T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BÜYÜK DİL MODELLERİ KULLANILARAK ANAHTAR KELİME ÜRETİMİ VE KONU MODELLEME : MÜŞTERİ GERİ BİLDİRİMLERİNDEN İÇGÖRÜ EDİNİMİ

Aleyna ER

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Banu DİRİ

Temmuz, 2024

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BÜYÜK DİL MODELLERİ KULLANILARAK ANAHTAR KELİME ÜRETİMİ VE KONU MODELLEME : MÜŞTERİ GERİ BİLDİRİMLERİNDEN İÇGÖRÜ EDİNİMİ

Aleyna ER tarafından hazırlanan tez çalışması 22.07.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Banu DİRİ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Banu DİRİ, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. İsim Soyisim, Üye
.... Üniversitesi _____

Prof. Dr. İsim Soyisim, Üye
.... Üniversitesi _____

Danışmanım Prof. Dr. Banu DİRİ sorumluluğunda tarafımca hazırlanan Büyük Dil Modelleri Kullanılarak Anahtar Kelime Üretimi Ve Konu Modelleme : Müşteri Geri Bildirimlerinden İçgörü Edinimi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Aleyna ER

İmza

Bilim Kadını olmak isteyen küçük Aleyna'ya

TEŞEKKÜR

Bu çalışmayı gerçekleştirirken bilgi ve desteğini hep hissettiğim, her zaman anlayışla yaklaşan değerli danışmanım Prof. Dr. Banu DİRİ'ye, akademik çalışmalarımı destekleyen ve yüreklendiren sayın Meltem TURHAN YÖNDEM ve Erdal ŞEKERCİ başta olmak üzere beraber çalışmaktan keyif aldığım yeteneklerimi geliştirmemde büyük katkısı olan Setur ekibine, süreç boyunca sevgi ve inançla hep yanımda olan ailem ve arkadaşlarıma gönülden teşekkür ediyorum.

Aleyna ER

İÇİNDEKİLER

KISALTMA LİSTESİ vii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xi

ABSTRACT xiii

1 GİRİŞ 1

1.1 Literatür Özeti . . . 3
1.2 Tezin Amacı . . . 7
1.3 Hipotez . . . 9

2 VERİ SETİ 10

2.1 Veri Seti İstatistikleri . . . 11
2.1.1 Anahtar Kelime Üretme Veri Seti İstatistikleri . . . 11
2.1.2 Konu Modelleme Veri Seti İstatistikleri . . . 13

3 YÖNTEMLER 15

3.1 Anahtar Kelime Üretimi için Ön Eğitimli Dil Modellerinde İnce Ayar Yapılması . . . 15
3.2 Anahtar Kelime Üretimi için Büyük Dil Modellerine İstem Verme . 17
3.3 Anahtar Kelime Kümeleme ile Konu Modelleme . . . 19
3.4 Sinirsel Konu Modelleme Yöntemleri (BERTopic) . . . 21
3.5 İstatistiksel Konu Modelleme Yöntemleri (LDA, NMF) . . . 22
3.6 Değerlendirme Metrikleri . . . 23
3.6.1 Anahtar Kelime Üretme Performansını Ölçme . . . 23
3.6.2 Konu Modelleme Performansını Ölçme . . . 25

4 DENEYSEL SONUÇLAR 27

4.1 Anahtar Kelime Üretimi Deneyleri . . . 27
4.1.1 Anahtar Kelime Üretimi Doğrulama Veri Seti Sonuçları . . 28

4.1.2 Anahtar Kelime Üretimi Test Veri Seti Sonuçları 32
4.1.3 Anahtar Kelime Üretme Çapraz Geçerlilik Sonuçları 33
4.2 Konu Modelleme Deneyleri 33
4.2.1 Anahtar Kelime Kümeleme ile Konu Modelleme Sonuçları . 34
4.2.2 Sinirsel Konu Modelleme (BERTopic) Sonuçları 41
4.2.3 İstatiksel Konu Modelleme Sonuçları 43

5 SONUÇ **46**

KAYNAKÇA **48**

A ANAHTAR KELİME ÜRETİMİ & KONU MODELLEME SONUÇLARI **52**

TEZDEN ÜRETİLMİŞ YAYINLAR **57**

KISALTMA LİSTESİ

AK	Anahtar Kelime
PLM	Pre-trained Language Model
LLM	Large Language Model
BERT	Bidirectional Encoder Representations from Transformers
T5	Text-to-Text Transformer
BART	Bidirectional and Auto Regressive Transformers
GPT	Generative Pre-trained Transformer
PCA	Principal component analysis
UMAP	Uniform Manifold Approximation and Projection
GMM	Gaussian Mixture Modelling
HDBSCAN	Hierarchical Density-Based Spatial Clustering
LR	Learning Rate
GAS	Gradient Accumulation Steps
TF-IDF	Term Frequency-Inverse Document Frequency
RNN	Recurrent Neural Network
GPT	Generative Pre-trained Transformer
MLM	Masked Language Modeling
KBIR	Keyphrase Boundary Infilling with Replacement
YAKE!	Yet Another Keyword Extraction
KEA	Keyphrase Extraction Algorithm
t-SNE	t-distributed Stochastic Neighbor Embedding
AHC	Agglomerative Hierarchical Clustering
CHI	Calinski-Harabasz Index

RP	Random Projection
DMM	Dirichlet Multinomial Mixture
BoW	Bag of Words
NPS	Net Promoter Score

ŞEKİL LİSTESİ

Şekil 1.1 Proje akışı 3
Şekil 1.2 AK üretimi örneği [6] 4
Şekil 2.1 Etiketleme örneği 11
Şekil 2.2 AK'leri oluşturan kelime sayısı dağılımı 13
Şekil 2.3 Üretilen AK'lerin kelime sayısı dağılımı 14
Şekil 3.1 T5 metinden metine formatı [17] 15
Şekil 3.2 T5 ön eğitim formatı [17] 16
Şekil 3.3 BART ön eğitim formatı [18] 17
Şekil 3.4 GPT4 modeli ile LLM istem örneği 18
Şekil 3.5 Yorumlar ile AK'leri kümeleme 19
Şekil 3.6 AK kümeleme ile konu modelleme akışı 19
Şekil 3.7 BERTopic konu modelleme akışı [46] 22
Şekil 3.8 LDA çalışma akışı [48] 23
Şekil 4.1 İnce ayar yapılan modellerin kayıp grafikleri 30
Şekil 4.2 PCA ile açıklanan varyans 34
Şekil 4.3 MT5-OpenAI-PCA-HDBSCAN konu modeli istatistikleri 36
Şekil 4.4 MT5-BERTurk-PCA-HDBSCAN konu modeli istatistikleri 37
Şekil 4.5 GPT4-OpenAI-PCA-HDBSCAN konu modeli istatistikleri 38
Şekil 4.6 GPT4-BERTurk-PCA-GMM konu modeli istatistikleri 39
Şekil 4.7 MT5-BERTurk-UMAP-GMM konu modeli istatistikleri 41
Şekil 4.8 BERTopic-PCA-GMM konu modeli konu dağılımları 42
Şekil 4.9 LDA konu modeli konu dağılımları 44
Şekil 4.10 NMF konu modeli konu dağılımları 45

TABLO LİSTESİ

Tablo 2.1 Konaklama sonrası anketleri açık uçlu soruları 10
Tablo 2.2 Veri seti istatistikleri 12
Tablo 2.3 Veri seti için anahtar kelime etiketleme kalitesi 13
Tablo 2.4 Konu modelleme veri seti istatistikleri 14
Tablo 2.5 Konu modelleme veri seti anahtar kelime üretme kalitesi 14
Tablo 3.1 İstem stilleri 18
Tablo 4.1 Model konfigürasyonları doğrulama sonuçları 29
Tablo 4.2 İnce ayar yapılan en iyi modellerin performans kararlılığı 29
Tablo 4.3 Modellerin doğrulama seti üzerindeki performansı 31
Tablo 4.4 Modellerin test seti üzerindeki performansı 32
Tablo 4.5 İnce ayarlı modellerin k-katlı çapraz geçerlilik doğrulama kayıpları 33
Tablo 4.6 MT5 ile AK kümeleme sonuçları 35
Tablo 4.7 MT5-OpenAI-PCA-HDBSCAN konu modeli 35
Tablo 4.8 MT5-BERTurk-PCA-HDBSCAN konu modeli 36
Tablo 4.9 GPT4& İstem2 ile AK kümeleme sonuçları 37
Tablo 4.10 GPT4-OpenAI-PCA-HDBSCAN konu modeli 38
Tablo 4.11 GPT4-BERTurk-PCA-GMM konu modeli 39
Tablo 4.12 MT5-BERTurk-UMAP-GMM konu modeli 40
Tablo 4.13 BERTopic sonuçları 41
Tablo 4.14 BERTopic-PCA-GMM konu modeli 42
Tablo 4.15 İstatistiksel konu modelleri sonuçları 43
Tablo 4.16 LDA modeli 43
Tablo 4.17 NMF modeli 44
Tablo A.1 AK Üretimi ÖRNEK-1 52
Tablo A.2 AK Üretimi ÖRNEK-2 53
Tablo A.3 AK Üretimi ÖRNEK-3 54
Tablo A.4 AK Üretimi ÖRNEK-4 55
Tablo A.5 Konu modelleme örneği 56

ÖZET

Büyük Dil Modelleri Kullanılarak Anahtar Kelime Üretimi Ve Konu Modelleme : Müşteri Geri Bildirimlerinden İçgörü Edinimi

Aleyna ER

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Banu DİRİ

Turizm sektöründe faaliyet gösteren Setur'un müşterilerinden anketler ile topladığı geri bildirim yorumları doğal dil işleme ile analizi yapılarak sunulan hizmet ile ilgili nelerden memnun kalındığı ve hangi noktaların gelişime açık olduğu tespit edilmek istenmiştir. Önerilen yöntemde, ilk aşamada yorumları temsil eden anahtar kelimeler (AK) oluşturulur, bu kelimeler ile metindeki önemli kelimeleri yakalamak amaçlanır. Yorumlardaki önemli kavramları yakalayabilmek için, Türkçe ve çok dilli veri setleri ile eğitilmiş ön eğitimli TRT5, MT5, TRBART ve MBART dil modellerine ince ayar yapılarak öğrenme transferi gerçekleştirilmiş ve GPT3.5, GPT4, Gemini Pro büyük dil modellerine one-shot ve few-shot stilinde istemler vererek anahtar kelimeler üretilmiştir. Üretilen AK'lerin etiketli AK'ler ile ne kadar örtüştüğü, yoruma benzerlikleri ve birbirlerine göre anlamsal çeşitlilikleri incelenmiştir. Aynı zamanda üretilen AK'lerin uzunluğu yorumun uzunluğu ile kıyaslanarak metni ne kadar sıkıştırabildiği ölçülmüştür.

Konu modellemenin sonraki adımında, üretilen anahtar kelimeler ve yorumlarının vektörleri oluşturulur ve uç uca eklenir. Elde edilen yeni vektörler kümelenerek konu modellemesi yapılır, oluşturulan her bir küme bir konu olarak ele alınır. Konuları temsil etmek için o konudaki yorumlara en benzer anahtar kelimeler seçilir. Uygulanan konu modelleme yöntemi ile konu içeriklerinde kelime öbekleri bulunabilir, bu sayede anlamsal olarak daha belirgin yorumlanması daha kolay konular oluşturulabilmiştir. Aynı zamanda, yorum-AK ikilileri ile kümeleme

yapılması ile bir yorum için birden fazla konu atanabilmesi sağlanır.

Üretilen AK'lerin kümelenerek oluşturulan konu modellerinin performansı sinirsel konu modelleme örneği olan BERTopic ve istatistiksel yöntemlerden LDA ve NMF ile karşılaştırılmıştır. Oluşturulan konular incelenerek, müşterilerin hizmet sürecindeki hangi noktalar ile ilgili yorum yaptıkları hakkında bilgi sahibi olunmuş, en çok konuşulan konular saptanmıştır.

Anahtar Kelimeler: Anahtar kelime üretimi, konu modelleme, metin kümeleme, büyük dil modelleri, ön eğimli dil modelleri, öğrenme transferi, kullanıcı geri bildirimleri, turizm, doğal dil işleme

ABSTRACT

KEYWORD GENERATION AND TOPIC MODELING USING LARGE LANGUAGE MODELS: EXTRACTING INSIGHTS FROM CUSTOMER FEEDBACK

Aleyna ER

Department of Computer Engineering
Master of Science Thesis

Supervisor: Prof. Dr. Banu DİRİ

It is aimed to determine what the customers of Setur, which operates in the tourism sector, are satisfied with and what points are open to development through the analysis of feedback comments collected from surveys using natural language processing. In the proposed method, in the first stage, key words representing the comments are created, and it is aimed to capture important words in the text with these words. In order to capture important concepts in the comments, transfer learning is performed by fine-tuning the pre-trained TRT5, MT5, TRBART, and MBART language models trained with Turkish and multilingual data sets and giving prompts in one-shot and few-shot style to GPT3.5, GPT4, and Gemini Pro large language models to produce keywords. The produced keywords have been examined in terms of how much they overlap with labeled keywords, their similarities to the comments, and their semantic diversity, and their ability to compress the text has been measured by comparing their lengths with the length of the comment.

In the next step of topic modeling, vectors of the produced keywords and their comments are created and concatenated end-to-end. The new vectors obtained are clustered to do topic modeling, and each cluster created is considered as a topic. The keywords most similar to the comments on that topic are selected to represent the topics. With the applied topic modeling method, word clusters can be found in topic contents, thus making it easier to create topics that are more semantically

distinct and easier to interpret. At the same time, assigning multiple topics for a comment is made possible by clustering with comment-keyword pairs.

The performance of the topic models created by clustering the produced keywords has been compared with BERTopic, an example of neural topic modeling, and statistical methods LDA and NMF. By examining the topics created, information has been obtained about which points the customers commented on during the service process, and the most talked-about topics have been determined.

Keywords: Keyword generation, topic modeling, text clustering, large language models, pre-trained language models, transfer learning, customer feedback, tourism, natural language processing

1
GİRİŞ

Hizmet sektörü, e-ticaret gibi insan ile direkt teması olan sektörlerde kullanıcı görüşlerini dinlemek verilen hizmeti ve ürün portföyünü değerlendirebilmek için kilit rol oynamaktadır. Çoğu şirket anketler aracılığı ile bu ölçümlemeyi yapmakta, anketlerde evet-hayır, şıklı ve puanlandırmalı sorular ile kullanıcıdan yorum toplamaktadır. Buna ek olarak, açık uçlu sorular vasıtasıyla diğer kategorilerde yakalayamadıkları sorun/önerileri almayı amaçlamaktadırlar. Puanlandırmalı sorular, analiz kolaylığı bakımından daha çok tercih edilen ve Net Tavsiye Skoru (NPS) ile müşterilerin memnuniyetini hesaplama için daha uygun soru tipleridir. Her ne kadar puanlandırmalı sorular NPS ölçümü için kullanılsa da bu metrik yanıltıcı olabilmektedir. Bazı kullanıcılar puanlandırmalı sorulara yüksek puan verme eğilimde olsa da, eleştirel yorumlar yapabilmektedir. Bu sebeple açık uçlu sorular sormak, alınan geri bildirimler ile iş akışındaki geliştirmeye açık noktaları gözden kaçırmamak, müşterilerin önem verdiği konular ile tercihleri hakkında detaylı içgörüler sağlamak için önemlidir.

Setur, 1960 yılından beri turizm sektöründe gümrüksüz satış mağazaları (duty free shop) işletimi, otel ve tur satışları gibi birçok alanda hizmet veren, Türkiye'de konumlandırılmış bir Koç Holding şirketidir. Özellikle otel - tur satışları için oteller ve tur düzenleyici şirketler ile anlaşarak müşterilere ürün sağlamakta, aracılık yapmaktadır. Setur, hem portföyündeki ürünleri değerlendirmek, hem de sunduğu hizmet kalitesini arttırabilmek için müşterilerinden geri bildirim toplamaktadır. 2020 yılından beri, müşterilerine hizmet sonlandıktan sonra sistematik olarak anket göndermektedir. Bu çalışmada, yurt içi – yurt dışı otel ve tur satışı hizmetlerini değerlendirmek için müşterilere gönderilen "Konaklama Sonrası Anketleri" analiz edilerek müşteri yorumlarından içgörü çıkarılması hedeflenmiştir. Konaklama Sonrası Anketleri'nde, otel ve tur ile ilgili çeşitli puanlandırmalı, şıklı, evet-hayır ve açık uçlu sorular bulunmaktadır. Şıklı sorular ve evet-hayır soruları daha çok kategoriler hakkında bilgi almayı amaçlarken, puanlandırmalı sorular hizmet esnasındaki kategorize edilmiş adımlar hakkında müşterilerin ne kadar memnun

kaldığına dair derecelendirme yapmalarını istemektedir. Açık uçlu sorular ise hizmet ile ilgili öneri, şikayet, memnuniyet detaylarını toplamaktadır. Açık uçlu sorulara gelen cevapların insan tarafından analizleri yoğun iş yükü getirdiğinden, puanlandırmalı sorulara gelen cevaplara öncelik verilmekte, puan analizi düşük olan anketler için müşteri yorumları incelenmekteydi. Sonrasında bu yorumlardan da bilgi edinebilmek amacıyla yorumların duygu (sentiment), niyet (intent) ve konu (subject) sınıflandırmaları yapılmaya başlanmıştır. Edinilen bilgiler yorumun spektrumu hakkında detay verse de, içeriğin ne olduğu belirlenen sınırlı sayıda konu sınıfları olarak etiketlendiğinden sığ bilgi sağlamaktadır. Puanlandırmalı sorulara yüksek yanıtlar veren bazı müşterilerin, açık uçlu sorulara negatif yanıt vermesi de yorum detaylarına ihtiyacı arttırmaktadır. Öte yandan, turizmin insan odaklı bir sektör olması; dünyada gelişen ekonomik kriz, bulaşıcı hastalık/pandemi, siyasi anlaşmalar gibi olaylardan oldukça etkilenmesi sebebiyle yorum konuları zamanla değişiklik göstermektedir. Bu durum, trendleri yakalayabilmek için yorumun içeriği hakkında dinamik analizler yapmayı gerektirmektedir.

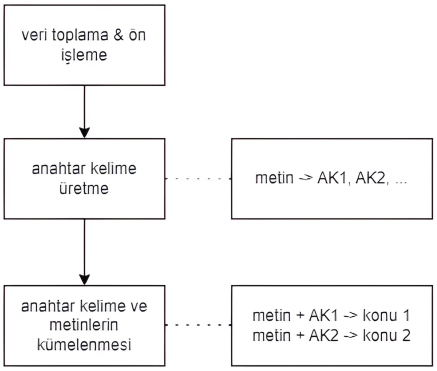
Konu modelleme (topic modeling), anahtar kelime (AK) çıkarma ve kelime bulutları (word cloud) oluşturma yöntemleri öngörülemeyecek konular ve trendleri gözlemlemek için elverişli bir yöntemdir. Kelime bulutları ile yorumlardaki sıkça kullanılan kelime grupları gözlemlenebilir, konu modelleme ile benzer yorumlar gruplanabilir ve grupları temsil eden kelimeler çıkarılabilir. Müşteri görüşlerinin öneminin bilincinde olarak benzer amaçlarla çeşitli çalışmalar yapılmıştır. Böyle bir çalışmada [1], elektronik cihazlarla ilgili çevrimiçi yorumların anahtar kelimeleri çıkarılmıştır. Skoghal ve arkadaşları [2] İsveç'teki bir otomotiv şirketinin müşteri deneyimi hakkında bilgi edinmek için müşteri geri bildirimlerinden anahtar kelime çıkarmaya yönelmiştir. Ayrıca konaklama alanında da en yeni trendleri ortaya çıkarmak için kullanıcı yorumlarından yararlanan araştırmalar bulunmaktadır [3],[4]. Kim ve arkadaşları [3] Google seyahat yorumlarından TF-IDF analizi yapmış ve kelime bulutları oluşturmuştur. Farklı bir çalışmada da [4] Londra 'daki Airbnb misafirlerinin tercihlerini anlamak için benzer bir yaklaşım benimsenmiştir.

Türkçe'nin biçimsel ve anlamsal zenginliği, anahtar kelimeleri yorumlamak için zorlayıcı olabilmektedir. Eş sesli kelimeler, farklı cümle yapıları ve dil kullanımının sektöre özgü değişmesi anahtar kelime çıkarma ve kelime bulutu oluşturarak metni analiz etmek için yetersiz kalabilmektedir. Örnek olarak bu çalışmada kullandığımız veri setlerinde, "arama" kelimesi hem çağrı merkezini aramak anlamında hem de web sitesi üzerinde gezinmek anlamında kullanılabilmektedir. Benzer şekilde "temsilci", "rehber", "görevli", "çalışan"kelimeleri de anlamsal olarak belirsizlik içermektedir. "Temsilci", "görevli" ve "çalışan" kelimeleri otel personelini, otelde bulunan Setur personelini, genel müdürlüğe bağlı Setur çağrı

merkezi çalışanını ya da Setur'a bağlı acente personelini temsil edebilmektedir. "Rehber" kelimesi ise hem tur rehberini belirtmek hem de otelde bulunan Setur temsilcisini (Setur otel rehberi) tanımlamak için kullanılabilmektedir. Anahtar kelimelerin direkt olarak kümelenmesi yahut gözlemlenmesi yanlış çıkarımlara sebep olabilir. Diğer seçenek olan konu modelleme yöntemi ise vaka analizlerinde yetersiz kalmaktadır. Konu modelleme algoritmalarının çıkardığı konuların ve konu içindeki kelimelerin yorumlanması zordur. Aynı zamanda, konu modelleme yöntemlerinde, bir metnin tek bir konuya ait olabilmesi gerçek hayat verilerinde işlevsel bulunmamaktadır, kullanıcı yorumları her ne kadar kısa metinler olsa da, çalıştığımız veri setinde müşteriler birden fazla konuya değinmektedir.

Bu çalışma, Setur'un müşterilerine gönderdiği "Konaklama Sonrası Anketleri" ne verilen geri bildirim metinlerinden anahtar kelime çıkarma ve bu kelimeleri ait oldukları metinler ile beraber kümeleyerek konu modelleme yöntemlerine yeni bir bakış açısı sunmaktadır (Şekil 1.1).

