



T.C.

ALANYA ALAADDİN KEYKUBAT ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

İŞLETME ANABİLİM DALI

**TÜRKİYE’DE SAYISAL YÖNTEMLERDE YAZILMIŞ YÜKSEK LİSANS
TEZLERİNİN METİN MADENCİLİĞİ İLE ANALİZ EDİLMESİ**

Yüksek Lisans Tezi

Ömer YILMAZ

**Danışman
Prof. Dr. Yakup AKGÜL**

ALANYA

2026

T.C.
ALANYA ALAADDİN KEYKUBAT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

TÜRKİYE’DE SAYISAL YÖNTEMLERDE YAZILMIŞ YÜKSEK LİSANS
TEZLERİNİN METİN MADENCİLİĞİ İLE ANALİZ EDİLMESİ

Yüksek Lisans Tezi

Ömer YILMAZ

Anabilim Dalı: İşletme Anabilim Dalı

Program Adı: İşletme

Danışman
Prof. Dr. Yakup AKGÜL

ALANYA
2026

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilemeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Alanya Alaaddin Keykubat Üniversitesi tarafından kullanılan “bilimsel intihal tespit programıyla tarandığını ve “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

Ömer YILMAZ

TEŞEKKÜR

Tez çalışmalarım boyunca beni her zaman destekleyen, bilgi birikimi ve tecrübeleri ile bana yol gösteren değerli hocam Prof. Dr. Yakup AKGÜL’e teşekkürü bir borç bilirim. Hayatımın her döneminde bana destek olan anneme, babama ve kardeşime teşekkürlerimi sunarım.



ÖZET

TÜRKİYE’DE SAYISAL YÖNTEMLERDE YAZILMIŞ YÜKSEK LİSANS TEZLERİNİN METİN MADENCİLİĞİ İLE ANALİZ EDİLMESİ

Ömer YILMAZ

İşletme Anabilim Dalı

Alanya Alaaddin Keykubat Üniversitesi, Lisansüstü Eğitim Enstitüsü,

Mart, 2026 (78 sayfa)

Bu tez çalışmasının temel amacı, sayısal yöntemler bilim alanının Türkiye’deki lisansüstü tez konularına yansımalarını ve gelişimini incelemektir. Bu amaçla, 2015-2025 yılları arasında İşletme Anabilim Dalında sayısal yöntemler alanında hazırlanan ve YÖK Ulusal Tez Merkezi web sitesi üzerinden erişilebilen 341 lisansüstü tez inceleme kapsamına alınarak Metin Madenciliği yöntemleri ile analiz edilmiştir. Çalışmada öncelikle metin madenciliği kavramı, uygulama alanları, süreçleri ve metin madenciliği ile ilgili genel bilgiler açıklanmıştır. Sonraki bölümde metin madenciliği yöntemleri kullanılarak yapılan çalışmalara yer verilmiştir. Analizlerde, programın doğası ve bünyesinde bulundurduğu eklentilerin avantajından dolayı verilerin kalitesinin ve sınıflandırmanın doğruluğunun artırılması için Orange Data minig Programı kullanılmıştır. YÖK Ulusal Tez Merkezi web sitesi üzerinden erişilen tezlerin verileri (başlıklar, anahtar kelimeler, özetler gibi) excel tablosuna aktarılıp Orange data mining programı aracılığıyla LDA algoritması kullanılarak konu modellemesi gerçekleştirilmiştir. Tez özetleri ve anahtar kelimelerin Orange veri madenciliği platformu ile analiz edilmesi ve Latent Dirichlet Allocation algoritması ile konu modellemesi uygulanması yoluyla, çalışma literatürün kavramsal yapısını ve tematik dağılımını belirlemiştir. Sonuçlar, tezlerde en sık kullanılan terimlerin analiz, veri, yöntem, model, karar ve sonuç gibi nicel araştırma uygulamalarıyla yakından ilişkili olduğunu ortaya koymuştur. Bu bulgular, bu alandaki lisansüstü araştırmaların deneysel analiz ve analitik modellemeye güçlü bir şekilde vurgu yaptığını göstermektedir. LDA analizi sonucunda, modelleme ve ürün odaklı yaklaşımlar, finansal ve teknolojik analizler, veri analizi ve uygulama çalışmaları, performans ve değer yönetimi, hizmet kalitesi ve müşteri odaklılık, yönetim ve iş süreçleri, karar verme ve problem çözme, değişken tabanlı ilişki çalışmaları, bilgi sistemleri ve analitik, pazar ve ekonomik analiz

dahil olmak üzere on baskın tema ortaya koyulmuştur. Bu temalar, sayısal yöntemler programları kapsamında yürütülen araştırmaların yalnızca metodolojik çalışmalarla sınırlı olmadığını, aynı zamanda işletme yönetiminin çeşitli fonksiyonel alanlarına da uzandığını göstermektedir. Çalışmada elde edilen bir diğer bulgu ise kavramların: stratejik, operasyonel/yönetimsel ve müşteri odaklı olmak üzere üç ana eksen etrafında kümelenmesidir. Bu yapı, işletme ve işletme yönetimi araştırmalarının çok boyutlu doğasını yansıtmaktadır. Yöntemsel açıdan, bu çalışma aynı zamanda büyük akademik belge kümelerinin analizinde metin madenciliği tekniklerinin yararlılığını da göstermektedir.

Anahtar Kelimeler: Metin Madenciliği, Sayısal Yöntemler, Orange Data Mining



ABSTRACT

ANALYSIS OF MASTER THESES WRITTEN IN QUANTITATIVE METHODS IN TURKEY BY USING TEXT MINING

Ömer YILMAZ

Department of Business Administration

Graduate School of Alanya Alaaddin Keykubat University,

March, 2026 (78 pages)

The main objective of this thesis is to examine the reflections and development of the field of quantitative methods in master thesis topics in Turkey. For this purpose, 341 master theses prepared in the field of quantitative methods in the Department of Business Administration between 2015 and 2025, accessible through the YÖK National Thesis Center website, were included in the review and analyzed using Text Mining methods. The study first explains the concept of text mining, its application areas, processes, and general information related to text mining. The following section presents studies conducted using text mining methods. In the analyses, the Orange DataMining Program was used to improve data quality and classification accuracy due to the nature of the program and the advantages of its included plugins. The data from the theses accessed through the YÖK National Thesis Center website (titles, keywords, abstracts, etc.) were transferred to an Excel worksheet, and topic modeling was performed using the LDA algorithm via the Orange DataMining program. By analyzing thesis abstracts and keywords using the Orange data mining platform and applying topic modeling with the Latent Dirichlet Allocation algorithm, the study determined the conceptual structure and thematic distribution of the literature. The results revealed that the most frequently used terms in thesis are closely related to quantitative research applications such as analysis, data, method, model, decision, and outcome. These findings indicate that master research in this field strongly emphasizes experimental analysis and analytical modeling. As a result of the LDA analysis, ten dominant themes emerged, including modeling and product-oriented approaches, financial and technological analyses, data analysis and application studies, performance and value management, service quality and customer orientation, management and business processes, decision-making and problem-solving, variable-based relationship studies, information systems and analytics, and market and

economic analysis. These themes demonstrate that research conducted within quantitative methods programs is not limited solely to methodological studies but also extends to various functional areas of business management. Another finding of the study is the clustering of concepts around three main axes: strategic, operational/managerial, and customer-oriented. This structure reflects the multifaceted nature of business and business management research. Methodologically, this study also demonstrates the usefulness of text mining techniques in the analysis of large academic document sets.

Keywords: Text Mining, Quantitative Methods, Orange Data Mining



İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI	i
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	ii
TEŞEKKÜR SAYFASI.....	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER SAYFASI	viii
ŞEKİLLER LİSTESİ	x
TABLolar LİSTESİ.....	xi
.....	
GİRİŞ	1
1. METİN MADENCİLİĞİ	6
2. METİN MADENCİLİĞİNE YÖNELİK MEVCUT DERLEME ÇALIŞMALARI....	8
3. YÖNTEM	38
3.1. Veri Toplama ve Arama Stratejisi.....	38
3.2. Durak Kelimeler (Stopwords)	39
3.3. Konu Modellemesi	40
3.4 Orange3 Text Mining: Özellikler ve Uygulama Alanları	42
4. ANALİZLER VE BULGULAR.....	44
4.1. Kelime Sıklığı	44
4.1.1 Kelime Bulutu Bulgularının Tartışılması ve Sonraki Analizlerle Entegrasyonu	45
4.1.2 Yöntemsel Katkı Açısından Değerlendirme	46
4.2. Hiyerarşik Kümeleme Analizi Bulguları	47
4.3 Küme Bazlı Tematik Yapı	50
Küme 2 (C2): Ürün, Yenilik ve Değer Yaratma:	50
Küme 3 (C3): Uygulama, Süreç ve Yetkinlik Temelli Yaklaşım:	50
Küme 4 (C4): İş Süreçleri ve Karar Verme Mekanizmaları:	50
Küme 5 (C5): Müşteri Deneyimi ve Algılanan Değer:	50
4.4. Konu Modellemesi (Topic Modeling) Analizi Bulguları.....	51
4.4.1 Konu Bazlı Tematik Yapı.....	52
Konu 1: Modelleme ve ürün odaklı yaklaşımlar	52
Konu 2: Finansal ve teknolojik analizler	52
Konu 3: Veri analizi ve uygulama çalışmaları	52
Konu 4: Performans ve değer yönetimi	52
Konu 5: Hizmet kalitesi ve müşteri odaklılık.....	52

Konu 6: Yönetim ve iş süreçleri.....	52
Konu 7: Karar verme ve problem çözme	53
Konu 8: Değişkenler ve müşteri ilişkileri.....	53
Konu 9: Bilgi sistemleri ve analitik.....	53
Konu 10: Pazar, verimlilik ve ekonomik analiz	53
4.5. Konu Modellemesi ve Hiyerarşik Kümeleme Bulgularının Entegrasyonu.....	53
4.5.1 Konu Modellemesi, Hiyerarşik Kümeleme ve MDS Bulgularının Bütünleşik Yorumu	55
4.6. Concordance (Bağlam Analizi) Bulgularının Tartışılması	57
4.6.1 Yöntemsel Bütünlük ve Katkı.....	58
4.7. t-SNE Tabanlı Konu Modellemesi ve Hiyerarşik Kümeleme Bulgularının Tartışılması.....	59
4.8. Heat Map Bulgularının Yorumu (Hiyerarşik Kümeleme Entegrasyonu)	61
1. Konu Temelli Dikey Kümelendirme (Sütunlar)	61
2. Doküman Temelli Yatay Kümelendirme (Satırlar)	62
3. Yoğunluk Desenleri ve Tematik Baskınlık	62
4. Konu Modelleme ve t-SNE ile Tutarlılık	62
5. TARTIŞMA	64
6. SONUÇ.....	66
7. KISITLAR VE ÖNERİLER	69
8. KAYNAKÇA.....	70
ÖZGEÇMİŞ	79

ŞEKİLLER LİSTESİ

Şekil 3.1. Detaylı arama	39
Şekil 3.2. Detaylı arama	39
Şekil 3.3. Orange model	42
Şekil 4.1. Kelime Bulutu	45
Şekil 4.2 Kelime bulutu- Abstractlarda Kelime Sıklığı.....	47
Şekil 4.3. Kelime Kümeleme Görselleştirmesi	48
Şekil 4.4. Orange yazılımında oluşturulan LDA konu modelleme çıktısı	51
Şekil 4.5. LDA ile oluşturulan 10 konunun MDS görselleştirmesi	55
Şekil 4.6. Quantitative” teriminin konkordans satırlarının analizine örnek.....	58
Şekil 4.7. t-SNE sonuçları (topic).....	61
Şekil 4.8. Heat map.....	61

TABLÖLAR LİSTESİ

Tablo 4.1. Analiz için kullanılan bileşenlerin detayları.....	49
--	----



GİRİŞ

Her gün farklı sorunları çözmek veya güncel bilgi sağlamak amacıyla farklı dergilerde ve konferanslarda birçok araştırma makalesi yayınlanmaktadır. Bu araştırma makaleleri, mevcut ve gelecekteki araştırmalar ve yeni bilgi üretimi için çok değerlidir. Bu yayınların çokluğu nedeniyle, belirli bir alandaki bilgiyi araştırmak amacıyla çok sayıda makaleyi manuel olarak okuma süreci çok zaman alıcı ve zordur. Bilgi patlaması çağında, birçok web platformu ve sosyal medya platformu her saniye milyarlarca bilgi üretmektedir ve bu 'Bilgi Okyanusundan' doğruluğu kesin olan bir damla bilgiyi çıkarmak, bilgi profesyonelleri için bile adeta doğa üstü güçlere sahip olmayı gerektirir (Das vd., 2020). Birey, toplum, siyaset ve özel sektör aktörlerine rehberlik edecek kesin ve doğru bilgiye erişim için yapılacak olan araştırmalar sosyal yapıların düzensizliği, kararsızlığı ve de tahmin edilemez oluşundan dolayı çeşitli metodolojik engellerle karşılaşmaktadır. Araştırmacılar, zamanla bu engelleri kolay ve hızlı şekilde aşmak için çeşitli yöntemler geliştirmişlerdir.

Dijitalleşmenin de etkisiyle metinsel verilerin çok hızlı üretilip yaygınlaşması, araştırmacıları büyük çaplı veri yığınlarını sistematik olarak analiz edebilen yöntemsel yaklaşımlar aramaya yöneltmiştir. Bu bağlamda, Nicel yöntemler kullanılarak yapılan incelemelerde, yapılandırılmış verilerin istatistiksel analizi ve objektif değerlendirme prensipleri temel alındığından sonuçların bilimsel geçerlilik ve güveni güçlendirdiği düşünülmektedir. Nicel yöntemler, araştırma bulgularının güvenilirliğini ve geçerliliğini artıran yapılandırılmış veri toplama, istatistiksel çıkarım ve deneysel genellemeyi vurgulamaktadır (Creswell ve Creswell, 2018). Nicel yöntemlerin avantajlarından biri de, büyük örneklemeleri analiz etme, istatistiksel analizler aracılığıyla ilişkileri test etme ve daha geniş kitlelere genelleştirilebilecek bulgular üretme becerisidir (Bryman, 2016). Nicel yöntemler kullanarak araştırmacılar, standartlaştırılmış prosedürler ve sayısal ölçümler kullanarak öznelliği ve önyargıyı azaltarak bulgularının güvenilirliğini ve doğruluğunu artırabilirler (Zyoud vd., 2024). Sonuç olarak, araştırma bulguları, çalışmanın tekrarlanmasına ve bağımsız araştırmacılar tarafından doğrulanmasına olanak tanıyan doğruluk nedeniyle daha güvenilirdir. Böylelikle araştırmacılar, büyük ölçekli veri kümelerinden objektif sonuçlara ulaşarak konuları kıyaslayabilir ve literatürdeki eğilimleri daha detaylı ölçüm yöntemleriyle belirleyebilirler. Metin madenciliği de, araştırmacıların çok miktarda lüzumsuz bilgi arasından yararlı bilgileri hızla keşfetmesini

ve çeşitli bilgiler arasındaki bağlantıları ortaya çıkarmasını sağlayan sayısal bir yöntem olarak öne çıkmaktadır (Song vd., 2022).

Mevcut elektronik metinsel bilgi miktarı sürekli artmaktadır (Dang ve Ahmad, 2015). Bilimsel literatür dahi gelişmekte ve gelecek nesiller için dijital formatlarda arşivlenmektedir (Nie ve Sun, 2017). Son yıllarda, akademik tezler, dergi makaleleri, kurumsal raporlar ve çevrimiçi yayınlar dahil olmak üzere dijital metinsel verilerin hızla yaygınlaşması, disiplinler arası araştırma manzarasını kökten dönüştürmüştür. Bu bağlamda yürütülen çalışmalar, bir dizi belgeden anlamlı bilgileri çıkarmak, analiz etmek ve bulmak için kullanılmaktadır (Sumathy ve Chidambaram, 2013). Metinsel bilginin artan ölçeği ve heterojenliği, geleneksel manuel içerik analizini kapsamlı, sistematik ve tekrarlanabilir sonuçlar üretmek için yetersiz hale getirmiştir. Bu zorluklara bir yanıt olarak, veri madenciliğinin bir alt dalı olarak da bilinen metin madenciliği (genellikle hesaplamalı metin analizi veya metin analitiği olarak birbirinin yerine kullanılır), araştırmacıların yapılandırılmamış metinden anlamlı örüntüleri, gizli konuları, kavramsal yapıları ve zamansal eğilimleri otomatik olarak çıkarmalarını sağlayan güçlü bir metodolojik çerçeve olarak ortaya çıkmıştır (Macanović ve Przepiorka, 2024; Farrell vd., 2024). Metin madenciliği teknikleri, geniş bir metin külliyatından gizli bilgileri keşfetmek için istatistiksel analiz uygulamaları ve bu da örüntüleri, eğilimleri, veri tekrarlarının bulunmasını ve yapılandırılmamış metnin kavramsal anlayışının geliştirilmesini sağlar (Thakur ve Kumar, 2022).

