategorisi Konu-Benzerlik Matrisi



Kaynak: Yazar tarafından derlenmiştir.

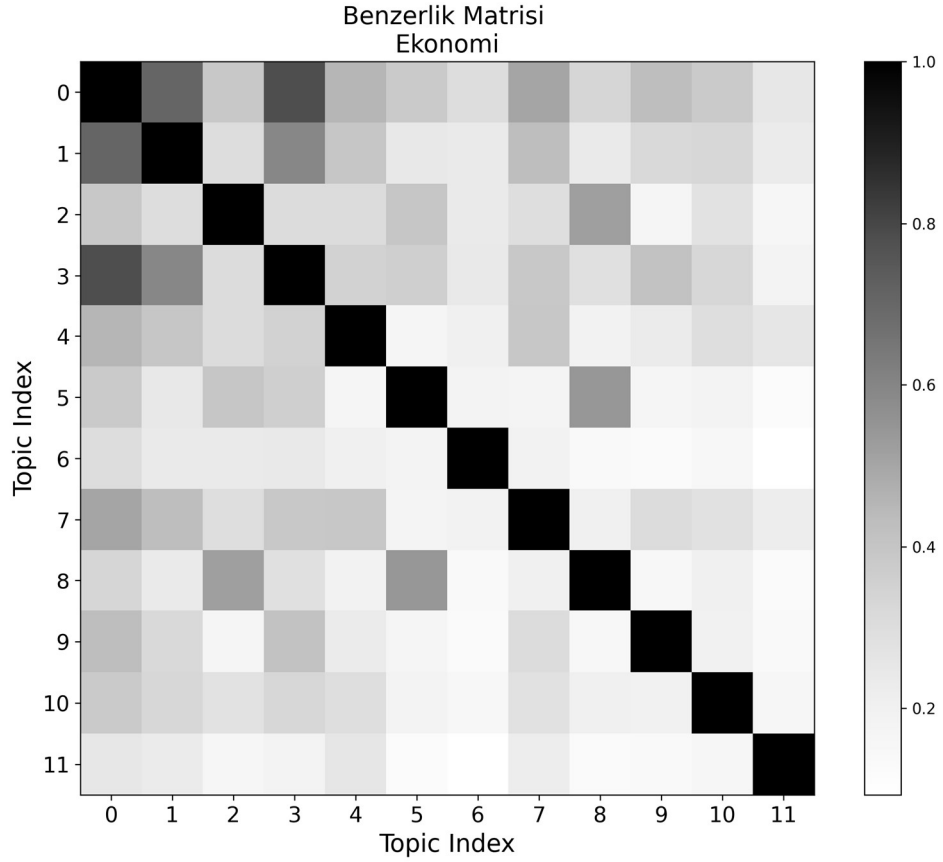
İkinci sırada doküman sayısına sahip “ekonomi” kategorisi için yapılan konu modelleme sonucunda hiç bir kümeye ait olmayan ve algoritma tarafından “-1” numaralı konu olarak tanımlanan ve kategorideki verilerin %4,91’ini içeren küme hariç 11 adet tema belirlenmiş ve bu temalar aşağıdaki tabloda görülebileceği gibi adlandırılmıştır. 2.682 dokümandan oluşan bu alt küme için konu modelleme sonuçları tablosu ve konu-benzerlik matrisi aşağıda verilmiştir. Bu kategorinin çeşitlilik skoru 0,6917 olarak hesaplanmıştır. Çeşitlilik skorunun siyaset kategorisine göre daha düşük olması ekonomi haberlerinin çok sayıda benzer kelime içermesi ve konuların doğan olarak birbirlerine oldukça yakın olmasından kaynaklanmaktadır. Dokümanlara atanan konuların manuel incelenmesi konuların birbirinden ayrışmasının anlamlı olduğunu göstermektedir. Tutarlılık skoru ise 0,5904 olarak hesaplanmıştır.

Tablo 21: Ekonomi Kategorisi Konu Modelleme Sonuçları

No	Verilen İsim	Oran %	Kısa İsim	Temsilci Kelimeler	Sayı
0	Türkiye için bir Pazar Olarak Afrika	22,11	Afrika Pazarı	'ticaret', 'ihracat', 'dolar', 'savunma', 'dolarlık', 'yatırım', 'ekonomik', 'dolara', 'hava', 'dolarla'	581
1	Türkiye - Afrika Ekonomik İşbirliği	21,92	Ekon. İşbr. TR+AFR	'ticaret', 'mutabakat', 'ekonomik', 'yatırım', 'işbirliği', 'ortak', 'alanında', 'eğitim', 'mutabakat zaptı', 'zaptı'	576
2	Enerji Zengini Afrika ve Enerji Krizleri	16,67	Enerji + Krizler	'petrol', 'elektrik', 'enerji', 'varil', 'gaz', 'ham', 'barajın', 'su', 'sahip', 'boru'	438
3	Afrika Tahıl Krizi	8,41	Tahıl Krizi	'tahıl', 'gıda', 'dolar', 'buğday', 'ton', 'ekonomik', 'para', 'tarım', 'dolarlık', 'merkez'	221
4	Türkiye Afrika Enerji İşbirliği	6,66	Enerji İşbr. TR+AFR	'enerji', 'gaz', 'petrol', 'bayraktar', 'doğal', 'doğal gaz', 'reis', 'oruç reis', 'oruç', 'sismik'	175
5	Ticaret Yollarının Kesişiminde Afrika	4,49	Ticaret Yolları	'kanalı', 'geminin', 'ever given', 'ever', 'given', 'deniz', 'gemi', 'kanal', 'idaresi', 'dev'	118
6	Afrika'nın Ekonomik Sorunları	4,34	AFR Ekon. Prob.	'para', 'ekonomik', 'küresel', 'sahip', 'dış', 'güvenlik', 'raporda', 'mücadele', 'dolar', 'ülkeler'	114
7	Afrika - Diğer Ülkeler Enerji İşbirliği	3,96	Enerji İşbr. AFR+Diğer	'gaz', 'petrol', 'doğal', 'doğal gaz', 'enerji', 'boru', 'doğalgaz', 'metreküp', 'hattı', 'noc'	104
8	Afrika - Çin Ekonomik İşbirliği	3,46	Ekon. İşbr. AFR+Çin	'işbirliği', 'focac', 'kalkınma', 'işbirliği forumu', 'küresel', 'ortak', 'forumu', 'işbirliğini', 'stratejik', 'ekonomik'	91
9	Afrika - Diğer Ülkeler Ekonomik İşbirliği	1,71	Ekon. İşbr. AFR+Diğer	'ton', 'ton ton', 'ticaret', 'ekonomik', 'dolar', 'petrol', 'darbe', 'yolu', 'usadf', 'doge'	45
10	Türkiye Afrika Ekonomik Kıyaslamaları	1,37	TR-AFR Kıyaslama	'puan', 'kredi', 'notu', 'değer', 'sırada', 'fitch', 'endeksi', 'le', 'lirası', 'gelir'	36

Kaynak: Yazar tarafından derlenmiştir.

Şekil 22: Ekonomi Kategorisi Konu-Benzerlik Matrisi



Kaynak: Yazar tarafından derlenmiştir.

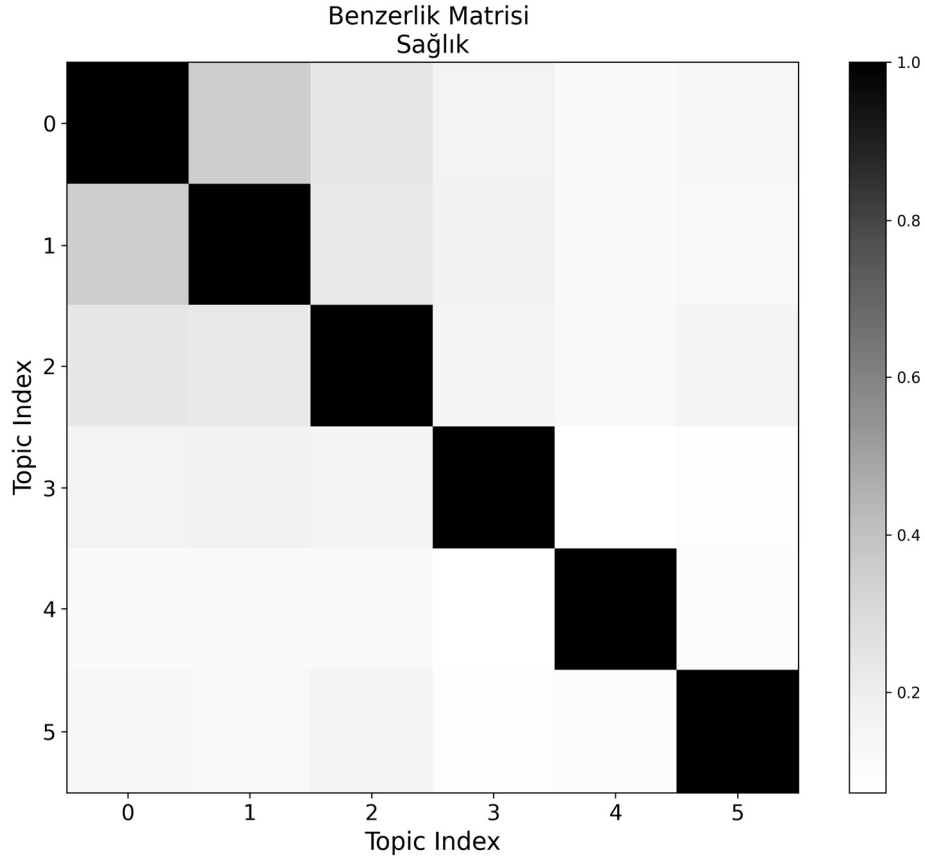
Üçüncü sırada doküman sayısına sahip “sağlık” kategorisi için yapılan konu modelleme sonucunda hiç bir kümeye ait olmayan ve algoritma tarafından “-1” numaralı konu olarak tanımlanan ve kategorideki verilerin %10,87’sini içeren küme hariç 5 adet tema belirlenmiş ve bu temalar aşağıdaki tabloda görülebileceği gibi adlandırılmıştır. 2.402 dokümandan oluşan bu alt küme için konu modelleme sonuçları tablosu ve konu benzerlik matrisi aşağıda verilmiştir. Bu kategorinin çeşitlilik skoru 0,85 olarak hesaplanmıştır. Çeşitlilik skorunun yüksek olması konuların birbirinden ayrışmasının anlamlı olduğunu göstermektedir. Tutarlılık skoru ise 0,5761 olarak hesaplanmıştır.

Tablo 22: Sağlık Kategorisi Konu Modelleme Sonuçları

No	Verilen İsim	Oran %	Kısa İsim	Temsilci Kelimeler	Sayı
0	Salgın Hastalıklar (Covid-19, M Çiçeği, Kolera)	54.87	Salgın Hastalık	'vaka', 'kovid', 'sağlık', 'tespit', 'sayısının', 'sayısı', 'virüs', 'toplam', 'koronavirüs', 'aşı'	1318
1	Doğal Afetler ve Kıtılık	12.68	Kıtılık + Afet	'gıda', 'sel', 'yardım', 'insani', 'şiddetli', 'yerinden', 'su', 'bölgede', 'karşıya', 'kuraklık'	522
2	Türkiye'nin Sağlık Alanında Afrika'ya Yardımları	5.66	Sağlık Yardımı	'sağlık', 'tıbbi', 'thy', 'tika', 'yardım', 'dostluk', 'hastanesi', 'alanında', 'tıbbi yardım', 'tedavi'	136
3	Afrika Sıcakları Kaynaklı Sağlık Riskleri	4.33	Sıcaklık + Risk	'hava', 'derece', 'bulutlu', 'sıcak', 'mevsim', 'meteoroloji', 'sıcak hava', 'sıcaklıkların', 'bekleniyor', 'normallerinin'	104
4	Yangın Kaynaklı Ölümler	2.54	Yangın	'orman', 'yangın', 'yangının', 'yangında', 'yangınında', 'alev', 'söndürme', 'yangınlarında', 'orman yangınlarında', 'binada'	61

Kaynak: Yazar tarafından derlenmiştir.

Şekil 23: Sağlık Kategorisi Konu-Benzerlik Matrisi



Kaynak: Yazar tarafından derlenmiştir.

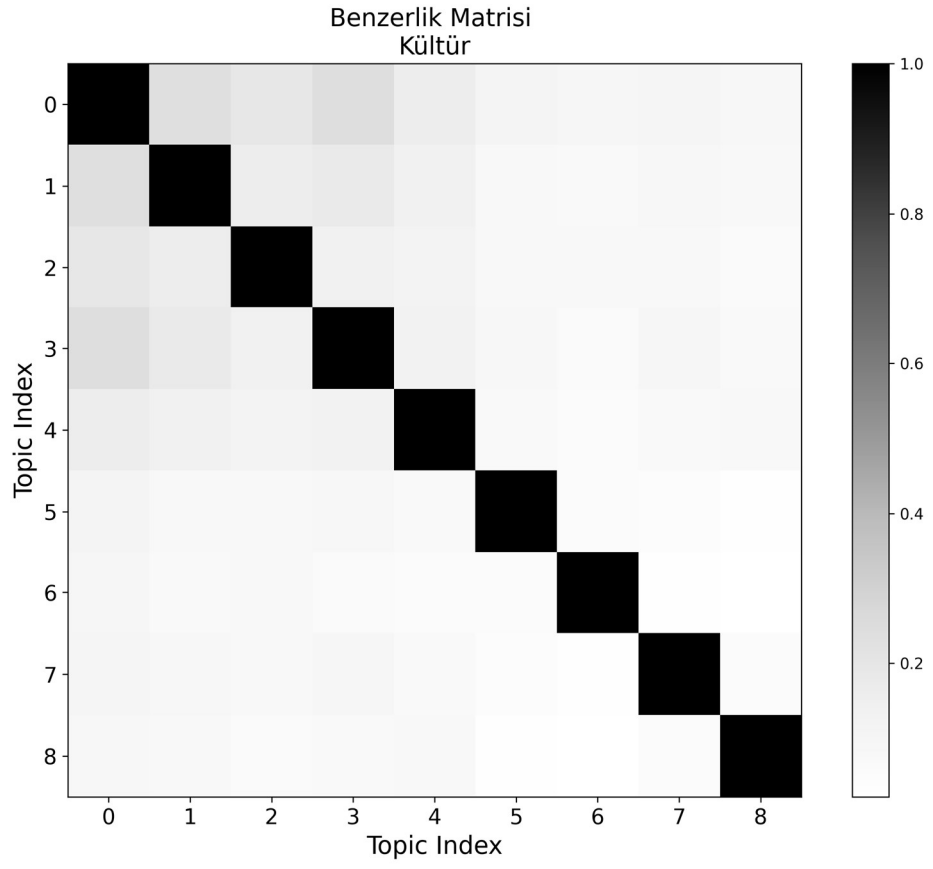
Dördüncü ve son sırada doküman sayısına sahip “kültür” kategorisi için yapılan konu modelleme sonucunda hiç bir kümeye ait olmayan ve algoritma tarafından “-1” numaralı konu olarak tanımlanan ve kategorideki verilerin %9,6’sını içeren küme hariç 8 adet tema belirlenmiş ve bu temalar aşağıdaki tabloda görülebileceği gibi adlandırılmıştır. 1.385 dokümandan oluşan bu alt küme için konu modelleme sonuçları tablosu ve konu-benzerlik matrisi aşağıda verilmiştir. Bu kategorinin çeşitlilik skoru 0,9444 olarak hesaplanmıştır. Çeşitlilik skorunun oldukça yüksek olması konuların birbirinden ayrışmasının anlamlı olduğunu göstermektedir. Tutarlılık skoru ise 0,5726 olarak hesaplanmıştır.