Metinleri temsil eden anahtar kelime örüntülerinin sinirsel dil modelleri tarafından öğrenilmesi, bu kelimelerin metinler ile birlikte kümelenmesi sonucunda konu modelleme kalitesinin arttırılması ve daha iyi içgörüler çıkarılması hedeflenmiştir.



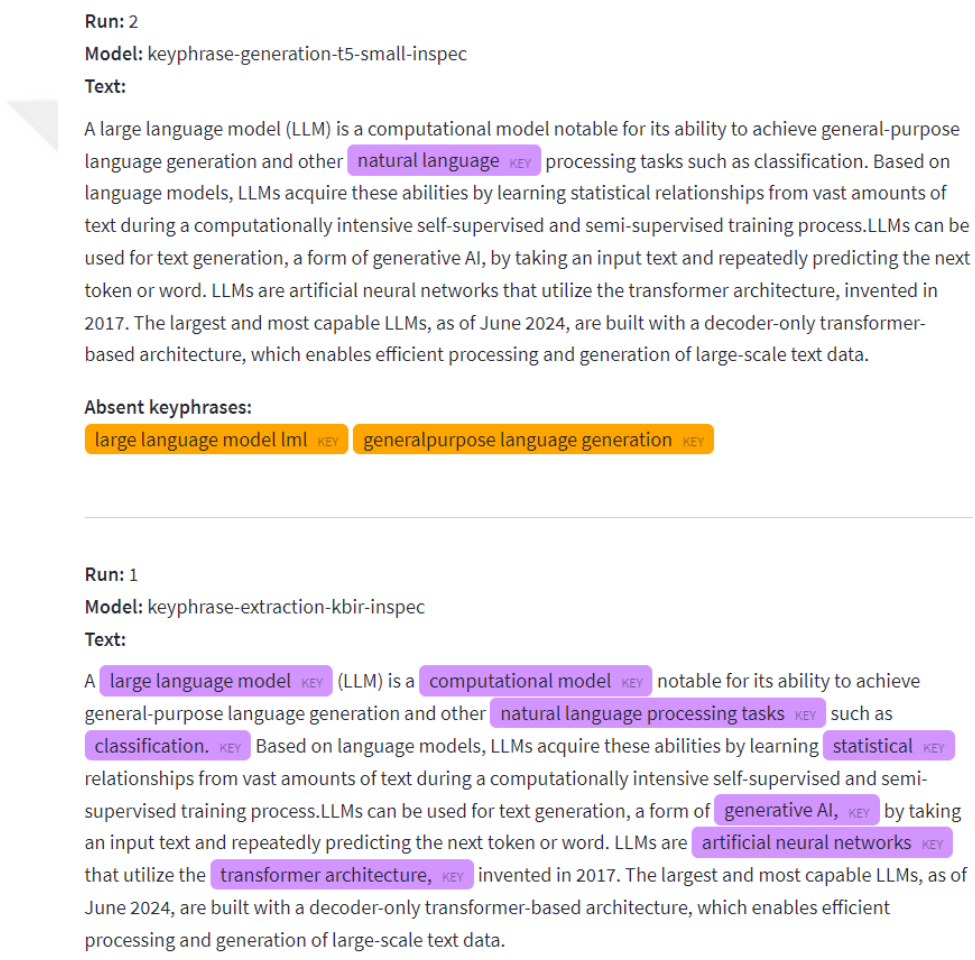
Şekil 1.1 Proje akışı

1.1 Literatür Özeti

Anahtar kelimeler metinlerin konusu hakkında fikir veren, metinle ilgili kısa ve öz bilgiler sunan kelime öbekleridir. Metinlerdeki ana bilgiyi özetlediklerinden dolayı, sınıflandırma, bilgi getirme (information retrieval) ve öneri sistemlerinde kullanılabilirler.

Metinlerde anahtar kelime bulmak için iki ana yaklaşım izlenmektedir: Çıkarıcı

(extractive) ve yorumlayıcı (abstractive). Çıkarıcı yaklaşımda, metinle ilgili bilgi veren kelime öbekleri metnin içinden seçilir, bunlara "present keywords" denir. Yorumlayıcı yaklaşımda ise, anahtar kelimeler metnin içinde birebir geçmez, “absent keywords“ olarak adlandırılırlar. Anahtar kelime üretimi ise (keyword generation) iki ana yaklaşımı harmanlar, hem metnin içinde direkt olarak bulunan hem de metinde geçmeyen anahtar kelimeleri bulmaya odaklanır [5]. Şekil 1.2’de AK üretme ve AK çıkarma örnekleri verilmiştir. Altta verilen Run:1 örneğinde AK çıkarma yaklaşımı metnin içindeki önemli kelimeleri seçerek AK’leri belirler. Üstteki Run:2 örneğinde ise hem metin içinde bulunan hem de bulunmayan AK’ler üretildiği görülmektedir.



Şekil 1.2 AK üretimi örneği [6]

Anahtar kelime bulma işleminde zaman içerisinde denetimli (supervised) ve denetimsiz (unsupervised) yaklaşımla çalışan birçok yöntem önerilmiştir. Metindeki önemli parçaları, anahtar kelimeleri belirleyen istatistiksel tabanlı yöntemlere TF-IDF [7] ve KP-MINER [8] örnek verilebilir. YAKE! [9]’de ise özellik tabanlı (feature-based) bir sistem önerilmiş, terimlerin birlikte görülme sıklığı (co-occurance) ve frekansları gibi yerel metin özellikleri ve istatistiksel

bilgiler de göz önünde bulundurulmuştur Graph-tabanlı çözümler PageRank [10], TextRank [11] geliştirilmiş, sinirsel yaklaşımlarda dil modelleri ile bağlamsal olarak önemli kelimelerin saptanması anlamsal benzerlik ölçülerek çalışılmıştır [12, 13]. Bu yaklaşımlarda POS örüntüleri anahtar kelimelerin filtrelenmesinde kullanılmış, graph ve dil modelleri ile birlikte çalışan hibrit yaklaşımlar önerilmiştir [14].

Denetimli yaklaşımlarda; KEA'da [15, 16] denetimsiz şekilde aday AK'lerin AK olup olmadığın sınıflandıran Naive Bayes modeli çalışılmış, bunu seq2seq (encoder – decoder) mimarisini kullanan RNN ve Transformer tabanlı yaklaşımlar izlemiştir. Pretrained dil modellerinin yaygınlaşmasıyla birlikte, T5 [17] ve BART [18] gibi transformer-tabanlı modellerin metin & AK ikilileri ile ince ayar yapılarak anahtar kelimeler elde edilmesi kolaylaşmış, anahtar kelime çıkarma terimi de anahtar kelime üretimi olarak ele alınmaya başlanmıştır. Wu ve arkadaşları çalışmasında PLM'lerin performansını karşılaştırmak için problemi hem AK çıkarma hem de üretme olarak ele almıştır [19]. Dil modellerini hem ince ayar yaparak, hem de sıfırdan eğiterek, dil modellerini eğitmenin görev üzerindeki etkisini gözlemlemişlerdir. Benzer bir diğer çalışma [20] ise, encoder-decoder tipindeki PLM'lerin performansını karşılaştırmak için problemi AK üretme olarak ele almıştır. BART ve T5 modellerine ince ayar yapılmış ve AK'ler için alfabetik ya da metinde görülme sırasına göre farklı sıralama yaklaşımları uygulanmıştır. İki çalışmada da PLM'lere ince ayar yapmanın önceki TF-IDF, YAKE, KEA yöntemlerine kıyasla daha başarılı olduğu gözlemlenmiştir.

Transformer-tabanlı bir diğer çalışma olan KBIR'de (Keyphrase Boundary Infilling with Replacement) [21], AK sınırlarını doldurma ve değiştirme amacıyla kullanılan yeni bir ön eğitim yöntemidir. AK'nin metinde bulunduğu lokasyonu göz önünde bulunduran bu yöntem, iki temel bileşen olan Keyphrase Boundary Infilling (KBI) ve Keyphrase Replacement Classification (KRC) görevlerini içerir ve RoBERTa üzerinde uygulanır. Maskelenmiş dil modeli (MLM) hedefiyle birlikte KBI ve KRC hedeflerini optimize eder.

Günümüzde GPT4 [22], Gemini [23] gibi büyük dil modelleri (LLM) ile herhangi bir eğitime gerek duymadan promptlar (istemler) ve örnekler vererek metinlerden AK üretilebilmektedir, bu örnekteki ilk çalışmalar [24, 25] zero-shot prompt stili ve farklı prompt metinleri ile bu yaklaşımları uygulamışlardır.

Araştırmacıların anahtar kelime üretme görevi için sıklıkla tercih ettiği veri setleri bilimsel makaleler, haber metinleri gibi yapısal olarak daha düzgün metinler içermektedir. En sık kullanılan veri setleri:

1. Inspec: 2000 makale özeti içerir
2. Krapivin: 2304 makale için metin ve anahtar kelimeleri içerir
3. NUS: 211 paper içerir
4. SemEval-2010: ACM Digital Library'den 288 paper içerir
5. KP20k: Bilgisayar bilimleriyle ilgili yaklaşık 20000 yayının başlıkları, abstractları ve keyphraselerini içerir
6. SemEval2017: Farklı alanlardaki bilimsel yayınlardan manuel olarak alınan paragrafları içerir

Ayan ve ekibinin çalışmasında [13] farklı olarak, Türkiye'de faaliyet gösteren seçili firmaların internet sitesinden metinler toplanmış, bu metinler için AK'ler etiketlenmiştir. Etiketli AK'leri vektörleştirmek için çok dilli veri setinde önceden eğitilmiş dil modelleri ile Word2Vec modeli kullanılmış, bu vektörler boyut azaltımı yapılmadan K-Means ve hiyerarşik kümeleme modelleri ile kümelenmiştir. Çalışmada ön eğitimli dil modellerinin Word2Vec'e göre vektörleştirme işleminde daha başarılı olduğu ve hiyerarşik kümeleme ile daha yüksek Silhoutte skorlarına ulaşıldığı belirtilmiştir. Bir diğer metin kümeleme çalışmasında [26], farklı veri setleri üzerinde yapılan deneylerde, vektörleştirme için TF-IDF ve BERT modellerinin yanı sıra OpenAI, Falcon ve LLaMA-2 büyük dil modelleri kullanılmış üretilen vektörler için PCA ve t-SNE teknikleri ile boyut indirgeme yapılmıştır. Elde edilen vektörlerin kümelenmesinde K-Means, AHC ve FuzzyCM modellerinin performansı Silhoutte ve CHI skorları ile kıyaslanmıştır. Kümeleme işlemlerinde, metinleri temsil etmek üzere özetleme işlemi yapılmış ve özetleri kullanma ile metinlerin orjinalini kullanmanın sonuca etkisi araştırılmıştır. Deney sonuçları incelendiğinde metin özetleme işleminin performans üzerinde bariz bir etkisi olmadığı ve en iyi sonuçlara ulaşan konfigürasyonlarda vektörleştirme için OpenAI modelinin kullanıldığı gözlemlenmiştir. Bu modeli BERT takip etmektedir.

Metin kümeleme yaklaşımını LDA ve DMM tabanlı konu modelleme yöntemleri ile karşılaştıran bir araştırma [27], görsel açıklamaları ve tweetler içeren veri setleri üzerinde çalışmış, metin kümeleme işleminde vektörleştirme için BERT, Fasttext ve TF-IDF yöntemlerini test etmiş kümeleme için ise gürültüye duyarlı olmaması sebebiyle HDBSCAN modelini tercih etmiştir. Vektörleştirme için BERT ile daha iyi sonuç aldıklarını ve HDBSCAN ile kümeleme işleminin temel konu modellerine kıyasla yüksek performans gösterdiğini vurgulamışlardır.

Gözetimsiz-yarı gözetimli makine öğrenmesi alanında araştırılan konu modelleme çalışmalarının bu örneğinde [28] yazarlar LDA ve NMF ile oluşturulan konuların, RP ve PCA'ye göre daha tutarlı olduğunu ve daha yüksek Coherence skorlarına ulaştığını raporlamışlardır. Kısa metinler ile konu modellemesi yapan [29] çalışmasında yazarlar kullandıkları konu modellemesi tekniklerini DMM tabanlı, kelimelerin birlikte görülmesi tabanlı (global word co-occurance) ve self-aggregation tabanlı olmak üzere 3 ana başlıkta toplamıştır. Bu teknikler sınıflandırma, kümeleme ve konu uyumluluğu çerçevesinde incelenmiş ve 6 adet gerçek veri seti üzerinde test edilmiştir. Konu modellemek için kelime vektörlerinden faydalanan teknikler, hız performansı açısından dezavantajlı olsa da doğruluk açısından daha yüksek performans göstermiştir. Kelime vektörleri ile geleneksel yöntemleri iyileştirmeyi hedefleyen bir çalışmada [30], araştırmacılar BoW yerine kelime vektörlerini kullanarak LDA ile konu modelleme yaparak daha yüksek Coherence (konu tutarlılığı) skorları almışlardır. Kelime vektörleri kullanarak sinirsel konu modelleme yapma ile geleneksel konu modellerini kıyaslayan bir diğer çalışma [31], sosyal medyadan topladığı yapısal olmayan metinler ile veri seti oluşturmuş, geleneksel LDA ve NMF yöntemleri ile sinirsel yaklaşımı yani vektör tabanlı konu modellemesi yapan Top2Vec ve BERTopic yöntemlerini karşılaştırırken, BERTopic ve NMF ile daha anlaşılır konular oluşturabilmiştir.

Yapılan çalışmalarda kısa metinler için DMM tabanlı yaklaşımlar ile iyi sonuçlar alınsa da bu yaklaşımda da her bir metin için sadece bir konu atanmaktadır. Kelime vektörleri ile performansın arttırılabilmesi, HDBSCAN ile gürültüye dayanıklı bir şekilde kümeleme yapılabilmesi deneylerde kullanılacak konfigürasyonları belirlemek için ilham verici olmuştur.

1.2 Tezin Amacı

Kullanıcı yorumları sunulan hizmeti ölçebilmek adına değerli bir kaynaktır, ancak analiz süreci oldukça zaman ve efor istemektedir. Verilen geri bildirimleri otomatik olarak analiz edebilmek, insanların ne konuştuğu hakkında anlamlı maddeler çıkarabilmek iş yükünü hafifletecektir. Bu hedefte, kullanıcı yorumlarını doğal dil işleme yöntemleriyle analiz edebilmek için farklı yöntemler denenmiştir. Çalışmanın amacı, Türkçe yorumları hem konu modelleme yöntemleriyle analiz ederken, yorumların ana fikrini taşıyan anahtar kelimeleri bulmak, hem de çıkarılan anahtar kelimeleri kümeleyerek daha anlamlı konular oluşturmaktır.

Yorumları temsil eden anahtar kelimeleri bulabilmek için, sinirsel dil modelleri

ile çalışılarak metindeki hangi kelimelerin önemli olduğunun dil modellerine öğretilmesi amaçlanmıştır. Öğrenim sürecinde anahtar kelimelerin bulunmasında, metinde hangi kavramlar için yorum yapıldığı cümle yapısından bağımsız olarak anlamsal örüntüler çıkararak modelleri eğitmek amaçlanmıştır. Çok dilli modeller ile Türkçe veri ile eğitilmiş modellere ince ayar yapmanın performansı nasıl etkilediğini incelemek için aynı model ailesindeki farklı versiyonlar kullanılmıştır. Az miktarda etiketli veri ile hatta hiç etiketli veri olmadan çalışmanın mümkün olup olmadığını gözemleyebilmek için büyük dil modellerine zero-shot ve few-shot yöntemler ile istem (prompt) vererek deneylerin yapılması, bu modellerin ön-eğitimli (pre-trained) dil modellerine ince ayar yaparak farkın ölçülmesi istenmiştir.

Metni temsil eden anahtar kelimeler üretildikten sonra, bu kelimeler kümelenerek konuların oluşturulması hedeflenmiştir. İstatistiksel ve sinirsel konu modelleme yöntemlerine kıyasla daha verimli konular oluşturabilmek için farklı bir yaklaşım önerilmektedir. Sunulan yaklaşımda öncelikle konuları oluşturan kelimeler dil modelleri ile üretilmiş anahtar kelimeler olacaktır. Bu kelimeleri kategorilerine ayırabilmek için her bir anahtar kelime vektörü temsil ettikleri metinlerin vektörleri ile birlikte kümelenecektir. İki vektörün işlenmesinde, vektörler hem uç uca eklenerek hem de ortalamaları alınarak performans kıyaslaması yapılacaktır. Kullanıcı yorumu vektörleri ve bu yorumlardaki anahtar kelimelerin vektörlerinin birleştirilmesi sonucu oluşan vektörler kümeleme algoritmalarıyla ayrıştırılacaktır. Böylece, hem bir yorum birden fazla konu ile temsil edilebilecek hem de kelimeler metinde kullanılan anlamlarını kaybetmeyeceklerdir. Bu çalışmada önerilen konu modelleme yaklaşımı üç adımda özetlenebilir:

1. Metinlerden anahtar kelimelerin üretilmesi,
2. Metindeki her anahtar kelime için anahtar kelime vektörü ve metnin vektörünün birleştirilerek yeni vektörlerin oluşturulması,
3. Oluşturulan anahtar kelime + metin vektörklerinin kümelenmesi

Burada önerilen yöntem ile hem istatistiksel ve sinirsel konu modelleme yöntemlerindeki metinlerin tek bir konuya dahil olabilmesi kısıtının aşılması hem de daha anlamlı konular oluşturulması hedeflenmiştir.

1.3 Hipotez

1. Metni temsil eden anahtar kelimeleri bulma görevinde, büyük dil modelleri kullanılarak daha az miktarda etiketli veri ile eğitim görece küçük dil modellerini daha çok veri ile ince ayar yapılarak eğitilmesinden daha başarılı sonuçlar verebilir.

2. Türkçe veri seti ile eğitilmiş dil modelleri, çok dilli versiyonlarına kıyasla Türkçe veri setinde daha iyi çalışabilir.

3. Konu modelleme yaparken, konuyu oluşturan kelimelerin metinlerdeki önemli kelimeleri yakalama örüntüsünü öğrenen modeller ile seçilmesi daha anlamlı konular oluşturmayı sağlayacak, aynı zamanda konu içindeki kelimeler de daha tutarlı olacaktır.

4. Metinler ile her bir anahtar kelimesinin vektörlerinin birleştirilerek beraber kümelenmesi ile bir metin birden fazla konuya dahil olabilecektir.

2
VERİ SETİ

Anahtar kelime üretme görevi için, Setur'un 2020-2022 yılları arasında müşterilerine gönderdiği 1000 anket rastgele olarak seçilmiştir. Bu anketlerdeki müşteriden hizmet hakkındaki yorumları istenen, önerileri şikayetleri sorulan açık uçlu üç farklı soru bulunmaktadır (Tablo 2.1). Ürün ve satış kanalına göre gönderilen anketlerdeki sorular değişmektedir, bu sebeple bazı anketlerde açık uçlu bu üç sorudan ikisi birlikte bulunabilmektedir. Ankette birden fazla açık uçlu soru bulunması durumunda, sorulara verilen yanıtlar birleştirilerek tek bir yorum olarak ele alınmıştır.

Verilen cevaplarda herhangi bir kavram üzerinde yorum yapılmıyorsa örneğin sadece teşekkür ediliyor, "hayır sağ olun", "yok teşekkürler" gibi kısa cevaplar veriliyorsa bu yorumların anahtar kelimesi "diğer" olarak işaretlenir. Yorum uzunluğu 3 ve daha az karakter olan yorumlar filtrelenerek veri setine dahil edilmemiştir.

Tablo 2.1 Konaklama sonrası anketleri açık uçlu soruları

	Soru başlığı
S1	Aldığınız hizmeti iyileştirmek için bize iletmek istediğiniz herhangi bir öneriniz/şikayetiniz var mı?
S2	Bize ek olarak iletmek istediğiniz öneri/şikayetiniz var mı?
S3	Setur ne yapmış olsaydı daha yüksek bir tavsiye puanı verirdiniz

Seçilen 1000 anket için, yorumlardaki anahtar kelimeler belirlenerek etiketleme yapılmıştır. Seçilen anahtar kelimeler hem metnin içinde geçen hem de geçmeyen ancak anlatılan kavramı belirten kelimelerdir. Anahtar kelimeleri yazarken [20] çalışmasında kullanılan sıralama yaklaşımlarından görünme sırasına benzer olarak cümledeki anlatım sırası göz önünde bulundurulmuştur, ancak metnin içinde geçmeyen (absent) AK'lerin başa eklenme durumu yoktur.

Etiketleme stili Şekil 2.1'de verilmiştir, anahtar kelimeler etiketlenirken ";" işareti ile ayrılmış ve yazımda küçük harfler kullanılmıştır. Anahtar kelime verisi üç ayrı

yorumlar	anahtar kelimeler
daha fazla otel ile anlaşma yapılması ve çağrı merkezine kısa sürede ulaşım	otel çeşitliliği; çağrı merkezi
yaşadığım soruna sahip çıksaydı	problem çözme
her şey çok güzeldi teşekkür ederiz	diğer
otel covid önlemleri yönünden yetersiz	covid önlemleri

Şekil 2.1 Etiketleme örneği

veri seti şeklinde kullanılmıştır: 50 örnek test için, kalan 950 örneğin %80'i anahtar kelime modelleri eğitimi %20'si ise bu modellerin doğrulaması için kullanılmıştır.

Konu modelleme görevi için, Setur'un müşterilerine 2023 yaz döneminde (1 Haziran 2023 – 1 Ekim 2023) gönderdiği rastgele seçilen 500 anketteki açık uçlu sorular kullanılmıştır. Anahtar kelime veri setine benzer şekilde, eğer aynı ankette birden fazla açık uçlu soru bulunuyorsa bunlara verilen cevaplar birleştirilerek tek bir yorum gibi ele alınmıştır. Konu modelleme verisi için herhangi bir etiketleme yapılmamış, yaklaşımların çıkardığı konu kümelerini incelemek için kullanılmıştır.