Metinsel verilerde saklı muazzam değeri ortaya çıkarmak için güçlü bir yöntem olarak ortaya çıkan metin madenciliği (metin analitiği veya doğal dil işleme (NLP) olarak da bilinir), çok sayıda kaynaktan veri toplamak, analiz etmek ve çözümlemek için çeşitli teknik ve prosedürleri kapsamaktadır. Metin madenciliğinin bazı ek uygulamaları arasında metin özetleme, bilgi alma, bilgi kümeleme, metin kategorizasyonu, bilgi çıkarma, dil tanımlama, yazarlık ilişkisi tanımlama, ifade yapılarını belirleme ve anahtar ifadeleri tespit etme, moleküller, bileşikler, gen dizileri, adlar, tarihler, kısaltmalar gibi "varlıkları" çıkarma ve kısaltmaları bulma yer almaktadır (Witten, Don, Dewsnip ve Tablan, 2004). Doğal dil işleme (NLP), makine öğrenimi ve istatistiksel modelleme tekniklerini uygulayarak, metin madenciliği büyük belge koleksiyonlarını analiz edilebilir sayısal temsillere dönüştürerek büyük ölçekli ve veri odaklı ampirik araştırmaları kolaylaştırır (Aleqbqie, Sfoq ve Albeer, 2024).

Metin madenciliğinin avantajları ölçeklenebilirliğin çok ötesine uzanır. Metin madenciliği, farklı alanlarda araştırma eğilimlerini izlemek, kavramsal evrimi haritalamak ve büyük ölçekli literatür külliyatlarını analiz etmek için yaygın olarak kullanılmaktadır. Metin madenciliği, iş, sağlık, sosyal medya analizi ve bilimsel çalışma dahil olmak üzere çeşitli alanlarda kullanılmaktadır. İşletmeler yapılandırılmamış metinlerde saklı gizli bilgilerden yararlanmaya çalışırken, metin madenciliği, eyleme geçirilebilir içgörüler ortaya çıkarma ve veri odaklı karar almayı geliştirme kapasitesiyle ilgi odağı haline gelmiştir. Sosyal medya araştırmalarındaki bibliyometrik çalışmalar, metin madenciliğinin araştırma yörüngelerini şekillendiren uzunlamasına örüntüleri ve tematik kümeleri ortaya çıkarabileceğini göstermektedir (Sandu, Cotfas, Stănescu ve Delcea, 2024). Benzer şekilde, sağlık bilimlerindeki geniş ölçekli analizler, otomatik metin analizinin akademik topluluklar içindeki metodolojik tercihlerdeki veya kavramsal odaklardaki değişimleri nasıl izleyebileceğini göstermektedir ki Oner, Hakli ve Zengul (2023) tarafından yapılan bir çalışmada, alanın on yıllar boyunca değişen önceliklerini ortaya çıkarmak için 15.000’den fazla hemşirelik araştırma makalesi incelenmiştir. Diğer sistematik incelemeler, NLP destekli metin madenciliğinin, ilgili literatürün otomatik olarak alınmasını, taranmasını ve sentezlenmesini sağlayarak sistematik inceleme iş akışlarını geliştirdiğini doğrulamaktadır (Sundaram ve Berleant, 2022). Smith (2020) yaptığı çalışmada, kamuoyunu anlamak ve piyasa trendlerini tahmin etmek için sosyal medya verilerini analiz etmede gelişmiş metin madenciliği tekniklerinin faydasını ortaya koyarken, Zeng vd. (2019)’nin çalışması, metin madenciliğinin sağlık hizmetlerinde klinik karar desteği ve biyomedikal araştırmaları geliştirmek için nasıl kullanılabileceğini göstermiştir. Bu referanslar, metin madenciliğinin artan önemini ve etkisini vurgulayarak, bu dinamik ve hızla gelişen alanda daha fazla keşif ve inovasyon için zemin hazırlamaktadır. Huang ve Liu (2020), borsa trendlerini tahmin etmek için sosyal medya verilerini analiz etmede metin madenciliği algoritmalarının kullanımını incelemiştir. Sonuçlar, duygu analizinin diğer metinsel özelliklerle birleştirilmesinin borsa hareketleri hakkında değerli bilgiler sağlayabileceğini göstermiştir. Zhao, Xu ve Wang (2019) tarafından yapılan bir başka çalışma, ürün öneri sistemleri için çevrimiçi incelemeleri analiz etmek amacıyla metin madenciliği tekniklerinin uygulanmasına odaklanmıştır. Araştırma, ürün önerilerinin doğruluğunu artırmada duygu analizi ve konu modellemesinin önemini vurgulamıştır.

Son metodolojik alıřmalar, hesaplamalı yaklařımların manuel kodlamada bulunan znel nyargıyı azalttıđını, tekrarlanabilirliđi artırdıđını ve geleneksel nitel prosedrlerle imkansız olacak řekilde analizi hızlandırdıđını vurgulamaktadır (Macanović ve Przepiorka, 2024). Otomatik sınıflandırma, kümeleme, konu modelleme ve duygu analizi, arařtırmacıların binlerce belgedeki temel yapıları tespit ederek, manuel inceleme sırasında gizli kalabilecek tematik dađılımları veya kavramsal iliřkileri ortaya ıkarmalarına olanak tanır. Dahası, karřılařtırmalı deđerlendirmeler, modern metin madenciliđi araçlarının dođru řekilde yapılandırıldıđında, zellikle zetler, sosyal medya gönderileri veya çevrimiçi yorumlar gibi kısa metinlerle alıřırken eřitli deđerlerde insan kodlama dođruluđuna yaklařabildiđini göstermektedir (Farrell vd., 2024). Metin madenciliđi akademisyenler arasında giderek daha fazla ilgi görmekte ve uygulaması birok alıřma alanında katlanarak artmaktadır.

Bu metodolojik olgunluk ve verimlilik göz önüne alındıđında, lisansüstü arařtırmalarda da metin madenciliđinin uygulanması yalnızca uygun olmakla kalmayıp, zellikle de büyük miktarda akademik literatürün mevcut olduđu ancak kaynak kısıtlamaları nedeniyle yeterince analiz edilmediđi ulusal bađamlarda giderek daha da önemli hale gelmektedir. Bu alıřmada metin madenciliđi yöntemleri kullanılarak son 10 yılda (2015-2025) İřletme alanında sayısal yöntemler kullanılarak yazılan lisansüstü tezlerde hangi konuların daha ok iřlendiđini bulmak amaçlanmıřtır. Ayrıca alıřma birkaç temel sorunun da yanıtını arařtırmaktadır:

-Türkiye’de Sayısal Yöntemler üzerine yapılan lisansüstü tezlerde ele alınan baskın konular nelerdir?

-Metin madenciliđi sürecinde Orange programından nasıl yararlanılabilir?

-Metin madenciliđinde kullanılan LDA algoritması ile oluřturulan konu modellemesi sonuçları nasıl yorumlanabilir?

Veri kaynađı olarak bu tezlerin ingilizce zet kısımları ve anahtar kelimeleri kullanılmıřtır. Analiz için konu modelleme yönteminden faydalanılmıřtır. Konu modelleme, makine đreniminin, metinlerden gizli konuları keřfetmek için tasarlanmıř önemli bir dalını oluřturur. Konu Modellemesi, nicel ve nitel metodolojinin bir birleřimidir ve sosyal bilimler metodolojisinin tüm yelpazesine dayanır. Bu alıřmada konu modelleme iřlemi Orange programı ve gerekli kütüphanelerin eklenmesi ile yapılmıřtır.

Konu modellemesi için farklı algoritmalar mevcuttur: örneğin, Gizli Dirichlet Tahsisi (LDA), Gizli Anlamsal İndeksleme (LSI) ve Hiyerarşik Dirichlet Süreci (HDP). Bu çalışma, ilk olarak 2003 yılında David Blei, Andrew Ng ve Michael Jordan tarafından tanımlanan Gizli Dirichlet Tahsisi'ni kullanmaktadır. LDA, esas olarak doğal dil işleme alanında kullanılan olasılıksal bir modeldir. Uzun bir metnin konusunu hızlı bir şekilde belirlemeye yardımcı olur. LDA, metinlerdeki konular hakkında, birlikte geçen kelimelerin sıklığına dayanarak tahminlerde bulunur.

Çalışma kapsamında yoktez.gov.tr üzerinden yayınlanmış olan yüksek lisans tezlerinden işletme bölümü sayısal yöntemler anabilim dalı temalı yapılan taramalar sonucunda elde edilen tezler belirlenmiş ve yapılan analizler sonucunda en çok tekrarlanan kelimeler ilgili yıllarda yazılan tezlerde o konunun daha çok işlendiğini göstermektedir. Yapılan bu çalışmanın sayısal yöntemler alanında yapılacak çalışmalara katkı sağlayacağı düşünülmektedir.

1. METİN MADENCİLİĞİ

Metin madenciliği, çeşitli kaynaklardan metin çıkarma, bunu yapılandırılmış bilgiye dönüştürme, aralarındaki ilişkileri belirleme ve ardından bu ilişkileri analiz ederek örüntüler bulma, sorunları çözme ve aynı önermelerin faydalı uygulamalarını oluşturma sürecidir. Metin madenciliği (metin analitiği olarak da bilinir), doğal dil işleme (NLP) kullanarak, belgeler ve veritabanları içindeki yapılandırılmamış metni, analiz için veya makine öğrenme algoritmalarını eğitmek için kullanılacak normalleştirilmiş, yapılandırılmış verilere dönüştürür (Tyagi, 2021). Doğal dil işleme (NLP), insan iletişimi ile bilgisayar anlayışı arasındaki boşluğu doldurmak için geliştirilen bilgisayar bilimi ve hesaplamalı dilbilim de dahil olmak üzere birçok disiplini bünyesinde barındırır (Mnasari, 2019). Metin analizi, mevcut olan daha büyük miktardaki metinsel verinin içinden bilgi çıkarma süreci olarak tanımlanır. Metin madenciliği ile doğal dil işleme arasında yakın bir ilişki vardır. Doğal Dil İşleme (NLP), teknikleri dilbilim esaslı yöntemler kullanır. Temel olarak iki tür NLP tekniği vardır: Sözdizimsel analiz ve anlamsal analiz. Sözdizimsel analiz veya ayrıştırma, bir cümle yapısını tespit etmek, kelimeleri düzenlemek ve bunların nasıl bağlandığını anlamak için temel dil bilgisi ilkelerini kullanarak metni inceler. Anlamsal analiz, metin yakalamaya dayanır ve her terimin önemini (sözcüksel anlambilim) inceler. Son olarak bulgular anlamsal olarak yorumlanarak metinden bilgi çıkarmak amaçlanır (Kao & Poteet, 2007). Metin madenciliği yaklaşımı, değerli bilgiler elde etmek için büyük miktarda metin verisini incelemek için kullanılır. Farklı türden problemleri çözmek için birçok metin madenciliği tekniği vardır. Metinsel verileri sayısal verilere dönüştürmek için kullanılan bir takım ön işleme teknikleri vardır:

Tokenleştirme, materyalin işlenmesini kolaylaştırmak için metni token adı verilen daha küçük parçalara (ifadeler veya kelimeler olabilir) bölmeyi içerir.

Kök bulma(Stemming), analizi kolaylaştırmak için çekimli ifadeleri yapım ve çekim eklerinden arındırarak temel biçimlerine indirgemekten oluşur.

Kelimenin görevini belirleme(part-of-speech tagging): Kelimelerin anlamını belirlemeye yardımcı olur (örneğin, "yüz" terimi, fiil (yüzmek) veya sıfat (100)).

Ağırlıklandırma: Madencilik görevlerinde terim katkısını en aza indirmek veya en üst düzeye çıkarmak için kullanılır. Örneğin, iyi bilinen bir teknik TF-IDF'dir (Terim Sıklığı-Ters Belge Sıklığı)

Durdurma kelimesi(stopwords) eleme: Bazı terimler madencilik görevleriyle ilgili değildir ve bu adımda elenebilir. Genellikle bu terimler, herhangi bir belgede yaygın olarak görünen kelimelerdir (artikeller, yardımcı fiiller ve edatlar)

Metinsel veriler sayısal veriler haline getirildikten sonraki aşama sınıflandırma ve metin madenciliği tekniklerinin uygulanmasıdır (Cecchini vd., 2010 ; Glancy & Yadav, 2011).



2. METİN MADENCİLİĞİNE YÖNELİK MEVCUT DERLEME ÇALIŞMALARI

Web of Science ve Scopus gibi veri tabanlarında yapılan taramalarda özellikle son yıllarda araştırmacılar ve uygulayıcıların metin madenciliği, veri madenciliği, doğal dil işleme ve duygu analizi yöntemlerini kullanarak birçok çalışma gerçekleştirdikleri görülmüştür. Örneğin;

Çelikbağ (2025) bu çalışmada, 1997-2023 yıllarını kapsayan kapsamlı bir Scopus tabanlı bibliyometrik analiz yoluyla yükseköğretimde öğrenme analitiğinin evrimini, metodolojik eğilimlerini ve araştırma odağını incelemektedir. Çalışmanın amacı, öğrenme analitiği araştırmalarındaki kavramsal yapıyı, etkili yazarları, baskın yayın yerlerini, tematik kümeleri ve kronolojik değişimleri haritalamaktır. Nicel bir bibliyometrik metodoloji kullanan çalışma, ana araştırma temalarının merkeziliğini, yoğunluğunu ve gelişimsel olgunluğunu belirlemek için eş atıf analizi, anahtar kelime eş zamanlı eşlemesi, yayın eğilimi analizi ve stratejik diyagramlar kullanmaktadır. Veri seti, öğrenme analitiği ile ilgili Scopus endeksli yayınların çıkarılması ve eğitim veri madenciliği ve yükseköğretim bağlamlarıyla ilişkisini sağlamak için filtrelenmesiyle oluşturulmuştur. Bulgular, öğrenci performansı tahmini, öğrenme yönetim sistemi veri analizi, kişiselleştirilmiş öğrenme ve öz-düzenlemeli öğrenmeye güçlü bir tematik vurgu ile 2015'ten sonra öğrenme analitiği araştırmalarında hızlı bir genişleme olduğunu göstermektedir. Sonuçlar ayrıca, Amerika Birleşik Devletleri, Çin ve Avrupa ülkelerinin yayın çıktısı ve etkisi açısından alanda lider olduğunu göstermektedir. Çalışma, öğrenme analitiğinin gelişmekte olan bir alandan olgun bir araştırma alanına dönüştüğü sonucuna varmıştır, ancak yine de daha derin teorik temellere ve makine öğrenimi yöntemlerinin pedagojik çerçevelerle daha güçlü bir şekilde bütünleştirilmesine ihtiyaç duyulmaktadır.

Musunuru (2025) bu çalışmada, bilgi sistemlerinin Sosyo-Çevresel Denetim Uygulamaları (SAP) üzerindeki etkisini değerlendirmeyi amaçlamaktadır. Bu nedenle, işletme faaliyetlerinin toplum ve çevre üzerindeki etkisinin denetlenebileceği ve bu tür denetim uygulamalarının verimliliğinin bilgi sistemlerine bağlı olduğu varsayılmaktadır. Veri kümeleri, metin madenciliği metodolojisi kullanılarak Scopus ve Web of Science'tan (WoS) çıkarılmış ve zorlukları ve teknolojileri belirlemek için Yazışma Analizi (CA); Altta yatan faktör yapılarını belirlemek için Keşifsel Faktör Analizi (EFA); teknolojiler ve zorluklar arasındaki neden-sonuç ilişkilerini değerlendirmek için Doğrulayıcı Faktör Analizi (CFA) yöntemleri kullanılarak analiz edilmiştir. Büyük veri,

blok zinciri, bulut bilişim çözümleri ve makine öğrenimi gibi birkaç yeni teknolojinin SAP üzerinde hem olumlu hem de olumsuz bir etkisi olduğu görülmektedir, ancak bunlar çoğunlukla önemsizdir. Şeffaflık, hesap verebilirlik ve sürdürülebilirlik gibi zorlukların SAP ile ilişkili olduğu bulunmuştur. Blockchain teknolojileri sosyal ve çevresel faktörleri olumlu yönde etkilerken, makine öğrenmesi EFA'yı olumsuz yönde etkilerken CFA üzerindeki etkisi olumlu ve anlamlıdır.

Eltaiba vd., (2025), bu sistematik çalışmada, Orta Doğu ve Kuzey Afrika (MENA) genelinde yükseköğretimde teknoloji destekli öğrenmenin (TEL) kapsamını, faydalarını ve zorluklarını araştırmaktadır. Yazarlar, 2015 ve 2024 yılları arasında yayınlanmış çalışmaları kapsayan yapılandırılmış bir arama stratejisi kullanarak Scopus ve Web of Science'tan veri toplamıştır. PRISMA seçim prosedürleri uygulandıktan sonra 85 ampirik çalışma dâhil edilmiştir. Veri analizinde, trend analizi, coğrafi dağılım ve MENA'da TEL'in faydalarının, zorluklarının ve uygulamalarının VOSviewer, Tableau ve R istatistik yazılımlarında Kavram Haritalama ve Metin Madenciliği gibi yöntemlerle tematik sentezi kullanılmıştır. Sonuçlar, TEL'in öğrenme erişilebilirliğini, katılımı ve dijital beceri gelişimini iyileştirdiğini; ancak dijital uçurum, sınırlı altyapı, tutarsız politika çerçeveleri ve yetersiz öğretim üyesi eğitimi nedeniyle benimsenmesinin bölge genelinde eşit olmadığını göstermektedir. Daha zengin, ileri teknoloji ülkeleri TEL entegrasyonuna öncülük ederken, düşük gelirli bölgeler sistemik engellerle karşı karşıyadır. İnceleme, TEL'in MENA yükseköğretimi için dönüştürücü bir potansiyele sahip olmasına rağmen, sürdürülebilir ve adil bir uygulamanın hedefli politika desteği, mesleki gelişim ve altyapı yatırımı gerektirdiği sonucuna varmıştır.