Tablo 23: Kültür Kategorisi Konu Modelleme Sonuçları

No	Verilen İsim	Oran %	Kısa İsim	Temsilci Kelimeler	Sayı
0	Türkiye Afrika Kültürel İlişkileri	33.57	Kültürel İlişk.	'kültür', 'eğitim', 'osmanlı', 'evi', 'kültürel', 'trt', 'efendi', 'tika', 'medya', 'kültür evi'	465
1	Ülke Profilleri ve Turistik Destinasyonlar	18.27	Ülke Profili	'para', 'saat', 'nüfusu', 'başkenti', 'renk', 'para birimi', 'birimi', 'farkı', 'nedir', 'kırmızı'	253
2	Afrika Doğal Hayatı ve Kaçak Avcılık	15.67	Doğal Hayat	'aslan', 'koruma', 'fil', 'güzellik', 'miss', 'polis', 'kaçak', 'kuş', 'gergedan', 'köpek'	217
3	Antik Mısır - Gizem	11.12	Antik Mısır	'antik', 'eski', 'inşa', 'piramit', 'piramitleri', 'piramidi', 'firavun', 'giza', 'piramitlerin', 'tarihi'	154
4	Afrika Mutfağı	4.04	Mutfak	'kaşığı', 'bardağı', 'su', 'su bardağı', 'yemek', 'tarifi', 'lif', 'çay kaşığı', 'çay'	56
5	Afrika Sıcakları	2.67	Sıcaklar	'çay', 'narenciye', 'mandalar', 'manda', 'zarar', 'sıcak', 'tanyıldızı', 'aşırı', 'hasat', 'verim'	37
6	Afrika Elmasları	2.53	Elmaslar	'elmas', 'karatlık', 'lirası', 'elmasın', 'bitcoin', 'dünyanın', 'organize', 'elmaslar', 'lucara', 'lirası karşısında'	35
7	Afrika Eserlerinin İadesi	2.53	Eser İadesi	'iade', 'eserlerin', 'kraliçe', 'kaçırılan', 'kraliyet', 'prens', 'meghan', 'british museum', 'british', 'museum'	35

Kaynak: Yazar tarafından derlenmiştir.

Şekil 24: K lt r Kategorisi Konu-Benzerlik Matrisi



Kaynak: Yazar tarafından derlenmiřtir.

SONUÇ

Türkiye’de yayınlanan Afrika haberlerinin kıtayı ve bu kıtadaki ülkeleri ne şekilde ele aldıklarının tespitini amaçlayan bu çalışma kapsamında öncelikle Media Cloud API yardımı ile aramalar yapılmış, Afrika ve Afrika ülkelerine dair haberlerin bağlantılarına ulaşılmıştır. Daha sonra bu bağlantılar kullanılarak web kazıma tekniği ile haber metinleri çekilmiştir. Veri temizleme aşamaları sonucunda haber metinlerinden 19.968 satırlık veri seti oluşturulmuştur. Bu veri seti öncelikle ince ayarlı bir BERTurk modeliyle kategorilere ayrılmıştır. Tespit edilen siyaset, ekonomi, sağlık, kültür ve spor kategorilerinden spor hariç her bir kategori için ayrı parametreler belirlenerek BERTopic tekniği ile veriye konu modelleme uygulanmıştır.

Elde edilen temaların analiz edilebilmesi amacı ile çalışmada, her bir kategori içerisindeki temaların oranları ve dağılımları yoğunluk haritası ile görselleştirilmiştir. Konu modelleme yöntemiyle elde edilen bulguların genel görünüşüne bakıldığında, Türkiye’de yayınlanan Afrika haberlerinin bu çalışmada incelenen süre zarfında bazı tematik örüntüler etrafında şekillendiği görülmektedir. Genel analiz sonuçları, Afrika kıtasının çoğunlukla sıcak, uzak, doğal afetler ve siyasi istikrarsızlıkların sık rastlandığı bir bölge olarak çerçevelendiğini, ancak aynı zamanda ekonomik fırsatlar ve ticaret potansiyeli ile de ilişkilendirildiğini göstermektedir. Bu tematik yapının zamansal dalgalanmalar gösterse de, ana hatlarıyla çalışılan dönem boyunca görece sabit kaldığı gözlemlenmiştir.

Politik temalar ile ekonomi ve sağlık kategorileri arasında zaman zaman belirgin kesişimler görülmektedir. Örneğin tahıl krizi ya da altyapı yetersizliklerinden kaynaklanan sağlık sorunları ve ölümler hem ekonomik hem de insani yardım söylemleriyle ilişkilendirilmiştir. Bu durum, haber konularının çoğunlukla çok boyutlu bir yapıya sahip olduğunu ve olayların yalnızca tek bir perspektiften sunulmadığını göstermektedir. Öte yandan genel olarak temaların ele alınış oranında artış ve düşüş ikili ilişkilere çok bağımlı seyretmekte ve Türkiye’nin yakın coğrafyası yahut sıkı siyasi ilişkilerinin bulunduğu ülkelerin gündemleri veyahut uluslararası siyasetteki durumları temalarda belirleyici olmaktadır.

Konu modelleme sonuçları Türkiye’deki medya organlarının Afrika’yı belli kalıp temalarla sunma eğiliminde olduğunu, ancak bu temaların zamana, olaylara ve ikili ilişkilere bağlı olarak değişkenlik gösterdiğini ortaya koymaktadır. Medyanın hangi haberleri yayımlayacağına dair tercihleri ve tematik vurgularında seçici bir davranış sergilediği, bazı temaların (örneğin adli olaylar veya magazinsel içerikler) belirli medya

kuruluşlarına özgü kaldığı anlaşılmaktadır. Bu durum, haber söyleminin hem kurum politikaları hem de kamuoyu algısıyla şekillendiğine bir işaret olarak değerlendirilebilir.

Tablo 24: Siyaset Kategorisi İlk 5 Konudan Örnekler

No	Konu İsmi (Kısaltılmış)	Başlık	Haber Metni (Kısaltılmış)
0	Diplomasi ve Uls. Güvenlik	Türkiye-Mısır müzakerelerinde ikinci tur ve ilişkilerin geleceği	Türkiye-Mısır ilişkilerinin yumuşayabileceğine ilişkin ilk sinyallerin kamuoyuna yansımalarının üzerinden bir yıl geçti. İstihbarat düzeyinde başlayan...
1	Darbe	Burkina Faso'da darbe girişimi başarısız oldu	Darbe söylentileri sonrası çok sayıda göstericinin sokağa inmesine sebep olan olaylarla ilgili açıklama yapan hükümet, Burkina Faso istihbarat ve güvenlik servisleri...
2	İç Karışıklıklar	BM: Sudan kabusun içinde sıkışıp kalmış durumda	Rosemary DiCarlo, El Cezire eyaletinde Hızlı Destek Kuvvetleri'nin (HDK) düzenlediği saldırı dalgasının, son 18 ayda yaşanan en şiddetli eylemler arasında...
3	Doğal Afetler	Güney Afrikalı STK'lar Türkiye'ye insani yardım ve arama kurtarma ekibi gönderdi	Gift of Givers ve SARZA tarafından oluşturulan 32 kişilik şehir arama kurtarma ekibi ile 1 tona yakın arama ekipmanı ve Al İmdat Vakfı tarafından sağlanan insani yardım...
4	Sav. San. İşbrl.	Bakan Akar, Cezayir Bağımsızlık Günü resepsiyonunda	Savunma Bakanı Hulusi Akar, "Türkiye ve Cezayir arasındaki iş birliği sadece ülkelerimize, silahlı kuvvetlerimize değil aynı zamanda bölge barışına da önemli...

Kaynak: Yazar tarafından derlenmiştir.

Siyasi haberlerde çerçeveler büyük oranda iç karışıklıklar, darbeler, düzensiz göç ve lider değişimleri gibi kriz temaları etrafında toplanmıştır. Diplomatik ilişkiler ve uluslararası güvenlik çerçeveleri ise daha çok Türkiye'nin kıta ülkeleriyle geliştirdiği siyasi temaslar ve resmi ziyaretler sırasında öne çıkmaktadır. "Siyaset" kategorisi için yapılan konu modelleme sonucunda "Diplomatik Görüşmeler, Ortaklıklar ve Uluslararası Güvenlik Sorunları" konusunun ağır bastığı ve %40,67 oranla birinci konu olarak algoritma tarafından belirlendiği görülmektedir. İncelenen verinin haber verisi olması nedeniyle özellikle diplomatik temaslar ve ikili ilişkiler konusunun ön planda olması olağan olarak değerlendirilmiştir. Ancak, bu konu altında gruplanan haber metinleri incelendiğinde bu metinlerin ağırlıklı olarak uluslararası güvenlik temasına yoğunlaştıkları görülmektedir. Bunun yanı sıra ikinci sırada %18,68'lik oranla "Darbeler ve Askeri Müdahaleler" çerçevesi ön plana çıkmıştır. Bu da özellikle

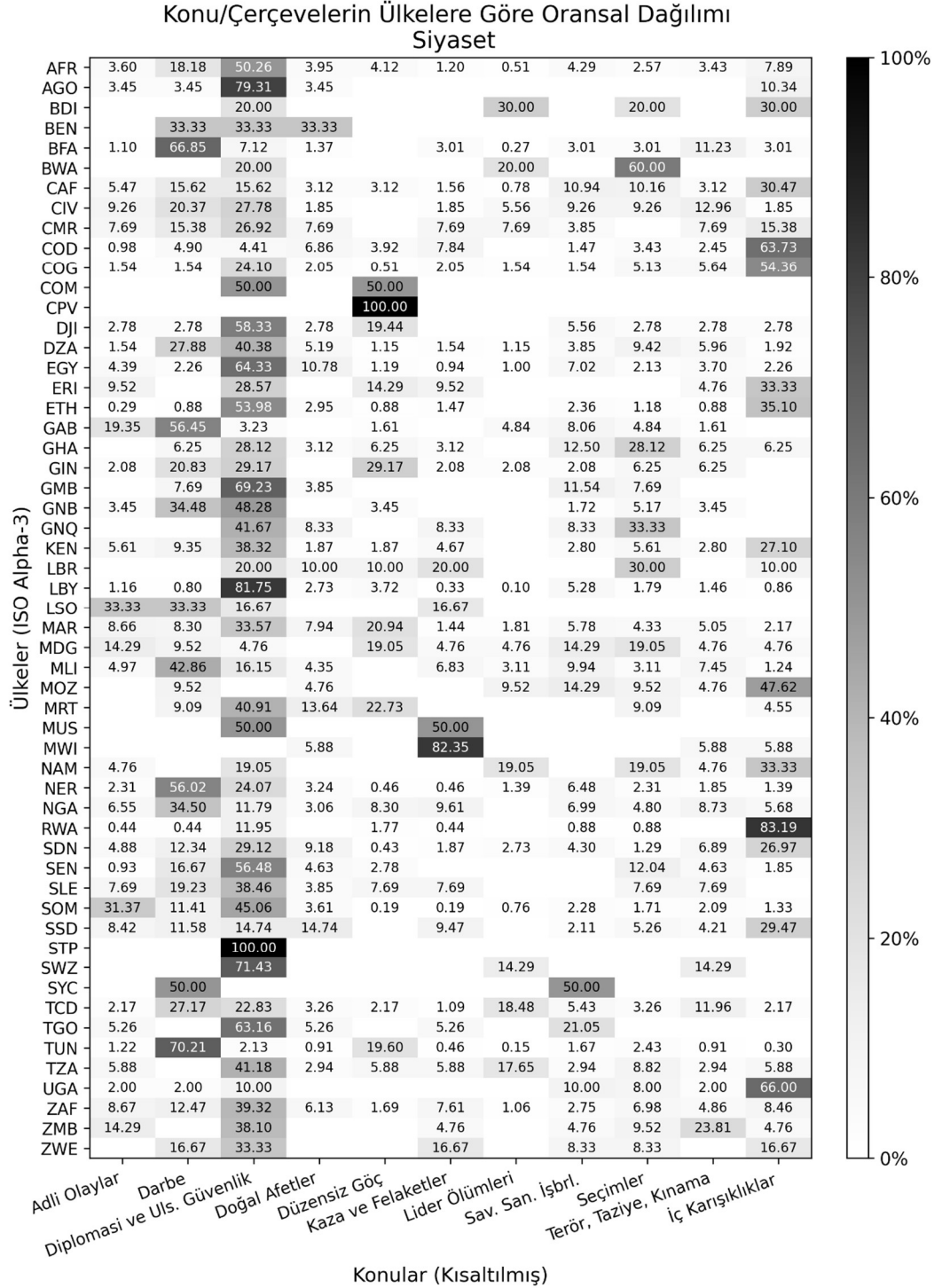
Afrika'da darbeler kuşağı olarak adlandırılan bölgede 2020'lerde görülen askeri darbe dalgasının doğrudan sonucu olarak değerlendirilebilir.

Bir başka ilginç bulgu ise “Siyasi Yöne Sahip Adli Olaylar” olarak adlandırılan çerçevedir. Türkiye-Afrika ilişkilerinin gelişmesi sebebi ile iki ülke arasındaki etkileşimin artması, özellikle Türk basınında Afrika ülkelerinin vatandaşlarının ya da Afrikalı siyasilerin ve ailelerinin karıştıkları adli olaylar bir çeşit sorgulama ile birlikte kendisine oldukça geniş yer bulmaktadır. “Siyaset” kategorisi olarak etiketlenmiş %3,76'lık verinin bu tema altında algoritma tarafından kümelenmesi bunun bir göstergesi olarak değerlendirilebilir.

2020'lerden itibaren hem Avrupa hem de Türkiye için gittikçe daha fazla önem kazanan bir diğer konu olan “Düzensiz Göç ve Göçmen Krizi” olarak adlandırılan çerçevenin de algoritma tarafından %3,34'lük pay ile ön plana çıkarıldığı görülmektedir. Bu konuya ait haberler incelendiğinde bu konunun bir insani kriz ve güvenlik meselesi kesişiminde medya tarafından değerlendirildiği görülmektedir. Bir diğer ilgi çekici konu ise son yıllarda Türkiye'nin giderek artan derecede gündeminde olan savunma sanayindeki gelişmelerin ve bu gelişmelerin Türkiye - Afrika ikili ilişkilerine yansımaları haberleştiren verilerin yer aldığı temadır. “Savunma Sanayii İşbirliği” olarak adlandırılan bu küme, genellikle Türkiye'nin Afrika'ya savunma sanayii ürünleri ihracatını ve Afrikalı ülkelerin bu ürünlere olan ilgisini konu almaktadır.