2.1 Veri Seti İstatistikleri

Hazırlanan veri setleri üzerine uzunluk ve etiketleme kalitesi bakımından ölçümlemeler yapılmıştır. Kullanıcı yorumlarının içerdiği kelime ve karakter sayıları ile bunları temsil eden anahtar kelimelerin kelime ve karakter sayıları kıyaslanmıştır. Anahtar kelimeler ile metinleri temsil etmek, metin özetlemeye benzer şekilde kısa ve öz şekilde anlatılan ana fikri yansıtmaya çalışmaktır. Bu sebeple, anahtar kelimelerin hem kullanıcı yorumuna kıyasla daha kısa, hem de kullanıcı yorumu ile yüksek anlamsal benzerlikte olması gerekmektedir. Anahtar kelime üretme ve anahtar kelimeler ile metinleri kümeleme görevleri için oluşturulan veri setlerinin istatistikleri Bölüm 2.1.1 ve Bölüm 2.1.2'de verilmiştir.

2.1.1 Anahtar Kelime Üretme Veri Seti İstatistikleri

Anahtar kelime üretimi için hazırlanan ve AK'leri etiketlenen 1000 örneklik veri seti, üçe bölünerek kullanılmıştır. Önce verinin %5'lik kısmı model performanslarını test edip kıyaslamak için ayrılmış, kalan kısmın %80'i modellere ince ayar yaparak eğitmek, %20'si ise doğrulama için ayrılmıştır.

Tablo 2.2'de veriye ait istatistikler verilmiştir. Eğitim setinde bulunan kullanıcı yorumları yaklaşık 144 karakter ve 20 kelime uzunluğunda, AK'ler ise yaklaşık 20 karakter ve 3,4 kelimeden oluşmaktadır (Tablo 2.2a). Doğrulama kısmında, kullanıcı yorumları ortalama 24 kelime uzunluğundadır, bunu temsil eden AK'ler ise 3,5 kelime bulundurmaktadır (Tablo 2.2b). Test setindeki 50 örnek farklı

yaklaşımlarla eğitilen modelleri kıyaslamak için ayrılmıştır. Tablo 2.2c'e bakıldığında, kullanıcı yorumları ortalama 21 kelimeden, temsil için etiketlenen AK'ler ise 3.3 kelimeden oluşmaktadır. Test setinde ise eğitim setine göre daha uzun yorumların bulunduğu ve daha az AK kullanılarak temsil edildiği görülmektedir.

Tablo 2.2 Veri seti istatistikleri

(a) Eğitim seti istatistikleri

İstatistikler	**Yorum**	**AK**
Toplam Karakter Sayısı	109478	18174
Toplam Kelime Sayısı	15257	2598
Ortalama Karakter Sayısı	144,05	23,91
Ortalama Kelime Sayısı	20,07	3,41

(b) Doğrulama seti istatistikleri

İstatistikler	**Yorum**	**AK**
Toplam Karakter Sayısı	33296	4628
Toplam Kelime Sayısı	4577	672
Ortalama Karakter Sayısı	175,24	24,35
Ortalama Kelime Sayısı	24,08	3,53

(c) Test seti istatistikleri

İstatistikler	**Yorum**	**AK**
Toplam Karakter Sayısı	7458	1160
Toplam Kelime Sayısı	1045	168
Ortalama Karakter Sayısı	149,16	23,20
Ortalama Kelime Sayısı	20,90	3,36

Oluşturulan veri setindeki etiketleme kalitesi üç metrik ile ölçülmüştür: Bert-Similarity, Diversity ve Compressing. Bu metrikler AK'ler ile metinlerin benzerliğini ölçme, aynı metni temsil eden AK'lerin birbirinden benzersizliği ve metni temsil eden AK'lerin metnin boyutunu ne kadar sıkıştırabildiği bakımından kullanılmaktadır. Detaylı açıklamaları Bölüm 3.6.1'de verilmiştir.

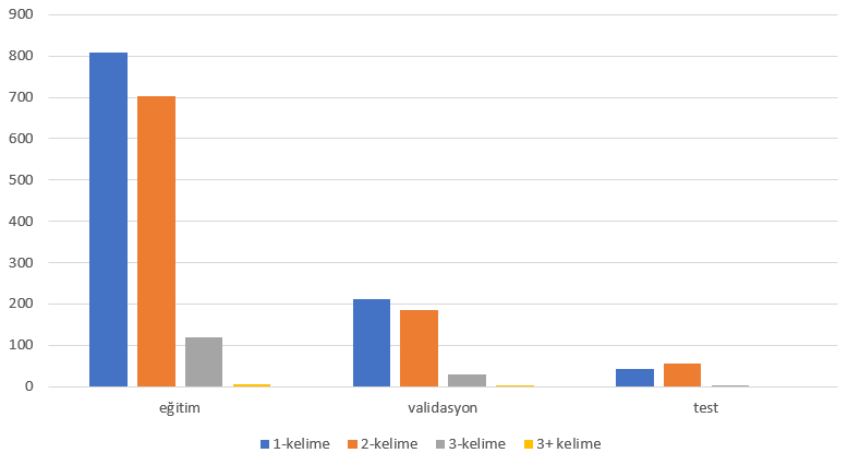
Tablo 2.3'te verildiği üzere, metne en benzer AK'ler eğitim setinde bulunmaktadır ve bu setteki AK'ler aynı zamanda birbirinden anlamsal olarak uzaktır. Diğer bir deyişle anlamsal olarak tekrara düşülmemiştir. Metnin uzunluğunu en iyi sıkıştırabilen, az karakter ile metni özetleyebilen kısım ise test seti olmuştur.

Metne karşılık gelen AK'lerin kaç kelimeden oluştuğunun grafiği Şekil 2.2'de verilmiştir. Eğitim ve doğrulama setindeki AK'ler çoğunlukla tek kelimeden

Tablo 2.3 Veri seti için anahtar kelime etiketleme kalitesi

Metrikler	Bert-Similarity	Diversity	Compressing
Eğitim seti	0,7477	0,2607	0,6982
Doğrulama seti	0,7330	0,2481	0,7455
Test seti	0,7085	0,2442	0,7849

oluşmaktadır, test setinde ise 2 parçadan oluşan AK'ler yoğunluktadır. Üçten fazla kelimeden oluşan AK'lerin sayısı ise oldukça azdır.



Şekil 2.2 AK'leri oluşturan kelime sayısı dağılımı

2.1.2 Konu Modelleme Veri Seti İstatistikleri

Anahtar kelime kümeleyerek konu modelleme görevi için 500 örneklik bir veri seti hazırlanmıştır. Bu veri setinde farklı olarak AK'ler için elle etiketleme yapılmamış, bunun yerine önerilen iki farklı yaklaşımın (ince ayar yapma ve prompt verme) en iyi modelleri kullanılarak AK'ler ürettirilmiştir. Seçilen iki modelin performansa etkisi Bölüm 4.2.1'de verilmiştir.

Tablo 2.4, veri seti hakkındaki istatistikleri göstermektedir. Kullanıcı yorumları ortalama yaklaşık 105 karakter ve 14 kelimeden oluşmaktadır. Bu yorumları temsil eden AK'ler için; GPT4 modeli ile üretilen AK'ler ortalama 36 karakter ve 5 kelimeden oluşmakta; ince ayar yapılan MT5 modeli ile üretilen AK'ler ise 20 karakter ve 3 kelimeden oluşmaktadır.

Veri seti için üretilen AK'lerin kalitesine baktığımızda, GPT4 ile üretilen AK'lerin metne anlamsal olarak daha çok benzediği, Bert-Similarity skorunun daha yüksek olduğu görülmektedir (Tablo 2.5). Aynı zamanda GPT4 ile üretilen AK'lerin birbirine göre anlamsal olarak daha uzaktır ve Diversity skorları yüksektir. MT5 modeli ile üretilen AK'ler ise metni sıkıştırma bakımından daha başarılıdır.

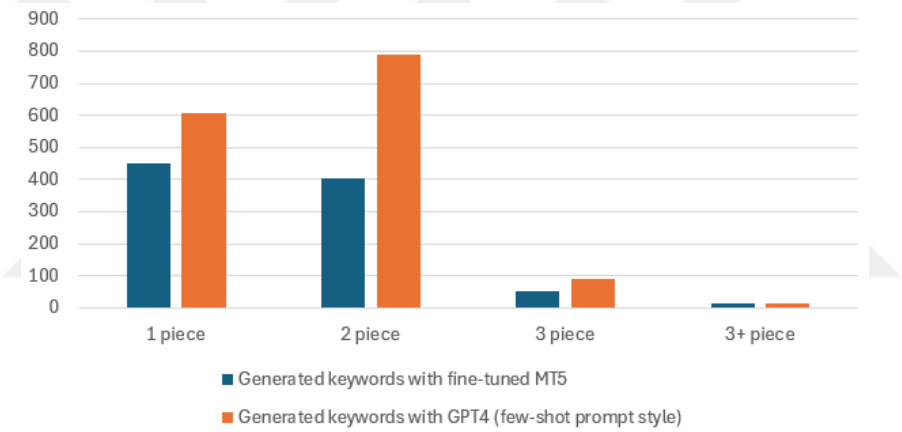
Tablo 2.4 Konu modelleme veri seti istatistikleri

İstatistikler	Yorum	MT5 ile üretilen AK	GPT4 ile üretilen AK
Toplam Karakter Sayısı	52723	10158	18185
Toplam Kelime Sayısı	7091	1483	2515
Ortalama Karakter Sayısı	105,44	20,31	36,37
Ortalama Kelime Sayısı	14,18	2,96	5,03

Tablo 2.5 Konu modelleme veri seti anahtar kelime üretme kalitesi

Metrikler	Bert-Similarity	Diversity	Compressing
MT5 ile üretilen AK	0,6712	0,2511	0,6556
GPT4 ile üretilen AK	0.7078	0,2612	0,2316

GPT4 modeli ile üretilen AK'ler çoğunlukla iki kelimeden oluşmaktadır, MT5 ise yoğunlukla tek ve iki kelimeden oluşan AK'ler üretmiştir (Şekil 2.3). Her iki model de az da olsa 3'ten fazla parçadan (kelimeden) oluşan AK'ler üretmiştir.



Şekil 2.3 Üretilen AK'lerin kelime sayısı dağılımı

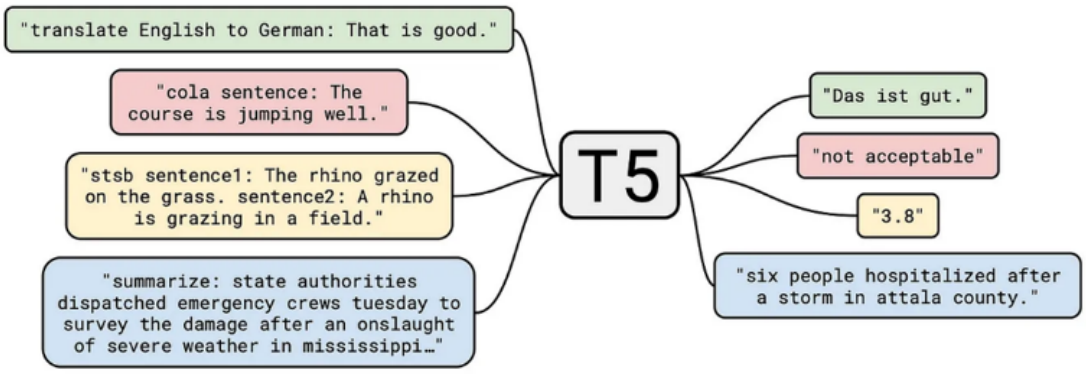
3
YÖNTEMLER

Bu bölümde kullanıcı yorumlarını analiz etmek üzere yapılan AK üretimi, bu AK'lerin kümelenerek konu modellemesi çalışmaları ve karşılaştırılan diğer konu modelleme yaklaşımları incelenmektedir. AK üretimi için ön eğitimli dil modellerine ince ayar yapma yaklaşımı Bölüm 3.1'de ve büyük dil modellerine farklı prompt verme stilleri ile AK üretme yaklaşımı Bölüm 3.2'de verilmiştir. Bölüm 3.3, 3.4 ve 3.5'te kullanılan konu modelleme yaklaşımları anlatılmaktadır.

3.1 Anahtar Kelime Üretimi için Ön Eğitimli Dil Modellerinde İnce Ayar Yapılması

AK üretimi için seçilen ilk yöntem dil modelleri kullanılarak girdi metin (kullanıcı yorumu) verilerek çıktı metin (AK) üretmektir. Bu hedefte, (seq2seq) model mimarisi kullanan ön-eğitimli T5 ve BART modellerine ince ayar yapılmıştır.

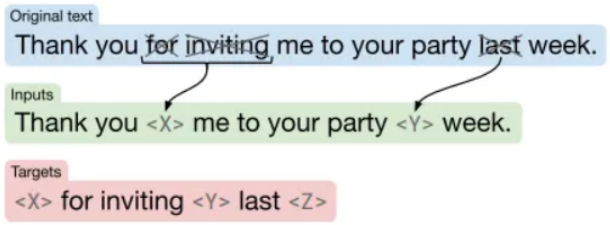
Google Research ekibi tarafından geliştirilen T5 modeli [17], görevleri "metinden metine" dönüşüm problemi olarak ele alır. Girilen ön ek (prefix) yardımıyla problem tanımı yapılabilir (Şekil 3.1), böylece model birden fazla görev için kullanılabilir.



Şekil 3.1 T5 metinden metine formatı [17]

Modelin ön eğitim aşamasında, Google Research ekibi dil işleme için farklı objektifler denemiştir. Otoregresif yaklaşım ile bir sonraki kelimeyi/kelimeleri tahmin etme, BERT benzeri yaklaşım ile girdi metindeki bazı kelimeleri maskeleyerek orjinal metni bulma, girdi metindeki kelimelerin yerlerini değiştirerek orjinal metni bulma, girdi metindeki bazı kelimeleri maskeleyerek/silerek çıktı olarak gizlenen kelimeleri üretme gibi farklı yöntemler ile ön eğitim yaparak performansı kıyaslamışlardır.

Denetimsiz ön eğitim aşamasında en iyi sonucu, girdi sekansından rastgele olarak %15 oranında token seçerek buradaki ardışık aralıkları "sentinel token" adı verilen tek bir token ile değiştirme ve bu değiştirilen tokenları modelin çıktısı olarak ürettirerek elde etmişlerdir. Google ekibinin T5 modeli için önerdiği yöntem Şekil 3.2 de verilmiştir. Bu yaklaşım ile ön eğitimi tamamlanan T5 modeli, girdi olarak metin alarak ve çıktı olarak metin üreterek metinden metne dönüşüm formatı ile görevleri yerine getirir.

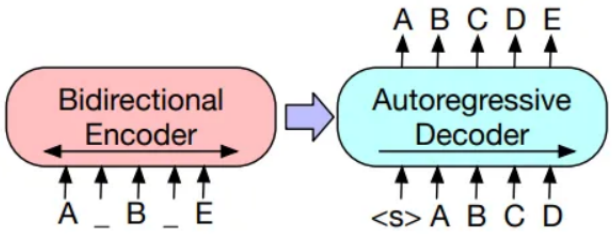


Şekil 3.2 T5 ön eğitim formatı [17]

Çalışılan veri setinin Türkçe olması sebebiyle, T5 model ailesindeki çok-dilli veri setinde ön eğitimi yapılan T5 versiyonu olan MT5 [32, 33] ve Türkçe veri setinde ön eğitimi yapılan TRT5 [34] modelleri ince ayar yapmak üzere kullanılmıştır. Modellere ince ayar yapılırken, ön ek olarak "Generate Keyphrase:" kalıbı verilmiştir, bu ön ek aynı zamanda modellere tahminleme yaptırılırken de kullanılmıştır.

Facebook ekibi tarafından geliştirilen BART [18] modeli, T5 modeline benzer olarak dönüştürücü (transformer) mimarisine dayanır ve kodlayıcı-kod çözücü (encoder-decoder) yapısını kullanır. BART modelindeki kodlayıcı, tıpkı BERT [35] modelindeki gibi metni çift yönlü olarak işlerken, kod çözücü GPT [36] modeline benzer olarak metni sol yönlü (otoregresif) olarak üretir (Şekil 3.3). Ön eğitim aşamasında, girdi metin belirli kısımları maskeleme, silme ve karıştırılma yoluyla bozulur. T5 modelindeki bozulan/maskelenen kısımları üretmek yerine, BART modelinden çıktı olarak orjinal metni üretmesi istenir. Yapısı bozulan girdi metni çift yönlü kodlayıcı tarafından kodlanır, ardından otoregresif kod çüzücü ile orjinal metin tahminlenmeye çalışılır.

BART modeline ince ayar yapılırken, metnin yapısını bozmadan hem kodlayıcıya hem de kod çözücüye girdi olarak verilir, kod çözücünün son gizli durumundaki (hidden state) temsiller kullanılır.



Şekil 3.3 BART ön eğitim formatı [18]

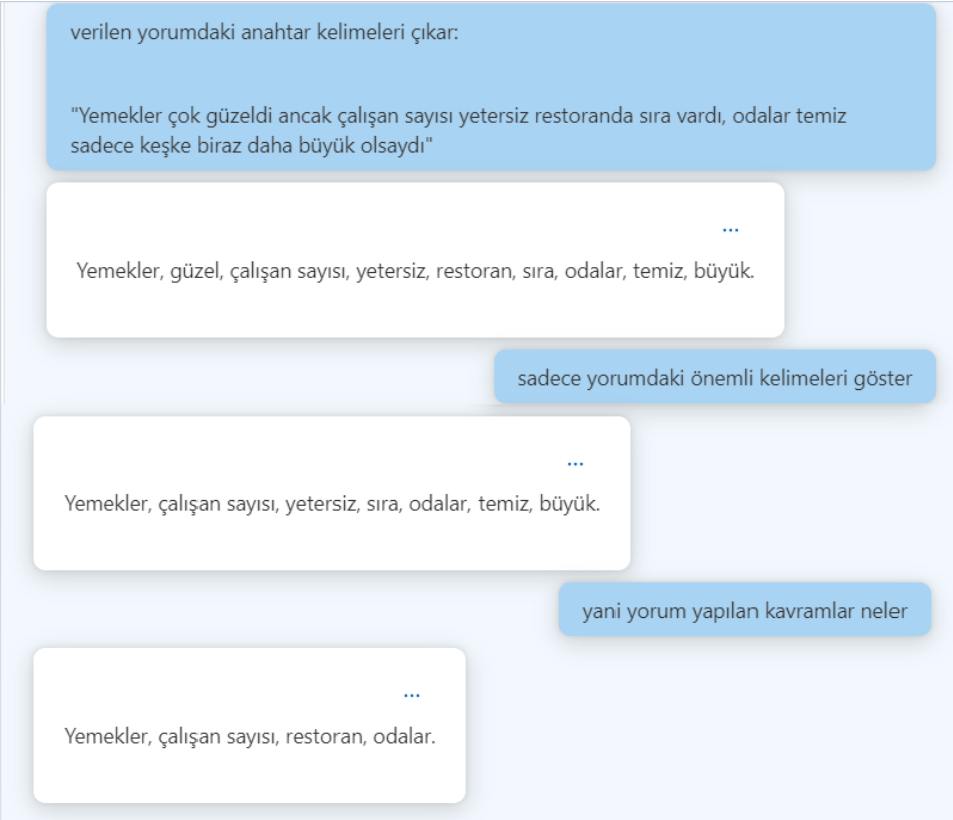
Çalışmada BART'ın çok dilli veri seti ile ön eğitimi yapılan versiyonu MBART [37, 38] ile Türkçe veri seti ile eğitilen TRBART [39, 40] modeli kullanılmıştır. İki farklı yaklaşım ile metin işleyen ve metin üreten bu modellerin, hem çok dilli hem de sadece Türkçe veri ile ön eğitimi yapılan versiyonları kullanılarak anahtar kelime üretme görevi için karşılaştırılmıştır.

3.2 Anahtar Kelime Üretimi için Büyük Dil Modellerine İstem Verme

Daha az veri kullanarak ya da hiç veri kullanmadan LLM'lere istem (prompt) vererek AK üretimi yapılabilir. İstem ile, LLM'e yapması gereken görev anlatılır. Farklı istem verme stilleri ile modelin ürettiği metinler iyileştirilebilir, en yaygın stiller zero-shot, one-shot ve few-shot yaklaşımlarıdır. Zero-shot stilinde, verilen görev ile ilgili hiçbir örnek vermez, sadece görev tanımı yapılır. One-shot stilinde tanımlanan görev için sadece bir adet örnek verilirken, few-shot'ta birkaç örnek istem metnine eklenir. Çalışmada istem verme stillerinin performansa etkisini gözlemlemek üzere zero-shot ve few-shot stilleri, seçilen GPT3.5, GPT4 [22] ve Gemini Pro [23] büyük dil modelleri ile birlikte kullanılmıştır. Yapılan ilk deneylerde, her ne kadar büyük dil modellerinin kapabilitesi yüksek olsa da, görev tanımına ek olarak görevle ilgili örneklere ve daha iyi açıklamalara ihtiyaç duyulduğu gözlemlenmiştir (Şekil 3.4).

OpenAI'ın Kasım 2022'de çıkardığı GPT3.5 modeli ile Mart 2023'te çıkardığı GPT4 modeli Eylül 2021 tarihine kadar olan veriler ile eğitilmiştir. Modellerin daha sonrasında çıkarılan versiyonlarının eğitiminde çıkış tarihine yakın veriler de bulunmaktadır [41].