Okoye vd., (2025) bu çalışma ile eğitimde dijital teknolojilerin kullanımının pedagojik etkisini ve fırsatlarını belirlemeyi amaçlamaktadırlar. Yazarlar dijitalleştirilmiş eğitimin kullanımı ve benimsenmesi önündeki potansiyel engelleri, zorlukları ve fırsatları belirleyerek mevcut literatürü araştırmak için PRISMA metodolojisini uygulamış ve pandemi öncesi, pandemi içi ve pandemi sonrası dönemi ve kırsal eğitim gelişimini dikkate almışlardır. Yazarlar, PRISMA metodolojisini kullanarak Scopus ve Web of Science'tan 548 makale belirlemiş, bunları sistematik dahil etme ve hariç tutma kriterleriyle tarayıp ve 2017-2023 yılları arasında yayınlanmış 324 çalışmadan oluşan nihai veri setini analiz etmişlerdir. Veri toplama süreci, dijital teknoloji, COVID-19, kırsal eğitim ve sürdürülebilir kalkınma ile ilgili anahtar kelimeler kullanılarak veritabanı aramalarını, ardından tam metin taramasını ve tematik sentezi içermektedir. Veri analizi,

trend incelemesini, veri setinin coğrafi ve bağlamsal çeşitliliğini ve COVID-19 ve kırsal eğitim bağlamında dijital teknolojilerin faydalarının, zorluklarının ve uygulanmasının tematik sentezini içermektedir. Bu işlem, VOSviewer, Tableau ve R Statistics yazılımları gibi veri görselleştirme ve analiz araçlarından yararlanılarak Kavram Haritalama ve Metin Madenciliği gibi yöntemler kullanılarak yapılmıştır. Çalışma, dijital (öğrenme) teknolojilerine ve çevrimiçi eğitim platformlarına veya kaynaklarına erişim eksikliği veya sınırlılığına atfedilen "dijital uçurum"u, özellikle uzak veya kaynakları kısıtlı bölgelerdeki eğitimciler tarafından dijitalleştirilmiş eğitimin etkili bir şekilde benimsenmesi ve kaliteli eğitimin sağlanması ve sunulması önündeki temel zorluk olarak belirlemiştir. Çalışma, temel bulguların teorik ve pratik çıkarımlarına ışık tutmakta ve özellikle eğitim toplulukları ve/veya kalkınma odaklı bir eğitim toplumu genelindeki eşitsizliklerin azaltılması da dahil olmak üzere eğitim kalitesiyle doğrudan ilişkili olan BM'nin sürdürülebilir kalkınma hedefleri (SKH) doğrultusunda ölçeklenebilir ve sürdürülebilir bir eğitim ve uygulamalara ışık tutmaktadır.

Sormanen, Hietajärvi ve Wilska (2025) tarafından yapılan çalışma, gençlerin açıkça politik olmayan bir çevrimiçi gençlik forumunda sivil ve politik tartışmalara nasıl katıldıklarını incelemektedir. Amaç, bu tür dijital alanlarda politik konuların ortaya çıkıp çıkmadığını, hangi temaların baskın olduğunu ve gençlerin politik katılımını hangi ifade kalıplarının karakterize ettiğini belirlemektir. Çalışma verileri, Finlandiyalı bir gençlik basılı dergisi olan Demi tarafından barındırılan ve yönetilen, çoğunlukla 13-19 yaş arası kadın katılımcıları hedefleyen tartışma forumu Demi.fi'den (aktif 1998-2021) alınmıştır. Yazarlar, geniş bir forum gönderisi koleksiyonu toplamış ve politik konuları sınıflandırmak, anlamsal ağları belirlemek ve çevrimiçi ifade biçimlerini kategorize etmek için otomatik metin madenciliği tekniklerini manuel içerik analiziyle birleştirmiştir. Sonuçlar, politik içeriğin toplam tartışmaların mütevazı bir kısmını oluşturmasına rağmen, genç kullanıcıların geleneksel politik temaların yanı sıra küresel olarak öne çıkan konularla da önemli ölçüde etkileşimde bulunduğunu göstermektedir. Anlamsal ağ analizi, gençlik politik söyleminde yeni yapılar ortaya koyarken, kaynak talepleri ve iletişim yoğunluğu bakımından farklılık gösteren altı farklı politik ifade biçimi, çeşitli etkileşim biçimlerini vurgulamaktadır. Çalışma, apolitik çevrimiçi ortamların yine de gençlerin sivil ifadesi için anlamlı alanlar olarak işlev gördüğü, çevrimiçi politik katılım modellerine teorik uzantılar sunduğu ve akran temelli dijital iletişimin siyasi sosyalleşmedeki önemini vurguladığı sonucuna varmıştır.

Keijzer ve van Lange (2025) tarafından yapılan çalışma, 1935 ve 1950 yılları arasında yazılmış Hollanda savaş mektuplarının büyük ölçekli sayısallaştırılması ve verileştirilmesine odaklanarak, tarihsel olarak önemli ancak erişilemeyen el yazısı bir külliyatı dijital beşeri bilimler araştırmaları için uygun makine tarafından okunabilir verilere dönüştürmeyi amaçlamaktadır. Projenin amacı, tarihsel, dilbilimsel ve hesaplamalı analizi kolaylaştırmak için yüksek çözünürlüklü taramaları, yarı otomatik transkripsiyonları ve yapılandırılmış meta verileri birleştiren kapsamlı bir veri kümesi üretmektir. Bunu başarmak için yazarlar çok aşamalı bir veri toplama ve işleme iş akışı kullanmışlardır: 160.000’den fazla arşiv belgesi profesyonel olarak taranarak 1.087 sayfalık bir Zemin Gerçeği alt kümesi kitle kaynak kullanımı yoluyla manuel olarak transkribe edildikten sonra El Yazısı Metin Tanıma (HTR) modelleri Transkribus’ta eğitilm ve yinelemeli olarak iyileştirilmiştir. Ve bağlamı, tarihleri ve muhabirleri tanımlayan meta veriler sistematik olarak CSV formatına derlenmiştir. Karakter Hata Oranı (CER) hesaplaması ve yinelemeli model yeniden eğitimi gibi değerlendirme teknikleri, transkripsiyon doğruluğunu ve sağlamlığını sağlamıştır. Bulgular, ortaya çıkan veri setinin ölçeklenebilir metin madenciliğini mümkün kıldığını, makine öğrenimi modellerinin geliştirilmesini ve kıyaslanmasını desteklediğini ve sosyal ağlar, dil çeşitliliği, savaş zamanı iletişimi ve arşiv kaynakları üzerine disiplinler arası çalışmalar için yeni yollar açtığını göstermektedir. Yazarlar, seçkin olmayan savaş zamanı yazışmalarının dijitalleştirilmesinin ve yapılandırılmasının yalnızca akademik erişimi artırmakla kalmayıp aynı zamanda kişisel verilerle ilgili etik hususları da gündeme getirdiği ve nihayetinde hem akademik araştırmayı hem de tarihsel materyallerle kamuoyunun etkileşimini güçlendiren değerli bir kaynak sağladığı sonucuna varmaktadır.

Weissenfels vd., (2025) yaptıkları çalışmada metin madenciliği yaklaşımını kullanarak, dijital sağlık teknolojileri genelindeki sağlık konularını bir bütün halinde incelemek için önde gelen bilgi teknolojileri dergilerinden 484 makaleyi analiz etmişlerdir. Çalışma, temel sağlık konularını ve eğilimlerini ortaya çıkarmakta, bunların zaman içindeki evrimini izlemekte ve güncel bilgi teknolojileri literatüründeki bilgi boşluklarını belirlemektedir. Bulgularda, bağlam duyarlı sağlık hizmetlerinde yeni teknolojilerin ve ek paydaşların önemi vurgulanmaktadır. Çalışmanın veri toplama (dergi seçimi gibi), seçilen metodoloji ve öznel yargılar ya da yorumlamadan kaynaklanan konu etiketlemesindeki olası önyargılar gibi sınırlamaları olsa da disiplinler arası iş birliğini teşvik ederek, yeterince keşfedilmemiş alanları ele almanın ve yeni teknolojilerin

entegrasyonunu artırmanın önemini vurguladığı düşünülmektedir. Bu yaklaşımın, yalnızca Bilgi Teknolojileri (BİT) alanını ilerletmekle kalmayacağı, aynı zamanda daha geniş sağlıkla ilgili BİT araştırmalarıyla uyumunu da destekleyerek gelecekteki çalışma yönleri için bir temel oluşturacağı düşünülmektedir.

Zhao vd., (2026), yaptıkları çalışmada literatürden çok sayıda anahtar terimleri ve konulara ilişkin içgörülerini çıkaran Gizli Dirichlet Tahsisi (LDA) konu modellemesini kullanarak veri odaklı içgörüler aracılığıyla dönüştürücü dayanıklılığı tanımlamayı ve konularını ve alanlarını haritalayarak ilgili araştırmalardaki eğilimleri keşfetmeyi amaçlamaktadır. Atıflar arasındaki bağlantılara dayanarak 1951'den 2024'e kadar araştırma konularının gelişimini incelemişlerdir. Elde edilen veriler, dergi makalelerinin, kitap bölümlerinin ve konferans bildirilerinin başlıklarını, özetlerini ve anahtar kelimelerini kapsayan "dönüştürücü dayanıklılık" sorgu terimi kullanılarak Scopus veri tabanından toplamda 3110 ilgili literatür parçası temin edilmiştir. Dönüştürücü dayanıklılık, sosyo-politik değerler ve politikalarda bir değişime eşlik eden sistem yapısındaki bir değişiklik olarak tanımlanmıştır. LDA'nın sonuçları, dönüştürücü dayanıklılık araştırmalarının yedi alanda geliştiğini göstermiştir: İnovasyon ve Ekonomik Politika, Tedarik Zinciri ve Teknolojik Müdahale, Sağlık ve Sosyal Refah, Tarım ve Çevresel Etki, Topluluk Dayanıklılığı ve Afet Kurtarma, Sürdürülebilirlik Teorisi ve Uygulaması ve Kentsel Yönetişim ve Planlama. Konu evrimi ilişkilerine dayanarak, bağlam perspektifine dayalı konular giderek daha uzmanlaşmış ve üretken hale gelirken, teori perspektifine dayalı olanlar durgunlaşmış veya yeterince araştırılmamıştır. Bulgulara bakıldığında, dönüştürücü dayanıklılık çalışmalarının, mevcut araştırmaları etkili bir şekilde entegre etmek için daha fazla disiplinlerarası bağlantı kurması ve mekanizasyon merkezli yaklaşımları bağlam merkezli dayanıklılık oluşturmaya odaklanması gerektiğini göstermektedir.

Zhang vd., (2025) Çin genelinde bilim ve teknoloji inovasyon politikalarını değerlendirmek için 2013-2023 yılları arasında yayınlanan 861 politika metnini incelemişlerdir. Metinler analiz edilip yüksek frekanslı bir düğüm ağı oluşturulduktan sonra Gephi programıyla modüler sınıflandırma yapılmıştır. İlk çok boyutlu politika uygulama etki değerlendirme endeks sistemini oluşturmak için bilim ve teknoloji politikasının temel ilişkilerini çevreleyen etki faktörleri seçilmiştir. Sonrasında bilimsel bir değerlendirme yapmayı daha verimli kılmak için faktör analizi ile bu endeks sistemi uygun hale getirilmiştir. Nihai olarak entropi metodu ile birleştirilip Çin genelinde bilim

ve teknoloji inovasyon politikasının genel olarak inişli çıkışlı yükselme trendi göstermiş olduğu gözlemlense de salgın döneminde güçlü bir dayanıklılık sergilediği sonucuna ulaşılmıştır. Buna dayanarak, Çin'in bilim ve teknoloji sistem yönetiminin etkinliğini güçlendirmesi, bilim ve teknoloji inovasyonuna yönelik finansal hizmetlerin desteğini artırması ve iç ve dış piyasalardaki hızlı değişimlere yanıt olarak politika uyarlanabilirliğini ve diğer karşı önlem ve önerileri geliştirmesi önerilmiştir.

Mostafa vd., (2025) 24.983 özetini kapsayan ve toplamda iki milyondan fazla kelimeden oluşan yirmi dokuz iletişim dergisinin külliyatına dayanarak, alanın son yüzyıldaki entelektüel yapısını analiz etmek ve izlemek için yapısal konu modellemesini (STM) kullanmışlardır. Elde edilen sonuçlar, iletişim teorisi, medya analizi, sağlık iletişimi, retorik teori, kişilerarası etkileşimler, küçük grup iletişimleri, siyasi tartışma ve konuşma eğitimi dahil olmak üzere alanda çok çeşitli önemli araştırma temaları olduğunu göstermektedir. Diyakronik olarak, sonuçlar "konuşma eğitimi" ve "siyasi tartışma" gibi bazı konuların zamanla azaldığını, "anlatı/söylem analizi" ve "küresel politika değişikliği" gibi diğer konuların ise son zamanlarda iletişim bilimcileri tarafından daha fazla ilgi gördüğünü ortaya koymaktadır. Bu bulgular yalnızca iletişim araştırmalarının entelektüel genişliğini ve tarihsel evrimini vurgulamakla kalmayıp aynı zamanda alandaki temel paradigma değişimlerini de ortaya koymaktadır. Çalışma, hesaplamalı metin analizinin akademik disiplinler içerisinde meta-teorik anlayış ve stratejik planlamaya nasıl katkı sağlayabileceğini göstermektedir.

Cogburn vd., (2025) küresel ilerlemeyi değerlendirmek, bölgesel farklılıkları belirlemek ve başarılı uygulamayı etkileyen faktörleri incelemek için mevcut tüm CRPD Eyalet Raporlarını (N = 170) metin madenciliği, Doğal Dil İşleme ve GenerativeAI araçları kullanarak analiz etmişlerdir. Bulgular, CRPD'nin yaygın bir şekilde uygulandığına, sosyal adalet ve ekonomik kapsayıcılığa artan desteğe ve sivil toplum katılımının önemine dair kanıtlar ortaya koymaktadır. Bu çalışmanın hibrit veri analizi yaklaşımı, dünya çapında anlaşma taahhütlerinin gerçekleştirilmesini ilerletmek için metinsel verilerin gücünden yararlanmak için umut verici bir çerçeve sunduğu düşünülmektedir.

Aubakirov vd., (2025) çalışmalarında GreedSum algoritmasında min-df (minimum belge frekansı) parametresini dinamik olarak optimize ederek çıkarımsal metin özetlemeyi iyileştirmeyi amaçlamaktadır. Bunu başarmak için için üç yöntem

önerilmiştir: geometrik bir yaklaşım, yüzdelik tabanlı bir eşik ve kümeleme tabanlı bir uyarlama stratejisi. Bu yöntemler, arXiv ve PubMed'den 17.038 bilimsel makaleden oluşan geniş ölçekli bir veri kümesine uygulanmıştır. Analizler, sabit min-df ayarlarına göre ROUGE-1 F-ölçümünde %2'lik bir iyileştirme göstererek %45'lik bir tepe ROUGE-1 F1 puanına ulaşmıştır. Bu performans, birkaç yerleşik çıkarımsal ve hibrit başlangıç çizgisini aşmaktadır. Çalışmanın literatüre katkıları arasında dinamik ayarlama stratejilerinin karşılaştırmalı değerlendirilmesi ve bunların uyarlanabilir özetleme için etkinliğinin gösterilmesi yer almaktadır. Bu bulgular, gelecekteki araştırmaların derin öğrenme entegrasyonu ve gerçek zamanlı, çok modlu özetleme üzerine odaklanmasıyla, dinamik min-df optimizasyonunun doğru ve verimli özetler üretme potansiyelini doğrulamaktadır

Kılınç vd., (2025) Kentsel Isı Adası (UHI) etkisinin bilimsel literatürde nasıl ele alındığını sistematik olarak analiz etmek için Web of Science veri tabanından toplam 9061 araştırma makalesine ulaşmış, bu makaleler ön işleme tabi tutulmuş ve analiz edilmiştir. Gelişmiş bir doğal dil işleme (NLP) aracı olan BERTopic kullanılarak, çalışma geniş bir akademik literatür külliyyatından baskın temaları çıkarmakta ve zaman içindeki evrimlerini takip etmektedir. BERTopic, anlamsal olarak ilişkili konuları kümelemiş ve zamansal dinamiklerini ortaya koyarak, ortaya çıkan ve düşüşte olan araştırma alanlarına ilişkin içgörüler sunmuştur. Sonuçlar, kaldırım malzemeleri ve kentsel bitki örtüsünün en çok incelenen konular arasında olduğunu ve yüzey malzemelerinin ve yeşil altyapının UHI'yi azaltmadaki önemini vurguladığını göstermektedir. Çalışma, özellikle yansıtıcı yüzeyler ve havalandırma koridorları olmak üzere kentsel soğutma stratejilerine artan bir ilgi olduğunu tespit etmektedir. Amacına uygun olarak, çalışma UHI literatürüne kapsamlı bir genel bakış sunmakta, mevcut boşlukları eleştirel bir şekilde belirlemekte ve gelecekteki araştırmalar için net yönler önermektedir. Kent planlamacılarına, politika yapıcılara ve araştırmacılara, UHI etkilerini azaltmak ve kentsel iklim direncini artırmak için veri odaklı stratejiler geliştirmelerinde destek sağlamaktadır.