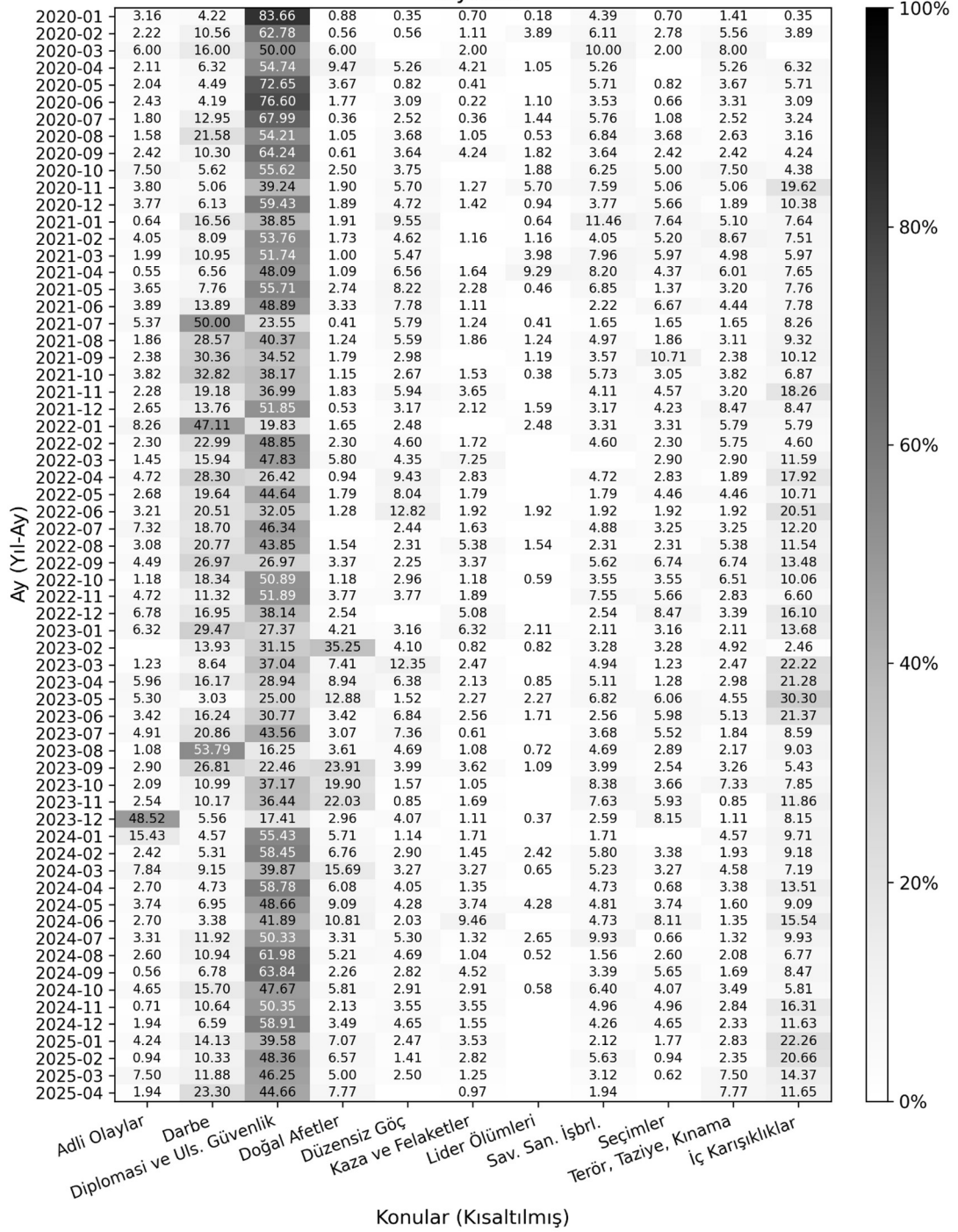
Bu kategori özelinde çalışmanın alt araştırma sorularından biri olan konu-ülke kısıtlımı incelendiğinde, Türkiye'yi yakından ilgilendiren Orta Doğu'daki çatışma ortamının Mısır, Libya iç savaşının ise Libya ile ilgili haberlerde dominasyonu olmasından hareketle, bu ülkelerde diplomatik ilişkiler ve uluslararası güvenlik çerçevesinin baskınlığı dikkat çekmektedir. Darbe ve iç karışıklıkların olduğu ülkeler de, konu modelleme analizinin verdiği sonuçlar ile doğrudan tablodan tespit edilebilmektedir. İncelenen tarih aralığında Somali Cumhurbaşkanı'nın oğlunun Ankara'da karıştığı trafik kazası neticesinde bir kişinin ölümüne sebep olması, Sudan Cumhurbaşkanı'nın oğlunun ise yine bir trafik kazasında hayatını kaybetmesi “adli olaylar” çerçevesinin bu iki ülke özelinde yüksek oranda temsilini açıklamaktadır. Botswana'da 2024 yılında yapılan seçimlerde ilk kez iktidar değişikliğinin yaşanmasının görece sıklıkla medyada yer alması da “seçimler” çerçevesinin bu ülke ile ilgili haberlerde baskın çıkmasına neden olmuştur.

Şekil 25: Siyaset Kategorisi Ülke/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Şekil 26: Siyaset Kategorisi Zaman/Konu Dağılımı Isı Haritası
Konuların Zamana Göre Oransal Dağılımı
Siyaset



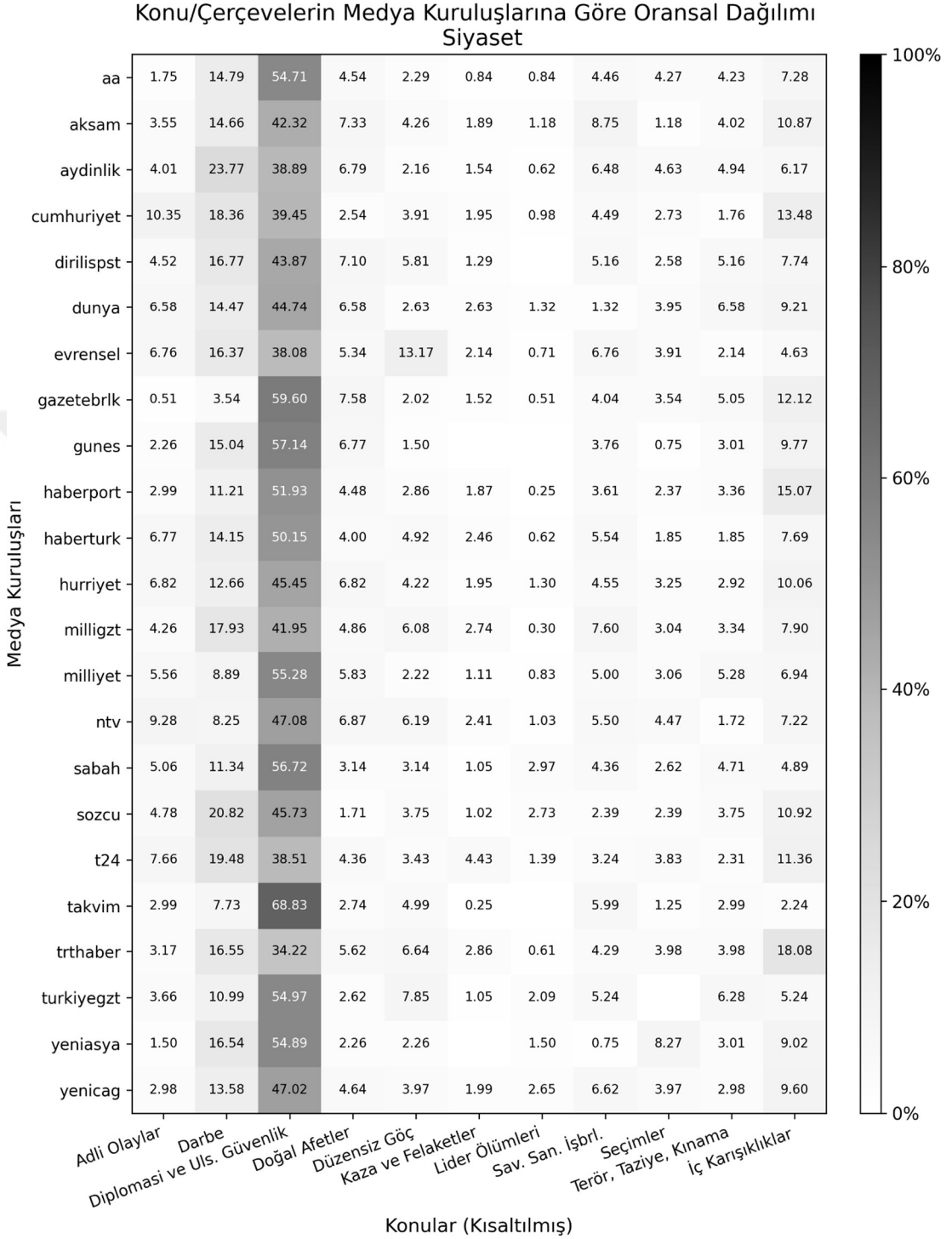
Kaynak: Yazar tarafından derlenmiştir.

Konu-zaman kırılımı incelendiğinde siyaset kategorisi özelinde gündelik siyasetin medyanın konuları ele alış sıklığında ve şeklinde doğrudan etkisi olduğunu söylemek mümkündür. Özellikle lider ziyaretleri ve ikili diplomatik görüşmelerin olduğu

tarihler, yahut bir Afrika ülkesinde iç karışıklık/darbe/doğal afet gibi olayların yaşandığı zamanlar doğrudan çerçevelere ve gündeme taşınan konulara etki etmektedir. İncelenen veri haber verisi olduğu ve asıl amacının “haber iletmek” olduğu göz önüne alındığında, bu zaman konu dağılımının bir hayli olağan olduğu değerlendirilebilir. Ancak bir diğer açıdan bakıldığında, medyanın kıtaya ve tek tek ülkelere olan ilgisinin bu ülkelerin ya da kıtanın Türkiye’nin radarına girmesi ile doğrudan ilgisi olduğu değerlendirilebilir.

Konu-medya kuruluşu kırılımı incelendiğinde, “darbe” ve “diplomasi - uluslararası güvenlik” temalarının birçok haber kuruluşunda baskın olduğu gözlenmektedir. “Düzensiz göç” temasının, az sayıda haber ile veri kümesinde yer almasına rağmen Evrensel Gazetesi’nde baskın temalardan biri olarak ön plana çıktığı dikkat çekicidir. Bunun yanı sıra “adli olaylar” temasının belli başlı bazı medya organlarında diğerlerine nazaran daha fazla yer aldığı açıkça görülmektedir. AA ve TRT Haber gibi resmi medya organlarının bazı temaları ön plana çıkardıkları tablodan okunabileceği gibi, bu iki devlet kaynaklı medya kuruluşunda “iç karışıklıklar” çerçevesinin farklı oranlarda ele alınmış olması ilgi çekicidir.

Şekil 27: Siyaset Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Ekonomi kategorisinde elde edilen bulgular, Afrika'nın yoğun bir biçimde bir ticaret ortağı ve ihracat pazarı olarak vurgulandığını ortaya koymaktadır. Türkiye ile Afrika ülkeleri arasındaki ticari işbirlikleri, enerji anlaşmaları ve altyapı projeleri bu kapsamda öne çıkan alt temalar arasında yer almaktadır.

Tablo 25: Ekonomi Kategorisi İlk 5 Konudan Örnekler

No	Konu İsmi (Kısaltılmış)	Başlık	Haber Metni (Kısaltılmış)
0	Afrika Pazarı	Türkiye ve Mısır arasında yaşanan olumlu gelişmeler önemli fırsatlar doğurabilir	Türkiye ile Mısır arasında son dönemde yaşanan olumlu siyasi gelişmeler in sürmesi iki ülke arasında ekonomik ve ticari ilişkilerin geliştirilmesi açısından yeni ve önemli fırsatlara kapı açabilir. AA muhabirinin derlediği bilgilere göre...
1	Ekon. İşbr. TR+AFR	Müsiad Konya Zimbabwe yerel yönetimler bakanını ağırladı	Bir dizi temaslarda bulunmak üzere kente gelen Zimbabwe Yerel Yönetimler Bakanı Hon July Moyo ve beraberindeki heyet MÜSİAD Konya Şubesi'ni ziyaret etti. Zimbabwe Yerel Yönetimler Bakanı Hon July Moyo, Parlamenter ve Madencilik...
2	Enerji + Krizler	Uganda nükleer enerji üretimine 2031'de başlayacak	Uganda Enerji ve Doğal Kaynaklar Bakanı Ruth Nankabirwa Ssentamu, yaptığı açıklamada, Çin Ulusal Nükleer Şirketi (CNNC) ortaklığında inşa edilecek Buyende Nükleer Enerji Santrali'nin 2031 itibarıyla ilk aşamada 1000 megavat...
3	Tahıl Krizi	Rusya'da Afrika zirvesi: Putin, Afrikalı liderlere bedava tahıl sözü verdi	Rusya Devlet Başkanı Vladimir Putin, Afrikalı liderlere, Batı yaptırımlarına rağmen on binlerce ton tahıl hediye edeceğini söyledi. St. Petersburg kentindeki...
4	Enerji İşbr. TR+AFR	Enerji ve Tabii Kaynaklar Bakanı Alparslan Bayraktar, "Cumhurbaşkanımız Sayın Recep Tayyip Erdoğan'ın Cezayir ziyareti kapsamında gerçekleşen Yüksek Düzeyli İş Birliği Konseyi İkinci Toplantısı...	Enerji ve Tabii Kaynaklar Bakanı Alparslan Bayraktar, "Cumhurbaşkanımız Sayın Recep Tayyip Erdoğan'ın Cezayir ziyareti kapsamında gerçekleşen Yüksek Düzeyli İş Birliği Konseyi İkinci Toplantısı...

Kaynak: Yazar tarafından derlenmiştir.