GPT3.5 ve GPT4 modeline rakip olarak Google tarafından çıkarılan Gemini Pro



Şekil 3.4 GPT4 modeli ile LLM istem örneği

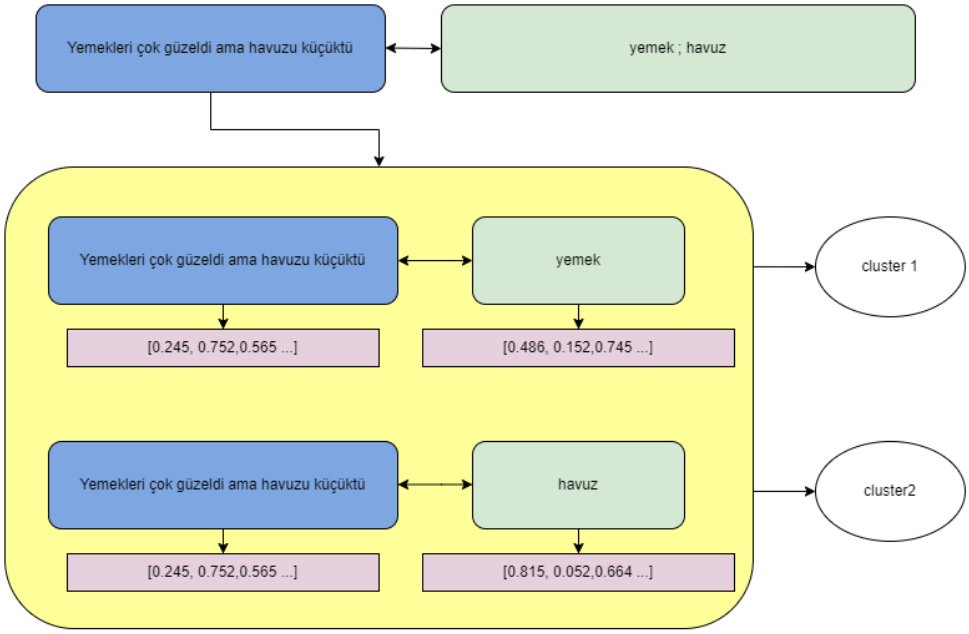
1.0 modeli ise Şubat 2024'te yayınlanmıştır. Çalışmada AK üretme görevi için LLM'leri istem vererek çalıştırmanın ince ayar yapılan modeller (T5, BART) ile kıyaslamasında bu üç model, iki farklı istem verme stili ve görev açıklaması verilerek kullanılmıştır. AK üretme görevi için verilen açıklama "generate keyword from given text" (verilen metinden anahtar kelimeler üret)'tir. İki stilde de görev açıklaması yapılmaktadır. Farklı olarak, few-shot stilinde görev ile ilgili 10 örnek verilmiştir (Tablo 3.1).

Tablo 3.1 İstem stilleri

Zero-shot Yaklaşım İstem1	Few-shot Yaklaşım İstem2
generate keywords from given text: output format: keyword1;keyword2	generate keyword from given text: output format: keyword1;keyword2 examples: feedback1: keyword1;keyword2... feedback2: keyword1;keyword2... ..... feedback10: keyword1;keyword2...

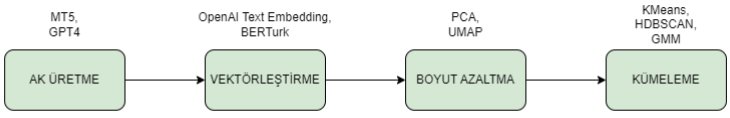
3.3 Anahtar Kelime Kümeleme ile Konu Modelleme

Konu modelleme görevi için önerilen bu yöntemde, ilk adımda her bir kullanıcı yorumu için AK'ler üretilir. AK üretimi için, ince ayar yapma ve istem verme yaklaşımlarındaki en iyi sonuç veren modeller seçilmiştir. Sonraki adımda, yorumlar ve üretilen AK'leri ayrı ayrı vektörleştirilir ve uç uça eklenir. Yorumlar ile her bir AK'sini birlikte alarak işlem yaparak, aynı yorum için farklı konuların bulunmasına olanak sağlanır (Şekil 3.5).



Şekil 3.5 Yorumlar ile AK'leri kümeleme

Yorum ve AK'ler vektörleştirildikten sonra, büyük vektörler ile işlem yapma zorluğundan dolayı boyut azaltma yapılır. Oluşturulan yeni vektörler kümelenerek konular oluşturulmaktadır.



Şekil 3.6 AK kümeleme ile konu modelleme akışı

Oluşturulan konuları temsil eden kelime parçacıklarını seçmek için, her bir AK ile yorumunun anlamsal benzerliği ölçülür. Anlamsal benzerliği ölçmek için yorum

ve AK'ler Türkçe metinde eğitilmiş BERTurk [42] modeli ile vektörleştirilir ve bu vektörlerin kosinüs benzerliği hesaplanır. Her bir konudaki benzerliği en yüksek 10 AK o konuyu temsil etmek üzere seçilir.

Şekil 3.6, yöntemdeki adımları özetlemektedir. AK üretme adımında performansı en yüksek olan ince ayar yapılmış MT5 modeli ile few-shot istem verme stilini kullanarak GPT4 modeline ürettirme yaklaşımları seçilmiştir. Deneylerde, kümelemede kullanılmak üzere metinleri vektörleştirmek için BERTurk ve OpenAI Text Embedding modelleri, boyut azaltma için PCA ve UMAP kullanılmıştır. Boyut azaltımı yapılan vektörler KMeans, HDBSCAN ve GMM algoritmaları ile kümelenerek performans ölçümü yapılmıştır.

Vektörleştirme aşamasında, çok dilli metin kümesinde eğitilen ve MTEB [43] metin vektörleştirme sıralamasında yüksek performans gösteren OpenAI'ın text-embedding-3-large [44] modeli ile Türkçe veri setinde eğitilen BERTurk modeli kıyaslanmıştır. OpenAI Text Embedding modeli metinler için 3072 boyutlu vektörler oluştururken, BERTurk modeli 768 boyutlu vektörler oluşturmaktadır.

Boyut indirgeme tekniklerinden PCA (Temel Bileşenler Analizi), sıkça tercih edilmektedir. Değişkenler arasındaki ilişkiyi belirlemek ve bu değişkenleri "temel bileşenler" (principal components) olarak adlandırılan daha az sayıda değişken ile temsil etmek için kullanılır. İlk olarak girdi değerleri için veri standardizasyonu yapılır ve verilerin belirli bir ölçekte olması sağlanır. Sonrasında kovaryans matrisi hesaplanarak değişkenler arasındaki ilişkiler tanımlanır. Matris, verinin temel bileşenlerini ve yönlerini belirleyen özvektörler ve her bileşenin açıkladığı varyans miktarını gösteren özdeğerler olarak iki bileşene ayrılır. Toplam varyansın önemli kısmını açıklayabilen belirlenen adette temel bileşenler seçilir. Orijinal verilerden, azaltılmış boyuta (bileşen sayısına) sahip dönüştürülmüş veri kümesini elde etmek üzere projeksiyon yapılır.

Verinin lokal ve global yapısını koruyarak boyut azaltımı yapmaı hedefleyen UMAP ise, graph tabanlı bir yaklaşımdır ve stokastik gradyen iniş (stochastic gradient descent) kullanır. PCA varyansı maksimize ederek doğrusal (linear) dönüşüm uygulayıp global yapıyı korumaya çalışırken, UMAP hem global hem de lokal yapıyı korumaya çalışır. UMAP, karmaşık ve doğrusal olmayan (non-linear) verilerde boyut azaltımı için tercih edilmektedir.

Kümeleme görevi için en çok kullanılan algoritmalardan biri olan K-Means, veriyi önceden belirlenen "k" sayıda kümeye ayırır. Her veri noktası, rastgele seçilen k tane küme merkezine yakınlıklarına göre gruplanır. Her bir küme merkezi, kümedeki noktaların ortalaması olacak şekilde güncellenir. Küme sayısının

önceden belirlenmesi ve uç değerlere hassas olması dezavantajdır.

Öte yandan, HDBSCAN için başlangıçta küme sayısını belirtmek gerekmez. Yoğunluk tabanlı olan bu kümeleme algoritması, veri yoğunluğunun yüksek olduğu alanları bir küme olarak belirlerken, yoğunluğu düşük olan alanları gürültü olarak ele almaktadır. Algoritma belirlenen minimum komşu nokta sayısına göre her veri noktası için en yakın komşu noktalara mesafeyi hesaplar, yoğunluk tabanlı mesafe matrisini oluşturur. Bu matris daha sonra hiyerarşik kümeleme ağacının oluşturulmasında kullanılır.

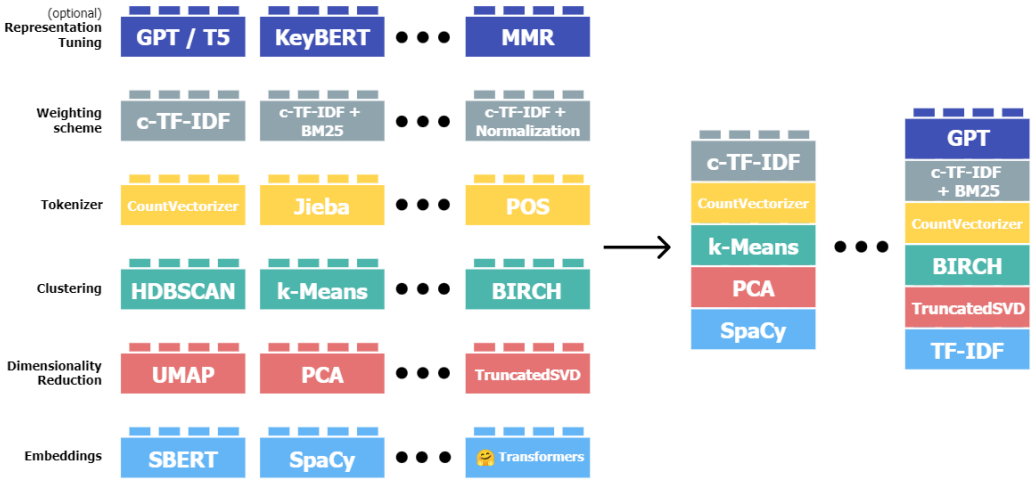
Verilerin belirli bir sayıda Gauss dağılımının karışımı olarak modelleyen GMM algoritmasında, her Gauss dağılımı bir küme olarak kabul edilir. Bu kümeleme algoritmasının avantajı, bir veri noktasının birden fazla kümeye ait olabileceği yumuşak kümeleme yapmasıdır. Böylece bir örneğin birden fazla kümeye ait olma olasılığı bulunmaktadır. K-Means'e benzer şekilde küme sayısını önceden belirlemek gerekmektedir.

Konu modelleme aşamasındaki farklı opsiyonları değerlendirebilmek için opsiyonların kombinasyonları deneylerde kullanılmış ve oluşturulan konu modellerinin performansı ölçülmüştür. Seçilen yöntemlerin karşılaştırılması Bölüm 4.2.1'de verilmiştir.

3.4 Sinirsel Konu Modelleme Yöntemleri (BERTopic)

Metinleri kümeleyerek konu modelleri oluşturmayı hedefleyen BERTopic [45], öncelikle metinleri sinirsel dil modellerini kullanarak vektörleştirir. Vektörler üzerinde boyut indirgeme yapılır ve kümeleme işlemi gerçekleştirilir. Çıkan kümeler konu olarak ele alınır ve her konuyu temsil eden kelimeler TF-IDF tabanlı c-TF-IDF yaklaşımı ile terim frekansları kullanılarak seçilir. Bu yaklaşımdaki c terimi, belirli bir kümeye atanan tüm dokümanların birleştirilerek tek bir doküman olarak ele alındığını temsil eder, diğer bir deyişle küme-1 olarak belirlenen tüm metinler birleştirilir ve tek bir metin olarak TF-IDF işlemi yapılır. Sınıf tabanlı bu yaklaşım sayesinde kümedeki önemli kelimeler bulunur.

Şekil 3.7'de verildiği üzere BERTopic kütüphanesi konu modelleme yaparken farklı modüller kullanmaya imkan sağlamaktadır. Sentence-bert [47] kullanarak BERT modelleri ile metin vektörleştirme yapılabileceği gibi, bu işlem için farklı model ve kaynaklar da kullanılabilmektedir. Boyut indirgeme adımında varsayılan olarak UMAP algoritması bulunsa da, isteğe bağlı olarak PCA'yi de entegre edilebilmekte, kümeleme aşamasında farklı algoritmalar çalıştırılabilmektedir.



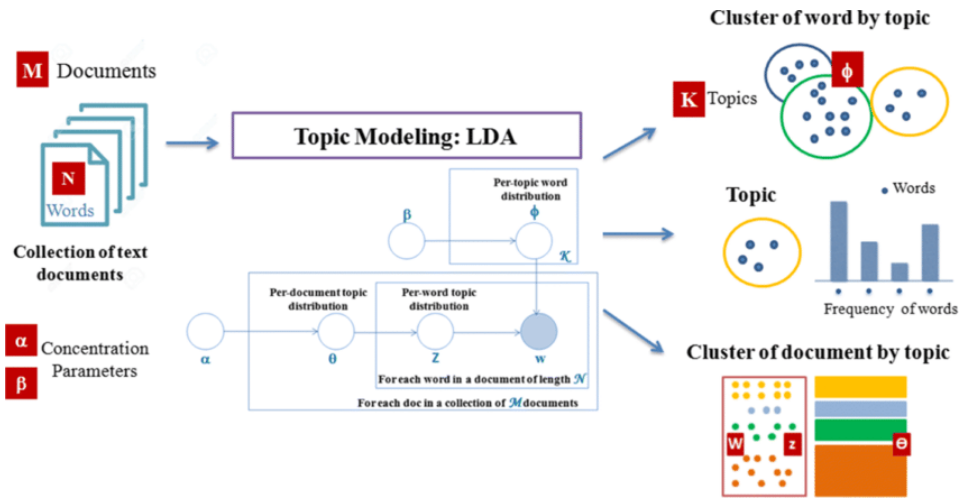
Şekil 3.7 BERTopic konu modelleme akışı [46]

3.5 İstatistiksel Konu Modelleme Yöntemleri (LDA, NMF)

Gizli Dirichlet Ayrımı (LDA), metinler ve konular arasındaki bu ilişkileri belirlemek için bir çeşit olasılık modellemesi kullanır. Her metin belirli olasılıklarla çeşitli konulardan oluşurken, her konu da belirli olasılıklarla çeşitli kelimelerde oluşmaktadır ve Dirichlet dağılımı ile modellenirler.

- M metin sayısı
- N belirli dokümandaki kelime sayısı
- W belirli bir kelime
- α belge başına konu dağılımını kontrol eden hiperparametre
- β konu başına kelime dağılımını kontrol eden hiperparametre olarak gösterilmektedir.
- Her metin için metinin konu dağılımı θ d Dirichlet(α) dağılımından örneklenir.
- Her konu için konunun kelime dağılımı ϕ k Dirichlet(β) dağılımından örneklenir.
- Her metindeki için her kelime pozisyonu için: Konu z, metinin konu dağılımından θ d örneklenir. Kelime w, seçilen konu z'ye karşılık gelen ϕ z dağılımından örneklenir (Şekil 3.8).

Negatif olmayan matris faktörizasyonu (NMF): Kelime frekanslarını tutan metin-kelime matrisini, metin-konu matrisi ve konu-kelime matrisi olarak iki ayrı



Şekil 3.8 LDA çalışma akışı [48]

matrise böler. NMF, dokümanların ve kelimelerin konularla nasıl ilişkili olduğunu belirlemek için bu matrisleri kullanır. Algoritmanın amacı, W ve H için negatif olmayan başlangıç koşullarıyla başlayarak, negatif olmayan V için güncelleme kurallarının iterasyonlarıyla, amaç (objective) fonksiyonunun yerel maksimumuna yakınsayan yaklaşık bir faktorizasyon bulmaktır [49].

3.6 Değerlendirme Metrikleri

AK üretme ve konu modelleme çalışmalarında kullanılan farklı tekniklerin performansını ölçmek için benzer çalışmalarda kullanılan metrikler incelenmiştir.

AK üretme aşamasında incelenen metrikler, üretilen AK'lerin etiketlenen AK'lere ne kadar benzediğini, çıkarıldığı kullanıcı yorumuna ne kadar benzediğini ölçmeyi amaçlar. Böylece üretilen AK'lerin yorumları ne kadar iyi temsil edebildiği karşılaştırılabilir.

Konu modelleme aşamasında, önerilen AK'leri kümeleyerek konu modelleri üretme yönteminin başarısını ölçebilmek için benzer yaklaşım kullanan BERTopic ve geleneksel yöntemlerden LDA ve NMF karşılaştırılmıştır. Yöntemlerin ürettiği konu kalitelerini kıyaslamak için konu tutarlılığı (Coherence) skorları göz önünde bulundurulmuş ve elle kontroller yapılmıştır.

3.6.1 Anahtar Kelime Üretme Performansını Ölçme

İnce ayar yapılmış dil modelleri ile farklı istem verme yaklaşımları ile kullanılan büyük dil modellerinin ürettiği AK'leri kıyaslamak için şu metrikler kullanılmıştır:

- Precision@k, Recall@k, F1@k: Etiketlenen AK'ler ile üretilen AK'lerin eşleşip eşleşmediğini ölçmek için kullanılan metriklerdir. Kesinlik (precision) metriği seçilen k adet üretilen AK için, etiketli AK'ler ile ne kadar eşleştiğini hesaplar.

$$Precision@k = \text{Üretilen ilk k AK içindeki doğru AK sayısı}/k \quad (3.1)$$

Duyarlılık (Recall) değeri ise üretilen k adet AK'nin etiketli AK'ler içinde kaç tanesini kapsadığını anlatır.

$$Recall@k = \text{Üretilen ilk k AK içindeki doğru AK sayısı}/\text{Tüm etiketli AK sayısı} \quad (3.2)$$

F1@k skoru ise, bu iki metriğin harmonik ortalamasını hesaplamaktadır. Metriklerde, Tablo 2.2'de verilen veri istatistiklerinde AK'lerdeki kelime sayısının yaklaşık 3 olması sebebiyle kullanılan "k" değeri 3 seçilmiştir.

- Rouge-1: Üretilen AK'ler ile etiketli AK'ler arasında yazılış benzerliklerini eşleşen tek karakter (unigram) sayısı ile ölçmektedir. Üretken yaklaşımlar için bu skorun düşük olduğu görülmektedir, çıkarıcı (extractive) yaklaşımlarda daha çok tercih edilir.

- Bert-Similarity (Anlamsal Benzerlik): Kullanıcı yorumları ve üretilen AK setleri arasındaki anlamsal benzerliği ölçmek için metinler Sentence-Bert ve BERTurk modeli kullanılarak vektörleştirilir. Vektörler arasındaki kosinüs benzerliği [20] çalışmasındaki gibi hesaplanır, 0-1 arasında çıkan bu değerler 1'e yaklaştıkça metinlerin anlamsal olarak daha çok benzediği kabul edilir. Tüm veri setinde metinler ve AK'leri arasındaki benzerliği ölçmek için Eşitlik 3.3 kullanılır.

$$AvgSim = \frac{Sim_{metin1\&AKset1} + Sim_{metin2\&AKset2} + ... + Sim_{metini\&AKseti}}{i} \quad (3.3)$$

- Diversity (Çeşitlilik) : Aynı yorum için üretilen AK'lerin birbirini tekrar etmemesi gerekmektedir. İlk deneylerde, özellikle prompt vererek LLM'lere AK ürettirildiğinde bu modellerin eş anlamlı kelimeleri birlikte verdiği gözlemlenmiştir. Yorumları kaliteli bir şekilde temsil edebilmek için üretilen AK'lerin anlamsal olarak çeşitli olması, az miktarda kelime ile asıl yorumdaki anlamı koruyabilmesi gerekmektedir. Bu amaçla LLMTake [25] eş anlamlılık kontrolü için SpaCy ile metinleri vektörleştirmektedir, bu çalışmada ise Bert-Similarity metriğindeki model vektörizasyon için kullanılmış ancak bu sefer kosinüs uzaklıkları kullanılmıştır. Mesafe 0-1

arasında değer almaktadır ve 1'e yakın değer AK'lerin anlamsal çeşitliliğe sahip olduğunu göstermektedir. Uzaklık hesaplanırken 1'e 1 (one-to-one) yaklaşımı kullanılmış, her AK için kombinasyonlara bakılmıştır.

Dist, kosinüs uzaklığını, C ise AK'lerin kombinasyonunu temsil etmektedir; yorumdan üretilen i adet AK için AK'lerin uzaklıkları aşağıdaki Eşitlik 3.4 ile ölçülür.

$$AvgDist = \frac{Dist_{AK1\&AK2} + Dist_{AK1\&AK3} + ... + Dist_{AKi\&AKj}}{C(i,2)} \quad (3.4)$$

Tüm veri setinde anlamsal çeşitliliği ölçmek için n adet metin olduğunu varsayarak Eşitlik 3.5 kullanılır:

$$TotalAvgDist = \frac{AvgDist_{metin1} + AvgDist_{metin2} + ... + AvgDist_{metinn}}{n} \quad (3.5)$$

- Compressing (Sıkıştırma): Oluşturulan AK'lerin uzunluğunun yorum uzunluğuna oranı hesaplanır. Metnin uzunluğu olarak karakter sayısı kullanılmıştır. Elde edilen değerin 1'e tümleyeni sıkıştırma puanını verir; AK uzunluğu yoruma kıyasla ne kadar kısaysa puan o kadar yüksektir. Skorun 0'dan düşük olması, AK uzunluğunun yorumdan daha uzun olduğunu gösterir. Eşitlik 3.6 ile hesaplanır:

$$CompressingScore = 1 - \frac{Length_{AKset}}{Length_{metin}} \quad (3.6)$$

Tüm veri seti için yapılmak istendiğinde n adet veri için Eşitlik 3.7 kullanılır.

$$TotalAvgComp = \frac{CompressingScore_{metin1} + ... + CompressionScore_{metinn}}{n} \quad (3.7)$$

Özellikle LLM'ler tarafından üretilen metinler uzun olabileceğinden, modellerin karşılaştırılmasında Compressing skoru bir metrik olarak dahil edilmiştir. Aynı zamanda, deneylerde ince ayar yapılmış T5 ve BART modellerinin AK üretirken bazen kelimeleri tekrarladıkları bir hata gözlemlenmiştir.

3.6.2 Konu Modelleme Performansını Ölçme

Önerilen AK kümeleme konu modelleme yönteminin, diğer konu modellerine göre performans olarak iyileştiğini ölçebilmek için LDA, NMF ve BERTopic ile konu tutarlılığı skorları ölçülmüştür. Böylece önerilen yöntemin ürettiği konuların kalite olarak nasıl bir fark yarattığı karşılaştırılmıştır. AK'leri kümelerken kullanılan

K-Means ve GMM algoritmaları parametre olarak küme sayısı istemektedir. Küme sayısını belirlemek için 2-20 adet küme için bu modellerin Silhouette skorları hesaplanmıştır. Silhouette skoru, tüm veri noktalarının Silhouette değerlerinin ortalaması alınarak hesaplanır. Veri noktasının kendi kümesine iyi uyduğunu ve komşu kümelerle kötü eşleştiği durumda skor yüksek çıkmaktadır. Bu skor kümeleme sonucunun genel kalitesini gösterir, skor 1'e yaklaştıkça küme yapısının iyi, düşük değerler için ise zayıf olduğu kabul edilir.