Galluccio vd., (2025) İtalyan çevrimiçi haberlerinin turizm sürdürülebilirliği anlatısını incelemek için, Avrupa ve dünya çapında hukuki görüşler, kamu kayıtları, haberler ve iş bilgileri toplayan bir çevrimiçi platform olan LexisNexis haber veritabanında yer alan gazetelerin bir analizini gerçekleştirmiştir. 1 Ocak 2017 ile 31 Mayıs 2020 tarihleri arasında dört İtalyan gazetesinde (Corriere della Sera, Il Resto del Carlino, Il Giornale ve La Nazione) çevrimiçi olarak yayınlanan turizm sürdürülebilirliği

hakkındaki haber makaleleri toplanıp analiz edildiğinde Louvain algoritması tarafından belirlenen konuların “İtalya’da 2017-2022 Dönemi Turizm Stratejik Kalkınma Planı”nda özetlenen temalarla uyumlu olduğunu göstermektedir. yerel, ulusal ve ulusüstü politikalarla ilgili konular tüm yıllar boyunca önemli olmuş ve stratejik planın da yaptığı gibi, siyasetin ve kurumların daha sürdürülebilir turizmin geliştirilmesine olan ilgisini, söz konusu dönemden önce bile yansıtmaktadır. Ek olarak, 2020’nin ilk beş ayını kapsayan verilerden ortaya çıkan konular, kötü turizm gibi önceki yıllardan önemli ölçüde farklılaşarak özellikle dikkat çekicidir. İlginç bir şekilde, belirlenen konular, yıllar içinde gelişen anlatının sonraki hükümet planlarını ve hedeflerini nasıl etkilediğini de vurgulamaktadır.

Zhang vd., (2025) tarafından metin madenciliği ve istatistiksel yöntemler kullanılarak, 2008-2025 yılları arasında Hong Kong, Makao ve Tayvan hariç, merkezi hükümet ve eyalet hükümetleri tarafından çıkarılan toplam 223 plastik kısıtlama politikası üzerinde metin analizi gerçekleştirilmiştir. Sonuçlar, politika sürümlerinin 2008 ve 2020’de zirve yaptığını göstermektedir. Doğu Bölgesi’ndeki her ilin düzenlediği ortalama poliçe sayısı 6,8 ile en yüksektir. Politika içeriği açısından, A3, B3 ve C3 sırasıyla en belirgin politika hedefleri, araçlar ve koruma önlemleridir. Politika içeriğinin sinerjisi, piyasa odaklı politikaların daha güçlü bir uygulamaya sahip olduğunu göstermektedir, örneğin, A2 ve B2 arasındaki sinerji gücü 0,7467’dir. Yasama ve uygulama politikaları arasındaki sinerji yetersizdir. Örneğin, A1 ve B1 arasındaki sinerji gücü yalnızca 0,2903’tür. Bölgesel sinerji güneybatıdan kuzeydoğuya doğru azalmaktadır. Kuzeydoğu bölgesi ise önemli ölçüde geride kalmaktadır. Zaman açısından, 2021 ve 2023 yılları arasındaki politika metinlerinin benzerliği sabit kalmıştır (0,3256-0,3666). Üç boyutlu çerçeve, politika sinerjisi değerlendirmesine bir yaklaşım sunmaktadır. Bu çalışmanın, plastik kirliliğinin azaltılması gibi küresel çevre koruma açısından da olumlu bir değere sahip olduğu düşünülmektedir.

Abbasiantaeb vd., (2025) çalışmalarında Patent-makale bağlantıları üzerine yapılan araştırmaların çoğunlukla ön sayfa referanslarına yoğunlaşmış ve daha karmaşık metin içi referansları sıklıkla ihmal ettiğini belirlemişler bu nedenle, metin içi patent referanslarını bilimsel yayınlarla eşleştirmek ve çıkarmak için üç aşamalı bir makine öğrenmesi hattı geliştirmişlerdir. Bu hatlar şu görevleri gerçekleştirir: 1)patent metinlerinden referans dizelerini çıkarma, 2) bu referans dizelerinden alanları ayrıştırma ve 3) referansları Web of Science (WoS) veritabanındaki yayınlarla

eşleştirme. Rastgele seçilen 392 (ve 319) EPO (ve USPTO) patentinden 3.900 (ve 3.901) adet elle ek açıklamalı referanstan oluşan bir eğitim veri seti geliştirilmiştir. İlk aşama olan referans çıkarmada, %98,9 hassasiyet ve referans düzeyinde %97,7 geri çağırma ile neredeyse mükemmel sonuçlar elde edilmiştir. Genel olarak, boru hattı, benzersiz patent-kağıt çifti düzeyinde %96,8 hassasiyet ve %91,9 geri çağırma oranıyla güçlü bir performans göstermiştir. Bu boru hattı 1990 ile 2022 yılları arasında verilen EPO ve USPTO patentlerine uygulandığında, 492.469 (ve 1.449.398) EPO (ve USPTO) patentinden 5.438.836 (ve 20.432.189) referans belirlenmiş ve bunların 2.763.779'u (ve 11.069.995'i) WoS yayınlarıyla eşleştirilmiştir. Bu kapsamlı veri setinin, bilim-teknoloji bağlantılarını incelemek için değerli bir kaynak olduğu düşünülmektedir.

Vestrelli vd., (2025) yaptıkları çalışmada yedi yüksek gelirli Avrupa ülkesinde farklı enerjiyle ilgili temaların (ERT) haber medyası açısından önemi ile yenilenebilir kaynaklardan elektrik üretim düzeyi arasındaki ilişkiyi incelemişlerdir. Hızlı teknolojik ilerleme ve artan çevre bilinci çağında, medya kamuoyunu şekillendirmede ve kamu eylemini etkilemede önemli bir rol oynamaktadır. Yapılan analizler, 2014'ten 2023'e kadar yedi yılda yayınlanan 260.000'den fazla makaleyi inceleyerek Avrupa çevrimiçi haberlerindeki temel ERT'lerin semantik önemini ölçmektedir. ERT'lere medya ilgisinin yenilenebilir elektrik üretimi üzerindeki potansiyel nedensel etkisini belirlemek için medya raporlamasında potansiyel içsellik ele alan bir araçsal değişken yaklaşımı benimsenmiştir. Metin madenciliği yöntemleri kullanılarak haberlerde ERT'lerin medya önemini nicelleştirmek için, farklı ERT'lerin medya açısından artan öneminin sonraki aylarda artan yenilenebilir elektrik üretimiyle ilişkili olduğu sonucuna ulaşılmıştır. Bu gecikmeli etki, medya etkisinin yenilenebilir enerji benimsenmesini etkili bir şekilde yönlendirebileceği optimum bir pencerenin varlığını göstermektedir.

Hicks ve Lobato, (2025) çalışmalarında, 1960'tan 2010'a kadar hem ana akım davranış bilimi dergilerinde hem de akademik olmayan dergi Mankind Quarterly'de ırk bilimi söyleminin görünümünü izlemek için bilimölçüm ve metin madenciliği yöntemleri kullanmışlardır. Analiz sonucunda, ırk bilimi söyleminde iki ayırt edilebilir çizgi bulunmuştur. Biri Mankind Quarterly ile güçlü bir şekilde ilişkiliyken, diğerinin -ırk ve zekayı vurgulayan- ana akım psikoloji dergilerinde, özellikle Intelligence ve Personality and Individual Differences'ta daha yaygın olduğu görülmüştür. Psikolojinin ırk bilimine verdiği yanıtlar ile diğer alanların yanıtları karşılaştırılmış ve bilim insanlarının ırk bilimiyle ilgili sosyal sorumluluğu tartışılmıştır.

Liu ve Yao, (2025)'nin çalışmaları, Çin'in resmi ingilizce gazetesi China Daily'de otizm spektrum bozukluğunun (OSB) temsilini incelemektedir. Çalışmada, Çerçeveleme Teorisi'ne dayanarak, 1981'den 2023'e kadar 93 makaleden (409.400 jeton) oluşan bir korpus üzerinde konu modelleme destekli bir söylem analizi kullanılmıştır. Analiz, China Daily'nin otizmi tıbbi bir eksiklik olarak değil, ailevi özveri, profesyonel müdahale ve kolektif toplumsal destek anlatılarıyla ele alınan bir sosyal entegrasyon zorluğu olarak stratejik olarak çerçevelediğini ortaya koymaktadır. Batı medyasında yaygın olan eksiklik odaklı anlatılarla keskin bir tezat oluşturan bu yapıcı tasvir, sorunları sosyal olarak tanımlayarak, nedenleri kolektif çabaya atfederek ve eğitim ve kapsayıcılığa odaklı çözümleri teşvik ederek temel çerçeveleme işlevlerini yerine getirdiği düşünülmektedir. Çalışmanın, küresel bir izleyici kitlesi için umut verici bir kamu sağlığı söylemi oluşturmada devlet bağlantılı medyanın stratejik rolüne dikkat çektiği ve çerçeve seçimini kültürel ve kurumsal bağlamlarla ilişkilendirerek Çerçeveleme Teorisini zenginleştirdiği düşünülmektedir.

Elitaş vd., (2025) tarafından yapılan çalışmada, 1980-2024 yılları arasında sinema alanında yapılan 16.891 akademik yayının bibliyometrik analiz ve konu modelleme yöntemleri kullanılarak değerlendirilmesi amaçlanmıştır. Web of Science (WOS) ve Scopus veri tabanlarından elde edilen veriler kullanılarak, yayınların yıllara göre dağılımı, makale başına düşen yıllık atıf sayısı ve oranı, alanda en üretken yazarlar, yazarların zaman içindeki üretim durumları, yazarların ülkeleri ve yayınladıkları makale sayıları ve en çok yayın yapan dergiler gibi bibliyometrik bulgular elde edilmiştir. Ayrıca Web of Science (WOS) ve Scopus veri tabanlarından elde edilen veriler kullanılarak metin madenciliği ve Latent Dirichlet Allocation (LDA) konu modellemesi kullanılarak makalelerde öne çıkan kelime grupları ve temalar belirlenmiştir. Analiz sonucunda kelime-metin ilişkileri ve yayın ağırlıklarına göre 12 ana tema ortaya çıkmıştır. Bulgular, sinema çalışmalarının yıllar içinde artan bir ivmeyle geliştiğini ve belirli konulara giderek daha fazla odaklanıldığını göstermektedir. Çalışma, sinema çalışmaları literatürünün gelişimini betimsel içerik analizi ve LDA konu modellemesi yoluyla sistematik olarak incelediğinden, sinema alanındaki akademik üretimin yapısal ve tematik dönüşümünü sistematik olarak ortaya koyması ve gelecekteki araştırmalar için teorik ve metodolojik bir temel sunması bakımından önemli olduğu düşünülmektedir. Ayrıca, sinema çalışmalarında giderek önem kazanan dijital eğilimleri, kültürel temsilleri ve

disiplinlerarası gelişmeleri ortaya koyması açısından disipline güncel ve çok boyutlu bir katkı sağladığı söylenebilir.

Cotrina vd., (2025) yaptıkları çalışmada, yenilenemeyen kaynaklar bağlamında Hartwick kuralının uygulanmasına ilişkin 2024 yılına kadar yayınlanmış tüm araştırmaların bibliyometrik ve sistematik bir incelemesini yapmayı amaçlamaktadır. Hartwick kuralı, yenilenemeyen kaynakların kullanımından elde edilen gelirlerin gelecek nesillere fayda sağlayacak diğer sermaye biçimlerine yeniden yatırılması durumunda bir ekonominin sürdürülebilir olabileceğini belirtir. Çalışmada bibliyometrik analiz, metin madenciliği ve nitel içerik analizini birleştiren karma yöntemli bir yaklaşım kullanılmıştır. İnceleme, PRISMA protokolü kullanılarak seçilen 71 hakemli makaleye dayanmaktadır ve önde gelen akademik veri tabanlarında özel bir arama ile tamamlanmaktadır. Bulgular, Hartwick kuralının teorik bir kavramdan, özellikle petrol, mineral veya gazı büyük ölçüde bağımlı ülkelerde kamu politikalarına rehberlik etmek için pratik bir araca nasıl evrildiğini göstermektedir. Literatür üç ana tematik alan etrafında yapılandırılmıştır: kuralın teorik ve ekonomik temelleri; yenilenemeyen kaynakların yönetiminde uygulanması ve sürdürülebilirliği için gereken varsayımlar ve koşullar. Çalışmanın, Hartwick kuralının yenilenemeyen kaynak yönetimindeki evrimi ve uygulamasına ilişkin yapılandırılmış bir genel bakış sunarak, sürdürülebilirlik açısından önemini vurguladığı söylenebilir.

Kurday ve Vladova (2025) bu çalışmada, 2014-2023 yılları arasında gerçekleştirilen Öğrenme Analitiği (ÖA) ve Eğitim Veri Madenciliği (ÖA) araştırmalarının detaylı bir bibliyometrik analizini ve seçilen makalelerin özetlerine dayalı içerik analizlerini yürütmektedir. Yazarlar, ÖA ve EDM alanlarının gelişimini, çeşitli uygulamalarını ve araştırmada kullanılan veri kaynaklarını incelemektedir. LA ve EDM uygulamalarına özel olarak odaklanan 356 yayın inceleme için seçilmiştir. Bibliyometrik araçlar kullanarak, araştırma ortamını haritalamak ve temel akademik ağları belirlemek için anahtar kelimeleri, iş birliklerini ve atıfları analiz edilmiştir. Ayrıca, LA ve EDM uygulamalarını ve kullanılan veri kaynaklarını sistematik olarak kategorize etmek için metin analizi amacıyla büyük dil modeli (LLM) olarak ChatGPT 4.0 kullanılmıştır. Bu amaçla, inceleme makalelerinin özetlerini değerlendirmek için yarı otomatik bir yöntem geliştirilmiş ve uygulanmıştır. Sonuç olarak, bu yayın, LA ve EDM araştırmalarındaki son gelişmelerin yapılandırılmış bir sentezini sunarak, araştırma boşluklarını belirlemek ve gelecekteki çalışmalar için daha derin konuları keşfetmek için

sağlam bir temel sunmaktadır. Ayrıca, geliştirilen üç adımlı yarı otomatik LLM analiz yaklaşımıyla, makale araştırmacılara içerik analizi yapmak için yeni bir yöntem sunmaktadır.

Huang ve Jiang (2024) yaptıkları çalışmada, Gizli Dirichlet Tahsisi (LDA) konu modelleme yaklaşımını kullanarak ikinci dil (L2) yazma araştırmalarında sözdizimsel karmaşıklığa (SC) genel bir bakış sunmaktadır. WoS Çekirdek Koleksiyonu'ndan (2005–2023) ve Scopus'tan (1978–2023) alınan 470 özet analizi eden bu araştırma, L2 yazımında SC'deki geniş yelpazedeki araştırma temalarını incelemektedir. Şaşkınlık ve tutarlılık testlerini kullanarak çalışma, modeldeki optimum konu sayısını beş olarak belirlemiştir: yazma kalitesi, yeterlilik, türler, SC gelişimi ve göreve dayalı koşullar. Bulgular şunları ortaya koymaktadır: Kesitsel çalışmalarda, yazma kalitesi, yeterlilik düzeyleri ve türler arasındaki etkileşim genellikle bir değişkenin diğerlerine bağımlı olarak incelenmesini içerir. Ek olarak, ifade yapılarını daha güvenilir karmaşıklık göstergeleri olarak görme eğilimi artmaktadır. Boylamsal araştırmalar ağırlıklı olarak SC'nin evrimini vurgular, ancak devam yazma görevlerinde araştırma kıtlığı vardır. Ölçüm metrikleri açısından, daha geniş SC değerlendirmelerinden daha karmaşık değerlendirmelere bir geçiş açıktır, ancak bazı çalışmalarda, özellikle devam görevlerinde kapsamlı incelemeler eksiktir.

Kim vd., (2024), evsel kırsal alan konusundaki akademik makalelere metin madenciliği uygulayarak kırsal alan araştırmalarındaki eğilimleri zaman içinde analiz etmeyi amaçlamaktadırlar. Çalışmada, veri toplama için Kore Atıf Endeksi'nde (KCI) arama anahtar kelimesi 'kırsal alan' olarak belirlenmiş ve 2003-2024 yılları arasında 1.581 akademik makale toplanıp uygunsuz olanlar çıkarıldıktan sonra toplam 473 makale seçilmiştir. Veriler beş döneme ayrılarak analiz gerçekleştirilmiştir. Ardından, büyük veri analitiği platformu Textom aracılığıyla ön işleme ve kelime sıklığı, TF-IDF ve LDA konu modelleme analizleri gerçekleştirilmiştir. Analizin sonuçları, beş dönem boyunca birkaç temel eğilimi ortaya koymuştur. İlk olarak, kırsal alan araştırmalarının hacmi zamanla artmış ve bu da alana olan ilginin arttığını göstermiştir. İkinci olarak, kırsal alan araştırmalarının paradigmasında, araştırma önceliklerinin değişimini yansıtan anahtar kelimelerdeki değişiklikler gözlemlenmiştir. Üçüncü olarak, kırsal alan araştırmalarının giderek kapsamlı bir yaklaşıma dönüştüğü ve ampirik araştırmalara doğru genişlediği keşfedilmiştir. Bu genişleme, her dönemden çıkarılan üç tema ile karakterize edilmektedir: 'köy düzeyinde kalkınma', 'peyzaj ve çevre' ve 'hizmet geliştirme ve değer

yaratma'. Sonuç olarak bulgular, kırsal alanın kendine özgü bir özelliği olan 'kırsallığın' korunmasına artan bir odaklanma olduğunu ve dikkatin peyzajın korunması, ekoloji ve çevreye kaydığını göstermiştir. Bu eğilim, kırsal alana yönelik beklenti ve değerlerin yanı sıra sürdürülebilir kalkınmaya değer veren toplumsal taleplerle de örtüşmekte ve araştırmaların daha fazla genişletilmesi ve canlandırılması potansiyelini göstermektedir.