%22,11'lik oranla en baskın temanın "Afrika Pazarı" olarak elde edilmiş olması ve çok büyük çoğunluk ile Türkiye'nin ve Türk şirketlerin Afrika'ya ihracat potansiyeline odaklı haberlerden oluşması, ekonomik düzlemde Afrika-Türkiye ilişkilerinin önemini göstermektedir. %16,67'lik enerji ve enerji krizleri teması ise, Afrika kıtasının dünyada gittikçe büyüyen enerji ihtiyacına paralel olarak sıklıkla gündeme geldiğini göstermektedir. Konu 4 ve 7'de de (bkz. Tablo 21) Afrika ile enerji işbirliğinin hem Türkiye hem de diğer ülkeler özelinde medyada yer bulduğu

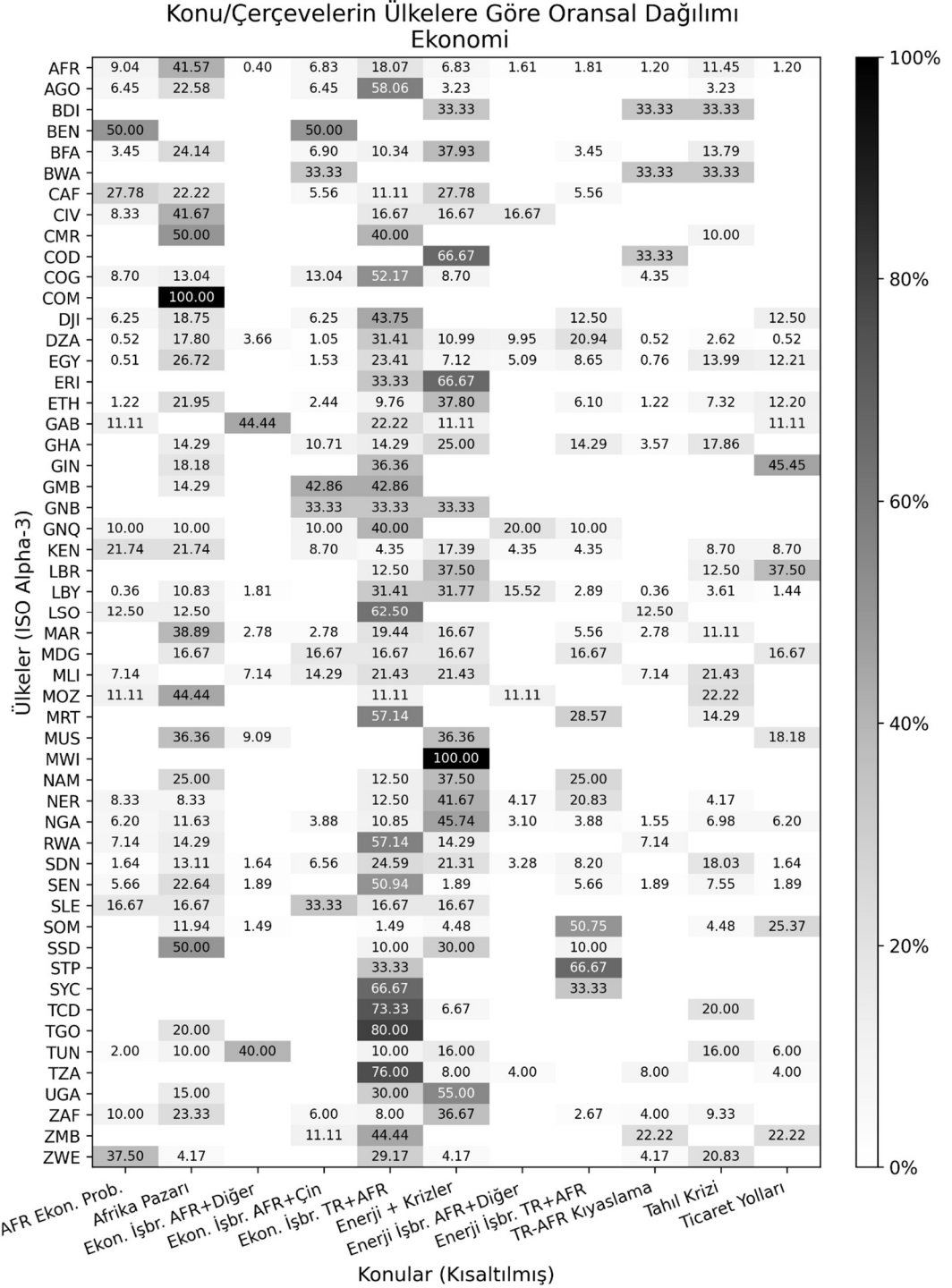
gözlenmektedir. Genel enerji temasını temsil eden genel konu olan konu 2'nin altında bulunan haberler, Afrika kıtasının ve özellikle petrol gibi önemli enerji kaynaklarına sahip ülkelerin potansiyeline vurgu yapmakta, ancak aynı zamanda da Afrika'nın yaşadığı enerji krizlerine dair haberleri içermektedir. Bu iki temanın aynı konu altında algoritma tarafından değerlendirilmesi, benzer kelimelerle ancak birbirinden bazı noktalarda ayrılan konuları ayırt etmede, en azından kullanılan cümle gömme algoritmasının başarısının sınırlı olduğuna işaret etmektedir.

Ekonomik işbirliği teması, hem Türkiye-Afrika (konu 1, %21.92), hem Çin-Afrika (konu 8, %3.46) hem de diğer ülkeler-Afrika (konu 9, %1.71) ekseninde oldukça sık ele alınan bir konudur. Konunun bu denli ön plana çıkmasındaki temel sebeplerden biri Afrika'nın ekonomik potansiyelinin hem Türkiye, hem de diğer ülkeler tarafından değerlendirilmek istemesinden kaynaklanmaktadır. Bu konu başlıklarına giren haberler karşılaştırmalı olarak incelendiğinde medya organlarının hemen hepsinde bir rekabet vurgusunun olduğu gözlemlenmektedir.

2022 yılında başlayan Rusya-Ukrayna savaşının yol açtığı tahıl krizinin ve covid-19 ile gittikçe artan ekonomik problemlerin de veri kümesinde (sırasıyla konu 3, %8.41 ve konu 6, %4.34) temsil edildiği görülmektedir. Ancak ekonomi olarak etiketlenmiş haberlerdeki en küçük küme olmakla birlikte belki de en ilgi çekici olanı "Türkiye Afrika Ekonomik Kıyaslamaları" (konu 10, %1.37) çerçevesidir. Bu konu altında yer alan haberler, son yıllardaki ekonomik zorlukların da etkisi ile Türkiye'nin ekonomik durumunu, Türk Lirasının değerini ve diğer ekonomik göstergeleri Afrika ülkeleri ile karşılaştıran içerikler barındırmakta ve genellikle Afrika ve Afrika ülkelerini "Türkiye'nin layık olmadığı alt bir lig" olarak çerçevelemektedirler. Bu çerçevenin çalışmanın uyguladığı yöntem ile ortaya konması, konu modellemenin kimi gizli kalıp ve çerçeveleri ortaya çıkarmadaki başarısına bir işaret olarak değerlendirilebilir.

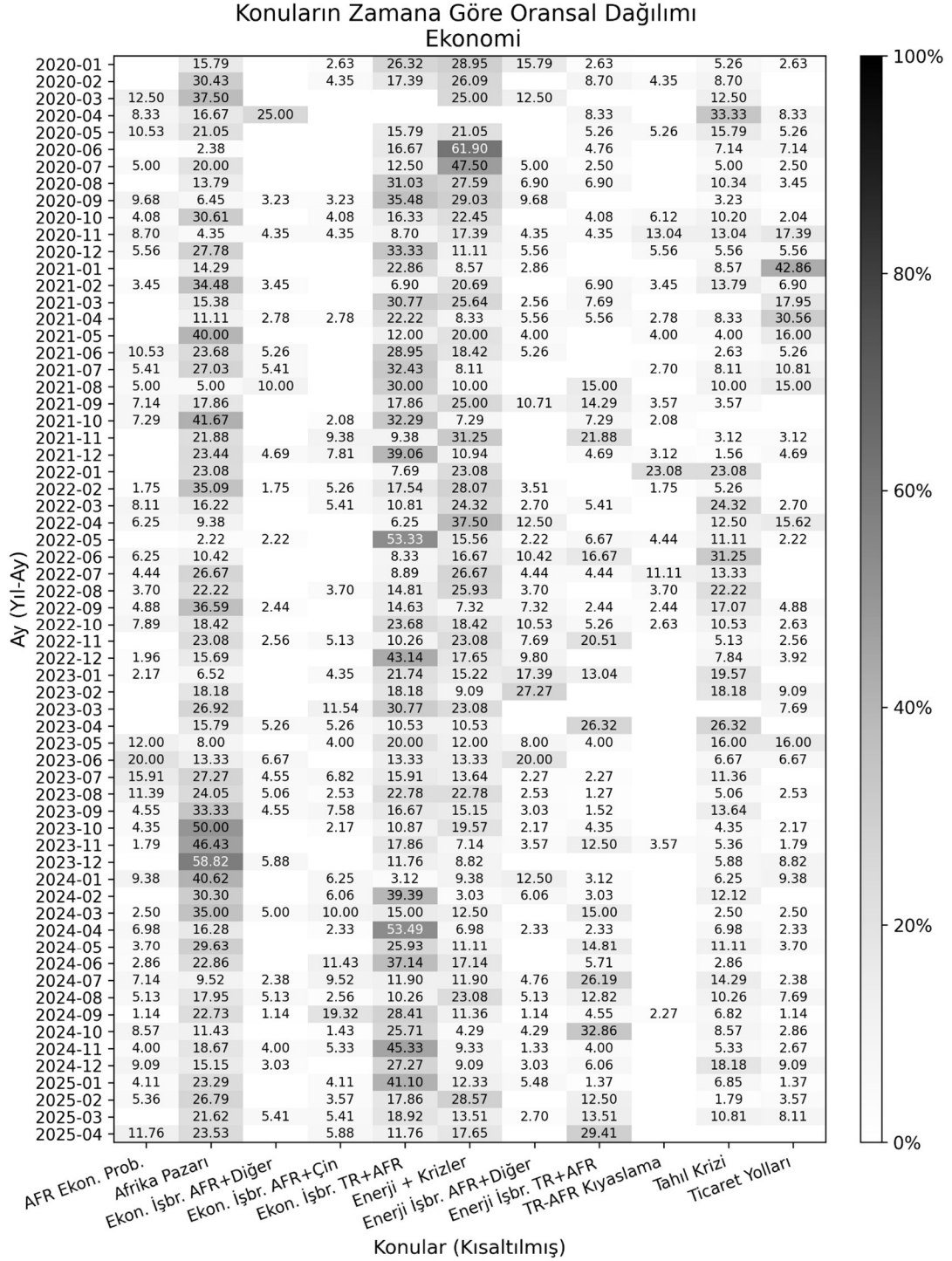
Kategorinin ülke-konu alt kırılımına bakıldığında ülke bazında dikkate değer anlamlı farklar görülmemekle birlikte, medya-konu kırılımında Dünya gazetesinin ekonomi ağırlıklı haberler yayınlaması sebebi ile "Afrika Pazarı" çerçevesine daha fazla yer verdiğini görmek mümkündür. Zaman-konu analizi ise, özellikle Türkiye-Afrika ekonomik kıyaslamaları çerçevesinin Türkiye'deki ani döviz hareketlerinin (2021 Aralık ve 2022 Haziran gibi) hemen sonrasında (2022 Ocak ve 2022 Temmuz) artış gösterdiğine işaret etmektedir. Benzer bir biçimde Tahıl Koridoru anlaşmasının hemen öncesi dönem olan Haziran 2022'den başlayarak anlaşmanın sona erdiği 2023 Sonbaharına kadar tahıl krizi temasının yoğunluğunun arttığı zaman kırılımı ısı haritasından gözlemlenebilmektedir.

Şekil 28: Ekonomi Kategorisi Ülke/Konu Dağılımı Isı Haritası



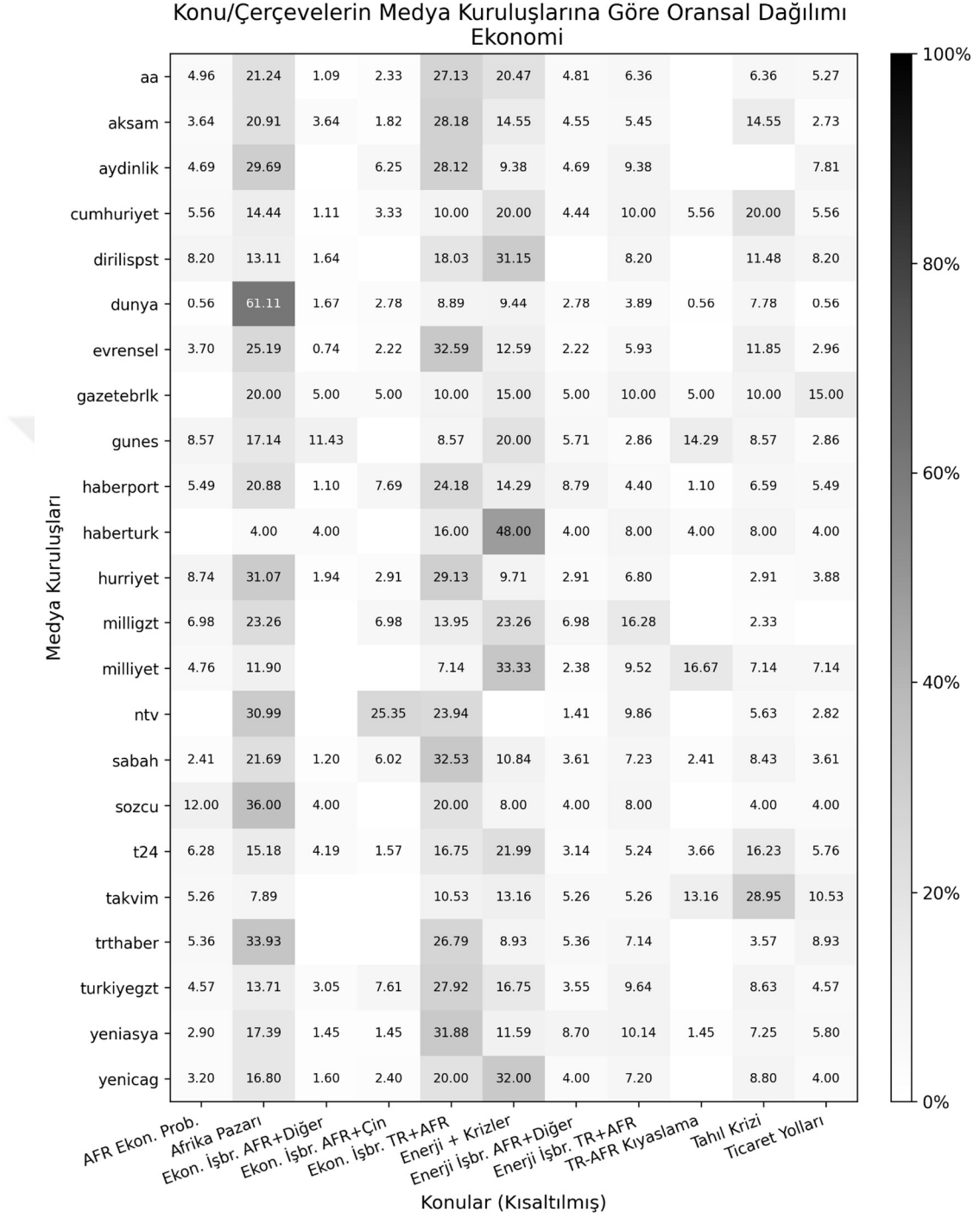
Kaynak: Yazar tarafından derlenmiştir.