Yüksek Silhouette skoruna sahip olan küme sayısı kümeleme algoritmalarına parametre olarak verilir. Sonrasında çıkan kümeleme sonuçları için kümeleri temsil eden AK'ler yorumlarına benzerliklerine göre seçilir. Her küme için Bert-Similarity'si en yüksek 10 AK temsil için seçilir ve Coherence skorları hesaplanır.

Konu modellemesi algoritmaları konuları oluşturmakta, bu konular ise bir dizi kelimeden oluşmaktadır. Konu tutarlılığı (Coherence) skoru kelimelerin bir konu etrafında ne kadar iyi gruplandığını ölçmek için kullanılır. Her bir kelimenin diğer kelimelerle olan benzerliği ölçülür. Bu ölçüm, genellikle kelimelerin birlikte ortaya çıkma sıklığı (co-occurance) veya benzer bir metinde birlikte ortaya çıkma olasılığı üzerinden yapılır. Coherence skoru ne kadar yüksek olursa, konunun kelimeleri birlikte iyi çalışır ve konu modeli o kadar iyi olur. Çıkarılan konular karşılaştırılırken Coherence skoruna ek olarak elle kontroller de yapılmıştır. Bu kontrollerde konuları oluşturan kelimelerin tutarlılığı, bir konuya atanan metinlerin anlamsal benzerliği göz önünde bulundurulmuştur.

4
DENEYSEL SONUÇLAR

Çalışmanın ilk aşamasında, AK üretimi için T5 versiyonlarından MT5 ve TRT5 modelleri ve BART versiyonlarından MBART ve TRBART modelleri ince ayar yapılmak üzere seçilmiştir. Sonrasında bu modellerin performansını LLM'ler ile karşılaştırmak için GPT3.5, GPT4 ve Gemini Pro modelleri seçilmiş, bu modellerden hem zero-shot yaklışımı kullanan İstem1 hem de few-shot yaklaşımı kullanan İstem2 ile AK üretmesi istenmiştir (Bölüm 3.2). AK üretme başarısının veri seti karakteristiğinden ne kadar etkilendiğini görebilmek için en başarılı modeller ile çapraz geçerlilik (cross validation) yapılmış ve sonuçlar raporlanmıştır.

İkinci aşamada konu modelleme veri setindeki kullanıcı yorumları ile AK kümeleme, BERTopic, LDA ve NMF teknikleri kullanılarak konular oluşturulmuş ve başarıları ölçülmüştür.

Bölüm 4.1'de AK üretme deney sonuçları verilmiştir, deney sonuçları veri setindeki farklı bölümler için (test, doğrulama, çapraz geçerlilik) başarılarına göre farklı başlıklar altında toplanmıştır. Konu modelleme veri seti ile yapılan farklı yaklaşımların kullanıldığı konu modelleme deney sonuçları ise Bölüm 4.2 alt başlıklarında verilmiştir.

4.1 Anahtar Kelime Üretimi Deneyleri

AK üretimi görevi için ince ayar yapılan MT5, TRT5, MBART ve TRBART modelleri eğitilirken en iyi konfigürasyonu bulabilmek amacıyla öğrenme oranları (learning rate) ile gradyan birikim adımları (gradient accumulation steps) parametreleri için hiperparametre optimizasyonu yapılmıştır. LR için 1e-3 (0,003), 1e-5 (0,00001) ve 4e-5 (0,00004) değerleri, GAS değerleri için ise 1, 2 ve 4 değerleri verilerek bu değerlerin kombinasyonu ile modeller kurulmuştur, diğer parametreler ise sabit tutulmuştur (epoch, batch size, optimizer vb). Toplamda 4 dil modeli için, 9 konfigürasyon denenmiş ve kurulan 36 model arasında her dil modelinin en iyi performans gösteren modeli seçilmiştir. En iyi modeli seçerken

doğrulama kaybı değerinin en düşük olduğu model göz önünde bulundurulmuştur.

Modellerin eğitimde bir adet A100 GPU kullanılmıştır, modeller 10 epoch boyunca eğitilmiş ve optimizer olarak "Adam" kullanılmış, batch size değeri ise 8 seçilmiştir. Batch size değeri kaç adet örneğin aynı anda işlendiğini göstermektedir. Bu değerin yüksek olması daha kararlı gradyan tahminlemeleri sağlayabilmektedir, ancak zaman ve bellek bakımından maliyetlidir. Hesaplama kısıtlmalarından dolayı deneylerde 8 olarak tutulmuştur ancak GAS ile bu değeri yükseltmek mümkündür. Daha büyük bir batch size oluşturmak için birkaç küçük batch üzerinde gradyanları biriktirir, model ağırlıklarını hemen güncellemez. GAS değeri 4 ve batch size değerinin 8 olduğu durumda, 4 adımda gradyanlar biriktirilir ve 4 adım sonunda güncellenir. Böylece 8*4'ten 32 örnek işlenmiş olunur. Kullanılan GPU'nun VRAM kısıtından dolayı GAS kullanılarak batch size en fazla 32'ye çıkarılabilmiştir (8*4 GAS).

Modellerin ince ayar işleminde Bölüm 2.1.1'de detaylı anlatılan anahtar kelime üretme veri setinin eğitim kısmı, doğrulama kayıplarının ölçümünde ise doğrulama kısmı kullanılmıştır.

Diğer AK üretme yaklaşımında ise, görev ile ilgili örnek verilmeyen zero-shot stilindeki İstem1 ve 10 adet örnek verilen few-shot stilindeki İstem2 ile GPT3.5, GPT4 ve Gemini (Pro) modelleri kullanılmıştır. İstem2'de verilen örnekler, AK üretme verisinin eğitim kısmından rastgele seçilmiştir. Bu yaklaşımda kullanılan LLM'ler varsayılan parametre konfigürasyonları ile kullanılmış, sadece GPT modelleri kullanılırken üreteceği maksimum token sayısı 200 olarak sabitlenmiştir.

İki yaklaşımda elde edilen sonuçların kıyaslanması için AK üretme veri setinin doğrulama ve test kısımları kullanılmıştır.

4.1.1 Anahtar Kelime Üretimi Doğrulama Veri Seti Sonuçları

Dil modellerine ince ayar yapma yaklaşımında ilk aşamada eğitim verisi kullanılarak MT5, TRT5, MBART ve TRBART modellerine ince ayar yapılmıştır. Çeşitli konfigürasyonlar ile eğitilen modellerde, her dil modeli için seçilen konfigürasyonların doğrulama kaybı değeri Tablo 4.1'te verilmiştir. Kayıp değerlerine baktığımızda, ince ayar yapılan modellerden en iyi MT5 modeli LR'si 4e-5 ve GAS'ı 1 olan versiyonken diğer dil modellerindeki durum sırasıyla TRT5 için LR:4e-5 GAS:2, MBART için LR: 1e-5 GAS:2 ve TRBART için LR:4e-5, GAS 2'dir. MBART modeli verilen LR'ler arasında en küçük adımları kullanarak başarıya ulaşmakta, MT5 ve TRBART modeli daha küçük batch size değerleri ile (8*1) daha iyi sonuçlar vermektedir.

Tablo 4.1 Model konfigürasyonları doğrulama sonuçları

parametreler / modeller	MT5	TRT5	MBART	TRBART
{'LR': 1e-03, 'GAS: 1}	2,5528	2,8770	4,5650	0,2749
{'LR': 1e-03, 'GAS: 2}	3,8393	2,6231	3,9762	0,2738
{'LR': 1e-03, 'GAS: 4}	2,1161	2,5153	3,8476	0,2727
{'LR': 4e-05, 'GAS: 1}	1,4875	1,5679	2,4117	0,0644
{'LR': 4e-05, 'GAS: 2}	1,6278	1,5481	2,0377	0,0640
{'LR': 4e-05, 'GAS: 4}	2,3338	1,9715	1,7903	0,1026
{'LR': 1e-05, 'GAS: 1}	2,5275	2,3051	1,6124	0,0840
{'LR': 1e-05, 'GAS: 2}	5,9638	2,8431	1,6030	2,1045
{'LR': 1e-05, 'GAS: 4}	7,9884	3,8044	1,9618	5,2128

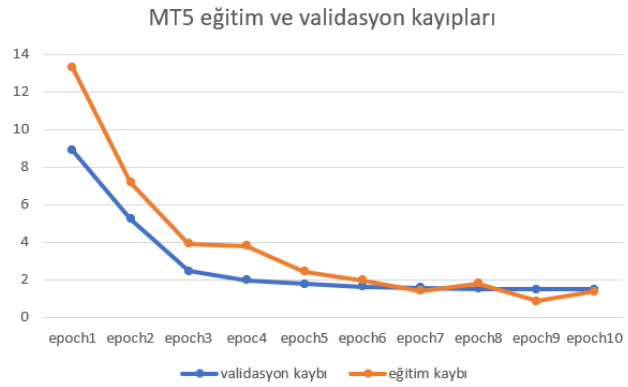
Her dil modeli için koyu sarı renk ile verilen en iyi modeller seçilmiş ve bu modellerin farklı zamanlarda 2'şer kez tekrar eğitilerek performanslarının kararlılığı ölçülmüştür (Tablo 4.2). Toplamda 3 kez ayrı ayrı eğitilen dil modellerinde, standart sapmanın en düşük olduğu modeller TRBART ile MT5'tir. Ek olarak, bu modellerin diğer ince ayar yapılmış modellere göre daha kararlı olduğu söylenebilir.

Tablo 4.2 İnce ayar yapılan en iyi modellerin performans kararlılığı

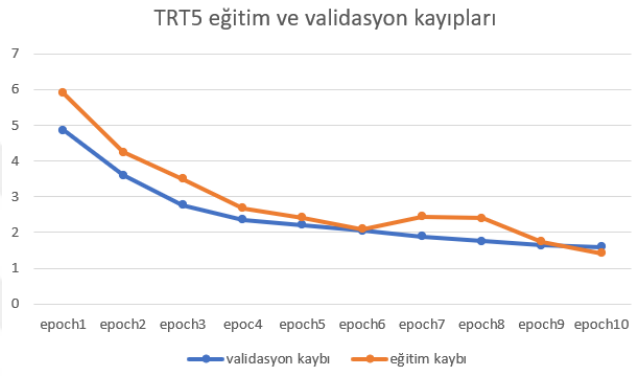
	MT5	TRT5	MBART	TRBART
eğitim-1	1,4875	1,5481	1,603	0,064
eğitim-2	1,4755	1,59	1,6179	0,0636
eğitim-3	1,5158	1,6217	1,4647	0,0635
standart sapma	0,0168	0,0301	0,0689	0,0002
ortalama	1,4929	1,5866	1,5618	0,0636

En iyi konfigürasyonlar ile ince ayar yapılmış modellerin eğitim sürecine dikkatle baktığımızda, Şekil 4.1d'te TRBART modelinin 2.ve 3.epochda doğrulama kaybının hızlı bir şekilde düştüğünü, son epochlarda dengelendiğini ve eğitim kaybı ile paralellik gösterdiği görülmektedir. Ancak eğitim sonlandığındaki doğrulama kaybının çok düşük olması şüphe uyandırıcıdır. Modelin aşırı öğrenmiş (overfit) olabileceğini gösterir, çapraz geçerlilik yöntemi ile veri farklı yerlerden bölünerek modelin genelleme yeteneği daha detaylı incelenebilir.

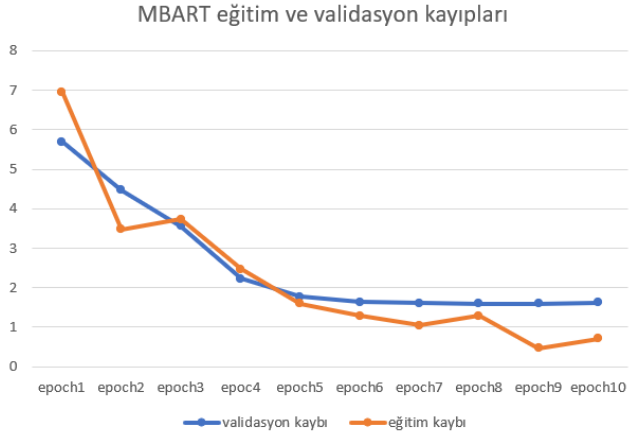
Şekil 4.1c MBART modelinin epochlar boyuncaki eğitim ve doğrulama kayıplarını, Şekil 4.1b TRT5 modeli kayıplarını ve son olarak Şekil 4.1a ise MT5 modeli kayıplarını göstermektedir. Modellerin doğrulama kayıplarının hızla düşerek bir süre sonra dengelenmesi modellerin erken aşamalarda hızlı bir öğrenim sağladığını gösterir. Eğitim kayıplarının düşüş gösterirken zaman zaman geçici olarak artması LR değerinin büyük olduğunu ve bu sebeple minimumu bulmaya çalışırken sapmaların olduğu şeklinde açıklanabilir.



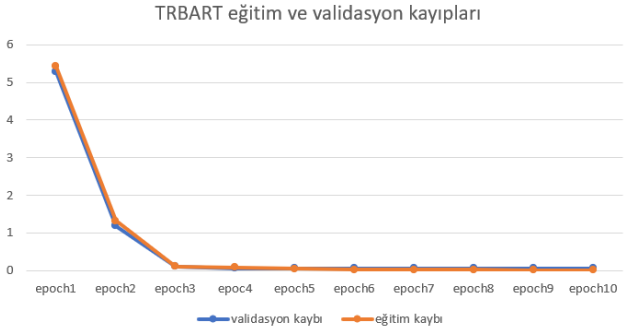
(a) MT5 kayıpları



(b) TRT5 kayıpları



(c) MBART kayıpları



(d) TRBART kayıpları

Şekil 4.1 İnce ayar yapılan modellerin kayıp grafikleri

İnce ayar yapılmış modellerin başarısını LLM'ler ile kıyaslamak için Bölüm 3.6.1'de anlatılan metrikler kullanılmıştır. GPT3.5, GPT4, Gemini dil modellerine İstem1 (zero-shot) ve İstem2 (few-shot) verilerek üretilen AK'ler ile ince ayar yapılmış TRT5, MT5, TRBART ve MBART modellerinin ürettiği AK'lerin kalitesi Tablo 4.3'te verilmiştir. Bu modellere doğrulama verisindeki kullanıcı yorumları verilerek bu yorumlar için AK'ler ürettirilmiş, AK'lerin yoruma ve etiketli AK'lere benzerliği ölçülmüştür.

Tablo 4.3 Modellerin doğrulama seti üzerindeki performansı

model	precision@3	recall@3	f1@3	rouge-1	bert-similarity	compressing	diversity
GPT3.5 & Prompt1	0,1561	0,2454	0,1782	0,0625	0,7897	-1,0785	0,2389
GPT4 & Prompt1	0,1474	0,2244	0,1658	0,0897	0,8221	-0,2379	0,2515
Gemini & Prompt1	0,1614	0,2423	0,1802	0,1292	0,7963	0,4764	0,2338
GPT3.5 & Prompt2	0,1982	0,319	0,2278	0,1597	0,7834	-0,1229	0,2534
GPT4 & Prompt2	0,2175	0,3667	0,2553	0,2695	0,8021	0,5844	0,2559
Gemini & Prompt2	0,2088	0,3457	0,243	0,2743	0,775	0,6777	0,2506
TRT5	0,2158	0,3609	0,2510	0,3110	0,7350	0,7421	0,2084
MT5	0,2316	0,3555	0,2587	0,3131	0,7481	0,7116	0,2380
TRBART	0,1895	0,3227	0,2217	0,2968	0,7484	0,7077	0,2162
MBART	0,2017	0,3259	0,2315	0,2609	0,7262	0,7565	0,2368

İnce ayar yapılmış modellerin sonuçları incelendiğinde, TRT5 ve MT5 modellerinin f1@3 ve rouge metriklerindeki başarısının yüksek olması bu modellerin ürettiği AK'lerin etiketli AK'ler ile daha çok eşleştiğini, yazılış olarak benzediğini göstermektedir. Doğrulama setindeki kullanıcı yorumlarına anlamsal olarak en benzeyen AK'leri üreten ince ayar yapılan modeller MT5 ve TRBART iken, bu iki model metni sıkıştırma konusunda TRT5 ve MBART'ın gerisinde kalmıştır. MT5 ve MBART modellerinin ürettiği AK'ler anlamsal olarak daha çeşitlidir.

Diğer yaklaşımda LLM'lere verilen İstem1 ve İstem2 ile üretilen sonuçlara bakıldığında görevle ilgili örnekler vermenin (İstem2) dil modellerinin daha doğru AK'ler üretmesini sağladığı, F1@3 ve Rouge skorlarının arttığı, metni sıkıştırma oranı (compressing) ile AK çeşitliliğinin (Diversity) iyileştiği gözükmektedir. Diğer yandan bu istem stilini kullanmanın modellerin üretkenliğini azalttığı, üretilen AK'lerin yoruma anlamsal benzerliğinin düştüğü gözlemlenmiştir.

Yaklaşımları karşılaştırdığımızda, LLM'ler ile alınan en iyi sonuçlara GPT4 ve Gemini modellerini İstem2 ile çalıştırırken, dil modellerine ince ayar yapmanın en iyi sonuçlarına ise TRT5 ve MT5 modeli ile ulaşılmıştır. İnce ayar yapılan modellerde eğitim için kullanılan veri miktarı, LLM'lerde isteme örnek olarak verilen veri miktarından katbekat fazladır. Bu durum, ince ayarlı modellerin LLM'lere kıyasla daha istenen şekilde AK üretmesini sağlamaktadır. İnce ayarlı modellerin ürettiği AK'ler etiketli AK'ler ile daha çok eşleşmekte ve metni daha iyi sıkıştırmaktadır. Ancak, LLM'ler ürettiği AK'ler ile metinleri anlamsal olarak daha iyi temsil etmektedir.

4.1.2 Anahtar Kelime Üretimi Test Veri Seti Sonuçları

AK üretme eğitim verisi ile ince ayar yapılan modeller ile LLM'lerin doğrulama veri seti üzerindeki performansları ölçülmüştür. Test veri seti karateristiği doğrulama setine göre farklılıklar göstermektedir. AK'ler doğrulama setinden farklı olarak çoğunlukla iki kelimeden oluşmaktadır, ayrıca veri setindeki yorum ve AK'ler daha kısadır (Tablo 2.2.)

Tablo 4.4 Modellerin test seti üzerindeki performansı

model	precision@3	recall@3	f1@3	rouge-1	bert-similarity	compressing	diversity
GPT3.5 & Prompt1	0,1533	0,2252	0,1615	0,0469	0,8096	0,1707	0,2528
GPT4 & Prompt1	0,16	0,2472	0,1755	0,1006	0,8424	0,3447	0,2781
Gemini & Prompt1	0,1667	0,2687	0,1869	0,0697	0,7942	0,5041	0,2585
GPT3.5 & Prompt2	0,2	0,3162	0,2222	0,0768	0,8014	-0,1190	0,2640
GPT4 & Prompt2	0,2533	0,4102	0,2872	0,1468	0,8141	0,5540	0,2634
Gemini & Prompt2	0,2	0,3207	0,2289	0,1166	0,7638	0,6868	0,2790
TRT5	0,1867	0,2878	0,2081	0,2435	0,7319	0,7686	0,2041
MT5	0,1533	0,2111	0,1625	0,2142	0,7473	0,7511	0,2609
TRBART	0,1467	0,2228	0,1622	0,1914	0,7333	0,7795	0,2249
MBART	0,1867	0,2545	0,1944	0,2329	0,7379	0,7727	0,2305

Tablo 4.4'te LLM'ler ile ince ayar yapılmış modeller karşılaştırılmıştır. Doğrulama sonuçlarına benzer olarak LLM'lerin daha fazla metin üretme eğiliminde olduğu bu sebeple Sıkıştırma (Compressing) skorlarının düşük olduğu gözlemlenmiştir. Yorumları en iyi sıkıştıran ve kısaltan modellerin ince ayar yapılan modeller olduğu söylenebilir. Üretilen AK'lerin anlamsal çeşitliliği incelendiğinde, LLM'lerin bu konuda daha başarılı olmuş ve daha yüksek diversity skorlarına ulaştığı görülmektedir.

Precision@3, Recall@3 ve F1@3 skorlarına baktığımızda, LLM'lerin İstem2 ile diğer modellere göre daha başarılı olduğu, ardından da TRT5 ve MBART modellerinin geldiği görülmektedir. LLM'ler İstem1 ile MT5 ve TRBART modellerinin performansına yakın sonuçlar vermiştir. Özellikle GPT4&İstem2 ile en doğru anahtar kelimelerin üretildiğini ve bu metrikler açısından diğer modellere kıyasla en başarılı model olduğu görülmektedir. Ayrıca, doğrulama sonuçlarındaki gibi LLM'lere istem verirken görevle ilgili örnekler vermek başarısını artırmış; böylece modeller daha tutarlı AK'ler üretilmiştir. Rouge-1 başarısı gözetildiğinde ise, gerçek AK'lere yazımsal olarak en benzer AK'leri üreten modellerin TRT5 ve MBART olduğu görülmektedir. İnce ayar yapılan modellerin bert-similarity açısından performanslarının birbirine yakın olduğu görülmektedir; bu metrikte en başarılı ince ayar yapılan model olan MT5 (0,7433) bile LLM'ler arasında en başarısız model olan Gemini & İstem2 (0,7638) modelinin gerisindedir.

4.1.3 Anahtar Kelime Üretme Çapraz Geçerlilik Sonuçları

İnce ayar yapılan dil modellerinin başarısında, eğitildiği veri setinin etkisini incelemek üzere AK üretme veri setinin Bölüm 2.1.1'de anlatıldığı gibi önce %5'i test için ayrılmıştır. Kalan kısım için 5-katlı çapraz doğrulama yapmak üzere %80'i eğitim ve %20'si doğrulama için 5 kez bölümleme yapılmış ve dil modelleri en iyi konfigürasyonları ile eğitilmiştir.