Raaymann ve Spinler (2024) yaptıkları çalışma, otomotiv endüstrisinde tedarik zinciri dayanıklılığının (SCR) nasıl ölçüleceğine dair cevaplar sunmaktadır. COVID-19 pandemisinin küresel tedarik zincirlerini altüst etmesinden yaklaşık üç yıl sonra, şirketler bir sonraki bozulmaya karşı dayanıklı tedarik zincirleri oluşturmaya giderek daha fazla odaklanmıştır. Birçok araştırmacı tedarik zincirlerindeki riski değerlendirmek için birden fazla çerçeve ve nicel model geliştirmiş olsalar da, tedarik zinciri dayanıklılığının (SCR) temel performans göstergeleri (KPI'lar) ile nasıl ölçüleceği sorusu cevapsız kaldığı gözlemlenmiştir. Bu çalışmada, tedarik zinciri dayanıklılığı (SCR) hakkında yayınlanmış 195 makale üzerinde metin madenciliği yoluyla literatürün bakış açısı araştırılmış ve endüstrinin bakış açısıyla karşılaştırılmıştır. En uygun temel performans göstergeleri (KPI) kombinasyonunu bulmak için Analitik Hiyerarşi Süreci metin madenciliği sonuçlarına uygulanmıştır. Endüstri verileri için yapılan görüşmelerde, sıralı regresyon yoluyla analiz edilen bir ortak yöntem uygulanmıştır. Araştırma, literatüre göre en önemli temel performans göstergeleri teslim süresi değişimi, OTIF (Tam zamanında) ve tedarikçilerin hacim esnekliği olduğunu ortaya koymaktadır. Aynı zamanda, sektör en büyük katkıyı OTIF ve hacim esnekliğine ve yüksek riskli parçaların stok seviyesine atfetmektedir. Bu çalışma ayrıca, OEM'lerin, 1. ve 2. kademe tedarikçilerin SCR ölçümündeki farklı önceliklerini de incelemektedir. Dayanıklılığın nasıl ölçüleceğine dair bakış açıları, sektör içinde ve sektör ile akademi arasında farklılık göstermektedir. Bu çalışma, sektör ve akademi arasında daha fazla etkileşimin yanı sıra dayanıklılık KPI'ları ve bunların sektördeki uygulamaları hakkında daha yapısal bir tartışmaya ihtiyaç olduğunu ortaya koymaktadır.

Wu vd., (2024) yaptıkları çalışmada, metin bölümlendirme, TF-IDF (terim sıklığı- ters belge sıklığı) yöntemi ve karmaşık ağ analizi dahil olmak üzere metin madenciliği teknikleri yoluyla 617 geleneksel ve 289 modern parçanın özelliklerini incelemektedir. Tekerlemeler, bir kültürün geleneklerine, inançlarına ve değerlerine dair içgörüler sunarak onu bir topluluğun mirasının ayrılmaz bir parçası haline getirdiğinden Hoklo halkının somut olmayan kültürel mirasının temsili dil kaynakları olan Minnan

tekerlemeleri (MNR'ler), Çin anakarasını, Tayvan Boğazı'nı ve denizaşırı Çin topluluklarını birbirine bağlamada önemli bir rol oynadığı düşünülmektedir. Önceki çalışmalara göre daha geniş bir ölçekte, küçük bir grup elle notlandırılmış örnek kullanarak şarkı sözlerinin sıklığının ve duygusal saflığının incelenmesinden ziyade anahtar terimleri düğümler ve düğümler arasındaki birlikte bulunma ilişkilerini bağlantılar olarak ele alarak, oluşturulan MNR ağlarının genel, bireysel, çekirdek-çevre yapılarını değerlendirerek MNR modelleri analiz edilmiştir. Araştırma, hem geleneksel hem de modern MNR ağlarındaki terimlerin heterojen doğasını ortaya koymaktadır. Ayrıca, topluluk tespit yöntemiyle, MNR'lerde sunulan beş temel görüntü özelliği belirlenmiştir. Geleneksel MNR'ler, Minnan kültürel mirasının kalıcı yönlerini yansıtan aile ilişkilerine, halk kültürüne ve yemek kültürüne vurgu yaptığından, modern MNR'ler, gelişen toplumsal değerlerin göstergesi olan çocuk duyguları ve doğal manzara temalarına odaklanır. Bu çalışma, MNR'lerin ilk büyük ölçekli karmaşık ağ analizini temsil etmekte olup, yerleşik Minnan kültürüne değerli içgörüler sunmakta ve bu değerli MNR kaynaklarında yansıtılan toplumsal dinamikler üzerine daha fazla araştırma için bir temel oluşturmaktadır.

Zhu ve Liu (2024) çalışmalarında, bilimsel etki geleneksel olarak atıf tabanlı metriklerle değerlendirildiğinden araştırma makalesi başlıklarının dilsel özellikleri ile çevrimiçi dikkat çekme arasındaki ilişkiyi inceleyen çok az araştırma olduğuna dikkat çekmişlerdir. Bu sorunu ele almak için, yazarlar mevcut çalışmada bir dizi başlık özelliği ile altrimetrik dikkat puanları arasındaki ilişkiyi incelemişlerdir. Veri seti 8.658 adet Bilim makalesi başlığından oluşmaktadır. Yazarlar başlık korpusundan altı özelliği çıkarmışlardır (yani, ortalama kelime uzunluğu, sözcüksel karmaşıklık, sözcük yoğunluğu, başlık uzunluğu, sözdizimsel bağımlılık uzunluğu ve duygu puanı). Yazarlar, bu özellikler ile çevrimiçi etki arasındaki korelasyonları analiz etmek için Spearman sıralaması analizlerini gerçekleştirmişlerdir. Daha sonra, makalelerin çevrimiçi etkisini öngörmek için aşamalı bir geriye dönük çoklu regresyon analizi yapılmıştır. Bulgular: Korelasyon analizleri, altı başlık özelliğinin tümü ile altrimetrik dikkat puanları arasında zayıf ancak anlamlı korelasyonlar olduğunu ortaya koymuştur. Regresyon analizi, başlıkların dört dilsel özelliğinin (ortalama kelime uzunluğu, sözcüksel karmaşıklık, başlık uzunluğu ve duygu puanı) araştırma makalelerinin çevrimiçi etkisi üzerinde mütevazı öngörücü etkilere sahip olduğunu göstermiştir. Sosyal medya ve çevrimiçi platformların yaygınlaştığı internet çağında, araştırmacıların araştırma

değerlendirmesinin değişen bağlamına uyum sağlaması giderek daha önemli hale gelmektedir. Bu çalışmanın, makale başlıklarının yazımında akademisyenlerin dikkatini çeken çeşitli dilsel özellikleri belirlediği düşünülmektedir. Ayrıca, akademik yöneticiler için pratik çıkarımlar ve akademik yazım dersleri eğitmenleri için pedagojik çıkarımlar içerdiği söylenebilir.

Özmantar vd., (2024) yaptıkları çalışmada, konu modelleme analizini kullanarak öğretmen eğitimcileri alanındaki araştırma temalarını belirlemiş ve zaman içindeki eğilimlerini araştırmışlardır. Scopus'tan 754 bilimsel yayından oluşan bir veri kümesi, Gizli Dirichlet Tahsisi yaklaşımıyla analiz edilmiştir. Bulgular, pedagoji/bilgi, sosyal adalet, iş birliği, kimlik, mesleki gelişim, politika, uygulama, kalite standartları, İngilizce dili ve teknoloji yeterliliğini kapsayan on ayrı araştırma temasını ortaya koymuştur. Tematik araştırma eğilimlerinin, 2010'lardan bu yana periyodik ve ardışık örüntüler sergilediği görülmüştür. Bu çalışma, eğitim dergilerinin etkisi, eğitim politikaları ve reform girişimleri ile Covid-19 pandemisini göz önünde bulundurarak mevcut inceleme çalışmalarına göre belirlenen temaları ve eğilimleri ele almıştır.

García ve Teichert (2024) yaptıkları çalışmada sosyal bilimlerde gıda ve sosyal medya ile ilgili mevcut literatürü analiz ederek, keşfedilen ana araştırma akımlarını ve temaları belirlemektedir. Bu amaçla, mevcut makalede faktör ve sosyal ağ analizi yoluyla 1356 dergi makalesini analiz etmek için bibliyometrik yöntemler kullanılmıştır. Pazarlama ve işletme alanında çalışan akademisyenler arasında gıda ve çevrimiçi iletişime olan ilginin hızla arttığını belirterek bu ilginin yalnızca bu alanlarla sınırlı kalmadığına değinmişlerdir. Farklı disiplinlerden araştırmacıların araştırmalarını farklı kavramlar, hedef kitleler, yaklaşımlar, metodolojiler ve teorik arka planlar üzerine odakladığını, bunun da büyüyen bilgi birikimini daha zengin hale getirdiğini, ancak aynı zamanda analiz edilmesini zorlaştırdığını dile getirmişlerdir. Çalışma, sosyal bilimlerde 11 baskın araştırma akımını içeren 4 kümeyi ortaya çıkararak, ana araştırma söylemleri arasındaki bağlantıları belirleyerek ve gelecekteki yeni araştırma konuları önererek literatüre katkıda bulunduğu düşünülmektedir.

Queiroz vd., (2024) yaptıkları çalışmada, Lizbon Metro Bölgesi'ndeki ebeveynlere odaklanarak 2017-2018 ve 2018-2019 okul yıllarını kapsayan 448 haneyi kapsayan iki anketin yanıtlarını ve saha müdahaleleri öncesi ve sonrası analiz ederek, hanelerin toplu taşıma hizmetlerini iyileştirmeye yönelik eleştirilerini ve önerilerini

değerlendirmek için LDA'yı kullanmışlardır. Çalışma, açık uçlu anket yanıtlarından değerli bilgiler çıkarmak ve ebeveynlerin çocuklarının PT aracılığıyla okula gidiş gelişine ilişkin beklentilerini aydınlatmak için Gizli Dirichlet Tahsisi (LDA) kullanan bir yöntem önermektedir. Bulgular, ilk anketteki genel eleştirilerden, sonraki ankette proaktif önerilere doğru bir geçiş olduğunu ve bu geçişin, fizik tedaviyle daha sürdürülebilir bir okul ulaşımını teşvik etme pazarlama çabalarıyla uyumlu olduğunu göstermektedir. Çalışma, deneysel olarak, LDA'nın kapalı uçlu sorularla genellikle göz ardı edilen kullanıcı geri bildirimlerini yakalamadaki etkinliğini vurgulamaktadır. Metinsel verilerin etkili bir şekilde ön işlenmesi, saha müdahalelerinin kolaylaştırılmasını sağlamaktadır. Fizik tedavi politika yapıcılarını bilgilendirmek ve kullanıcı odaklı iyileştirmelerin dahil edilmesini teşvik etmek için kullanıcı merkezli içgörüler sunması genel olarak literatüre yapılan bir katkı olduğu söylenebilir.

Zheng ve Pee (2024) yaptıkları çalışmada, bilgi bilimi araştırmacılarının, bilimsel araştırmalarla üretilen bilginin akademi dışında nasıl kullanıldığını anlamaya giderek daha fazla ilgi duyduklarını belirterek bilgi kullanımının kavramsal düzeyleri iyi belirlenmiş olsa da, araştırmacılar tarafından bilgi kullanımı için kullanılan çeşitli bilgi uygulamalarına dair anlayışların sınırlı olduğunu dile getirmişlerdir. Çalışma, Birleşik Krallık Araştırma Mükemmelliği Çerçevesi kapsamında belgelenen 6637 vaka çalışmasını metin madenciliği yoluyla inceleyerek bu tür bilgi uygulamalarını belirlemiştir. Sonuçlar, araştırma bilgisi kullanımının teorik çerçevesine dayalı dokuz türden oluşan bir çerçeve geliştirmek için uzman görüşüyle desteklenmiştir. Üç yeni tür belirlenmiştir: müzakere, ortak yaratım ve öngörü. Bunlar, potansiyel yararlanıcıları araştırmanın değerini kullanma ve gerçekleştirme konusunda harekete geçirmek için sosyal medya ve analitik yeteneklerden yararlanan bilgi uygulamalarının yükselişini göstermektedir.

Gatzert ve Reichel (2024) yaptıkları çalışmada, 2013-2018 yılları arasında kamuya açık raporlarına yansıdığı şekliyle Avrupa ve ABD sigorta sektöründe sürdürülebilir yatırımın gelişimini incelemek için metin madenciliği analizi kullanmaktadır. Veri seti 77 firmanın, sürdürülebilirlik ve yatırım ile ilgili yıllık 1215 belgesinden oluşmaktadır. İlkeler, kriterler ve terminolojilerin yanı sıra stratejiler içeren bir sözlük geliştirilmiş ve raporların kalitesi arasında ayrım yapılmıştır. Sonuçlar, sürdürülebilir yatırıma atıfta bulunan firma sayısının ve ilgili kelime sayısının örneklem dönemi boyunca önemli ölçüde arttığını ve sürdürülebilir yatırım hakkında rapor veren

sigortacıların ortalama olarak önemli ölçüde daha fazla olduğunu göstermektedir. Ayrıca, örnekleme yer alan ABD’li sigortacılara kıyasla Avrupalı sigortacıların sürdürülebilir yatırım uygulamaları hakkında çok daha kapsamlı rapor verdiği tespit edilmiştir. 2018’de en alakalı olanlar genel ESG kriterlerine yapılan atıflar, ardından sorumlu yatırım ve Sürdürülebilir Kalkınma Hedefleri’ne yapılan atıflardır. Bahsedilen en önemli stratejiler ESG entegrasyonu ve etki yatırımdır; bu sayede sigortacıların zamanla tek bir stratejiden bahsetmekten birden fazla stratejiye doğru evrildiklerini gözlemlenmiştir. Son olarak, regresyon analizi, Tobin’in Q’su ile ilgili raporlamada sürdürülebilir yatırımla ilgili anahtar kelimelerin değer etkisini göstermemektedir. Bu durumun, yatırımın oldukça uzun vadeli bir perspektiften ele alınmasından kaynaklandığı söylenebilir.

Lee (2024) yaptığı çalışmada, teknolojik entegrasyon ile çalışanlar üzerindeki sonuçları arasındaki karmaşık ilişkiyi araştırmış ve 1983’ten 2022’ye kadar ileri teknolojilerin kurumsal operasyonlar ve iş gücü dinamikleri üzerindeki gelişen etkilerine odaklanmıştır. Metin madenciliği analizi ve sistematik bir literatür incelemesini birleştiren gelişmiş bir yöntem kullanılan araştırmada, literatürdeki teknolojiyle ilgili konuların ve anahtar kelimelerin ilerlemesini izlemek için 7.000 makaleden oluşan kapsamlı bir veri seti analiz edilmiştir. Bulgular, teknolojinin iş gücü üzerindeki etkisinin ikili bir doğasını ortaya koymaktadır. Bir yandan, teknolojik gelişmeler artan işten ayrılma niyetleri, kaygı ve sağlık sorunları gibi zorluklarla ilişkilendirilirken diğer yandan, teknoloji aracılı etkileşimler ile geleneksel yüz yüze etkileşimler arasındaki nüanslı farklılıkları anlamaya yönelik artan bir kayma olduğu gözlemlenmiştir. Çalışma, yalnızca üretkenliği artırmakla kalmayıp aynı zamanda çalışanların refahını da koruyan stratejik teknoloji entegrasyonunun gerekliliğini vurgulamaktadır. İş-yaşam dengesine yeniden tanımlanmış bir yaklaşımın gerekliliğini vurgulayan araştırma, kurumsal bağlamlarda teknolojinin çok yönlü etkilerinin gelecekteki araştırmalar için bir temel oluşturmakta ve özellikle bireysel bilgi davranışlarının ve teknolojik gelişmelere yanıt olarak sektöre özgü teknoloji etkilerinin daha derinlemesine incelenmesini önermektedir. Bu gelişmiş odak noktası, bu teknolojik entegrasyonların çeşitli kurumsal katmanlarda ve zaman içinde bilgi davranışlarını nasıl etkilediğini ele alarak bilgi bilimine daha doğrudan katkıda bulunmayı amaçlamaktadır.