Şekil 29: Ekonomi Kategorisi Zaman/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Şekil 30: Ekonomi Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Sağlık temaları bağlamında, kıta genellikle salgın hastalıklar, sağlık yardımı ve doğal afetler sonrası insani müdahaleler çerçevesinde ele alınmıştır. Salgın hastalık temasının ağırlığı şüphesiz veri setinin de zaman olarak örtüştüğü 2020'de

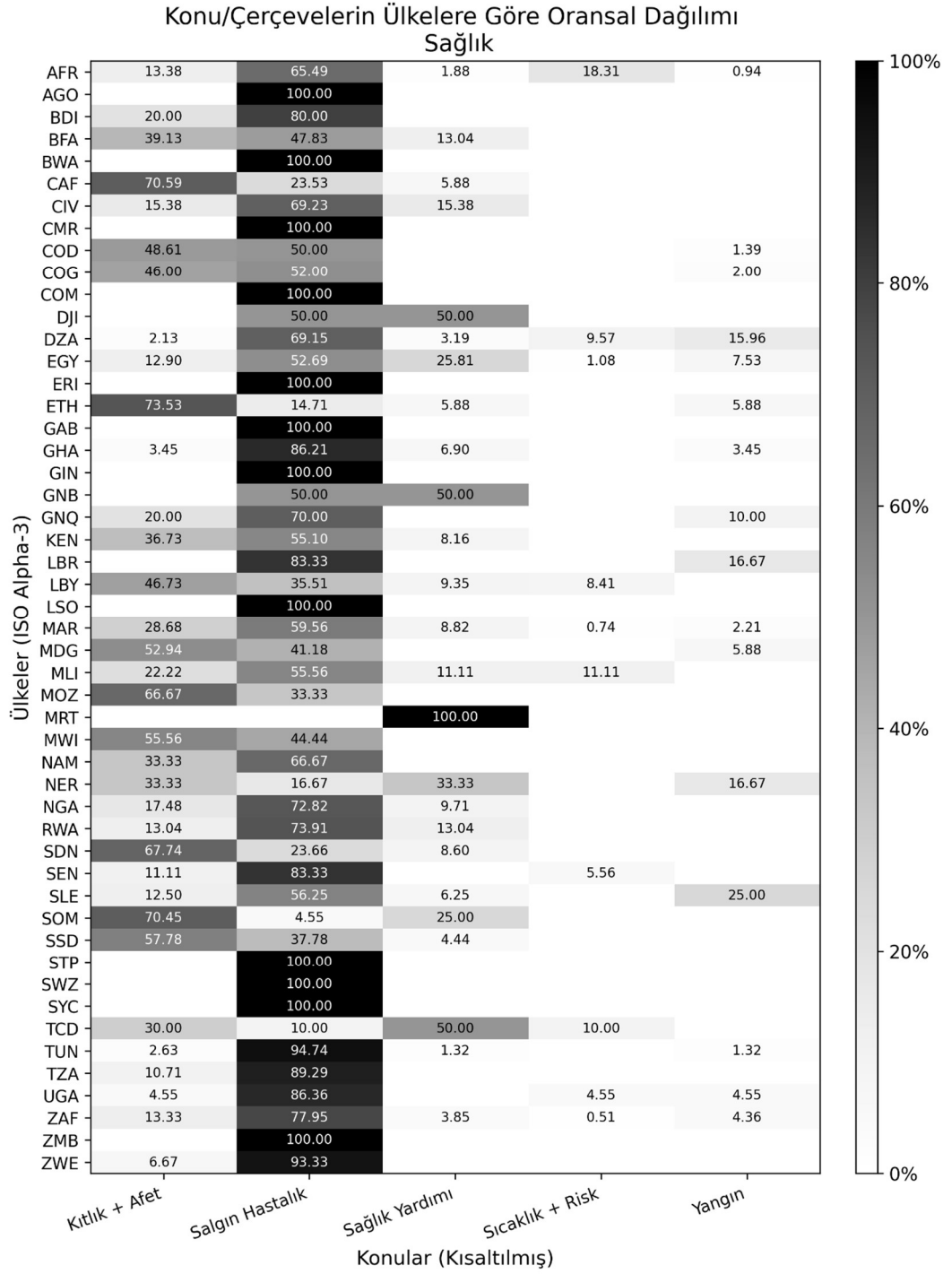
yaşanan koronavirüs pandemisidir. Ancak %54,87 gibi büyük bir oranla sağlık kategorisindeki haberlerin çoğunluğunu oluşturan bu tema sadece koronavirüs değil, kıtada yaşanan M çiçeği, ebola ve kolera salgınlarına dair haberleri de içermektedir. Salgın haberlerinin yapı ve söz dizimi açısından birbirlerine benzerliği, algoritmanın bu temaları bir arada değerlendirmesine neden olmuştur. Konu zaman kırılımı da incelendiğinde, bu salgınların, özellikle de koronavirüs pandemisinin etkilerinin yoğun yaşandığı zamanlarda bu temaya ait haberlerin büyük yer tuttuğu görülmektedir.

Tablo 26: Sağlık Kategorisi İlk 5 Konudan Örnekler

No	Konu İsmi (Kısaltılmış)	Başlık	Haber Metni (Kısaltılmış)
0	Salgın Hastalık	DSÖ'den koronavirüse karşı yeni hamle! Güney Sudan'da tesis kuruldu	DSÖ Twitter hesabından yapılan açıklamada, Afrika'daki birçok ülkede oksijen tedarikinde sıkıntılar yaşandığı hatırlatıldı. Açıklamada, Güney Sudan'a hükümet ve Afrika Kalkınma Bankasının...
1	Kıtlık + Afet	BM'den Somali açıklaması: Kuraklık nedeniyle göç eden kişi sayısı yüzde 231 arttı	BM İnsani Yardım Koordinasyon Ofisi 'nin (OCHA), BM Uluslararası Göç Örgütü (IOM) ve BM Çocuklara Yardım Fonu (UNICEF), Somali' deki kuraklık durumuyla ilgili ortak açıklama yayınladı. Açıklamada, ülkede 1-22 Haziran tarihlerinde kuraklık nedeniyle 112 bin 400'den fazla kişinin...
2	Sağlık Yardımı	Cumhurbaşkanı Erdoğan, Somali Cumhurbaşkanı Farmajo'ya solunum cihazı yardımıyla birlikte mektup gönderdi	Türkiye'nin, yeni tip koronavirüs (Kovid-19) salgınıyla mücadele için Somali'ye aralarında yerli yoğun bakım solunum cihazının da bulunduğu tıbbi yardım malzemelerini taşıyan A 400M tipi nakliye uçağı başkent Mogadişu'ya ulaştı...
3	Sıcaklık + Risk	Afrika sıcakları kapıda; pazar günü rekor sıcaklık bekleniyor!	İtibaren bu yağışlı sistem, yurdumuzu terk etti. Yerini Kuzey Afrika üzerinden gelen sıcak hava kütesinin etkisine bıraktı. Afrika üzerinden gelen yüksek basınç ve sıcak hava dalgası bölgemizi etkilemeye...
4	Yangın	Tunus da yanıyor	Tunus'ta dün akşam başlayan orman yangını kontrol altına alma çalışmaları devam ediyor. Başkent Tunus'un yakınlarında yer alan Ben Arus'taki Bukernin Dağı'nı kaplayan ormanlık...

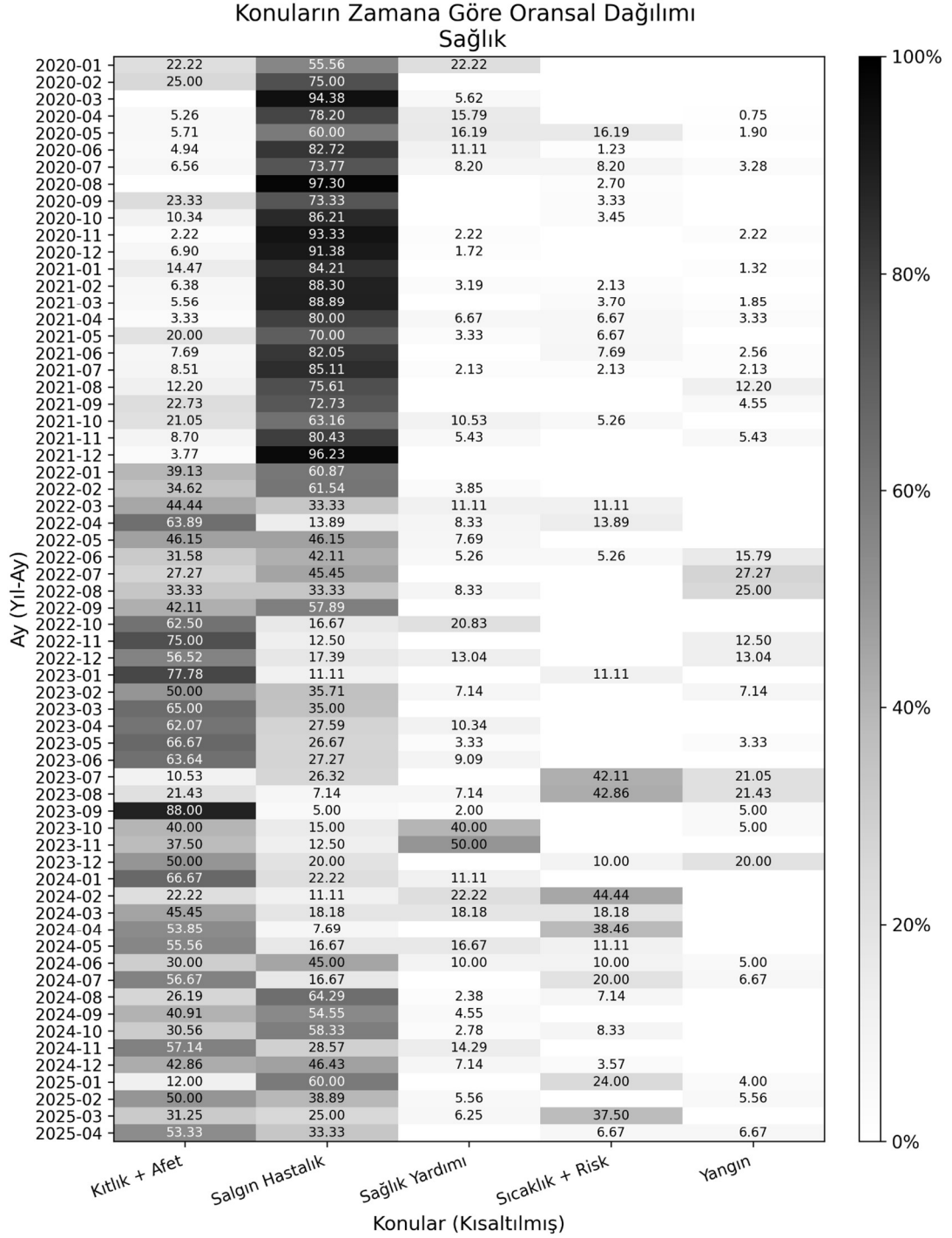
Kaynak: Yazar tarafından derlenmiştir.

Şekil 31: Sağlık Kategorisi Ülke/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Şekil 32: Sağlık Kategorisi Zaman/Konu Dağılımı Isı Haritası



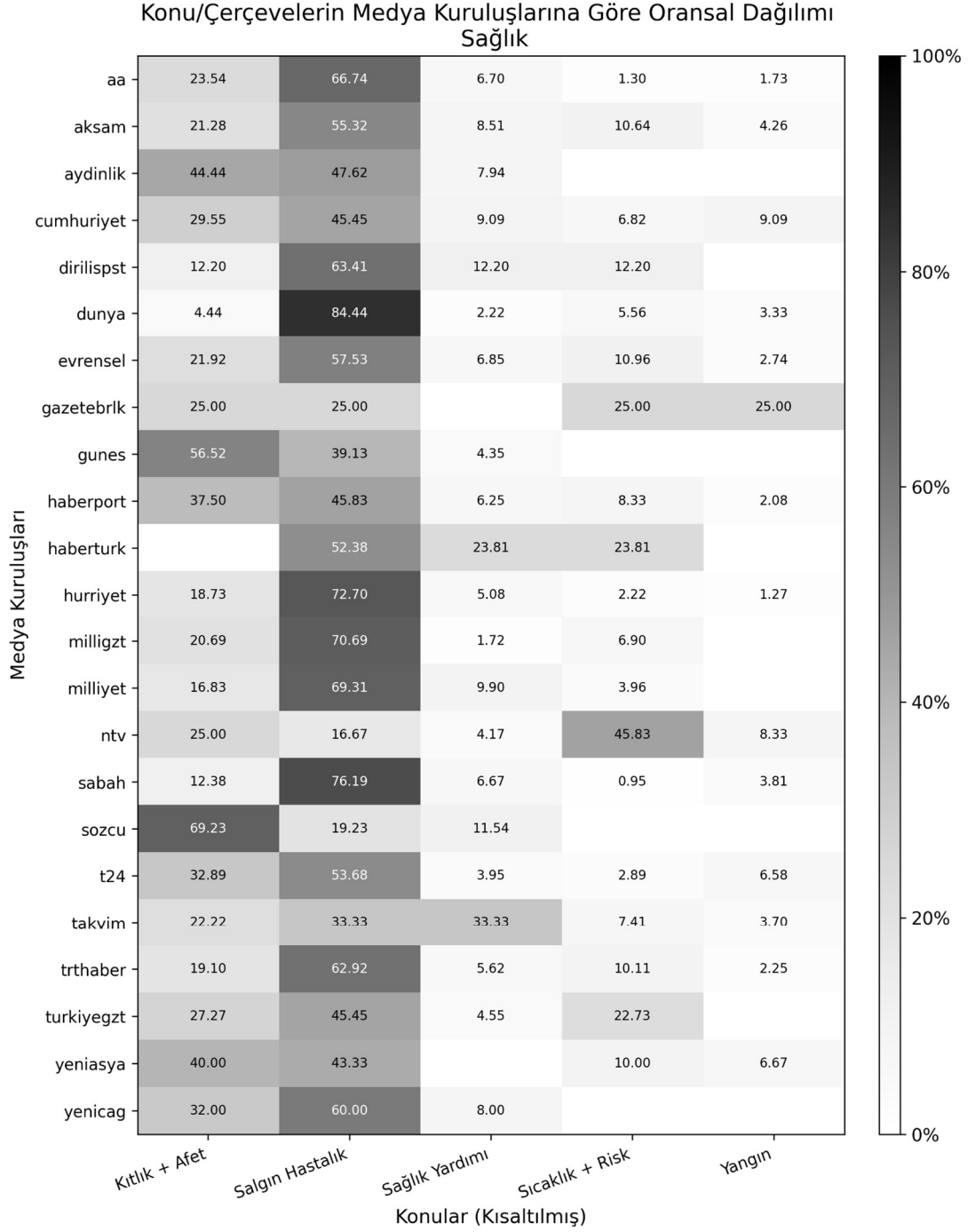
Kaynak: Yazar tarafından derlenmiştir.