Modellerin ince ayar başarısı Tablo 4.5'de incelendiğinde en düşük doğrulama kaybını TRBART ve MT5 modelinin verdiği görülmüştür. Bölüm 4.1.1'de alınan sonuçlara benzer şekilde, TRBART doğrulama kaybının çok düşük olması modelin aşırı öğrenme yapmış olabileceği ve genelleme konusunda yetersiz kaldığı söylenebilir. Bu model ayrı tutulduğunda, doğrulama kaybına bakarak en başarılı ve kararlı performansı MT5 modelinin gösterdiği, bunu MBART ve TRT5'in takip ettiği sonucuna varılmıştır.

Tablo 4.5 İnce ayarlı modellerin k-katlı çapraz geçerlilik doğrulama kayıpları

k-kat / model	**MT5**	**TRT5**	**MBART**	**TRBART**
kat-1	1,4443	1,5502	1,6527	0,0647
kat-2	1,5705	1,664	1,6766	0,0617
kat-3	1,5084	1,6548	1,4554	0,0585
kat-4	1,3962	1,6119	1,573	0,0546
kat-5	1,6223	1,8658	1,5187	0,0678
kayıpların ortalaması	1,5083	1,6693	1,5752	0,0614
kayıpların standart sapması	0,0818	0,1061	0,0822	0,0046

4.2 Konu Modelleme Deneyleri

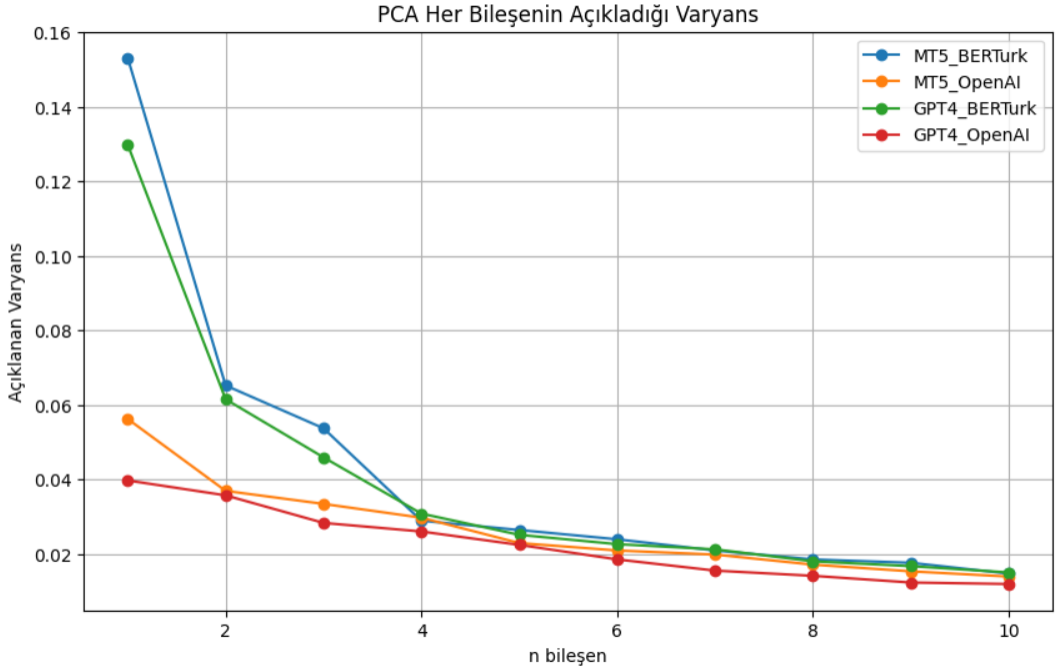
Müşterilerin sorulara verdiği yorumlar hakkında bilgi edinebilmek için konu modelleme yöntemleri ile bahsedilen konular ve bu konuları oluşturan kelimelerin analizi etkili bir yöntemdir. Çalışmada önerilen yöntem kullanıcı yorumlarından AK üretip bu AK'leri kümeleyerek konuları oluşturma ve konuları temsil etmek üzere üretildiği metine anlamsal olarak en benzer AK'leri seçmektir. Yöntemin oluşturduğu konuların tutarlılığını ölçmek için Coherence metriği kullanılmış, benzer konu modelleme yöntemi olan BERTopic ile ve istatistiksel yöntemlerden LDA ve NMF ile performans karşılaştırması yapılmıştır.

Yöntemlerin konu modelleme veri seti üzerindeki sonuçları alt başlıklarda verilmiştir.

4.2.1 Anahtar Kelime Kümeleme ile Konu Modelleme Sonuçları

Sunulan yöntemde, AK üretimi performansı konu modelleme sonuçlarını etkliyeceğinden, kıyaslama yapabilmek için MT5 ve İstem2 ile GPT4 modelini kullanma yaklaşımları test edilmiştir. Yapılan diğer konfigürasyonlarda metinlerin vektörleştirilmesi için BERTurk ve OpenAI modelleri seçilmiş, boyut azaltma için PCA ve UMAP yöntemleri denenmiş, kümeleme algoritmalarından K-Means, HDBSCAN ve GMM kullanılmıştır.

Vektörleştirme modellerinin farklı boyutlarda vektörler ürettiğinden bahsedilmişti, boyut azaltma adımında PCA'nin indirgeyeceği uzay belirlenirken bileşenlerin toplam varyansı ne kadar kapsayabildiği incelenmiştir. MT5 ve GPT4 modellerinin ürettiği AK'lerin BERTurk ve OpenAI vektörleştirme modelleri ile üretilen vektörleri için 10 bileşenin her birinin temsil ettiği varyanslar Şekil 4.2'de verilmiştir.



Şekil 4.2 PCA ile açıklanan varyans

Deneylerde kullanılmak üzere, AK üreten model fark etmeksizin 1536 boyutlu (768+768) BERTurk vektörleri için PCA bileşeni 3, 6144 boyutlu (3072+3072) OpenAI için 4 seçilmiştir. UMAP kullanıldığı tüm durumlarda ise veri 2 boyutlu uzaya indirgenmiştir. Kümeleme algoritmalarından K-Means ve GMM'e parametre olarak verilecek küme sayısı bilgisini seçebilmek için [2,3,4,5,6,7, 8,9,10,11,12,13,14,15,20,30,40] değerlerinde küme sayıları ile elde edilen Silhouette skorları gözlemlenmiştir. Silhouette değeri yüksek olan küme sayısı bilgisi seçilerek konu modelleme işlemine devam edilmiştir. Diğer kümeleme algoritması olan HDBSCAN ise küme sayısı bilgisine ihtiyaç duymamaktadır, min

cluster size=25, min samples=5 değerleri kullanılarak kümeleme yapılmıştır.

AK üretme modeli olarak MT5 seçilerek test edilen 12 konfigürasyon için, konu modelleme sonuçları Tablo 4.6'da verilmiştir. Vektörleştirme modeline göre en iyi konfigürasyonlar incelenmiştir, HDBSCAN ile oluşturulan konu modellerinin Coherence skorunun yüksek olduğu görülmüştür. Denenen konfigürayonlar içinde Coherence skoru en yüksek model OpenAI ile oluşturulan vektörlerin PCA ile boyutunun indirgenmesi ve HDBSCAN ile iki küme (konu) oluşturması ile elde edilmiştir. Ancak bu modelin Silhoutte skorunun görece düşük olması ve çok az sayıda konu oluşturması sebebiyle manuel elle ihtiyaç duyulmuştur.

Tablo 4.6 MT5 ile AK kümeleme sonuçları

AK modeli	Vektörleştirme modeli	Boyut azaltma	Kümeleme algoritması	Küme sayısı	Silhoutte skoru	Coherence skoru
MT5	BERTurk	PCA	K-Means	5	0,3874	0,4668
MT5	BERTurk	UMAP	K-Means	13	0,4936	0,4992
MT5	BERTurk	PCA	HDBSCAN	6	0,3077	0,5878
MT5	BERTurk	UMAP	HDBSCAN	11	0,276	0,4716
MT5	BERTurk	PCA	GMM	3	0,3767	0,5343
MT5	BERTurk	UMAP	GMM	13	0,4784	0,5563
MT5	OpenAI	PCA	K-Means	5	0,3236	0,4864
MT5	OpenAI	UMAP	K-Means	20	0,5020	0,4729
MT5	OpenAI	PCA	HDBSCAN	2	0,3199	0,7439
MT5	OpenAI	UMAP	HDBSCAN	3	0,2895	0,5962
MT5	OpenAI	PCA	GMM	8	0,2865	0,5446
MT5	OpenAI	UMAP	GMM	20	0,4423	0,4732

Manuel kontrol yapıldığında, en yüksek skoru veren konu modelini oluşturan kelimeler Tablo 4.7'te verilmiştir. Konuyu oluşturan AK'lere baktığımızda konunun baskın olarak satış sonrası hizmet (karşılama, istek, otele giriş) ile ilgili olduğu ve üretilen bazı AK'lerin yanlış veya eksik yazıldığı görülmüştür.

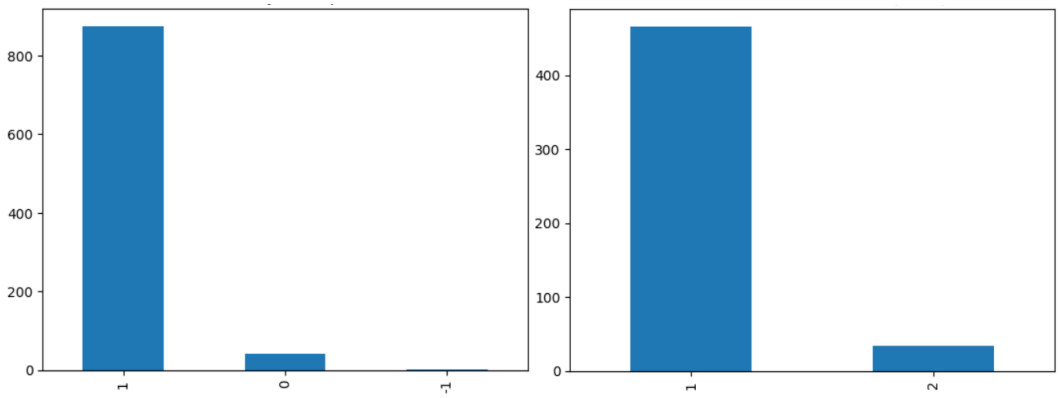
Şekil 4.3a, konu modelinin yorum-AK ikililerini nasıl ayırdığını göstermektedir. Model çok az bir kısmı gürültü (-1) olarak etiketlerken, verinin çoğunluğunu tek bir konuya atayarak dengesiz bir ayrım yapmıştır.

Tablo 4.7 MT5-OpenAI-PCA-HDBSCAN konu modeli

Konu-0	**Konu-1**
'otel', 'otel tipi'	'ücret', 'setur satis sonrasi etkisiz eleman', 'otel girisinde karsilama', 'rezervasyon sonrasi otele iletilen isteklerin karsilanma', 'seffaf', 'fiyatlandirma', 'otel karsilami', 'indirim kuponu', 'diğer', ' 'otele giriste serum görevlisi'

Yorumların konu dağılımlarına baktığımızda büyük çoğunluğunun tek bir konu ile

temsil edildiği tespit edilmiştir (4.3b).



(a) Yorum-AK ikilisi konu dağılımları **(b)** Yorumların konu dağılımları

Şekil 4.3 MT5-OpenAI-PCA-HDBSCAN konu modeli istatistikleri

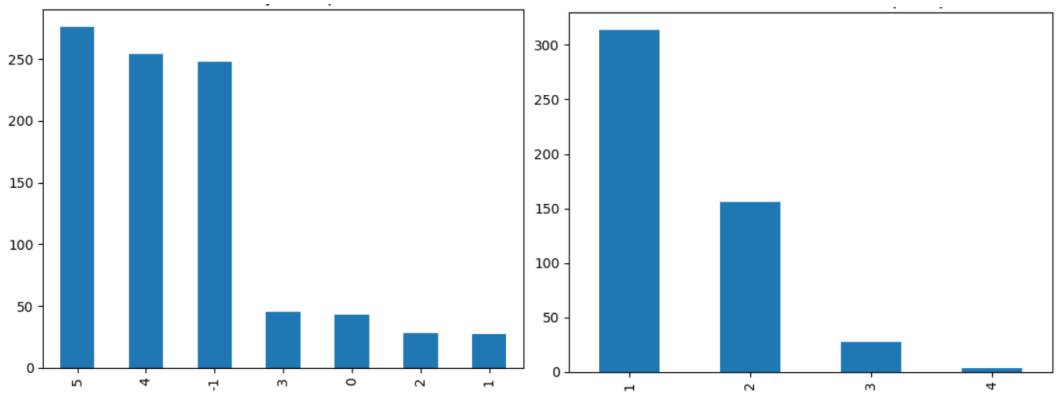
BERTurk vektörleştirme modeli kullanılarak en iyi sonucu veren konu modelinin içeriği Tablo 4.8'te verilmiştir.

Tablo 4.8 MT5-BERTurk-PCA-HDBSCAN konu modeli

Konu-0	**Konu-1**	**Konu-2**
'otel', 'konaklama'	'otel rezervasyonları', 'otel bilgisi', 'oda tercihi', 'otel problem', 'otel rezervasyonu', 'otel tavsiye', 'otel temsilcisi', 'oda bilgisi', 'oda tipi'	'ücret', 'diğer', 'planlama', 'indirim', 'taksit', 'kampanya', 'bilgilendirme', 'bilgi', 'ilgi'
Konu-3	**Konu-4**	**Konu-5**
'otel karsilayan', 'otelde yemek kalitesi', 'otel kalitesi', 'otel personeli', 'oteldeki rehber', 'otel seçimleri', 'otelde temizlik', 'otel tavsiyesi', 'otel uyarıları', 'otel sayısı'	'arama', 'cevap', 'ilgi', 'diğer', 'tur', 'ücret', 'rehber', 'yolcu', 'sorma', 'rezervasyon'	'otel çesitliligi', 'balayi konsepti', 'puan kisitlama', 'anket sonrasi taraf', 'yemek alanindaki kedi', 'otelde yapilan etkinlik', 'yemek alanindaki ürünler', 'yardimci olma', 'tur anlatilanlari', 'rezervasyon gunu'

MT5-BERTurk-PCA-HDBSCAN modeli ile 6 konu oluşturulmuştur. Konu atamalarının biraz daha dengeli olduğu görülmektedir (Şekil 4.4a), aynı zamanda bir yorumun birden fazla konu ile temsil edilme oranı da artmıştır (Şekil 4.4b).

Yorum-AK ikililerinin farklı konulara atanarak birleştirilmesi ile bir yorum, birden fazla konuya atanabilmektedir. Birden fazla kez aynı konuya atanmış ise tekrar gösterilmemiştir.



(a) Yorum-AK ikilisi konu dağılımları **(b)** Yorumların konu dağılımları

Şekil 4.4 MT5-BERTurk-PCA-HDBSCAN konu modeli istatistikleri

Konu modelleme yaparken GPT4 modelinin ürettiği AK'ler kullanılması ile konfigürasyonlarda alınan sonuçlar Tablo 4.9'da gösterilmiştir. Konu tutarlılığını veren Coherence skoru en yüksek modellerde, boyut indirgeme yöntemi olarak PCA ve kümeleme olarak HDBSCAN ve GMM modelleri kullanılmıştır.

Tablo 4.9 GPT4& İstem2 ile AK kümeleme sonuçları

AK modeli	Vektörleştirme modeli	Boyut azaltma	Kümeleme algoritması	Küme sayısı	Silhoutte skoru	Coherence skoru
GPT4	BERTurk	PCA	K-Means	5	0,4116	0,4366
GPT4	BERTurk	UMAP	K-Means	9	0,4495	0,4366
GPT4	BERTurk	PCA	HDBSCAN	6	0,2426	0,4706
GPT4	BERTurk	UMAP	HDBSCAN	4	0,0848	0,378
GPT4	BERTurk	PCA	GMM	5	0,3710	0,5149
GPT4	BERTurk	UMAP	GMM	5	0,4263	0,4094
GPT4	OpenAI	PCA	K-Means	5	0,3145	0,4351
GPT4	OpenAI	UMAP	K-Means	20	0,4039	0,3804
GPT4	OpenAI	PCA	HDBSCAN	3	0,1886	0,812
GPT4	OpenAI	UMAP	HDBSCAN	2	0,2288	0,3925
GPT4	OpenAI	PCA	GMM	5	0,2703	0,4566
GPT4	OpenAI	UMAP	GMM	13	0,3453	0,3868

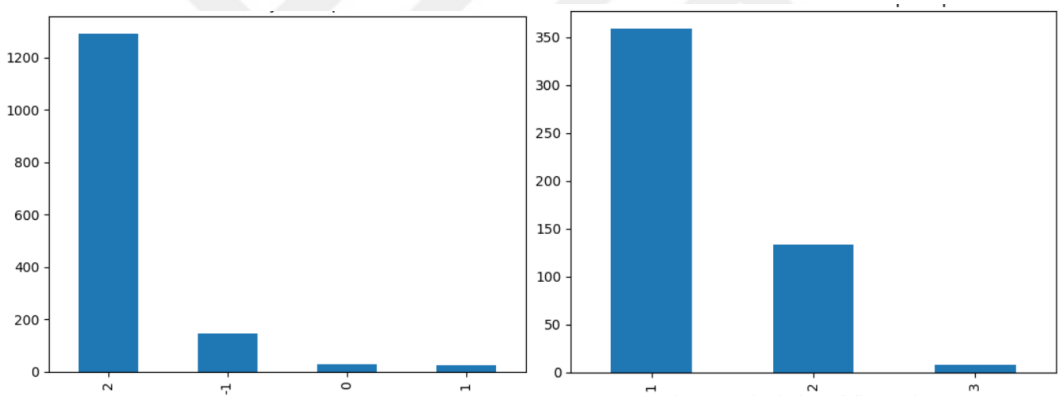
TRT5 ile üretilen AK'lerin kümelenmesine benzer şekilde, GPT4 ile üretilen AK'ler ile kurulan konu modellerinde de en iyi konfigürasyon OpenAI ile oluşturulan vektörlerin PCA ile boyutunun indirgenmesi ve HDBSCAN ile konu modellemesi olmuştur. Silhoutte skoru görece düşük olmasına rağmen en iyi Coherence skorunu vermiştir.

Konu modelinin oluşturduğu konular Tablo 4.10'da verilmiştir. Konu-1'de sadece "setur" olması ve veri incelendiğinde bazı yorumlar için AK olarak "setur" kelimesinin birkaç kez üretilmesi ile meydana gelmiştir. Hatalı AK üretimi, yanlış bir şekilde Coherence skorunun yüksek olmasına sebep olmuştur ve HDBSCAN modeli bunu bir gürültü olarak saptamamıştır.

Tablo 4.10 GPT4-OpenAI-PCA-HDBSCAN konu modeli

Konu-0	Konu-1	Konu-2
'otel setur', 'otel', 'otel transfer'	'setur'	'ücret', 'daha iyi oda', 'sehiriçi transfer', 'satis sonrasi destek', 'indirim kuponu', 'çözüm üretebilme kabiliyeti', 'diğer', 'çözüm üretme', 'otel oda fotosu', 'fiyat avantajı'

Şekil 4.5'da GPT4-OpenAI-PCA-HDBSCAN konu modeli ile oluşturulan konular için yorum-AK ikilileri konu dağılımları ve yorumların kaç konuya atandığının dağılımları verilmiştir. Bir yorumun en fazla üç konuya atanmış ve yorum-AK ikililerinde yaklaşık 200 değer gürültü olarak değerlendirilmiştir.



(a) Yorum-AK ikilisi konu dağılımları **(b)** Yorumların konu dağılımları

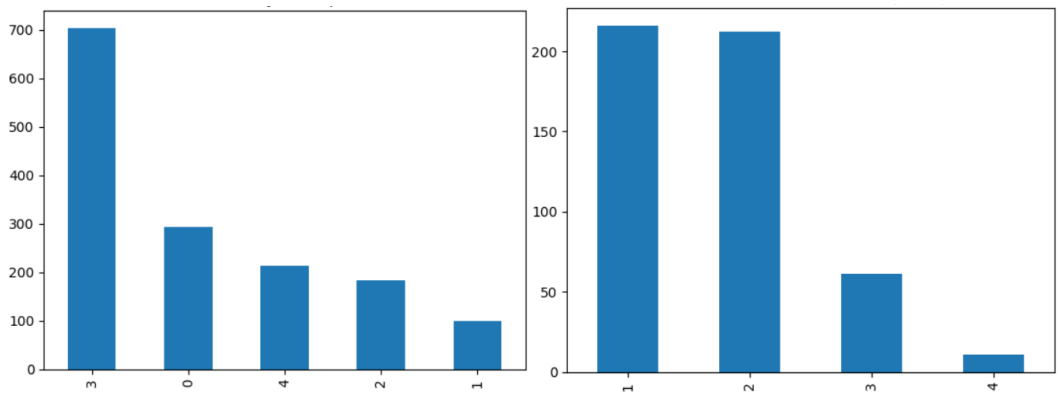
Şekil 4.5 GPT4-OpenAI-PCA-HDBSCAN konu modeli istatistikleri

GPT4 ile üretilen AK'ler BERTurk modeli kullanılarak vektörleştirildiğinde ulaşılan en iyi Coherence skoru GMM kümeleme algoritması ile alınmıştır. Konu modelinin oluşturduğu konular ve içerikleri Tablo 4.11'de verilmiştir. Tur ile ilgili Konu-1 (ekstra tur bedeli), Konu-2 (merkeze yakın otel)'de benzer AK'lerin bulunduğu, Konu-2'de tur sırasında konaklanan otel ile ilgili yorumlar olduğu görülmektedir. Konu-1'de de otel ile ilgili AK'ler bulunduğundan buradaki konuların iyi toparlanamadığı ve kümelemenin iyi yapılamadığı söylenebilir.

Şekil 4.6a'da yorum-AK ikililerinin çoğunlukla Konu-3'e atandığı görülmektedir. Yorumların konu dağılımlarına bakıldığında ise bu konu modeli ile bir yoruma birden fazla konu atandığı söylenebilir (Şekil 4.6b).