Wu, Deng ve Zhao (2024)’ya göre hesaplamalı iletişim, dijital verilerin, sosyal medya etkileşimlerinin ve iletişim süreçlerini şekillendiren algoritmaların analizini ele alır; ancak çok az çalışma, uyarlanabilir konularla ilgili metodolojik çerçevenin

çerçevesine ve iç yapısına odaklanır. Bu nedenle yazarlar, uluslararası bilimsel atıf veri tabanlarından 9795 yayını analiz etmek için metin madenciliği tekniklerini kullanmış ve ampirik araştırmalarda kullanılan yöntemleri tanımlamak için bir sınıflandırma çerçevesi sunmuşlardır. Bu çerçeve, geleneksel nicel yöntemleri ve yeni hesaplamalı yöntemleri vurgulamaktadır. İlki, orta ölçekli ve yapılandırılmış örneklemeler üzerinde istatistiksel analiz gerçekleştirirken, ikincisi kapsamlı veri analiziyle mikroskobik görünümeler sunmaktadır. Deneyisel sonuçlar, yöntem kategorilerinin tematik dağılımını, evrim aşamalarını ve konu sınırlarını göstermektedir. Çalışmanın, sosyal hesaplama metodolojisinin kapsamını genişlettiği ve zengin deneyisel içgörüler sunduğu düşünülmektedir.

Wang vd., (2024) yaptıkları çalışmada, nitel ve nicel yöntemlerin bir kombinasyonu yoluyla çevrimiçi edebiyatın dilsel özelliklerini incelemeyi amaçlamaktadır. Çalışma, en büyük Çin çevrimiçi edebiyat web sitesi Qidian'dan geniş ölçekli bir korpus oluşturmakta ve kelime sıklığı, sözcüksel zenginlik, sözcüksel yoğunluk, kelime vektörleri, noktalama dağılımı ve ortalama paragraf uzunluğu olmak üzere altı nicel dil göstergesine odaklanmaktadır. Ek olarak, üç metin madenciliği göstergesi - tematik yoğunluk, bilgi entropisi ve yazar bakış açısı - tanıtılmaktadır. Bulgular, çevrimiçi edebiyatın, yüksek sıklıkta işlev sözcükleri, daha kısa ortalama paragraf uzunlukları ve diyaloga önemli ölçüde bağımlılık ile karakterize edilen gündelik ve gayri resmi bir dil tarzına sahip olduğunu ortaya koymaktadır. Tematik yoğunluk genellikle düşüktür; bu, çevrimiçi eserlerin ayrıntılı açıklamalardan ziyade anlatı akışına vurgu yapmalarıyla ilişkilidir. Genel olarak bu çalışma, çevrimiçi edebiyatın dilsel özelliklerinin anlaşılmasını geliştirmek için deneyisel veriler sunmakta ve gelecekteki araştırmaların daha geniş ölçütler ve daha kapsamlı metin madenciliği teknikleri benimseyebileceğini ileri sürmektedir.

Naz vd., (2024) yaptıkları çalışmada sistematik bir literatür taraması (SLR) yoluyla yapay zekanın (YZ) yeniden üretimdeki (RM) önemini araştırmayı amaçlamışlardır. Seçilen araştırma alanındaki SCOPUS veritabanından yaklaşık 453 makale toplanarak RM'deki YZ üzerine seçilen belgelerden ortaya çıkan araştırma temalarını oluşturmak için Yapısal Konu Modellemesi (YK) kullanılmıştır. Ayrıca, RM'deki YZ alanındaki araştırma eğilimlerini incelemek için R paketi kullanılarak bibliyometrik bir analiz yapılmıştır. Çalışmanın bulguları, bu alandaki yıllık küresel makale üretimi sınırlı olduğundan, bu alanda araştırma için çok geniş bir kapsam

olduğunu göstermiştir. Ancak, gelişmekte olan bir alan olduğundan birçok araştırma iş birliği tespit edilmiştir. Çalışma, kapsamlı bir araştırma çerçevesi ve gelecekteki araştırma geliştirmeleri için öneriler sunmaktadır.

Xia vd., (2024) tarafından literatür zenginliğiyle öne çıkan turizm araştırma alanlarını incelemek için uygulanabilecek yeni bir hesaplamalı literatür inceleme çerçevesi geliştirilmiştir. Konu modellemesine sahip geliştirilen hesaplamalı literatür inceleme çerçevesi sayesinde bu çalışma, 1.130 destinasyon rekabet gücü çalışmasının kapsamlı bir incelemesini sunmaktadır. Sonuç olarak, gelecekteki destinasyon rekabet gücü çalışmalarında incelenebilecek birçok umut verici araştırma konusu, tema ve teorik bakış açısı belirlenmiştir.

Sonk ve Tunger (2024) tarafından yapılan çalışma, metin madenciliği ve konu modellemesinin trend izleme için gelecek araştırmalarına nasıl etkili bir şekilde entegre edilebileceğini, örnek olarak kentsel hareketliliği kullanarak göstermektedir. Web of Science'dan (1991–2021) alınan yaklaşık 4.400 yayından oluşan bibliyometrik bir veri setinden yararlanan yazarlar, hareketlilik, kentsel ve geleceğe yönelik anahtar kelimeleri birleştiren kapsamlı bir arama stratejisi uygulamış ve veri setini ilgili konu kategorileriyle sınırlamıştır. Korpusu özel bir durdurma sözcüğü listesiyle ön işleme tabi tuttukten sonra, 500 konu oluşturmak ve zamansal eğilimlerini belirlemek için Orange Veri Madenciliği ortamında Gizli Dirichlet Tahsisi (LDA) kullanmışlardır. Analiz, 211 olumlu trend konuyu ortaya koymuş ve otonom paylaşımlı hareketlilik, esnek/paylaşımlı hareketlilik sistemleri, enerji depolaması olarak e-hareketlilik, yapay zeka destekli otonom sürüş, şarj altyapısı geliştirme, hareketlilik-konut ilişkileri, toplu taşıma işlevselliği, gelişmiş raylı hareketlilik ve çevre odaklı hareketlilik gibi dokuz temel ortaya çıkan temayı vurgulamıştır. Çalışma, nicel konu modellemesinin nitel yorumlamayla birleştirilmesinin, büyük metin korpuslarındaki erken sinyalleri ve ortaya çıkan kalıpları tespit etme yeteneğini geliştirdiği, kentsel hareketlilikte geleceğe yönelik eğilimleri belirlemek ve daha nesnel öngörü uygulamalarını desteklemek için ölçeklenebilir ve tekrarlanabilir bir yöntem sunduğu sonucuna varmıştır.

Zheng vd., (2024) ROST CM tabanlı konu madenciliği ve duygu analizi kullanarak iki MOOC dersinden 511 inceleme metnini analiz ederek çevrimiçi lojistik eğitimindeki öğrenci deneyimlerini incelemiştir. İtme-Çekme-Demirleme çerçevesine dayanan analizleri, kelime sıklığı kalıplarını, anlamsal ağ yapılarını ve duygusal

eğilimleri entegre ederek öğrencilerin ağırlıklı olarak öğretim kalitesine, ders organizasyonuna ve içerik netliğine odaklandığını ortaya koymuştur. Duygu sonuçları, incelemelerin önemli bir çoğunluğunun olumlu olduğunu, daha küçük bir kısmının ise nötr veya olumsuz tutumlar ifade ettiğini ve pedagojik iyileştirmeye ihtiyaç duyan alanlara işaret ettiğini göstermiştir. Genel olarak çalışma, metin madenciliğinin öğrencilerin beklentilerini ve duygusal tepkilerini yakalamak için sistematik bir araç sağladığını ve böylece lojistik eğitiminde çevrimiçi ders tasarımını geliştirmek için veriye dayalı içgörüler sunduğunu göstermektedir.

Alawadh vd., (2024) otomatik puanlama doğruluğunu artırmak için söylem makro yapısal ve istatistiksel sözcüksel özelliklerini birleştiren bir metin madenciliği yaklaşımı kullanarak verimli bir Otomatik Deneme Notlandırma çerçevesi geliştirdiler. Yazarlar, 12.976 öğrenci denemesinden oluşan Kaggle ASAP veri kümesini kullanarak dil bilgisi ve yazım düzeltme, puan niceleme ve tutarlılık, bağdaşıklık, TF-IDF tabanlı sözcüksel örüntüler ve makro yapısal söylem özelliklerini yakalayan özellik çıkarma prosedürlerini uygulamışlardır. Farklı özellik kombinasyonlarını test etmek için üç deney boyunca doğrusal regresyon modelleri eğitilmiş ve sonuçlar, istatistiksel sözcüksel ve makro yapısal özelliklerin birleştirilmesinin en yüksek performansı verdiğini ve 0,9339'luk bir QWK puanı elde ettiğini göstermiştir; bu, önceki derin öğrenmeye dayalı AEG sistemlerini geride bırakmıştır. Çalışma, dikkatlice seçilmiş dilsel ve yapısal özelliklerin, otomatik değerlendirme görevleri için hesaplamalı olarak daha hafif ancak sağlam bir alternatif sunarak verimli, yüksek doğruluklu puanlama modelleri üretebileceği sonucuna varmıştır

Tran (2024), bilgi tabanlı işgücü piyasalarındaki yüksek kaliteli insan kaynakları alanının evrimini araştırarak bilimsel araştırma eğilimlerini haritalamayı ve bu alanı şekillendiren baskın temaları belirlemeyi amaçlamaktadır. Yazar, 1994 ile 2023 yılları arasında Web of Science veritabanında indekslenen 1.560 çalışmadan yararlanarak, VOSviewer ve RapidMiner kullanarak bibliyometrik analiz, anahtar kelime eş-geçiş haritalaması ve metin madenciliği tekniklerini kullanmaktadır. Analiz, yayınlarda önemli bir büyüme olduğunu (2019'da zirveye ulaştı) ortaya koymakta ve Amerika Birleşik Devletleri'nin bu literatüre en etkili katkıda bulunan ülke olduğunu göstermektedir. Bulgular beş temel araştırma kümesini vurgulamaktadır: (1) cinsiyet eşitsizliği, göç ve beyin göçü de dahil olmak üzere yüksek kaliteli işgücü piyasalarındaki yapısal zorluklar; (2) yetenekli çalışanların elde tutulmasını şekillendiren liderlik tarzı, iş tatmini ve iş-

yaşam dengesi gibi örgütsel faktörler; (3) profesyonel gelişimde dijital dönüşüm ve bilgi yönetiminin kesişimi; (4) işveren markası, kurumsal sosyal sorumluluk uygulamaları ve kişi-kurum uyumu yetenek çekiminin belirleyicileri olarak; ve (5) yüksek vasıflı çalışanların performans ve motivasyonunu şekillendirmede öz yeterliliğin rolü. Yazar, genel olarak, yüksek kaliteli insan kaynakları yönetiminin teknolojik, örgütsel ve sosyo-ekonomik dinamikleri ele alan bütünleşik stratejiler gerektirdiği sonucuna varırken, yetenek ve iş gücü kalitesine ilişkin teorik çerçeveleri geliştirmek için daha fazla platformlar arası ve karşılaştırmalı analize ihtiyaç duyulduğunu vurgulamaktadır.

Tian ve Chen (2024) yaptıkları çalışmada, 2016-2023 yılları arasında 334 kurumdan 7832 iş ilanı içeren bir veri kümesi üzerinde metin madenciliği teknikleri kullanarak Çin'de kütüphane ve bilgi bilimi alanında alternatif akademik kariyerlerin ortaya çıkan manzarasını incelemiştir. Bulgular, alternatif akademik pozisyonların akademik, kültürel, hükümet ve ticari kurumlar arasında yaygın olarak dağıldığını ve disiplinlerarası ve uygulamalı araştırma alanlarında giderek artan bir varlığa sahip olduğunu ortaya koymuştur. Sonuçlar, dört farklı işlevsel türün ortaya çıktığını vurgulamaktadır: kaynak oluşturma, akademik hizmetler, istihbarat araştırması ve sosyal işlevler. Alternatif akademik roller, dijital bilgi ortamının değişen taleplerini yansıtarak daha fazla profesyonelleşme, uzmanlaşma ve iş birliğine doğru evrilmiştir.

Shrestha (2024) bu çalışmada, metin madenciliği ve duygu analizinin otel operasyonlarında müşteri deneyimini ve hizmet kalitesini nasıl artırabileceğini, vaka çalışması olarak Finlandiya'daki bir otele odaklanarak incelemeyi amaçlamaktadır. Araştırma, kahvaltı kalitesi, oda konforu, personel duyarlılığı, giriş deneyimi ve genel misafir memnuniyeti gibi hizmet unsurlarına odaklanarak temel temaları ve duygusal kalıpları belirlemek için müşteri geri bildirimlerini incelemektedir. Nicel bir araştırma tasarımı kullanan çalışma, konu modellemesi için Gizli Dirichlet Tahsisi (LDA) ve misafir yorumlarını duygu kategorilerine ayırmak ve tekrarlayan kalıpları ortaya çıkarmak için BERT tabanlı duygu analizi kullanmaktadır. Veri seti, Kasım 2021 ile Ekim 2024 arasında konaklama sonrası anketler aracılığıyla toplanan dahili müşteri geri bildirimlerinden oluşmaktadır ve çıkarma işlemi sırasında tüm kişisel olarak tanımlanabilir bilgiler çıkarılmıştır. Bulgular ağırlıklı olarak olumlu misafir duygusunu ortaya koymakla birlikte, gürültü kontrolü, kahvaltı çeşitliliği ve oda konforu ile ilgili kalıcı sorunları da vurgulamaktadır. Çalışma, veriye dayalı analitik yaklaşımların otel yöneticilerine hizmet eksikliklerini belirleme, operasyonları optimize etme ve kaynakları

misafir memnuniyeti üzerinde en yüksek etkiye sahip alanlara yönlendirme konusunda önemli ölçüde destek olabileceği sonucuna varmıştır.

Dzhunushalieva ve Teuber (2024) bu çalışmada, Sürdürülebilir Kalkınma Hedefleri (SKH), İnovasyon ve Bilim, Teknoloji ve İnovasyon (BTİ) hakkındaki akademik literatürün entelektüel yapısını analiz etmeyi amaçlamaktadır. Scopus ve Web of Science'ın 2015 ve 2023 yılları arasında yayınladığı 544 İngilizce yayından oluşan bir veritabanını kullanarak, bibliyometrik analizler, SKH haritalama ve metin madenciliği tekniklerinden oluşan üç yönlü bir yaklaşım ile çalışma temellendirilmiştir. Bulgular, analizde tanımlanan bir kümedeki inovasyonların ekonomik, sosyal ve çevresel boyutlar sergilediğini göstermektedir. Ayrıca, literatürde inovasyonun temel rollerinin sürdürülebilir kalkınmayı teşvik etme, ekonomik büyümeyi sağlama, işletme performansını artırma ve politikaları güçlendirme ile ilişkili olduğu bulunmuştur. Örnek literatürde, 17 hedefin tamamı SKH Haritalayıcısı tarafından belirlenmiştir. 5P (İnsanlar, Gezegen, Refah, Barış ve Ortaklıklar) arasında Refah üzerine makalelerin belirgin bir üstünlüğü görülmüştür. Başlık ve özetlerde yapılan metin incelemeleri, "sti" teriminin SGD'lerle "inovasyon"dan daha az ilişkilendirildiğini göstermektedir. Ancak, daha geniş bir kitleye ulaşmak için başlık ve özetlerde "inovasyon" teriminin kullanıldığına dair bazı kanıtlar bulunmuştur.

Gulati, Sinhababu ve Chakravarty (2023) yaptıkları çalışmada, Google Akademik'ten "akıllı kütüphaneler" konulu 81 açık erişimli dergi makalesi, metin madenciliği teknikleri kullanılarak analiz edilmiştir. Makaleler, açık erişimli olmaları ve ana bileşenleri olarak akıllı kütüphaneler içermeleri kriterine göre seçilmiştir. Makaleleri analiz etmek ve alınan makalelerdeki ilginç metin kalıplarını keşfetmek için, açık kaynaklı bir metin madenciliği aracı olan Voyant Tool kullanılmıştır. Bu araç, metinlerde yakından ilişkili olan Corpus Eşdizimlerini bulmaya ve akıllı kütüphaneler alanındaki n-gramları (çevrelerindeki bağlamla birlikte ortaya çıkan kelime dizileri) belirlemeye yardımcı olur. Sonuçlar, corpusta en sık kullanılan kelimelerin akıllı(3533), bilgi (2253), veri (1753), teknoloji (1382) ve hizmet (1327) olduğunu göstermiştir. Ayrıca, bulgular en uzun belgenin 14210 kelime, en kısa belgenin ise 1012 kelime içerdiğini göstermiştir. Genel olarak, çalışmanın bulgularının akıllı kütüphaneler ekosistemini daha derinlemesine anlamamıza yardımcı olacağı düşünülmektedir.

Piwowar-Sulej vd., (2023) tam metin madenciliği tekniklerini uygulayarak 2000 ve 2020 yılları arasında insan kaynakları yönetiminde (İKY) uzun vadeli araştırma eğilimlerini tespit etmeyi amaçlamaktadır. Başlıklara, özetlere ve anahtar kelimelere dayanan geleneksel literatür incelemelerinin aksine, bu çalışma Scopus ve Web of Science'da indekslenen İKY dergi makalelerinin tam metinlerini analiz eder. Yazarlar, TF-IDF, kümeleme algoritmaları ve HDBSCAN kullanarak tematik kümeleri ve ortaya çıkan eğilimleri otomatik olarak belirleyerek araştırmacı yanlılığını en aza indirmiştir. Sonuçlar, stratejik İKY'ye, çalışan deneyimine, uluslararası İKY'ye ve teknoloji odaklı İK uygulamalarına daha fazla odaklanma gibi İKY araştırmalarında önemli değişimleri ortaya koymaktadır. Çalışma ayrıca sürdürülebilirlik, dijitalleşme ve küresel işgücü zorlukları gibi güncel temaları da belirlemektedir. Genel olarak yazarlar, tam metin madenciliğinin İKY araştırma evrimine dair daha kapsamlı ve nesnel bir anlayış sağladığı ve gelecekteki araştırma yönlerine dair değerli bilgiler sunduğu sonucuna varmıştır.