Yine bir diğer baskın konu olan ve %21.37 ile ikinci sırada yer alan “Doğal Afetler ve Kıtık” teması, hem iklim değişikliği hem de savaşlar nedeniyle oluşan kıtlığa dair haberleri, aynı zamanda kıtada yaşanan sel, deprem ve diğer doğal afetleri

içermektedir. Bu temanın zaman kırılımına bakıldığında, Sudan iç savaşı ve Rusya Ukrayna savaşının neden olduğu tahıl krizinin etkisi, ülke kırılımına bakıldığında ise yine Somali, Sudan, Etiyopya, Orta Afrika Cumhuriyeti gibi ülkelerin yaşadığı sıkıntılar görülebilmektedir. Bu tema içerisinde doğal afetler de bulunduğundan, Fas'ta 2023 Eylül ayında yaşanan deprem felaketi ve yine aynı zamanlarda Libya'da yaşanan sel felaketinin ağırlığı ülke ve zaman kırılımlarından okunabilmektedir.

Bu konulara ilaveten ilgi çekici bir başka tema ise “Afrika Sıcakları” temasıdır. Sağlık kategorisi ile birlikte kültür kategorisinde de gözlemlenen bu temanın, sağlık kategorisinde bulunan temsilci verilerine bakıldığında, Türkiye’de görülecek olan ve özellikle yaşlı, çocuk ve kronik sağlık sorunları bulunan dezavantajlı kişilerde sağlık sorunlarına yol açabilecek sıcak havaların tanımlanmasında kullanıldığı görülmektedir. %4,33'lük bir temsil ile ayrılan bu küme, sıcak havanın Afrika kıtasından gelen bir sağlık riski olarak çerçevelendiğini göstermektedir.

Şekil 33: Sağlık Kategorisi Medya Kuruluşu/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Kültür kategorisinde yapılan konu modelleme sonuçları, tematik çeşitliliği ve olumlu çerçevelmeleri ile diğer kategorilerden ayrılmaktadır. En yüksek orana sahip olan tema, %33,57 ile “Türkiye-Afrika Kültürel İlişkileri” başlığı altında toplanmıştır. Bu temada daha çok kültür alanındaki iş birlikleri, TİKA’nın faaliyetleri bulunmakla birlikte,

zaman zaman kültürel yakınlık, eğitim, Osmanlı geçmişine atıflar yapan haberlere de rastlanılmaktadır. Kimi haberlerde Ebubekir Efendi örneğinde olduğu gibi tarihsel figürler üzerinden kültürel etkileşim öne çıkarılmaktadır. Bu durum, Türk medyasının Afrika ile tarihsel ve kültürel bağları vurgulayarak olumlu bir temsiliyet sunduğuna işaret etmektedir.

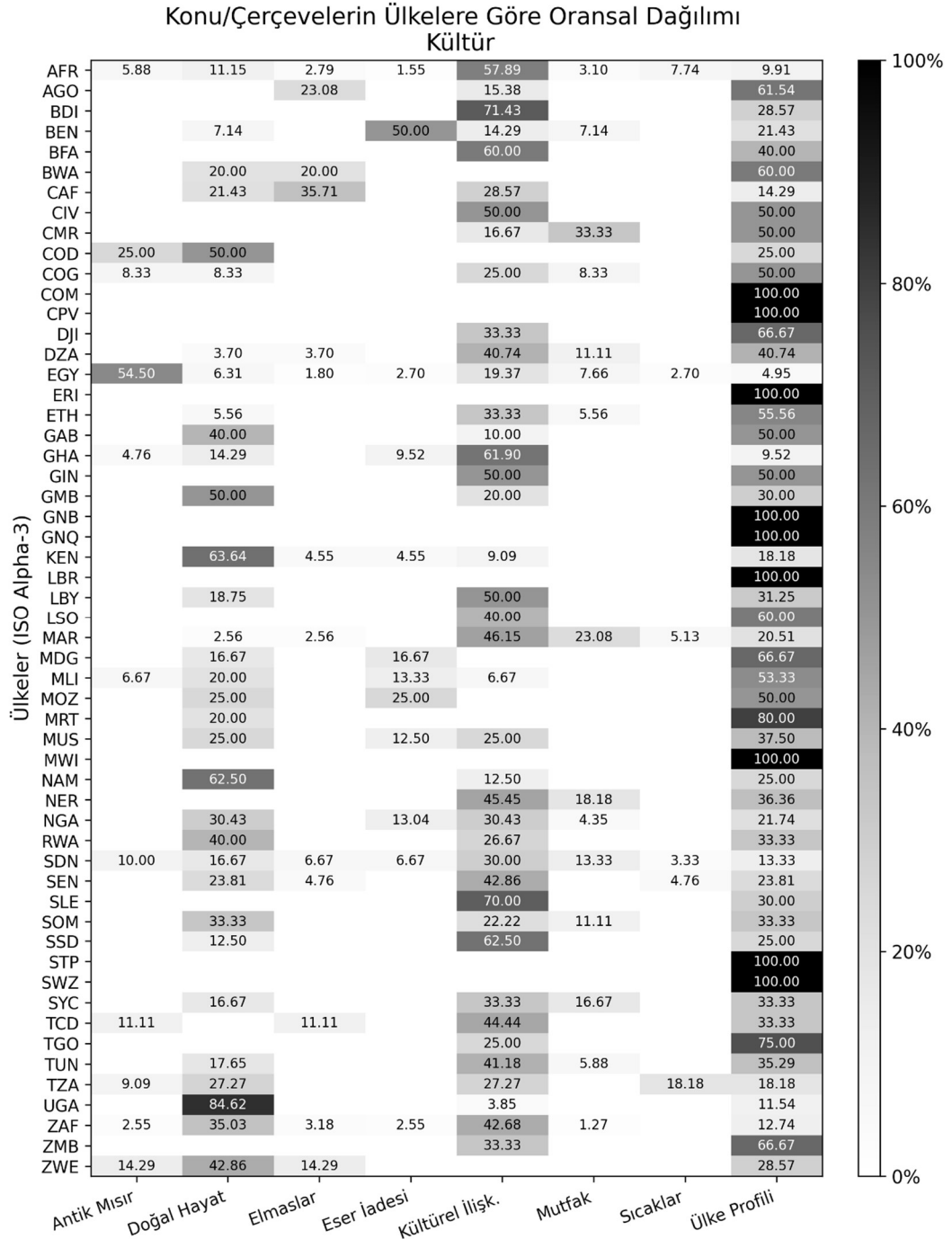
Tablo 27: Kültür Kategorisi İlk 5 Konudan Örnekler

No	Konu İsmi (Kısaltılmış)	Başlık	Haber Metni (Kısaltılmış)
0	Kültürel İlişk.	Ebubekir Efendi'nin Güney Afrika'daki etkisi memleketi Erzurum'da kutlanıyor	1814 yılında Osmanlı topraklarında doğan Seyit Ebubekir Efendi uzun seneler müderrislik yaptığı Erzurum şehrinde İslam ilimlerine ve Güney Afrika'daki Afrikaans dilinin gelişimine yaptığı katkılarla...
1	Ülke Profili	Güney Afrika'nın gözde doğa turizmi merkezi: Karoo	Güney Afrika tarihinde ve kültüründe önemli bir yere sahip olan Karoo, uçsuz bucaksız düzlükleri üzerine dizili düz tepeli kum taşı dağ sıralarından meydana gelen özgün jeolojik yapısıyla dikkati çekiyor...
2	Doğal Hayat	Güney Afrika'da aslanlar, artan sırtlanlar karşısında yaşam mücadelesi veriyor	Afrika savanının en baskın yırtıcısı aslanlar, geleneksel ilaç ve takı yapımında kullanılmak üzere çok yüksek fiyatlardan yoğun talep gören kemikleri ve vücut parçaları nedeniyle kaçak avcılarının en gözde avları arasında yer alıyor...
3	Antik Mısır	Piramitlerin İçinde Ne Var? Mısır Piramitlerinin İçi Boş Mu, Mezar Var Mı?	Mısır'daki piramitler, sadece firavunlar için değil, aynı zamanda önemli kişilerin de mezarlarıdır. Antik Mısırlılar, firavunların ölümden sonra tanrılarla birleşeceğine ve sonsuz bir yaşama geçeceğine...
4	Afrika Mutfağı	Malva pudding tarifi ve malzemeleri! Güney Afrika tatlısı	Malva pudding tarifi Güney Afrika'nın sevilen tatlıları arasında yer alıyor. Bu leziz tatlının ülkeye Hollanda mutfağından girdiği biliniyor. Marmelat ve kekin muhteşem uyumuyla ortaya çıkan malva...

Kaynak: Yazar tarafından derlenmiştir.

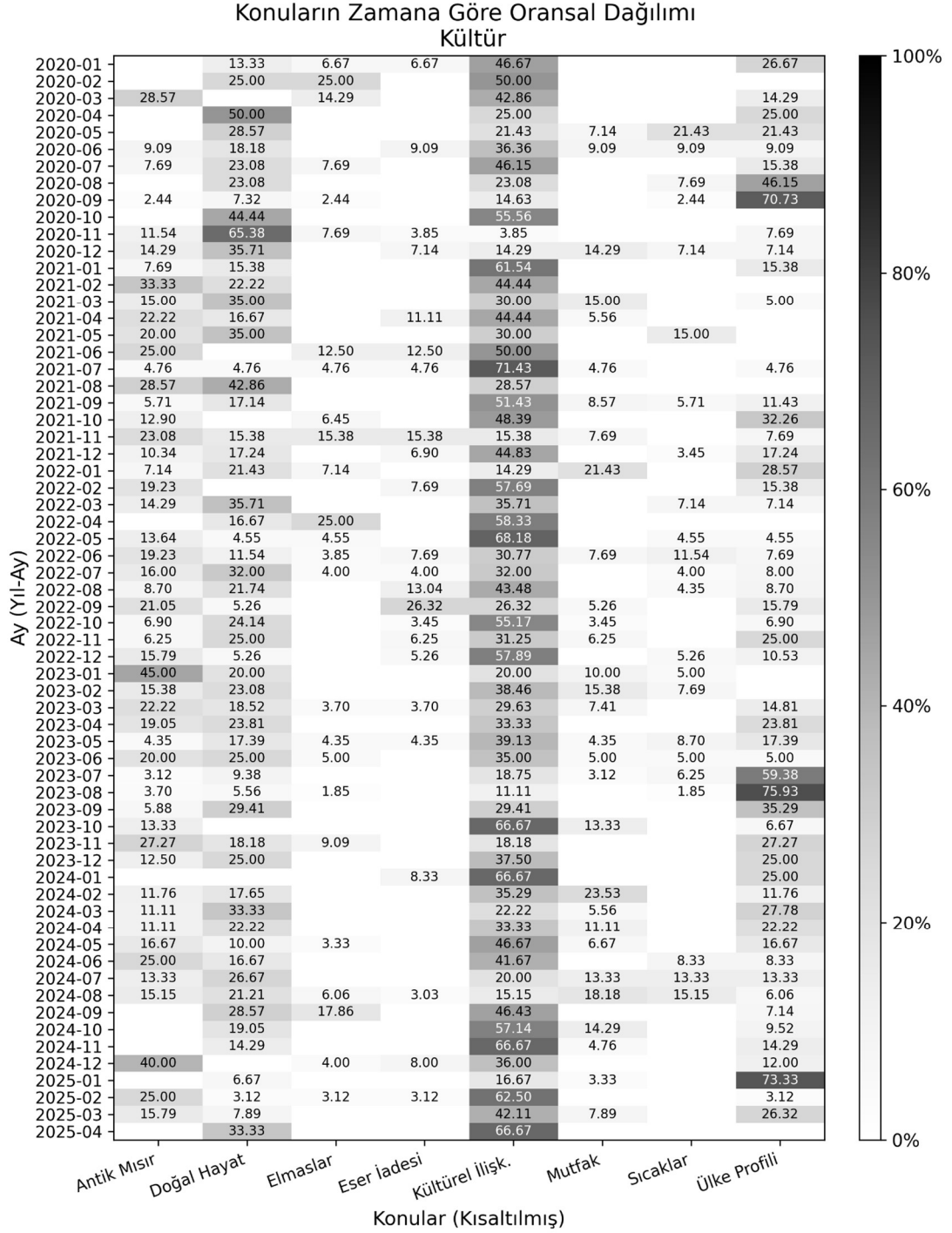
İkinci sıradaki tema, %18,27 oranla “Ülke Profilleri ve Turistik Destinasyonlar” şeklinde tanımlanmıştır. Bu başlık altındaki haberlerde, Afrika'daki ülkelerin tanıtımına yönelik içeriklerin baskın olduğu görülmektedir. Para birimi, nüfus, bayrak gibi yüzeysel bilgilerle şekillenen bu içerikler genellikle ansiklopedik bir üslupla sunulmuş, turistik cazibe merkezleri ile ilgili yüzeysel tanıtımlar da bu tema içerisinde yer almıştır. Bu durum, haberlerin bazı durumlarda bilgi verici olmaktan ziyade doldurucu içerikler barındırdığına işaret etmektedir.

Şekil 34: Kültür Kategorisi Ülke/Konu Dağılımı Isı Haritası



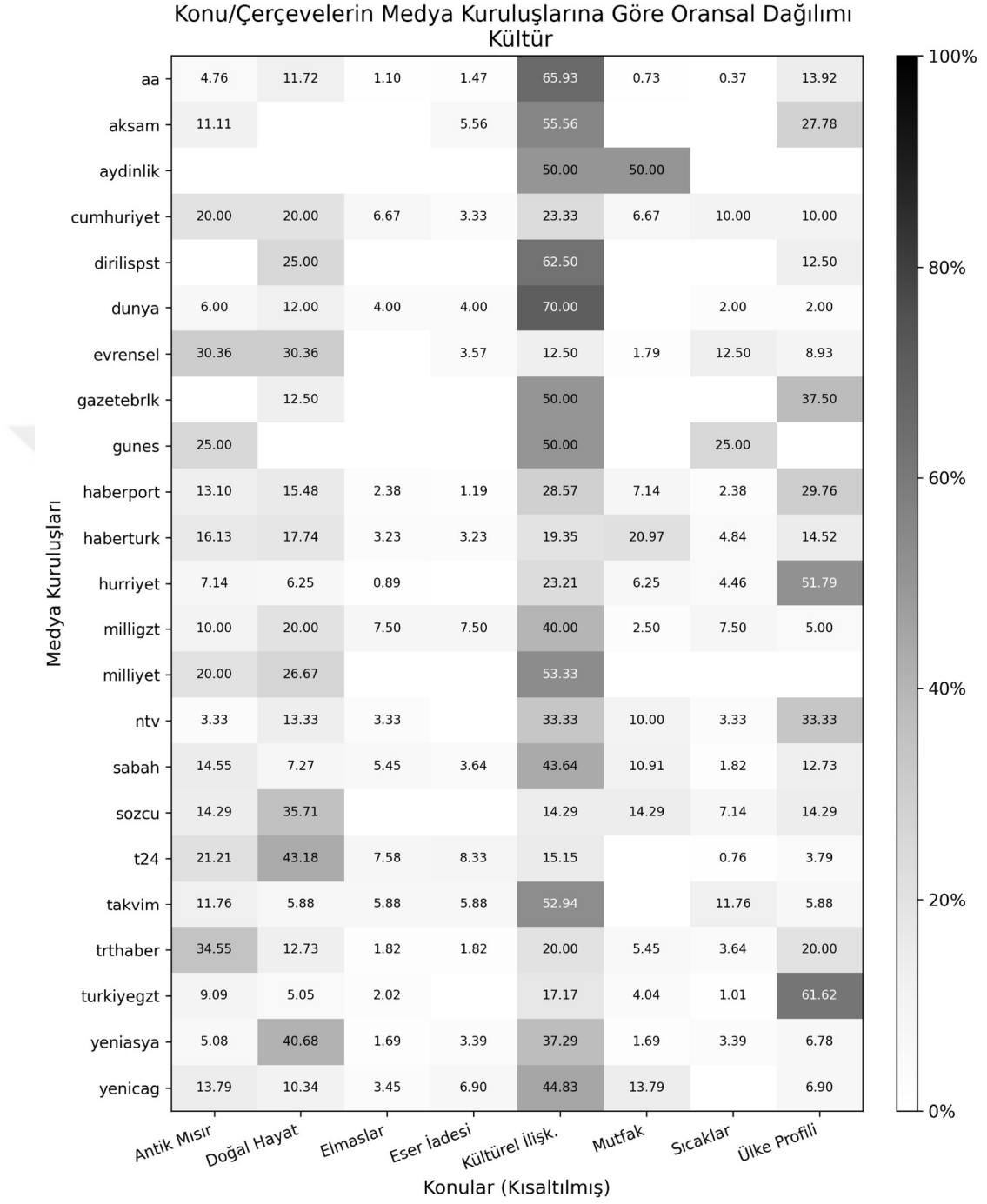
Kaynak: Yazar tarafından derlenmiştir.