Tablo 4.11 GPT4-BERTurk-PCA-GMM konu modeli

Konu-0	**Konu-1**	**Konu-2**
'arama', 'cevap', 'çalışanlar', 'diğer', 'sigorta', 'tur', 'rehber', 'bilgili', 'konum', 'rezervasyon'	'ekstra tur bedelleri', 'merkezi yerler', 'otel anlatımı', 'otel rezervasyonları', 'güzel hizmet', 'sorun çözümü', 'tur bilgisi', 'transfer hizmeti', 'rehberin kokartlı olmaması'	'daha iyi oda', 'otel oda fotosu', 'merkeze yakın oteller', 'otel karşılama', 'otel kalitesi', 'otel ilgilenme', 'konaklama', 'otel personeli', 'yemek kalitesi', 'otel girişi'
Konu-3	**Konu-4**	
'satis sonrasi destek', 'otelin olumsuz yanları', 'havuz başı havlu teslim alanı', 'karsilik bulmama', 'aynı gün iade', 'sonuç alamama', 'balayı konsepti', 'çocuk priz korumalığı', ' ekstra geziler'	'ücret', 'sehiriçi transfer', 'rehberlik hizmeti', 'indirim kuponu', 'çözüm üretebilme kabiliyeti', 'diğer', 'çözüm üretme', 'fiyat avantajı', 'satis sonrasi', 'rehberler'	



(a) Yorum-AK ikilisi konu dağılımları **(b)** Yorumların konu dağılımları

Şekil 4.6 GPT4-BERTurk-PCA-GMM konu modeli istatistikleri

AK kümeleme ile kurulan konu modelleri incelendiğinde, konfigürasyon seçmek için sadece Coherence skoruna bakmak yanıltıcı olmuştur, bazı modellerde bu skor yüksek olmasına rağmen çok az sayıda konu üretildiği ve yorumların konularının dengesiz dağıtıldığı fark edilmiştir. Hem Coherence skoru hem de Silhoutte skoru yüksek olan konfigürasyonlardan MT5 modelinin ürettiği AK'leri kümeleyen MT5-BERTurk-UMAP-GMM konu modelinin ürettiği konular Tablo 4.12'de verilmiştir. Tur ile ilgili başlıkların Konu-1 ve Konu-6 olmak üzere iki konuda toplandığı, az da olsa diğer konularda da tur ile AK'ler olduğu görülmektedir. Çevrimiçi kanallar ile ilgili AK'leri Konu-11'de, satış sonrası hizmet yorumları ise Konu-10'da toplanmıştır.

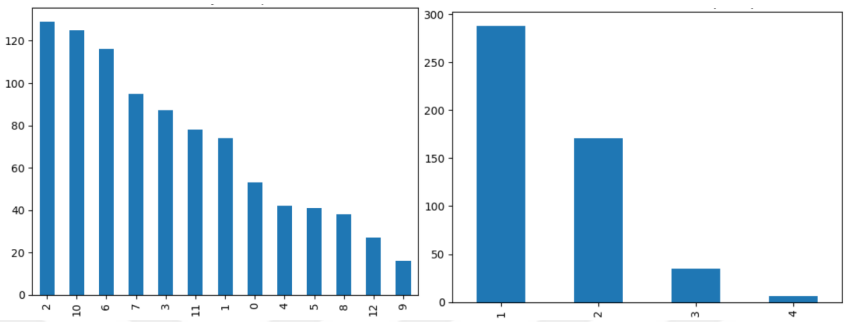
Tablo 4.12 MT5-BERTurk-UMAP-GMM konu modeli

Konu-0	**Konu-1**	**Konu-2**	**Konu-3**	**Konu-4**
!otel sartları', 'oteli ücretsiz sekilde degistime', 'otel rezervasyonları', 'transfer hizmeti', 'otel bilgisi', 'otel temsilcisi', 'iade sigortası', 'oda tercihi', 'otel listesi', 'konaklama yerleri'	'arti', 'rehber', 'hizmet', 'hizmet yerine getirme', 'diğer yolcular', 'zaman kullanımı', 'kalifiye', 'bin rota', 'tur', 'otel rehber'	'indirim kuponu', 'planlama', 'indirim', 'fiyat', 'taksit', 'organizasyon', 'hijyen sorunu', 'kampanya', 'uygulama', 'diğer'	'ilgi görme', 'arama', 'cevap', 'ilgi', 'bilgi', 'telefon arama', 'odak', 'diğer yüz', 'aranma'	'otel', 'konaklama'
Konu-5	**Konu-6**	**Konu-7**	**Konu-8**	**Konu-9**
'setur a dair ilgi', 'setur danismanı', 'setur yetkilisi', 'otel setur temsilcisi', 'setur görevli', 'setur yetkili', 'setur kalitesi', 'setur indirimi', 'otel setur çizgi', 'setur çalisani'	'fiyat avantajı', 'tur sahiplenmesi', 'tur detayları', 'ekstra indirim', 'ekstra tur bedelleri', 'tur org.', 'rehberlik hizmeti', 'seyahat öncesi bilgilendirme', 'ücret ödeyen tur sirketleri', 'tur rehberi'	'otel tercihi', 'konaklama oteli', 'otel kalitesi', 'konaklama', 'oteldeki firma temsilcisi', 'sehiri', 'otelde ilave avantaj', 'otel rehberi', 'oda', 'merkezi yer'	'ücret', 'seffaf', 'diğer'	'oda'
Konu-10	**Konu-11**	**Konu-12**		
'setur satis sonrasi etkisiz eleman', 'otel girisinde karsilama', 'rezervasyon sonrasi otele iletilen isteklerin karsilanma', 'fiyatlandirma', 'otel karsilami', 'otele giriste serum görevlisi', 'rezervasyon sonrasi hizmet', 'otelle iletisime geçme', 'otel çesitliligi', 'otelde danismani'	'kullanici dostu', 'bilgisizlik', 'telefonla iletisim', 'seçenek sunma', 'stabil çalismasi', 'cozum', 'sorun çözme', 'internet sitesi', 'destek verme', 'web sitesi'	'bilgilendirme', 'otobüs kalitesi', 'serbest zaman', 'personel bilgilendirme', 'yönlendirme', 'internet', 'uyarı'		

Modelin yorumlara atadığı konuların dağılımı Şekil 4.7a'da verilmiştir. Önceki konu modellerine benzer olarak konu dağılımında hala dengesizlik olsa da, modelin daha fazla konu üretmesi ile belirli bir konudaki yüksek yığılımın azaldığı görülmüştür. Şekil 4.7b'da bir yorumun atandığı konu sayılarının dağılımı gösterilmektedir. Öncekilere benzer şekilde çoğunluk tek bir konuya atansa da, yorumların yaklaşık %40'ı birden fazla konu ile temsil edilmektedir.

Önerilen yöntem ile konu modelleme işleminde AK üretimi adımı kilit rol oynamakta ve oluşturulan konuların kalitesini ciddi bir şekilde etkilemektedir. MT5 modelinin ürettiği AK'lerde yazımsal hatalar bulunsa da, GPT4 modeline göre daha az AK üreterek yorumu daha iyi sıkıştırmakta ve daha az gürültü oluşturmaktadır. Konu modeli veri setindeki 500 yorum için MT5 modeli ile 921 yorum-AK ikilisi oluşturulurken, GPT4 modeli için bu sayı 1494'tür. Karşılaştırılan konu

modellerinde, MT5-BERTurk-UMAP-GMM konu modeli ile üretilen konuların diğer konu modellerine göre daha iyi kırılımlar yaptığı ve daha detaylı bilgi sağladığı gözlemlenmiştir. Konu modelleri karşılaştırılırken Silhoutte ve Coherence skorlarını birlikte incelenmelidir. Silhoutte skorunun daha iyi olduğu modeller konuları daha iyi bölümlemektedir. Modellerin yaptığı konu dağılımlarına bakarak, konu modelleme veri setindeki yorumların önemli bir kısmının birden fazla konu ile temsil edilebileceği yargısına ulaşılmıştır.



(a) Yorum-AK ikilisi konu dağılımları **(b)** Yorumların konu dağılımları

Şekil 4.7 MT5-BERTurk-UMAP-GMM konu modeli istatistikleri

4.2.2 Sinirsel Konu Modelleme (BERTopic) Sonuçları

Sinirsel konu modelleme yöntemlerinden BERTopic yönteminde, önerilen yönteme benzer olarak dil modelleri, metinleri vektörleştirmek için kullanılır. Bu yöntemde, konu modelleme veri setindeki kullanıcı yorumları BERTurk modeli ile vektörleştirilmiş, boyut azaltma için PCA ve UMAP kullanılmış, kümeleme için HDBSCAN, K-Means ve GMM algoritmaları kıyaslanmıştır. Toplamda 6 farklı konfigürasyon kullanılarak BERTopic yaklaşımı ile konu modellemesi oluşturulmuş, konuları temsil eden kelimeler seçilirken dolgu sözcükleri (stopwords) kaldırılmıştır. AK kümeleme yöntemine paralel olarak, metinleri daha küçük uzaylara indirgerken PCA için 3, UMAP için 2 bileşen kullanılmıştır. Parametre olarak küme sayısının belirtilmesi gereken K-Means ve GMM algoritmalarında, en iyi Silhoutte skorunu veren küme sayısı değeri seçilmiş ve konu modellemeye devam edilmiştir.

Tablo 4.13 BERTopic sonuçları

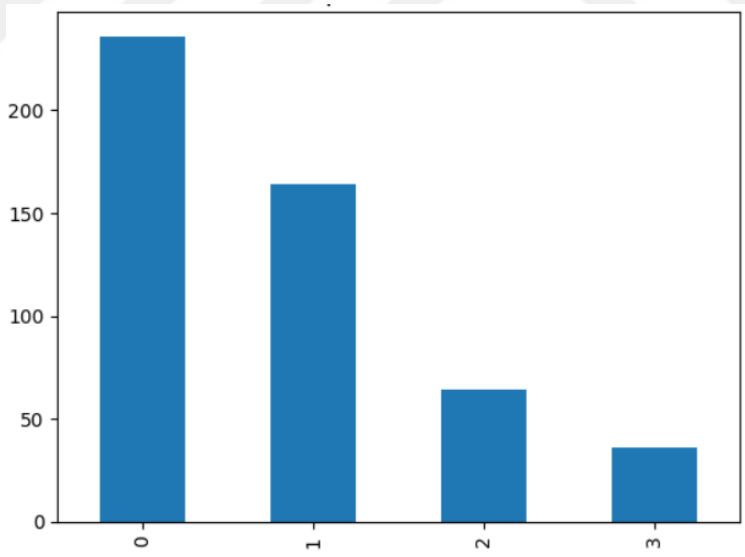
Vektörleştirme modeli	Boyut azaltma	Kümeleme algoritması	Küme sayısı	Coherence skoru
BERTurk	PCA	K-Means	2	0,3617
BERTurk	PCA	HDBSCAN	3	0,3955
BERTurk	PCA	GMM	4	0,4466
BERTurk	UMAP	K-Means	2	0,381
BERTurk	UMAP	HDBSCAN	13	0,4036
BERTurk	UMAP	GMM	2	0,355

BERTopic yaklaşımında kurulan konu modellerinin performansları Tablo 4.13'de verilmiştir. En iyi Coherence skorunu PCA ve GMM kullanılarak ulaşılmış, ikinci en iyi skor ise UMAP ve HDBSCAN ile daha fazla küme oluşturularak elde edilmiştir.

En iyi BERTopic konfigürasyonu BERTopic-PCA-GMM konu modelinin oluşturduğu konu ve içerikler Tablo 4.14'te verilmiştir. Şekil 4.8'de de yorumların hangi konulara atandığının dağılımı verilmiştir.

Tablo 4.14 BERTopic-PCA-GMM konu modeli

Konu-0	**Konu-1**	**Konu-2**	**Konu-3**
'setur', 'tur', 'otel', 'otelde', 'otele', 'gün', 'sadece', 'oda', 'ilgili'	'otel', 'setur', 'hakkinda', 'hizmeti', 'bilgi', 'otelde', 'otelleri', 'tur', 'ilgili', 'konusunda'	'otel', 'ilgili', 'tur', 'bilgilendirme', 'öncesi', 'setur', 'ilgilenseydi', 'rehberlik', 'otelleri', 'yetersiz'	'indirim', 'ilgi', 'konaklama', 'yüksek', 'güvenimi', 'sarsti', 'seffaf', 'sehiriçi', 'taksit', 'karsilamasi'



Şekil 4.8 BERTopic-PCA-GMM konu modeli konu dağılımları

Konu modellerinin temsilinde TF-IDF tabanlı yöntem kullanıldığından bahsedilmişti. Bu teknikte konu sınıfı için önemli kelimeler seçilmektedir. Seçilen kelimeler tek başına anlamsal olarak belirsiz olabilmekte ya da fiiller yoğunlukta olabilmektedir. En iyi ikinci konfigürasyon incelendiğinde bu duruma fazlaca rastlanılmıştır.

4.2.3 İstatiksel Konu Modelleme Sonuçları

Önerilen yöntem ile karşılaştırma yapmak üzere seçilen temel modeller LDA ve NMF modelleridir. İstatistikel tabanlı bu yöntemleri kullanırken, BERTopic yöntemindeki gibi dolgu kelimeler metinlerden kaldırılmıştır. Modeller için küme sayısı belirlenirken, 2-20 küme arası değerler denenerek Coherence skorunu maksimize eden değer seçilmiştir. İşlem sırasında her deney 3'er kez tekrarlanarak ortalamaları karşılaştırma için kullanılmıştır. İstatistiksel konu modellerinin en iyi konfigürasyonlarının başarısı Tablo 4.15'te verilmiştir.

Tablo 4.15 İstatistiksel konu modelleri sonuçları

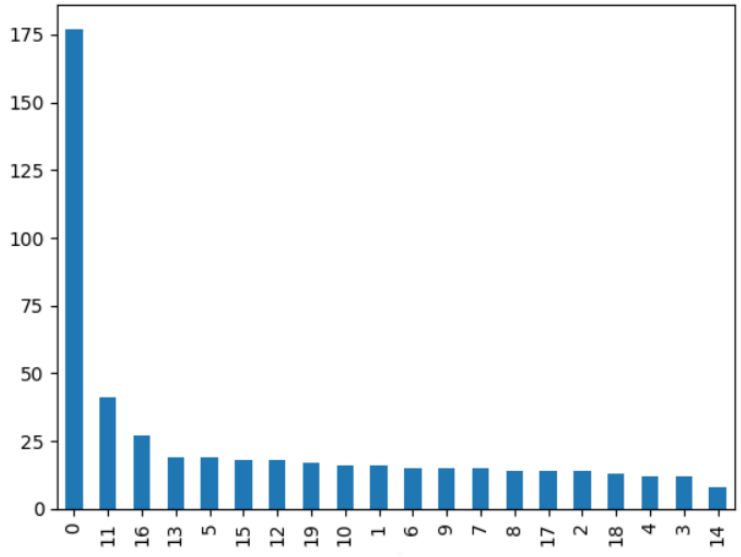
Konu modeli	Küme sayısı	Coherence skoru
LDA	20	0,3896
NMF	20	0,4086

Tablo 4.16'da en iyi Coherence skoruna 20 küme ile ulaşan LDA modelinin oluşturduğu konular incelendiğinde, çekim ekleri ile farklılaşan ancak aynı kavramı ifade eden kelimelerin bulunduğu (otel/otelde), bazı kelimelerin birçok konu içerisinde ortak olarak yer aldığı (setur, otel, tur) görülmektedir. Konuların keskin sınırlar ile bölümlenmemesinin yanı sıra konuları temsil eden kelimelerde anlamsal belirsizlikler bulunmaktadır. Dolgu kelimelerinin genişletilmesi, metinlerdeki yazım yanlışlarının düzeltilmesi ve kelime kökleri ile işlem yapılması bu modelin başarısını az miktarda da olsa arttırabilir.

LDA modeli ile yorumlara atanan konular Şekil 4.9'da verilmiştir. Model her ne kadar yüksek sayıda konu üretmiş olsa da konu dağılımlarında dengesizlik bulunmaktadır. Yorumların büyük çoğunluğu Konu-0'a atanırken diğer konuların temsil ettiği yorum sayıları görece dengelidir.

Tablo 4.16 LDA modeli

Konu-0	**Konu-1**	**Konu-2**	**Konu-3**	**Konu-4**
"otel","otelde", "yüksek","puan"," ilgili","ücret", "otellerde","hakkinda"	"otel","setur", "oda","otelde", "ragmen","konaklama", "oteli","otellerde"	"problem","otel", "otele","setur", "sonrasi","yemek", "bilgilendirme","adim", "hizmeti"	"setur","otel", "otele","yüksek", "sadece","olmasi", "giris","satis"	"sikayetleri","kere", "kaldigimiz","zorunda", "baska","bunun", "otelde","müsteri"
Konu-5	**Konu-6**	**Konu-7**	**Konu-8**	**Konu-9**
"tur","otel", "otelde","setur", "ilgili","olmasi", "baska","sizin"	"otel","dogru", "bilgi","tur", "ekstra","rehber", "hakkinda","oda"	"transfer","yüksek", "konusunda","ilgili", "otel","otele", "ilgilenseydi","olmasaydi"	"setur","tur", "otel","yardimci", "ilgili","ayni", "otelde","sekilde"	"tur","setur", "oteller","otel", "giris","olurdu", "kaldigimiz"
Konu-10	**Konu-11**	**Konu-12**	**Konu-13**	**Konu-14**
"gün","otel", "bile","setur", "bos","otelde" "tercih"	"sonra","oda", "otel","tur", "puan","aldiktan", "vermis","ilgili", "otele"	"ilgi","setur", "bilgilendirme","otelde", "otelleri","gün", "sonra","un"	"setur","otel", "kadar","sadece", "idi","hakkinda", "iletisime","otelin"	"setur","otel", "indirim","tur", "bunu","diger", "yüksek","iletisime"
Konu-15	**Konu-16**	**Konu-17**	**Konu-18**	**Konu-19**
"otel","tur", "rezervasyon","hizmeti", "rehberlik","olmali", "olmasi","setur"	"setur","otel", "tur","rezervasyon", "ilgili","transfer", "sitesi","özellikle"	"turu","tur", "setur","rehber", "ayni","ayrica", "yapilmadi"	"setur","tur", "otelde","otel", "ilgi","indirim", "müsterilerine","hizmet"	"otel","oteller", "tur","bilgilendirme", "rezervasyon","tatil", "kisi","ücret"



Şekil 4.9 LDA konu modeli konu dağılımları

Daha yüksek Coherence skoruna ulaşan NMF modelinin oluşturduğu konular ve içerikleri Tablo 4.17'de bulunmaktadır. LDA modeline benzer şekilde en iyi performansa 20 küme ile ulaşılmıştır. Modelin yorumlarda anlatılmak istenenleri LDA'ya göre daha iyi toparladığı görülmektedir. Örneğin Konu-12'nin çevrimiçi kanallar ile ilgili olduğu, tur ile ilgili yorumların Konu-9 ile temsil edildiği söylenebilir. Ayrıca bazı kelimelerin birçok kümede bulunması LDA'daki kadar yaygın değildir.

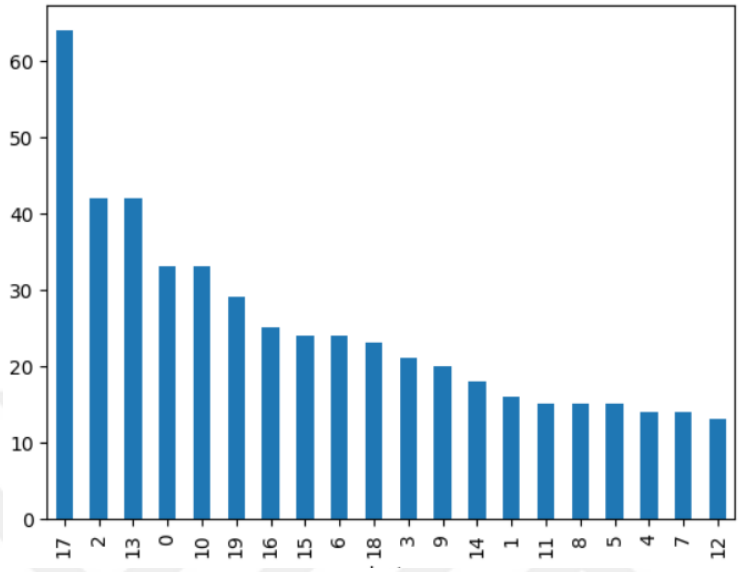
Tablo 4.17 NMF modeli

Konu-0	**Konu-1**	**Konu-2**	**Konu-3**	**Konu-4**
"otel", "organizasyonu", "tercih", "kalitesi", "yorumlari", "asla", "giderilmesi", "ilgilenseydi", "girisinde", "seçimi"	"otelde", "yetkilisi", "görevlisi", "danismani", "saglasaydi", "karsilama", "elemani", "idi", "bizi"	"tur", "uçak", "yetersiz", "olurdu", "tek", "katildigimiz", "tesiste", "rehberi", "rehberlerini", "ekstra"	"ilgi", "bilgilendirme", "yok", "sonrasi", "yapmasi", "alaka", "satis", "sifir", "gösterseydi", "müsterilerine"	"hizmeti", "rehberlik", "transfer", "verseydi", "rezervasyon", "berbatti", "tam", "sonrasi", "saglasaydi", "firma"
Konu-5	**Konu-6**	**Konu-7**	**Konu-8**	**Konu-9**
"otelleri", "yakin", "secseydi", "merkezine", "sehir", "vasat", "kaliteli", "sehire", "magdur", "bizde"	"otele", "rezervasyon", "giris", "giriste", "yapilan", "mesela", "sonrasi", "iletisime", "personeli", "tane"	"indirim", "kampanya", "kuponu", "fiyat", "üyelerine", "yüksek", "ayri", "acentalarla", "yapilabilirdi", "dört"	"bilgi", "hakkinda", "detayli", "vermeliydi", "vermesi", "verseydi", "sahibi", "otel", "esnasinda","ayrintili"	"rehber", "saglasaydi", "gerçek", "profesyonel", "gruba", "acente", "kullansaydi", "farkli", "ekstra", "rehberler"
Konu-10	**Konu-11**	**Konu-12**	**Konu-13**	**Konu-14**
"ilgili", "oteldeki", "yetkili", "çözüm", "bulsaydi", "olabilirlerdi", "eksikti", "yanlis", "otellerle", "otelle"	"oteli", "konaklama", "anlatabilirdi", "çikarmali", "tatil", "satmasaydi", "sikayet", "temizlik", "alan", "kadar"	"sitesi", "web", "olmali", "internet", "çalisiyor", "kullanici", "dostu", "çalismasi", "hizli", "islem"	"yüksek", "puan", "verirdim", "verebilirdim", "etmeseydi", "tavsiye", "diger", "skor", "fiyatlar", "skoru"	"oteller", "kaliteli", "tercih", "anlasmis", "soför", "idi", "otobüste", "anlasmalar", "alt", "yakin"
Konu-15	**Konu-16**	**Konu-17**	**Konu-18**	**Konu-19**
dogru", "bizimle", "bilgi", "otellerin", "verseydi", "yapmis", "isini", "öncesi", "rez", "gün"	"yardimci", "konusunda", "transfer", "iade", "sorun", "vize", "olabilirlerdi", "olan", "ulasim", "bilgilendirme"	"setur", "baska", "uygun", "bile", "yetkilisi", "sadece", "görmedim", "rezervasyon"	"oda", "satis", "güncel", "yapilirken", "aldigim", "tam", "sonrasi", "gruba"	"yapsaydi", "dönüs", "kendi", "müsteri", "yasadigim", "isini", "gidis", "hizmetlerini", ragmen", "aramama"

NMF modeli ile oluşturulan konu modellemesinde, yorumların konular ile temsili daha dengeli olmuştur (Şekil 4.10). Konu-17, 60'tan fazla yorumu temsil etmektedir.