Musilova vd., (2023) WoS ve Scopus veri tabanlarında indekslenen makalelere genel bir bakış sunarak, performans izleme makalelerinin iş stratejisiyle ilişkisini değerlendirmeyi amaçlamaktadır. Makalelerin ilk veri seti, dikkatlice geliştirilen anahtar kelimelere dayanarak elde edilmiştir. İkinci aşamada, seçimi 2019'dan önce İngilizce yazılmış işletme ve yönetim dergisi makaleleri ve incelemeleriyle sınırlandırılmıştır. Yazarlar, çalışmanın üçüncü aşamasındaki 571 makalenin tamamının özetlerini eş zamanlı olarak incelemişlerdir. Bunlar daha sonra yeniden okunup sınırlandırılmıştır; böylece, daha ileri analiz için nihai veri seti 1983 ile 2019 yılları arasında yayınlanmış 157 makaleyi içermektedir. En önemli sonuç, iş stratejisi ve performans alanının trendlere tabi olmamasıdır. İş stratejisi alanında literatür hâlâ Porter'a göre genel stratejiler ve Miles ve Snow'a göre strateji tipolojisi tarafından domine edilirken, performans ölçümü alanında pazar konumunu (satış hacmi, satış hacmindeki büyüme vb.) ve kârlılığı (kâr marjı, varlık getirisi vb.) baskın yöntem olarak belirlenmiştir. Orta Avrupa hedef kitlesi için çıkarımlar: Bibliyografik analize odaklanan önemli miktarda çalışmaya rağmen, strateji ve performans arasındaki bağlantı üzerine literatürde hala eksiklikler bulunmaktadır. Bu makale, performans alanındaki bilginin sistematikleştirilmesine ve iş stratejisi alanındaki eğilimlere dikkat çekerek Orta Avrupa hedef kitlesi için potansiyel bir değer sunmaktadır. Teorik çerçevelerin genel bir bakışı, mevcut durumun daha iyi anlaşılmasına katkıda bulunmakta ve gelecekteki araştırmalar için fırsatlar yaratmaktadır.

Belhekar ve Bhargava (2023) çalışmalarında Hintçe Kelime Derlemi (Hindi WordCorp) ve Marathi Kelime Derlemi'nin (Marathi WordCorp) gelişimi açıklanmakta ve kelime torbası modeli (BoW) kullanılarak ilgili dillerin yazılı metinlerinde kullanılan tek kelimelerin (1-gram) sıklığı raporlanmaktadır. Kelime sıklıkları on bir on yıl için verilmiştir (1920 öncesi, 1920-2020). Bu metinler kitapları (kurgu, kurgu dışı, tarih, otobiyografiler, vb.) ve dergileri içermektedir. Akademik ve referans kitapları kullanılmamıştır. Hintçe WordCorp ve Marathi WordCorp sırasıyla 640 ve 712 metin kullanmıştır. Kullanılan metinlerin, kelimelerin toplam sıklıklarının sıralamasını sabitlemek için yeterli olup olmadığını kontrol etmek amacıyla bir analiz yapılmıştır. Zipf ve Heaps yasası katsayıları, metinlerin yeterliliğini göstermiştir. Dilbilim, sosyal bilimler, metin madenciliği, makine öğrenimi vb. çeşitli alanlardaki araştırmacılar, veri setini dil ve kültürle ilgili araştırma sorularını yanıtlamak için kullanabilmektedirler. Veri setlerinin iki dilde kullanımına dair bazı örnek uygulamalar sunulmuştur. Veri seti, açık bir veri havuzunda erişime açıktır. Çalışma, Hindi ve Marathi WordCorp için veri seti oluşturma sürecini ele almaktadır. Bu nedenle, oluşturulan verilere dayanarak herhangi bir deneysel sonuç veya çıkarım yapılmamıştır. Kullanıcılar için Indian Languages Word Corpus (ILWC) adlı bir Web Uygulaması geliştirilmiştir. Metin madenciliği ve dil modelleri için gelecekteki yönelimler tartışılmıştır.

Vinayavekhin, Banerjee ve Caputo (2023) yaptıkları çalışmada sürdürülebilirlik araştırmalarının geniş ve karmaşık olduğuna değinerek bu durumun genel resmi kavramayı zorlaştırdığını belirtmişlerdir. Araştırmacıların alanı daha iyi anlamalarına yardımcı olmak için, atıf tabanlı kümeleme ve metin madenciliği teknikleri kullanılarak, Chartered Association of Business Schools (ABS) Akademik Dergi Rehberi'nde kategorilendirilen 22 akademik disiplinde sürdürülebilirlikle ilgili 25.737 makaleyi analiz ederek yönetim literatüründe sürdürülebilirliğe dair kapsamlı bir bilgi haritası oluşturulmuştur. 23 araştırma dalı belirlenerek bunlardan ilk üçü kurumsal sosyal sorumluluk, operasyonlar ve tedarik zinciri ve turizm olmuştur. Her araştırma dalının, farklı tanım kapsamlarına, analiz seviyelerine, disiplinlerarasılık seviyelerine ve 17 numaralı Birleşmiş Milletler Sürdürülebilir Kalkınma Hedefi (SKH) ile bağlantılara sahip olduğu görülmüştür. 9, 10, 12 ve 13 numaralı hedefler, 3, 14 ve 15 numaralı hedeflere göre daha fazla ilgi görmektedir. Bulgular, araştırmacılar arasında daha fazla disiplinlerarası iş birliğine ihtiyaç olduğunu göstermekte ve bunu başarmak için dört araştırma yönü önerilmektedir.

Kondo, Yoshida ve Hisano (2023) yaptıkları çalışmada 1998-2018 dönemini kapsayan ve bölge mahkemesinden yüksek mahkeme düzeyine kadar uzanan yaklaşık 110.000 Japonya mahkeme kararından elde edilen verileri kullanarak, mahkeme kararlarından böyle bir yapıyı kavrayan ve başlıca makine öğrenimi modellerinin güçlü ve zayıf yönlerini vurgulayan iki görev önermektedir. Bunlardan biri, metinsel bilgileri yasal kodlara ve geçmiş mahkeme emsallerine bağlayan maskeli dil modellemesine dayalı bir tahmin görevidir. Diğeri ise, hukuktaki gizli karşılıklı bağımlılık yapısının tahmin edildiği dinamik bağlantı tahmin görevidir. Sonuç olarak gelecekteki gelişmeler için içgörüler elde etmek amacıyla başlıca makine öğrenimi modelleri arasında nicel ve nitel karşılaştırmalar yapılmıştır.

Valverde-Berrocso vd., (2022) 2011 ve 2020 yılları arasında yayınlanmış çalışmaların sistematik bir literatür incelemesini yürüterek dezenformasyona karşı eğitimsel yanıtları incelemektedir. Yazarlar, vatandaşları sahte haberlere ve yanlış bilgilere karşı güçlendirmek için hangi pedagojik modellerin ve çoklu okuryazarlık yaklaşımlarının önerildiğini anlamayı amaçlamaktadır. Veriler, PRISMA yönergeleri kullanılarak Web of Science, Scopus ve ERIC'ten toplanmış ve 76 hakemli makaleden oluşan nihai bir örneklem oluşturulmuştur. Veri kodlama ve analizi için referans yönetimi, kavram haritalama ve metin madenciliği araçları (VOSviewer, NVivo) kullanılmıştır. İnceleme, özellikle medya okuryazarlığı ve bilgi okuryazarlığı olmak üzere baskın okuryazarlık çerçevelerini belirlerken, veri okuryazarlığı ve haber okuryazarlığının ortaya çıkan önemini de belirtmektedir. Bulgular üç ana pedagojik yaklaşımı vurgulamaktadır: yeterlilik temelli stratejiler, içerik odaklı öğretim ve vatandaşlık eğitimi. Yaygın öğretim uygulamaları arasında eleştirel düşünmeyi, işbirlikçi bilgi inşasını ve yurttaşlık değerlerini geliştirmeyi amaçlayan atölyeler ve yapılandırılmış ders planları yer almaktadır. Yazarlar, dezenformasyonla mücadelede çoklu okuryazarlık eğitiminin önemli olduğu ve gelecekteki araştırmaların daha disiplinlerarası ve pedagojik açıdan çeşitli yaklaşımlar benimsemesi gerektiği sonucuna varmaktadır.

Urru vd., (2022) bu çalışmada, hızla büyüyen COVID-19 bilimsel literatüründeki başlıca araştırma konularını belirlemeyi ve Kasım 2019 ile Aralık 2021 arasındaki evrimlerini analiz etmeyi amaçlamaktadır. Yazarlar, PubMed, Scopus ve Web of Science'da indekslenen COVID-19 ile ilgili yayınlara odaklanmaktadır. Otomatik veri çıkarma ve sıkı dahil etme kriterleri (DOI'li İngilizce makaleler) kullanılarak, veri seti dil tespiti, kök bulma ve belirteç azaltma yoluyla derlenmiş ve ön işleme tabi tutulmuştur.

Gizli tematik yapıları tespit etmek ve model seçimi için anlamsal tutarlılık ve ayrıcalık ölçütleriyle desteklenen zamansal eğilimleri incelemek için Yapısal Konu Modellemesi (STM) kullanılmıştır. Analiz, klinik sunum, eğitim, aşı geliştirme, epidemiyoloji, tele-tıp, ruh sağlığı, tahmin ve moleküler biyoloji dahil olmak üzere on ayrı konuyu ortaya koymuştur. Sonuçlar, araştırma ilgisinde zaman içinde kaymalar olduğunu göstermiştir; örneğin, aşılar ve epidemiyolojiye erken odaklanma ve daha sonra ruh sağlığı ve moleküler biyolojiye artışlar. Çalışma, COVID-19 araştırmalarının tematik olarak çeşitli ve dinamik olduğu ve kapsamlı literatür taraması için birden fazla veri tabanına güvenmenin önemli olduğu sonucuna varmıştır. Bulgular, pandemi döneminde sistematik ve hızlı incelemeler yapmak için değerli bir referans sağlamaktadır.

Ahadi vd., (2022) bu sistematik incelemede, 2000-2022 yılları arasında eğitim araştırmalarında metin madenciliği ve doğal dil işleme (NLP) uygulamalarındaki son durumu incelemektedir. Yazarlar, metinle ilgili, eğitimle ilgili ve yöntemle ilgili terimleri birleştiren genişletilmiş bir anahtar kelime stratejisi kullanarak Scopus ve Web of Science'tan yayınlar toplamıştır. 9.666 kayıt tarandıktan sonra, 981 yayından oluşan nihai set, Bibliometrix R paketi ile bibliyometrik yöntemler kullanılarak analiz edilmiştir. Analiz, öğrenme analitiği, MOOC veri analizi, otomatik yazma değerlendirmesi, eğitimde duygu analizi ve öğretim ve öğrenme için NLP araçlarının geliştirilmesi gibi ana temaları belirlemiştir. Çalışma, bazı alanların (örneğin, öğrenme analitiği, yazma analitiği) olgunlaştığını, bazılarının ise (örneğin, ontoloji tabanlı metin madenciliği, derin öğrenme yöntemleri) yeterince kullanılmadığını ortaya koymuştur. Yazarlar, metin madenciliğinin eğitim araştırmalarında giderek daha önemli bir rol oynadığı sonucuna varmış ve daha fazla araştırmayı hak eden birkaç yeni ortaya çıkan alana dikkat çekmişlerdir.

Dvorak vd., (2022) tarafından yapılan çalışma, iş stratejisi alanındaki bilgiyi sistematize etmeyi ve teorik ve ampirik gelişimindeki uzun vadeli eğilimleri belirlemeyi amaçlamaktadır. İş düzeyi ve rekabet stratejisi araştırmalarına odaklanan yazarlar, özellikle Porter'ın genel stratejileri, Miles ve Snow'un tipolojisi ve Mavi Okyanus Stratejisi olmak üzere temel stratejik çerçevelerin akademik yayınlarda nasıl kullanıldığını analiz etmektedirler. Bibliyometrik teknikler ve metin madenciliği ile tamamlanan sistematik bir literatür taraması kullanarak, Web of Science ve Scopus veri tabanlarından makaleleri taramışlar ve nihayetinde 1983 ile 2019 arasında yayınlanmış 430 çalışmadan oluşan rafine bir veri setini değerlendirmişlerdir. Bulguları, iş stratejisine

yönelik akademik ilginin giderek arttığını, az sayıda önde gelen dergide etkili yayınların güçlü bir şekilde yoğunlaştığını ve ampirik, nicel yönelimli araştırmaların açık bir şekilde baskın olduğunu göstermektedir. Analiz ayrıca, Porter ve Miles & Snow'un çerçevelerinin en etkili kavramsal temeller olmaya devam ettiğini, Mavi Okyanus Stratejisi gibi daha yeni yaklaşımların ise nispeten küçük bir rol oynadığını ortaya koymaktadır. Hızlı teknolojik değişime ve gelişen rekabet ortamlarına rağmen, çalışma, bu klasik teorilerin yerini temelde yeni bir stratejik paradigmanın almadığı sonucuna vararak hem alanın istikrarını hem de yeni kavramsal gelişmelere olan potansiyel ihtiyacı vurgulamaktadır.

Song, Wu ve Wang (2021), 1995 ve 2020 yılları arasında yayınlanmış 1.282 Web of Science belgesi toplamış ve R'deki Bibliometrix paketi aracılığıyla uygulanan performans analizi ve bilim haritalama tekniklerini (özellikle anahtar kelime eş zamanlılığı, ortak yazarlık ağları ve ortak atıf analizi) kullanmışlardır. Veri kümesi, genişletilmiş bir metin madenciliğiyle ilgili anahtar kelime kümesi kullanılarak oluşturulmuş ve yönetimle ilgili Web of Science kategorilerine göre filtrelendi. Sonuçlar, yönetimde metin madenciliğinin kullanımında üstel bir büyüme olduğunu, Amerika Birleşik Devletleri'nin yayın çıktısı ve atıf etkisinde lider olduğunu göstermiştir. Konu modelleme, duygu analizi ve doğal dil işleme baskın araştırma temaları olarak ortaya çıkmıştır. Makalenin geri çekilmesine rağmen, ilk sonuçları metin madenciliğinin yönetim araştırmalarında giderek daha etkili bir metodolojik yaklaşım haline geldiğini ve alandaki kavramsal, sosyal ve entelektüel yapıları şekillendirdiğini göstermiştir.

Mirzyńska vd., (2021) bu çalışmada, 2011-2018 yılları arasında Web of Science ve Scopus'tan 3.486 yayın ve 106.504 tweet ile ilgili dairesel ekonomi temaları hakkındaki tartışmaların eş zamanlı veya eşlik eden kavramlarındaki eğilimi incelemektedir. Metin madenciliği kullanılarak, Jaccard benzerlik endeksi yıllara bölünmüş olarak hesaplanmıştır. Sonuçlar, dairesel ekonomi tartışmalarına eşlik eden temalarda zaman içinde değişiklikler olduğunu ve genel kamuoyu tartışmalarında araştırmayla ilgili anahtar kelimelerin artan tanınırlık eğilimini ortaya koymaktadır; ağırlıksız benzerlik 2018'de %39,44'e ulaşmıştır. Twitter anahtar kelime araştırma perspektifimiz, tüketicilerin günlük aktiviteleriyle yakından ilişkili anahtar kelimeler aracılığıyla dairesel ekonominin gelişiminde tüketicinin rolünün dikkate alınması gerektiğini göstermektedir

Stansfield, O'Mara-Eves ve Thomas (2017) bu çalışmada, metin madenciliği tekniklerinin özellikle farklı terminolojilerle karakterize edilen alanlarda sistematik incelemelerde arama terimlerinin geliştirilmesini nasıl destekleyebileceğini incelemektedir. Makale, hassas ve hassas arama terimlerini belirlemenin zorluklarını tartışmakta ve bu bağlamda hassasiyeti artırma, hassasiyeti artırma, veri tabanları arasında çeviriye yardımcı olma, entegre arama ve tarama sistemleri ve nesnel arama stratejileri oluşturma dahil olmak üzere metin madenciliğinin beş ana uygulamasını incelemektedir. Literatür taraması, vaka çalışmaları ve TF-IDF, Termine, kümeleme algoritmaları ve kelime sıklığı analizi gibi araçların karşılaştırılmasının bir kombinasyonu yoluyla yazarlar, metin madenciliğinin ilgili terminolojiyi genişletmeye, alakasız tematik kümeleri tespit etmeye ve arama stratejilerini yinelemeli olarak iyileştirmeye yardımcı olabileceğini göstermektedir. Bulguları, bu araçların hem sistematik arama stratejilerinin hassasiyetini hem de hassasiyetini artırabileceğini ancak bu araçların etkinliği büyük ölçüde örneklem seçimine, araç yeteneklerine ve yorumlayıcı uzmanlığa bağlı olduğundan insan yargısının yerini tamamen alamadığını vurgulamaktadır. Çalışma, metin madenciliğinin arama stratejisi kalitesini artırmada anlamlı faydalar sağladığı sonucuna varmıştır, ancak optimum kullanımının geleneksel, uzman odaklı arama geliştirme yöntemleriyle dikkatli bir şekilde bütünleştirilmesini gerektirmektedir.