Şekil 35: Kültür Kategorisi Zaman/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

Şekil 36: Kültür Medya Kuruluşu/Konu Dağılımı Isı Haritası



Kaynak: Yazar tarafından derlenmiştir.

%15,67 oranla üçüncü sırada yer alan “Afrika Doğal Hayatı ve Kaçak Avcılık” teması ise egzotiklik çerçevesinin tipik bir örneği olarak öne çıkmaktadır. Bu tema altında yer alan haberlerde Afrika, vahşi doğanın hakim olduğu, egzotik ve uzak bir kıta olarak temsil edilmiştir. Bu tür anlatılar, Afrika’nın bir “belgesel coğrafyası”

şeklinde sunulmasına katkı sağlamakta ve sıklıkla Sahra-Altı Afrika ülkelerine atıf içermektedir.

%11,12'lik oran ile antik Mısır haberleri de konu modelleme algoritması tarafından ayrıştırılmıştır. Bu haberler özellikle içerdikleri gizem, mistisizm ve uzak ve farklı bir dünyaya ait olma çerçeveleri nedeni ile okuyucu ilgisi cezbetmek ve sayfa ziyaret sayılarını arttırmak amacı ile medyada yer buldukları öne sürülebilir.

Medya kuruluşu-tema dağılımına bakıldığında, Dünya Gazetesi ve AA'nın kültürel ilişkiler temasına daha fazla yer verdiği, antik Mısır ve doğal hayat çerçevelerinin ise hemen hemen her medya kuruluşunda yer aldığı gözlemlenmektedir. Zaman-tema ısı haritası incelendiğinde ise kültür kategorisindeki içeriklerin belirli dönemsel özellikler barındırmadığı söylenebilir.

Sağlık kategorisi ile aynı şekilde adlandırılan ve Afrika Sıcakları temasını barındıran haberler ise kültür kategorisi altında genelde Türkiye ile iklim ve doğal hayat bağlamında paralellik kuran haberler için tercih edilmiştir. Bu tema altında kimi manzaraların (hayvanların doğada bulunuş şekli vb) Afrika'daki doğal hayata benzetildiği haberler de bulunmaktadır.

Çalışmada elde edilen bulgular, BERTopic yönteminin geniş ve karmaşık metin veri setleri üzerinde anlamlı tematik kümeler oluşturmada önemli avantajlar sağladığını göstermektedir. Özellikle Uİ disiplini kapsamında, medya söylemleri gibi çok boyutlu ve bağlamsal çeşitlilik içeren veri kümelerinde, yöntemin bağlamsal gömme ve yoğunluk temelli kümeleme yaklaşımı sayesinde geleneksel tekniklere kıyasla daha anlamlı ve yorumlanabilir temalar üretebildiği görülmüştür.

Ancak çalışmanın sonuçları, tamamen denetimsiz biçimde uygulandığında BERTopic yönteminin bazı sınırlılıklara sahip olduğunu da ortaya koymuştur. Veri setinde benzer kelimeler ile anlatılan ancak farklılıklar barındıran haberlerin yüksek oranda bulunması, konu kümelerinin zaman zaman örtüşmesine neden olmaktadır. Bir başka deyiş ile, denetimsiz olarak uygulandığında BERTopic bu farklılıkları yakalayamamaktadır. Dahası, algoritmanın ürettiği konu temsillerinin anlamlı konu başlıkları oluşturma ve hatta başlıklarda düzenleme ile kullanılma kapasitesi oldukça düşüktür. Uzman müdahalesi ve yorumlaması olmadan haber metinleri gibi çeşitlilik ve derinlik içeren derlemlerde kullanılarak elde edilen konuların anlamlandırılması oldukça güç olarak nitelenebilir.

Çalışmada ayrıca, analiz edilen haber metinlerinin geniş tematik çeşitliliği ve olay yoğunluğu dikkate alındığında, belirli alt konulara yahut periyotlara (örneğin lider ziyaretleri, ya da belirli bir anlaşmanın/olayın öncesi sonrası gibi) odaklanıldığında

daha tutarlı ve anlamlı konu kümelerinin elde edilebileceği sonucuna ulaşılmıştır. Bu durum, veri setinin sınırsız ve denetimsiz biçimde analiz edilmesinin konu kümelerinde dağınıklığa ve bazı temaların aşırı ayrışmasına yol açtığını göstermektedir. Bu çerçevede, sınırlı veya hiyerarşik konu modelleme yaklaşımlarının özellikle büyük ve karmaşık Uİ verileri üzerinde daha başarılı sonuçlar verebileceği düşünülmektedir.

Çalışmanın bulguları, konu modellemenin yalnızca tematik yapıların ortaya konulmasına değil, aynı zamanda karar destek sistemleri açısından da değerli bilgiler ürettiğine işaret etmektedir. Zaman, ülke ve medya kuruluşu boyutlarında gerçekleştirilen analizler, temaların sadece içeriksel değil, aynı zamanda bağlamsal ve zamansal eğilimlerinin de sistematik biçimde izlenebildiğini ortaya koymuştur. Bu yaklaşım, Uİ disipliniinde metin analizi ile politika üretimi arasındaki bağı güçlendirmekte ve söylem verilerinden elde edilen tematik örüntülerin dış politika karar alma süreçlerine entegre edilebilir olduğuna işaretler. Çalışma, bu yönüyle yalnızca akademik alana değil, aynı zamanda yönetim bilişim sistemleri perspektifinden politikaların tasarımı, izlenmesi ve rafine edilmesi süreçlerine de somut katkılar sunmaya adaydır. Dolayısıyla Uİ ile Yönetim Bilişim Sistemleri alanlarının kesişiminde yapılacak interdisipliner çalışmaların hem akademik literatüre hem de politika yapım süreçlerine katkı sunma potansiyelinin yüksek olduğu değerlendirilmektedir.

Sonuç olarak, çalışmanın ortaya koyduğu konu modelleme süreci ve bulgular, Türkiye’de Afrika’ya yönelik medya söyleminin analizi bağlamında disiplinlerarası bir yaklaşımın hem araştırma hem de karar destek sistemleri alanında uygulanabilirliğine güçlü kanıtlar öne sürmüştür.

KAYNAKÇA

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*. 236(1): 197-213.

Alghamdi, R., Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications*. 6(1): 147-153.

Alsmadi, M. K., Alzaqebah, M., Jawarneh, S., ALmarashdeh, I., Al-Betar, M. A., Alwohaibi, M., Al-Mulla, N. A., Ahmed, E. A., AL Smadi, A. (2024). Hybrid topic modeling method based on dirichlet multinomial mixture and fuzzy match algorithm for short text clustering. *Journal of Big Data*. 11(1): 1-21.

Al-Zaman, Md. S., Rashid, M. H. O. (2025). The Humanitarian Crisis in the Media: Framing Analysis of Rohingya-Related International News Using BERTopic. *Journalism Studies*. 26(1): 1-23.

Amaratunga, T. (2021). What is Deep Learning?. *Deep Learning in Windows* (ss. 1-14). New York: Springer.

Angelov, D. (2020). *Top2Vec: Distributed Representations of Topics* (No. arXiv:2008.09470). arXiv preprint. <https://doi.org/10.48550/arXiv.2008.09470> (13.05.2025).

Arslan, M., Riaz, Z., Cruz, C. (2023). Revolutionizing management information systems with natural language processing for digital transformation. *Procedia Computer Science*. 225: 2835-2844.

Asnawi, M. F., Hanafi, M., Kurniawan, N. F., Suwondo, A., Nasrullah, A., Setyawan, C. (2024). Topic Modelling Analysis on Indonesian News Using BERT Topic Model. IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10903779>

Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*. 3(19-48): 5-1.

Baradad, V. P., Mugabushaka, A. (2015). Corpus Specific Stop Words to Improve the Textual Analysis in Scientometrics. ISSI. https://www.issi-society.org/proceedings/issi_2015/0999.pdf (13.05.2025).

Barrett, P., Hunter, J., Miller, J. T., Hsu, J.C., Greenfield, P. (2005). Matplotlib—A Portable Python Plotting Package. *Astronomical data analysis software and systems XIV* (347): 91 - 95.

Borčin, M., Jose, J. M. (2024). Optimizing BERTopic: Analysis and Reproducibility Study of Parameter Influences on Topic Modeling. *Advances in Information Retrieval* (ss. 147-160). İsviçre: Springer Nature.

Brookes, G., McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*. 21(1): 3-21.

Bull, H. (1966). International theory: The case for a classical approach. *World politics*, 18(3): 361-377.

Campello, B., Moulavi, D., Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. *Advances in Knowledge Discovery and Data Mining* (ss. 160-172). New York: Springer.

Choi, H., Kim, J., Joe, S., Gwon, Y. (2021). Evaluation of bert and albert sentence embedding performance on downstream nlp tasks. IEEE Xplore.. <https://ieeexplore.ieee.org/abstract/document/9412102> (13.05.2025).

Churchill, R., Singh, L. (2022). The Evolution of Topic Modeling. *ACM Computing Surveys*. 54(10):1-35.

Chomsky, N. (1955). Logical Syntax and Semantics: Their Linguistic Relevance. *Language*. 31(1): 36-45.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6): 391-407.

Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*. 1:4171-4186.

Dieng, A. B., Ruiz, F. J., Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*. 8: 439-453.

Eckhard, S., Patz, R., Schönfeld, M., Van Meegdenburg, H. (2023). International bureaucrats in the UN Security Council debates: A speaker-topic network analysis. *Journal of European Public Policy*. 30(2): 214-233.

Eisele, O., Heidenreich, T., Litvyak, O., Boomgaarden, H. G. (2023). Capturing a News Frame – Comparing Machine-Learning Approaches to Frame Analysis with Different Degrees of Supervision. *Communication Methods and Measures*. 17(3): 205-226.

El Naqa, I., Murphy, M. J. (2015). What Is Machine Learning?. *Machine Learning in Radiation Oncology* (ss. 3-11). New York: Springer International Publishing.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*. 43(4): 51-58.

Eyryce Tepeciklioğlu, E., Tepeciklioğlu, A. O. (2024). Turkey–Africa voting cohesion in the UN General Assembly. *Southeast European and Black Sea Studies*. Online First:1-22. <https://doi.org/10.1080/14683857.2024.2435136>

Eyryce Tepeciklioğlu, E., Tepeciklioğlu, A. O. (2021). Introduction: Contextualizing Turkey's Africa policy. *Turkey in Africa: A New Emerging Power?* (ss. 1-15). Londra: Taylor and Francis Inc.

Fan, J., Ma, C., Zhong, Y. (2021). A Selective Overview of Deep Learning. *Statistical Science*. 36(2): 264-290.

Franzosi, R., Dong, W., Dong, Y. (2022). Qualitative and quantitative research in the humanities and social sciences: How natural language processing (NLP) can help. *Quality & Quantity*. 56(4): 2751-2781.

Gillings, M., Hardie, A. (2023). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities*. 38(2): 530-543.

Grant, P., Sebastian, R., Allassonnière-Tang, M., Cosemans, S. (2021). Topic modelling on archive documents from the 1970s: Global policies on refugees. *Digital Scholarship in the Humanities*. 36(4): 886-904.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2203.05794> (13.05.2025).

Heidenreich, T., Lind, F., Eberl, J. M., Boomgaarden, H. G. (2019). Media framing dynamics of the 'European refugee crisis': A comparative topic modelling approach. *Journal of Refugee Studies*. 32(1): 172-182.

Hoffmann, L. B., Hameleers, M. (2024). Unequal Framing in Times of Hardship? How Newspapers from Germany, Spain, the United Kingdom and Switzerland Portray Syrian and Ukrainian Refugees – Evidence from a Deductive and Inductive Automated Content Analysis. *Mass Communication and Society*. 27(6): 1685-1716.

Jacobs, T., Tschötschel, R. (2019). Topic models meet discourse analysis: A quantitative tool for a qualitative approach. *International Journal of Social Research Methodology*. 22(5): 469-485.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications*. 78(11): 15169-15211.

Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. *Machine Learning: ECML-98* (ss. 137-142). Berlin: Springer .

Johri, P., Khatri, S. K., Al-Taani, A. T., Sabharwal, M., Suvanov, S., Kumar, A. (2021). Natural Language Processing: History, Evolution, Application, and Future Work. *Proceedings of 3rd International Conference on Computing Informatics and Networks* (ss. 365-375). Singapur: Springer.

Jumle, V., Makhortykh, M., Sydorova, M., Vziatysheva, V. (2024). *Finding frames with BERT: A transformer-based approach to generic news frame detection*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2409.00272> (13.05.2025).

Kacprzyk, J., Zadrozny, S. (2010). Modern data-driven decision support systems: The role of computing with words and computational linguistics. *International Journal of General Systems*. 39(4): 379-393.

Kaplan, M. A. (1966). The new great debate: Traditionalism vs. science in international relations. *World Politics*. 19(1): 1-20.

Kapočiūtė-Dzikienė, J., Salimbajevs, A., Skadiņš, R. (2021). Monolingual and Cross-Lingual Intent Detection without Training Data in Target Languages. *Electronics*, 10(12): 1-24.

Kaya, A., Gülbandılar, E. (2022). Konu modelleme yöntemlerinin karşılaştırılması. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*. 3(2): 46-53.

Khan, S., Ahmed, F., Mubeen, M. (2022). A Text-Mining Research Based on LDA Topic Modelling: A Corpus-Based Analysis of Pakistan's UN Assembly Speeches (1970–2018). *International Journal of Humanities and Arts Computing*. 16(2): 214-229.