İstatiksel modellerin oluşturduğu konular incelendiğinde, NMF modelinin daha iyi konular oluşturduğu ve görevde daha başarılı olduğu görülmektedir. BERTopic

modeline benzer şekilde, konuları oluşturan kelimelerin tek kelimeden oluşması zaman zaman anlamsal belirsizlik yaratabilmektedir. Yine benzer şekilde, fiiller ve tamamlayıcı sözcükler bazı konuların içeriğinde bulunabilmekte ve ilk bakışta konunun anlaşılmasını zorlaştırmaktadır.



Şekil 4.10 NMF konu modeli konu dağılımları

5
SONUÇ

Kullanıcı ile etkileşimi yüksek sektörler için müşterilerinin geri bildirimlerini incelemek sunulan hizmeti iyileştirip ulaşılan müşteri sayısını arttırmak ve var olan müşterileri kaybetmemek adına önem arz etmektedir. Çalışmada, turizm sektöründe faaliyet gösteren Setur'un müşterilerine verdiği hizmet sonlandıktan sonra gönderdiği "Konaklama Sonrası Anketleri" ile topladığı geri bildirimlerden içgörü çıkarmak amaçlanmıştır. Kullanıcı yorumlarında anlatılan konuları tespit edebilmek için geleneksel konu modellemeleri yerine, bu yorumları temsil edecek anahtar kelimeler üretmek ve bunları kümeleyerek konular oluşturmak önerilmiştir.

İlk aşama olan AK üretme görevinde iki farklı yaklaşım uygulanmıştır. Ön eğitimli dil modellerine ince ayar yapmanın, daha az veri kullanarak ya da hiç veri kullanmayarak büyük dil modellerine istemler vererek AK ürettirmenin performansı kıyaslanmıştır. Yapılan deneylerde ince ayar yapılan TRT5, MT5, TRBART ve MBART modellerinden en yüksek ve kararlı performans gösteren modelin MT5 olduğu sonucuna ulaşılmıştır. T5 ve BART versiyonları kıyaslandığında, performansta çarpıcı bir fark olmasa da T5 modelinin AK üretme görevi için daha uygun olduğu söylenebilir. Ön eğitimi Türkçe veri setinde yapılan dil modellerini (TRT5, TRBART) ince ayar yaparak kullanmanın modellerin çok dilli versiyonlarına kıyasla daha iyi çalışmadığı gözlemlenmiştir, bunun sebebi modellerin eğitiminde kullanılan veri setleri gösterilebilir. Öte yandan, LLM'lerden GPT3.5, GPT4 ve Gemini iki farklı istem stili ile çalıştırılmıştır. Görev ile ilgili örnek verilmediğinde bu modellerin herhangi bir kısıt gözetmeksizin AK ürettiği, yorumu kelimelerine ayrıştırdığı gözlemlenmiştir. Örnek verilen few-shot istem stilinin ise performansı iyileştirdiği ve gereksiz kelimeleri daha az ürettiği sonucuna varılmıştır. Dil modellerine ince ayar yapma ve LLM'leri istemler ile kullanma yaklaşımları kıyaslandığında, özellikle GPT4 ve Gemini modelleri few-shot stili ile kullanıldığında etiketli AK'ler ile daha çok tam eşleşme (F1@3) sağlandığı görülmektedir. GPT4 ve Gemini'nin few-shot istem ile ürettiği AK'ler yoruma anlamsal olarak daha çok benzemektedir (Bert-similarity), ayrıca bu AK'ler

birbirine anlamsal olarak daha uzaktır, anlamsal çeşitlilikleri yüksektir (Diversity). Diğer yandan, ince ayar yapılan dil modelleri yazımsal olarak etiketli AK'lere daha çok benzeyen (Rouge) AK'ler üretmekte ve az metin (AK) üreterek yorum temsilini yapabilmekte diğer deyişle metni sıkıştırabilmektedir (Compressing). İnce ayarlı dil modelleri, LLM'lerin örnek verilmeden kullanımına kıyasla daha yarışabilir performans gösterebilmekteyken 10 örnek gibi çok az etiketli veri verilerek LLM'ler kullanıldığında bu modellerin gerisinde kalmaktadır.

LLM'lerin fazla üretkenliğinin azaltılması Sıkıştırma (Compressing) skorlarının iyileştirilebilmesi ve daha doğru AK'ler üretebilmesi için daha iyi istemler yazılabilir. Verilen istemlerin iyileştirilmesi, görevin daha açıklayıcı bir dil ile yazılması ve verilen örneklerin daha fazla ya da genelleştirici olması bu modellerin başarısını arttırabilir. LLM'lerin 10 örnek ile ürettiği AK'lerin, 760 örnek ile tekrar eğitilen T5 ve BART versiyonlarının ürettiklerinde doğruluk bakımından önüne geçmiş olsa da; gelişime açık diğer konu, GPT3.5 veya açık kaynaklı dil modellerine ince ayar yapılmasıdır.

AK üretme yaklaşımlarının test setinde verdiği sonuçlar ekte verilmiştir (Bölüm A). Bazı yorumlar için, ince ayar yapılan modellerin bir kelimeyi üst üste ürettiği durumlar tespit edilmiştir. Benzer çalışmalarda bu durumu önleyebilmek için tekrar etme cezası (repetition penalty) kullanılması tavsiye edilir.

Önerilen konu modeli BERTopic, LDA ve NMF ile kıyaslandığında daha anlaşılır konular oluşturduğu ve daha iyi bölümleme yaptığı görülmüştür. BERTopic, LDA ve NMF'teki ortaklıklar konu içeriğindeki kelimelerin yeterince anlaşılır olmayışı ve veri temizliği aşamasındaki zorluklardır. Tablo A.5, örnek yorumlar için konu modellerinin sonuçları gösterilmektedir.

AK kümeleme ile daha iyi konu modelleri oluşturulabilmesi için yorum-AK ikililerinin kalitesi önemli rol oynamaktadır. Performans gelişimi için, farklı vektörleştirme modellerinin test edilmesi ya da oluşturulan vektörlerin farklı yaklaşımlarla birleştirilmesi (ortalamaları alınarak vb.) sonraki çalışmaların konusu olabilir.

KAYNAKÇA

[1] S. Park, H. M. Kim, "Improving the accuracy and diversity of feature extraction from online reviews using keyword embedding and two clustering methods," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, vol. 84003, 2020, V11AT11A020.

[2] T. Skoghäll, D. Öhman, *Summarization and keyword extraction on customer feedback data: Comparing different unsupervised methods for extracting trends and insight from text*, 2022.

[3] Y.-J. Kim, H.-S. Kim, "The impact of hotel customer experience on customer satisfaction through online reviews," *Sustainability*, vol. 14, no. 2, p. 848, 2022.

[4] C. K. H. Lee, Y. K. Tse, M. Zhang, J. Ma, "Analysing online reviews to investigate customer behaviour in the sharing economy: The case of airbnb," *Information Technology & People*, vol. 33, no. 3, pp. 945–961, 2020.

[5] R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, Y. Chi, "Deep keyphrase generation," *arXiv preprint arXiv:1704.06879*, 2017.

[6] ml6team, $Ml6team_keyphrase-extraction$, [Online], Available: `https://huggingface.co/spaces/ml6team/keyphrase-extraction`.

[7] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[8] S. R. El-Beltagy, A. Rafea, "Kp-miner: A keyphrase extraction system for english and arabic documents," *Information systems*, vol. 34, no. 1, pp. 132–144, 2009.

[9] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, A. Jatowt, "Yake! keyword extraction from single documents using multiple local features," *Information Sciences*, vol. 509, pp. 257–289, 2020.

[10] S. Brin, L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[11] R. Mihalcea, P. Tarau, "Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in natural language processing*, 2004, pp. 404–411.

[12] M. Grootendorst, *Keybert: Minimal keyword extraction with bert.* Version v0.3.0, 2020. doi: `10.5281/zenodo.4461265`. [Online]. Available: `https://doi.org/10.5281/zenodo.4461265`.

[13] E. T. Ayan, R. Arslan, M. S. Zengin, H. A. Duru, S. Salman, B. Bardak, "Turkish keyphrase extraction from web pages with bert," in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2021, pp. 1–4.

[14] J. K. Kumar S, A. Kumar M, "Unsupervised keyphrase extraction using graph and embedding techniques," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, 2023, pp. 1–8.

[15] I. H. Witten, G. W. Paynter, E. Frank, C. Gutwin, C. G. Nevill-Manning, "Kea: Practical automatic keyphrase extraction," in *Proceedings of the fourth ACM conference on Digital libraries*, 1999, pp. 254–255.

[16] N. Pala, I. Cicekli, "Turkish keyphrase extraction using kea," in *2007 22nd international symposium on computer and information sciences*, IEEE, 2007, pp. 1–5.

[17] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[18] M. Lewis *et al.*, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[19] D. Wu, W. U. Ahmad, K.-W. Chang, "Pre-trained language models for keyphrase generation: A thorough empirical study," *arXiv preprint arXiv:2212.10233*, 2022.

[20] A. Glazkova, D. Morozov, "Applying transformer-based text summarization for keyphrase generation," *Lobachevskii Journal of Mathematics*, vol. 44, no. 1, pp. 123–136, 2023.

[21] M. Kulkarni, D. Mahata, R. Arora, R. Bhowmik, "Learning rich representation of keyphrases from text," *arXiv preprint arXiv:2112.08547*, 2021.

[22] J. Achiam *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.

[23] G. Team *et al.*, "Gemini: A family of highly capable multimodal models," *arXiv preprint arXiv:2312.11805*, 2023.

[24] M. Song *et al.*, "Is chatgpt a good keyphrase generator? a preliminary study," *arXiv preprint arXiv:2303.13001*, 2023.

[25] R. Y. Maragheh *et al.*, "Llm-take: Theme-aware keyword extraction using large language models," in *2023 IEEE International Conference on Big Data (BigData)*, IEEE, 2023, pp. 4318–4324.

[26] A. Petukhova, J. P. Matos-Carvalho, N. Fachada, "Text clustering with llm embeddings," *arXiv preprint arXiv:2403.15112*, 2024.

[27] M. S. Asyaky, R. Mandala, "Improving the performance of hdbscan on short text clustering by using word embedding and umap," in *2021 8th international conference on advanced informatics: Concepts, theory and applications (ICAICTA)*, IEEE, 2021, pp. 1–6.

[28] R. Albalawi, T. H. Yeap, M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Frontiers in artificial intelligence*, vol. 3, p. 42, 2020.

[29] J. Qiang, Z. Qian, Y. Li, Y. Yuan, X. Wu, "Short text topic modeling techniques, applications, and performance: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 3, pp. 1427–1445, 2020.

[30] S. Limwattana, S. Prom-on, "Topic modeling enhancement using word embeddings," in *2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2021, pp. 1–5. doi: 10.1109/JCSSE53117.2021.9493816.

[31] R. Egger, J. Yu, "A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts," *Frontiers in sociology*, vol. 7, p. 886 498, 2022.

[32] Google, *mt5-base*, [Online], Available: https://huggingface.co/google/mt5-base.

[33] L. Xue *et al.*, "Mt5: A massively multilingual pre-trained text-to-text transformer," *arXiv preprint arXiv:2010.11934*, 2020.

[34] Turkish-NLP, *t5-efficient-base-turkish*, [Online], Available: https://huggingface.co/Turkish-NLP/t5-efficient-base-turkish.

[35] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[36] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, *et al.*, "Improving language understanding by generative pre-training," 2018.

[37] facebook, *mbart-large-cc25*, [Online], Available: https://huggingface.co/facebook/mbart-large-cc25.

[38] Y. Liu *et al.*, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020.

[39] mukayese, *transformer-turkish-summarization*, [Online], Available: https://huggingface.co/mukayese/transformer-turkish-summarization.

[40] A. Safaya, E. Kurtuluş, A. Göktoğan, D. Yuret, "Mukayese: Turkish nlp strikes back," *arXiv preprint arXiv:2203.01215*, 2022.

[41] OpenAI, *Openai platform*, Accessed on 05.06.2024. [Online]. Available: https://platform.openai.com/docs/models.

[42] dbmdz, *bert-base-turkish-cased*, https://huggingface.co/dbmdz/bert-base-turkish-cased.

[43] Huggingface, *Mteb leaderboard*, Accessed on 05.06.2024. [Online]. Available: https://huggingface.co/spaces/mteb/leaderboard.

[44] OpenAI, *Openai embeddings platform*, Accessed on 05.06.2024. [Online]. Available: https://platform.openai.com/docs/guides/embeddings.

[45] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[46] M. Grootendorst, *Bertopic webpage*, [Online], Available: `https://maartengr.github.io/BERTopic/algorithm/algorithm.html#visual-overview`.

[47] N. Reimers, I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[48] A. Amara, M. A. Hadj Taieb, M. Ben Aouicha, "Multilingual topic modeling for tracking covid-19 trends based on facebook data analysis," *Applied Intelligence*, vol. 51, pp. 3052–3073, 2021.

[49] D. D. Lee, H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *nature*, vol. 401, no. 6755, pp. 788–791, 1999.

A
ANAHTAR KELİME ÜRETİMİ & KONU MODELLEME SONUÇLARI

AK üretimi deneylerinde ince ayar yapılmış TRT5, MT5, TRBART ve MBART ile LLM'lere istem verme yaklaşımındaki dil modelleri GPT3.5, GPT4 ve Gemini modellerinin ürettiği AK'lerin incelenmesi için AK üretme veri setindeki test kısmında bulunan yorum-AK örnekleri verilmiştir.

Tablo A.5'da karşılaştırılan konu modelleme tekniklerinin sonuçları inceleyebilmek adına ilgili veri setindeki yorumlar ve yorumlara atanan konulara örnek verilmiştir.

Tablo A.1 AK Üretimi ÖRNEK-1

Yorum	**Etiketli AK**
oda 25m2 yazılmıştı ama kesinlikle çok daha küçüktü yanıltıcı olduğunu düşünüyorum o konuda hiç memnun kalmadım onun dışında güzel bir seyahattı ve iyi hizmet aldım	oda boyutu;hizmet
Model	**Üretilen AK**
GPT3.5 & Prompt1	oda; 25m2; küçük; yanıltıcı; olduğunu düşünüyorum; memnun kalmadım; seyahat; iyi hizmet.
GPT4 & Prompt1	oda; 25m2; yanıltıcı; memnun kalmamak; seyahat; iyi hizmet
Gemini & Prompt1	Yanıltıcı ilan;Küçük oda
GPT3.5 & Prompt2	oda boyutu;yetersiz;seyahat;iyi hizmet
GPT4 & Prompt2	oda büyüklüğü;yanıltıcı bilgi; memnuniyetsizlik;iyi hizmet
Gemini & Prompt2	oda büyüklüğü; yanıltıcı reklam; hizmet kalitesi
TRT5	oda;oda;hizmet
MT5	oda;yanıltıcı;hizmet
TRBART	oda ; yanıltıcı olduğunu düşünüyorum; hizmet
MBART	otel boyutu;hizmet

Tablo A.2 AK Üretimi ÖRNEK-2

Yorum	**Etiketli AK**
web sitenizin hızlanması gerekiyor ödeme kısımlarının uçtan uça iyi test edilmesi gerekiyor web sitesi daha kullanışlı ve hızlı olsaydı	hız;ödeme kısmı;test edilme; kullanışlılık
Model	**Üretilen AK**
GPT3.5 & Prompt1	web site, hız, ödeme, test, uçtan uca, kullanışlılık. Keywords: web site; hız; ödeme; test; uçtan uca; kullanışlılık.
GPT4 & Prompt1	web sitesi; hızlanma;ödeme kısımları; uçtan uca test; kullanışlı; hızlı
Gemini & Prompt1	web sitesi hızı;ödeme testi; kullanılabilirlik
GPT3.5 & Prompt2	web site hızı; ödeme testi; kullanışlılık
GPT4 & Prompt2	web sitesi hızı;ödeme testi;kullanışlılık
Gemini & Prompt2	web sitesi hızı;kullanıcı deneyimi; ödeme güvenliği
TRT5	web sitesi;ödeme
MT5	ödeme kısımları;ödeme kısımları;ödeme
TRBART	web sitesi ; ödeme kısımları ; hızlılık
MBART	hız; ödeme kısımları;kullanıcı

Tablo A.3 AK Üretimi ÖRNEK-3

Yorum	**Etiketli AK**
ekstra turlar dahil denmişti sonradan ekstra tur çıktı tabi katılım mecburi değildi rehberimiz ve otobüs şoförümüz çok ilgili insanlardı genel olarak memnun kaldım	ekstra tur;rehber;şoför
Model	**Üretilen AK**
GPT3.5 & Prompt1	ekstra turlar, katılım, mecburi değil, rehber, otobüs şoförü, ilgili, memnun. output: ekstra turlar; katılım; mecburi değil; rehber; otobüs şoförü; ilgili; memnun.
GPT4 & Prompt1	ekstra turlar; katılım; rehber; otobüs şoförü; ilgili insanlar; memnun kalmak
Gemini & Prompt1	ekstra turlar;katılım mecburi değil; rehber ilgi;şoför ilgi;genel memnuniyet
GPT3.5 & Prompt2	ekstra turlar; katılım; rehber; otobüs şoförü; ilgi; memnuniyet
GPT4 & Prompt2	ekstra turlar;rehber;otobüs şoförü; memnuniyet
Gemini & Prompt2	ekstra turlar;rehber;otobüs şoförü; memnuniyet
TRT5	ekstra tur;otobüs şoförü
MT5	ekstra tur;rehber;otobüs şoföri
TRBART	ekstra tur ; tur ; otobüs şoförü
MBART	tur dahil;takip;rehber;otobüs şoför

Tablo A.4 AK Üretimi ÖRNEK-4

Yorum	**Etiketli AK**
otel hakkında yeterli bilgisi yoktu	otel bilgisi
Model	**Üretilen AK**
GPT3.5 & Prompt1	otel; bilgi; yeterli değil.
GPT4 & Prompt1	otel; bilgi
Gemini & Prompt1	otel;bilgi eksikliği
GPT3.5 & Prompt2	otel bilgisi;yetersizlik
GPT4 & Prompt2	otel bilgisi
Gemini & Prompt2	otel bilgisi
TRT5	otel bilgisi
MT5	otel bilgisi
TRBART	otel bilgisi
MBART	otel bilgisi

Tablo A.5 Konu modelleme örneği

Yorum	MT5 AK	GPT4 AK	NMF	LDA	BERTopic	GPT4-OpenAI-PCA-HDBSCAN	GPT4-BERTurk-PCA-GMM	MT5-BERTurk-UMAP-GMM	MT5-BERTurk-PCA-HDBSCAN	MT5-OpenAI-PCA-HDBSCAN
Tur personeliniz ilgisizdi otelde herseyi kendimiz saga sola sorarak ögrendik	tur personeli; otel personeli	Tur personeli;otel; bilgi edinme	2	1	0	[2,0]	[3,2]	[6,7]	[-1,3]	1
Bünyesindeki otelleri seçerken daha dikkatli olsaydi bir yildizi bile hat etmeyen oteli dört yildizli diye satmasaydi.	otel seçimi; yıldizi;otel	otellerin seçimi; dört yıldızlı otel; otel kalitesi	11	13	0	2	2	[7,11,4]	[3,5,0]	[1,0]
Alacarte Restaurant sorunumuzu çözebilseydi ve otele giriste personeli olsaydi yüzyüze görüsme saglanabilseydi	otel personeli; yüzyüze görme	Alacarte Restaurant; personel;yüz yüze görüşme	6	2	0	[-1,2]	[3,0]	[7,11]	[3,5]	1
Tur rehberlerini daha iyi ve özenli tutarak, gezilen yerlerin hacmini daraltarak	tur rehberi	tur rehberi;kalite; gezilecek yerler	2	1	2	[-1,2]	4	6	-1	1
Diger tur firmalari ile birlestirmeden daha profesyonel is ortaklari ile çalissaydi...	tur firmaları; is ortakları	tur firmaları;profesyonel iş ortakları	2	1	0	2	3	6	5	1
sorun yasadigimda cozum uretmiyor	problem;cozum	sorun çözme	16	2	1	2	3	[3,11]	[-1,5]	1
Otele transfer konusunda yardimci olabilirlerdi.	transfer	otel transfer;yardım	16	18	1	[0,2]	[1,0]	2	4	1
Ordu acentesinin tutumu daha iyi olabilir.	acente	Ordu acentesi;tutum	0	0	2	2	4	6	-1	1
Daha iyi bir oda	oda	daha iyi oda	18	0	3	2	2	7	-1	-1

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. A. Er, B. Diri, M. Turhan Yöndem, "LLM Prompting versus Fine-Tuning PLMs: A Comparative Study on Keyword Generation from Customer Feedback", 20th International Conference on Artificial Intelligence Applications and Innovations (AIAI 2024), 27-30 June 2024.