Kirill, Y., Mihail, I. G., Sanzhar, M., Rustam, M., Olga, F., Ravil, M. (2020). Propaganda identification using topic modelling. *Procedia Computer Science*. 178: 205-212.

Kolini, F., Janczewski, L. (2017). *Clustering and topic modelling: A new approach for analysis of national cyber security strategies*. PACIS 2017 Proceedings. <https://core.ac.uk/download/pdf/301372894.pdf> (13.05.2025).

Koruyan, K. (2022). BERTopic Konu Modelleme Tekniği Kullanılarak Müşteri Şikayetlerinin Sınıflandırılması. *İzmir Sosyal Bilimler Dergisi*. 4(2): 66-79.

Krstovski, K., Smith, D. A., Kurtz, M. J. (2016). Online multilingual topic models with multi-level hyperpriors. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. <https://aclanthology.org/N16-1053.pdf>. (13.05.2025).

Lee, D. D., Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788-791.

Lee, R. S. T. (2024). Natural Language Processing. *Natural Language Processing* (ss. 3-17). Singapur: Springer Nature.

Liddy, E. D. (2001). *Natural language processing*. Encyclopedia of Library and Information Science (ss. 1-16) New York: Marcel Decker.,

Lock, I. (2020). Debating Glyphosate: A Macro Perspective on the Role of Strategic Communication in Forming and Monitoring A Global Issue Arena Using Inductive Topic Modelling. *International Journal of Strategic Communication*. 14(4): 223-245.

Lockwood, K. K., Salman A., Eyad Kourdi, P. (2024). *UN says no food has entered northern Gaza since start of October, putting 1 million people at risk of starvation*. CNN. <https://www.cnn.com/2024/10/11/middleeast/food-northern-gaza-starvation-un-intl/index.html> (13.05.2025).

Madrid-García, A., Freitas-Núñez, D., Merino-Barbancho, B., Pérez Sancristobal, I., Rodríguez-Rodríguez, L. (2024). Mapping two decades of research in rheumatology-specific journals: A topic modeling analysis with BERTopic. *Therapeutic Advances in Musculoskeletal Disease*. 16: 1-17.

Mandelbaum, A., Shalev, A. (2016). *Word Embeddings and Their Use In Sentence Classification Tasks*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1610.08229>. (13.05.2025).

Mao, K., Xu, J., Yao, X., Qiu, J., Chi, K., Dai, G. (2022). A text classification model via multi-level semantic features. *Symmetry*, 14(9): 1938-1950.

Marwala, T. (2023). *Artificial intelligence, game theory and mechanism design in politics*. New York: Springer Nature.

McInnes, L., Healy, J., Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1802.03426>. (13/05/2025).

McKinney, W. (2011). pandas: A foundational Python library for data analysis and statistics. *Python for high performance and scientific computing*. 14(9): 1-9.

Meleouni, C., Efthymiou, I. P. (2023). Artificial Intelligence and its Impact in International Relations. *Journal of politics and ethics in new technologies and AI*. 2(1): 1-12.

Métais, E. (2002). Enhancing information systems management with natural language processing techniques. *Data & Knowledge Engineering*. 41(2-3): 247-272.

Murshed, B. A. H., Mallappa, S., Abawajy, J., Saif, M. A. N., Al-ariqi, H. D. E., Abdulwahab, H. M. (2023). Short text topic modelling approaches in the context of big data: Taxonomy, survey, and analysis. *Artificial Intelligence Review*. 56(6): 5133-5260.

Nagy, M. (2024). *Testing an Inductive Mixed-method Computational Approach to News Frame Analysis: An analysis of Hungarian online reporting of the 2014 Russia-Ukraine conflict*. IFLA International News Media Conference 2024. <https://repository.ifla.org/items/9934907b-15d3-4b30-b5a1-e536aa84f97e>. (13.05.2025).

Nanayakkara, A. C., Kumara, B. T. G. S., Kapila Tharanga Rathnayaka, R. M. (2024). Cross-Platform Topic Analysis: Identifying Trends and Divergences in Social Media Discourse. *2024 8th SLAAI International Conference on Artificial Intelligence*. <https://doi.org/10.1109/SLAAI-ICAI63667.2024.10844977> (13.05.2025).

Ordun, C., Purushotham, S., Raff, E. (2020). *Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2005.03082> (13.05.2025)

Pöyhtäri, R., Nelimarkka, M., Nikunen, K., Ojala, M., Pantti, M., Pääkkönen, J. (2021). Refugee debate and networked framing in the hybrid media environment. *International Communication Gazette*. 83(1): 81-102.

Pradhan, N., Gyanchandani, M., Wadhvani, R. (2015). A Review on Text Similarity Technique used in IR and its Application. *International Journal of Computer Applications*. 120(9): 29-34.

Ramos, J. (2003). Using Tf-Idf to Determine Word Relevance in Document Queries. *Proceedings of the first instructional conference on machine learning*, 242(1): 29-48.

Reimers, N., Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1908.10084> (13.05.2025).

Roberts, H., Bhargava, R., Valiukas, L., Jen, D., Malik, M. M., Bishop, C. S., Ndulue, E. B., Dave, A., Clark, J., Etling, B. (2021). Media cloud: Massive open source collection of global news on the open web. *Proceedings of the International AAAI Conference on Web and Social Media*. 15: 1034-1045.

Ruder, S., Vulić, I., Søgaard, A. (2019). A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*. 65: 569-631.

Saleem, S., Mehrotra, M. (2023). An Analytical Framework for Analyzing Tweets for Disaster Management: Case Study of Turkey Earthquake 2023. *2023 14th International Conference on Computing Communication and Networking*

Technologies (ICCCNT). <https://doi.org/10.1109/ICCCNT56998.2023.10306894> (13.05.2025).

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2(3): 1-21.

Sawant, S., Yu, J., Pandya, K., Ngan, C.-K., Bardeli, R. (2022). An Enhanced BERTopic Framework and Algorithm for Improving Topic Coherence and Diversity. *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application* <https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00332>. (13.05.2025).

Saxe, A., Nelli, S., Summerfield, C. (2021). If Deep Learning is the Answer, What is the Question?. *Nature Reviews Neuroscience*. 22(1): 55-67.

Sayed, M. A., Braşoveanu, A. M. P., Nixon, L. J. B., Scharl, A. (2023). Unsupervised Topic Modeling with BERTopic for Coarse and Fine-Grained News Classification. *Advances in Computational Intelligence* (ss. 162-174). İsviçre: Springer Nature.

Sbalchiero, S., Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity*. 54(4): 1095-1108.

Schofield, A., Magnusson, M., Mimno, D. (2017). Pulling out the stops: Rethinking stopword removal for topic models. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. 2: 432-436.

Schrodt, P. A. (2019). Artificial intelligence and international relations: An overview. *Artificial intelligence and international politics* (ss., 9-31). New York: Routledge.

Schweter, S. (2020). *BERTurk—BERT models for Turkish*. Zenodo. <https://doi.org/10.5281/zenodo.3770924> (13.05.2025).

Shah, A. H. (2019). How episodic frames gave way to thematic frames over time: A topic modeling study of the Indian media's reporting of rape post the 2012 Delhi gang-rape. *Poetics*. 72: 54-69.

Singh, R., Singh, S. (2021). Text similarity measures in news articles by vector space model using NLP. *Journal of The Institution of Engineers (India)*. 102: 329-338.

Srivastava, A., Sutton, C. (2017). *Autoencoding Variational Inference For Topic Models*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1703.01488>. (13.05.2025)

Stewart, G., Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering Algorithm. *Applied Sciences*. 12(5): 1-21.

Stoykova, S., Shakev, N. (2023). Artificial intelligence for management information systems: Opportunities, challenges, and future directions. *Algorithms*. 16(8): 357 - 370.

Tang, C. (2024). Dynamics of Sino-American relations: A decade-long analysis using topic modeling. *Applied and Computational Engineering*. 93: 127-133.

Teh, Y., Jordan, M., Beal, M., Blei, D. (2004). Sharing clusters among related groups: Hierarchical Dirichlet processes. *Advances in neural information processing systems*. https://proceedings.neurips.cc/paper_files/paper/2004/hash/fb4ab556bc42d6f0ee0f9e24ec4d1af0-Abstract.html (13.05.2025).

Tepeciklioğlu, A. O., Tepeciklioğlu, E. E., & Karabıyık, C. (2024). (Exploring) the Impact of Turkey's Embassies on Trade with sub-Saharan Africa. *Turkey's Pivot to the African Continent* (ss. 72-89). New York: Routledge.

Tepeciklioğlu, E. E. (2016). African studies in Turkey. *Uluslararası İlişkiler Dergisi*, 13(50): 3-19.

Turing, A. M., (1951). Can Digital Computers Think. *The Essential Turing*. (ss. 476-486). Oxford: Oxford Academic.

Uğuz, S., Tülü, Ç. N. (2024). Topic Modeling Analysis in the Field of Large Language Models with BERTopic (2020–2024). *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE Xplore. <https://ieeexplore.ieee.org/abstract/document/10757169> (13.05.2025).

Ünver, H. A. (2019). Computational international relations what can programming, coding and internet research do for the discipline?. *All Azimuth: A Journal of Foreign Policy and Peace*. 8(2): 157-182.

Verbytska, A. (2024). Topic modelling as a method for framing analysis of news coverage of the Russia-Ukraine war in 2022–2023. *Language & Communication*. 99: 174-193.

Voskergian, D., Jayousi, R., Yousef, M. (2024). Topic selection for text classification using ensemble topic modeling with grouping, scoring, and modeling approach. *Scientific Reports*. 14(1): 1-19.

Wang, C., Blei, D., Heckerman, D. (2015). *Continuous Time Dynamic Topic Models*. arXiv Preprint. <https://doi.org/10.48550/arXiv.1206.3298> (13.05.2025).

Wendt, A. (1995). Constructing international politics. *International security*. 20(1): 71-81.

Wojciechowska, J., Sypniewski, M., Śmigielska, M., Kamiński, I., Wiśnios, E., Schreiber, H., Pielniński, B. (2023). *Deep Dive into the Language of International Relations: NLP-based Analysis of UNESCO's Summary Records*. arXiv Preprint. <https://doi.org/10.48550/arXiv.2307.16573>. (13.05.2025).

Yan, X., Guo, J., Lan, Y., Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd international conference on World Wide Web*. <https://doi.org/10.1145/2488388.2488514> (13.05.2025).

Zheng, C., He, G., Peng, Z. (2015). A Study of Web Information Extraction Technology Based on Beautiful Soup. *Journal of Computing*. 10(6): 381-387.



EK 1: Afrika Ülke Adları ve Kısaltmaları

Ülke İsmi	ISO-alpha3 Kodu	Ülke İsmi	ISO-alpha3 Kodu
Cezayir	DZA	Liberya	LBR
Angola	AGO	Libya	LBY
Benin	BEN	Madagaskar	MDG
Botsvana	BWA	Malavi	MWI
Burkina Faso	BFA	Mali	MLI
Burundi	BDI	Moritanya	MRT
Yeşil Burun Adaları	CPV	Mauritus	MUS
Kamerun	CMR	Fas	MAR
Orta Afrika Cumhuriyeti	CAF	Mozambik	MOZ
Çad	TCD	Namibya	NAM
Komor Adaları	COM	Nijer	NER
Kongo Cumhuriyeti	COG	Nijerya	NGA
Fildişi Sahili	CIV	Ruanda	RWA
Demokratik Kongo Cumhuriyeti	COD	Sao Tome and Principe	STP
Cibuti	DJI	Senegal	SEN
Mısır	EGY	Şeyseller	SYC
Ekvator Ginesi	GNQ	Sierra Leone	SLE
Eritre	ERI	Somali	SOM
Esvatini	SWZ	Güney Afrika Cumhuriyeti	ZAF
Etiyopya	ETH	Güney Sudan	SSD
Gabon	GAB	Sudan	SDN
Gambiya	GMB	Togo	TGO
Gana	GHA	Tunus	TUN
Gine	GIN	Uganda	UGA
Gine-Bissau	GNB	Tanzanya	TZA
Kenya	KEN	Zambiya	ZMB
Lesotho	LSO	Zimbabve	ZWE

EK 2: Kullanılan Kütüphaneler ve Versiyonları

accelerate==1.6.0	jupyter_client==8.6.3	pywin32==308
aiohappyeyeballs==2.6.1	jupyter_core==5.7.2	PyYAML==6.0.2
aiohttp==3.11.18	jupyterlab_widgets==3.0.13	pyzmq==26.2.0
aiosignal==1.3.2	kiwisolver==1.4.8	regex==2024.11.6
asttokens==2.4.1	llvmlite==0.43.0	requests==2.32.3
async-timeout==5.0.1	lxml==5.3.1	safetensors==0.4.5
attrs==25.3.0	MarkupSafe==3.0.2	scikit-learn==1.5.2
beautifulsoup4==4.12.3	matplotlib==3.10.0	scipy==1.14.1
bertopic==0.16.4	matplotlib-inline==0.1.7	sentence-
certifi==2024.8.30	mediacloud==4.3.0	transformers==3.3.0
charset-normalizer==3.4.0	mpmath==1.3.0	six==1.16.0
click==8.1.8	multidict==6.4.3	soupsieve==2.6
colorama==0.4.6	multiprocess==0.70.16	stack-data==0.6.3
comm==0.2.2	nest-asyncio==1.6.0	sympy==1.13.1
contourpy==1.3.1	networkx==3.4.2	tenacity==9.0.0
cycler==0.12.1	nltk==3.9.1	threadpoolctl==3.5.0
datasets==3.5.0	numba==0.60.0	tokenizers==0.20.3
debugpy==1.8.8	numpy==2.0.2	torch==2.6.0+cu118
decorator==5.1.1	openpyxl==3.1.5	torchaudio==2.6.0+cu118
dill==0.3.8	packaging==24.2	torchvision==0.21.0+cu118
et_xmlfile==2.0.0	pandas==2.2.3	tornado==6.4.1
evaluate==0.4.3	parso==0.8.4	tqdm==4.67.0
exceptiongroup==1.2.2	pillow==11.0.0	traitlets==5.14.3
executing==2.1.0	platformdirs==4.3.6	transformers==4.46.2
filelock==3.16.1	plotly==5.24.1	typing_extensions==4.12.2
fonttools==4.56.0	prompt_toolkit==3.0.48	tzdata==2024.2
frozenlist==1.6.0	procache==0.3.1	umap-learn==0.5.7
fsspec==2024.10.0	psutil==6.1.0	Unidecode==1.3.8
hdbscan==0.8.39	pure_eval==0.2.3	urllib3==2.2.3
huggingface-hub==0.26.2	pyarrow==20.0.0	wcwidth==0.2.13
idna==3.10	Pygments==2.18.0	widetsnbextension==4.0.1
ipykernel==6.29.5	pynndescent==0.5.13	3
ipython==8.29.0	pyparsing==3.2.1	wordcloud==1.9.4
ipywidgets==8.1.5	python-	xxhash==3.5.0
jedi==0.19.2	dateutil==2.9.0.post0	yarl==1.20.0
Jinja2==3.1.4	python-docx==1.1.2	
joblib==1.4.2	pytz==2024.2	