Alksher vd., (2016) bu çalışmada, fikir madenciliği olarak bilinen, yapılandırılmamış metinsel verilerden yenilikçi fikirler çıkarmak için kullanılan hesaplamalı yöntemlerin kapsamlı bir incelemesini sunmaktadır. Makale, fikir madenciliğinin kavramsal temellerini inceleyerek, karar alma ve inovasyon süreçlerini destekleyen yeni, eyleme geçirilebilir kavramları belirlemeye odaklanmasıyla onu geleneksel metin madenciliğinden ayırmaktadır. Metin ön işleme, desen çıkarma, vektör uzayı modelleme, kümeleme ve benzerliğe dayalı değerlendirme alanındaki önceki çalışmalardan yararlanan yazarlar, teknoloji, tıp ve sosyal bilimler gibi çeşitli alanlarda anlamlı fikir desenlerini belirlemek üzere tasarlanmış algoritmaları ve modelleri incelemektedir. Sentezleri, topluca yenilikçi metinsel desenlerin tespitini sağlayan durdurma kelimesi filtreleme, sözcük türü etiketleme, kök bulma, metin deseni oluşturma ve fikir tabanlı benzerlik ölçümleri gibi yöntemleri vurgulamaktadır. Bulgular, fikir madenciliği yaklaşımlarının alan bilgisiyle birleştirildiğinde ürün geliştirme, sinyal tespiti ve stratejik analiz gibi görevleri etkili bir şekilde destekleyebileceğini göstermektedir. Çalışma, fikir madenciliğinin hala gelişmekte olmasına ve evrensel olarak kabul görmüş

bir tanımının olmamasına rağmen, metin deseni oluşturma ve fikir-özgü değerlendirme ölçümlerindeki ilerlemelerin, büyük metin korpuslarından yararlı fikirleri otomatik olarak belirleme yeteneğini önemli ölçüde artırdığı ve gelecekteki araştırmalar için umut verici yönler sunduğu sonucuna varmıştır.

Bemke-Switilnik ve Drabek (2016) bu çalışmada, Polonya merkezli Merkez Madencilik Enstitüsü Madencilik ve Çevre Araştırma Raporları (RRCMIME) dergisinin on bir yıllık dönemdeki yayın etkinliğini ve atıf etkisini incelemektedir. Amaç, derginin Web of Science, Scopus, BazTech ve Google Akademik gibi birden fazla veri tabanında nasıl atıf aldığını değerlendirmek ve indekslenmemiş ulusal dergilerin bibliyometrik analizi için en doğru kaynakları belirlemektir. Manuel veri toplama stratejisi kullanan yazarlar, dergi başlığı atıf formatlarındaki tutarsızlıkları gidermek için yazar tabanlı arama prosedürleri uygulayarak 2002 ve 2012 yılları arasında yayınlanmış 1216 makale için atıf verisi derlemiştir. Analiz, BazTech'in en yüksek kapsamı sağladığı ve Google Akademik'in BazTech tarafından indekslenen atıfların yalnızca %84'ünü yakaladığı 800 benzersiz atıfı ortaya koymaktadır. Bulgular, atıfların 2008'den sonra önemli ölçüde arttığını, çoğu atıfın bildiri kitaplarından geldiğini ve yerli atıfların uluslararası atıflara göre baskın olduğunu göstermektedir. Çalışma, BazTech gibi ulusal veri tabanlarının yerel, indekslenmemiş dergileri değerlendirmek için daha güvenilir olduğu ve Polonya madencilik bilimindeki atıf uygulamalarının büyük ölçüde yerel kapsamda kaldığı sonucuna varmıştır.

Akılan (2015) bu çalışmada, hızla büyüyen yapılandırılmamış metinsel veri hacmini ele almakta ve metin madenciliğini, metinden anlamlı örüntüler çıkarmak için bilgi alma, doğal dil işleme, makine öğrenimi ve veri madenciliğini birleştiren çok disiplinli bir teknik olarak konumlandırmaktadır. Çalışmanın amacı, metin arıtma ve bilgi damıtımı olmak üzere iki ana bileşenden oluşan metin madenciliği için kavramsal bir çerçeve sunmak ve alandaki teknolojik gelişmeleri ve devam eden zorlukları incelemektir. Makale, yeni veri toplamak yerine önceki literatürü sentezlemektedir; metodolojik katkısı, ham metnin ara gösterimlere (belge tabanlı veya kavram tabanlı) nasıl dönüştürüldüğünü ve kümeleme, kategorizasyon, öngörücü modelleme ve gizli anlamsal dizinleme gibi analitik tekniklerin nasıl uygulanabileceğini açıklamaktır. Temel bulgular, anlamsal analiz, çok dilli metin işleme, büyük ölçekli depolama ve verimli ara formlara olan ihtiyaçtaki zorlukları vurgulamaktadır. Makale, gelecekteki araştırmaların ölçeklenebilir anlamsal işleme, çok dilli arıtma algoritmaları ve büyük gövdeli metin

madencilik uygulamalarını desteklemek için geliştirilmiş depolama altyapılarına odaklanması gerektiği sonucuna varmaktadır.



3. YÖNTEM

Bu çalışmada Türkiye’de İşletme anabilimdalı sayısal yöntemler bölümünde yazılmış tezlerin metin madenciliği yöntemleriyle analiz edilmesi amaçlanmaktadır. Bu bağlamda materyal olarak kullanılan veri setinin analizi için LDA algoritması ile konu modelleme yöntemi kullanılmıştır.

3.1. Veri Toplama ve Arama Stratejisi

Veri setini 2025 yılına kadar Sayısal Yöntemler alanında hazırlanan ve YÖK Ulusal Tez merkezi web sitesi üzerinden erişim sağlanan 341 yüksek lisans tezi oluşturmaktadır. Veri seti oluşturma süreci aşağıdaki gibidir:

YÖK Ulusal Tez Merkezi web sitesine erişim sağlanarak detaylı arama kısmı seçilmiş, ana bilim dalı kısmına ‘İşletme Anabilim Dalı’ ifadesi yazılmıştır. Bilim dalı kısmına önce ‘Sayısal Yöntemler Bilim Dalı’ yazılarak tarama yapılmış ardından ana bilim dalı kısmı aynı kalmak kaydıyla bilim dalı kısmına ‘Sayısal Yöntemler ve Yönetim Bilimi Bilim Dalı’ yazılarak arama işlemi yapılmıştır. Her seferinde tez türü yüksek lisans olarak işaretlenerek ‘bul’ bağlantısı ile sonuçlar listelenmiştir. Böylelikle ana bilim dalı işletme olmak kaydıyla bilim dalları ‘Sayısal Yöntemler Bilim Dalı’ ve ‘Sayısal Yöntemler ve Yönetim Bilimi Bilim Dalı’ olan yüksek lisans tezleri görüntülenmiştir. Yapılan aramaya dair görüntüler şekil3.1 ve şekil3.2 de gösterilmiştir.

YÖKSEKÖĞRETİM KURULU BAŞKANLIĞI
Tez Merkezi

Ana Sayfa Tarama Mevzuat İstatistikler SSS Yasal Uyarı Bize Ulaştın Yardım

Detaylı Tarama Gelişmiş Tarama Son Eklenen Tezler Hazırlanmakta Olan Tezler

Detaylı Tarama

Üniversite Seç

Enstitü Seç

Ana Bilim Dalı İşletme Ana Bilim Dalı

Bilim Dalı Sayısal Yöntemler ve Yönetim Bilimi Bilim Dalı

Konu Konu

Dizin Dizin

Tez Türü Yükle

İzin Durumu Seç

Durumu Seç

Dil Seç

Grubu Seç

Özet Özet

Bul Temizle

Şekil 3.1. Detaylı tarama

Detaylı Tarama Gelişmiş Tarama Son Eklenen Tezler Hazırlanmakta O

Üniversite Seç

Enstitü Seç

Ana Bilim Dalı İşletme Ana Bilim Dalı

Bilim Dalı Sayısal Yöntemler Bilim Dalı

Konu Konu

Dizin Dizin

Tez Türü Yükle

İzin Durumu Seç

Durumu Seç

Dil Seç

Grubu Seç

Özet Özet

Bul Temizle

Şekil 3.2. Detaylı arama

3.2. Durak Kelimeler (Stopwords)

Bilgi alma görevlerinde yer alan belgeler, araştırmacılar için yararlı olmayan çok büyük bir oranda veri içerir. Bu nedenle, bu tür verileri tanımlamak ve işleme başlamadan

önce veri kümesinden çıkarmak için otomatik bir yöntem geliştirilmesi arzu edilir. Bu tür veriler durak kelimeleri olarak bilinir.

Duraklama-durdurma-durak kelimeleri ilk olarak 1958’de H.P. Luhn tarafından tanıtılmıştır (Ladani ve Desai.,2020). Durak kelimeleri, bir belgede en sık geçen ve genellikle gerekli olmayan az miktarda bilgi içeren kelimelerdir (Kaur ve Buttar.,2018). Örneğin, İngilizce’de a, about, the, above, after, is, again, against, all, am, an, and, any, are, as, at, be, been, could, do, does, during, vb. gibi metinde en sık geçen ancak az bilgi içeren bazı kelimeler vardır. Bu kelimelere duraklama-durdurma kelimeleri denir. Bir durak kelimeleri kümesi "durak kelime listesi(Stopwordslist)" veya "durdurma kelime külliyyatı(Stopword corpus)" olarak bilinir.

Genel olarak araştırmacı, metin üzerinde çalışırken, anlamsal referans veya belgelerin anlamsal yapısını etkileyen, sözde durdurma sözcükleri de dahil olmak üzere çeşitli kelimelerle karşılaşır (Rani ve Lobiyal, 2022). İndeksleme, konu modelleme, metin sınıflandırma ve bilgi alma gibi NLP görevlerinin doğruluğunu ve verimliliğini sağlamak için (Blanchard, 2007; Wilbur ve Sirotkin, 1992), genellikle “duraklama sözcükleri” olarak adlandırılan bilgilendirici olmayan sözcüklerin ön işleme adımında kaldırılması gerekir. Duraklama sözcükleri, birçok farklı doğal dil belgesinde veya bir belgedeki metnin bölümlerinde sıklıkla görünür, ancak ait oldukları metin bölümü hakkında çok az bilgi taşırlar. Durdurma kelimelerinin ortadan kaldırılması, metin sayısını genellikle %35 ila %45 oranında azaltır ve metin madenciliği yöntemlerinin uygulanmasını daha etkili hale getirir (Jha vd., 2016). Dolayısıyla, durdurma kelimelerinin kaldırılması, yapılandırılmamış metinde sinyal-gürültü oranını azaltabilir ve böylece belirli bir görev için önemli olabilecek terimlerin istatistiksel önemini artırabilir.

3.3. Konu Modellemesi

Bu çalışmada, kelime kümeleri temel alınarak bir metin kümesindeki soyut konuları keşfetmek için konu modelleme yaklaşımı kullanılmıştır. Analiz süreci ‘Orange Data Mining’ programı kullanılarak yürütülmüştür.

Orange yazılımı, Slovenya’daki Ljubljana Üniversitesi tarafından sağlanan açık kaynaklı bir programdır. Orange, kullanıcılarının veri analizi yapmalarına, veri akışını görmelerine ve çok daha verimli olmalarına olanak tanıyan güçlü bir açık kaynaklı makine öğrenimi ve veri görselleştirme platformudur (Nuriu, 2023). Kodlama ya da

programlama alanında tecrübe sahibi olmayan kullanıcılar tarafından da kolaylıkla çalıştırılabilmesi araştırmada yürütülen analizlerde bu programın tercih edilme sebebidir.

Metin madenciliği için konu modellemesi, metin belgelerindeki kalıpları belirlemek için kullanılır. Belge koleksiyonlarından gizli konuları keşfetmek için kullanılan istatistiksel bir yaklaşımdır. Konu modelleri, konuların ağırlıklı karışımı olarak modellenen metin belgelerinden türetilir ve her konu, kelimeler üzerinde bir olasılık dağılımıdır. Konu modelleme teknikleri, haber makaleleri, web sayfaları ve sosyal haberler gibi farklı belgelerden temalar bulmak için kullanılır. Konu modelleme teknikleri, metin külliyatını aramak, taramak ve özetlemek için kullanılır. Konu modelleme, belgelerden gizli konuları keşfeder. Belgeler bu konularla etiketlenir ve bu etiketler, tahminlerden arama, özetleme ve düzenleme için kullanılır.

Konu modellemesi yapılırken alternatif algoritmalar mevcuttur. Öne çıkan konu modelleme yöntemleri arasında Latent Dirichlet Allocation (LDA) (Y. Guo vd., 2017), cümle-LDA (J. Li vd., 2020), yapısal konu modelleme (Shi & Chen, 2021), latent semantik analiz (Xu, 2018) ve dinamik konu modeli yer almaktadır. Bu çalışmada, ilk olarak 2003 yılında David Blei, Andrew Ng ve Michael Jordan tarafından tanımlanan LDA yöntemi kullanılmıştır.

LDA, metin verilerini analiz etmek için gelişmiş ve etkili bir yöntemdir. Denetimsiz bir öğrenme algoritması olarak, önceden tanımlanmış anahtar kelimeler gerektirmez. Gizli konuları çıkararak, LDA bir belge koleksiyonunun kapsamlı bir modelini sağlar, büyük ölçekli veri kümelerinde kapsamlı bir konu modelinin oluşturulmasını mümkün kılar ve çevrimiçi incelemelerin çoklu konu analizi için metodolojik bir garanti sunar (Maier vd., 2021). LDA'nın uygulamaları çeşitlidir ve çevrimiçi platformları, pazarlamayı, sosyal medyayı ve müşteri geri bildirim analizlerini kapsar (Larsen ve Thorsrud, 2019; Ye ve Xue, 2021). Son zamanlarda, Üretken Yapay Zeka uygulamalarının incelenmesinde metin analizi için LDA modelleme yöntemleri ilgi görmeye başlamıştır (Alkamli ve Alabduljabbar, 2024; Mujahid vd., 2023; Taecharungroj, 2023).

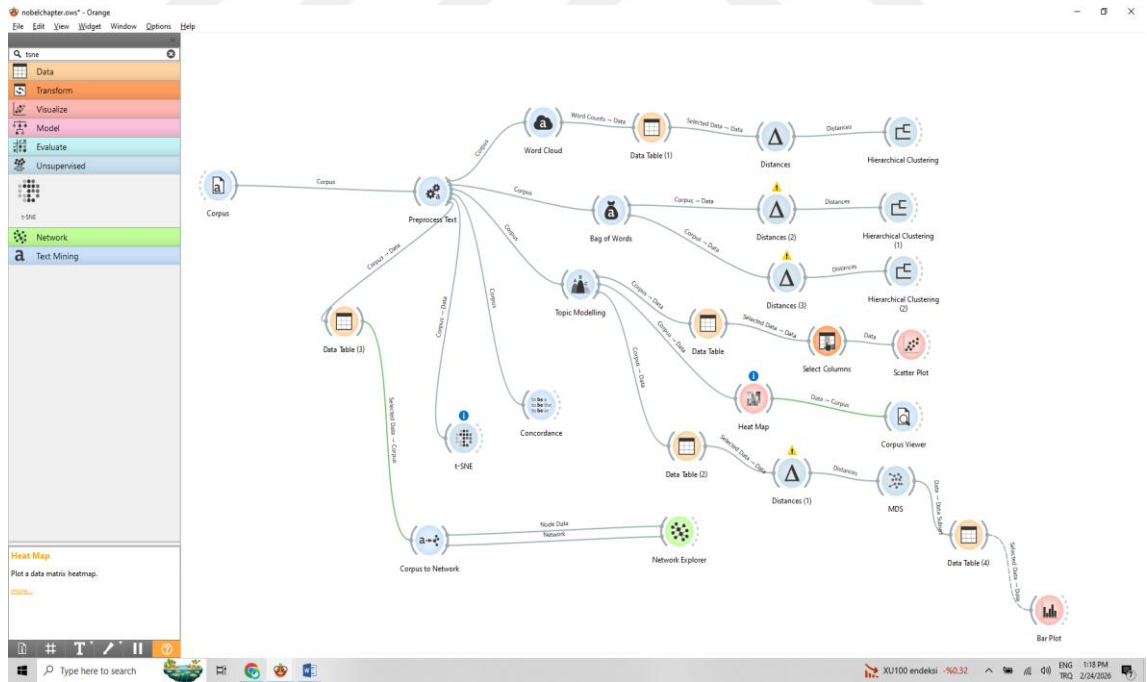
Orange, çok çeşitli standart ve standart olmayan makine öğrenimi ve veri madenciliği algoritmalarının yanı sıra veri girişi ve manipülasyonu için rutinler içeren, Python'da daha kolay yapılabilen bazı işlevleri uygulayan Python tabanlı modüllerden oluşan bir koleksiyondur (Demsar ve Zupan., 2021). Orange, makine öğreniminde

algoritmalarını geliştirmek ve test etmek isteyen ve mümkün olduğunca kodu yeniden kullanmak isteyen deneyimli kullanıcılar ve araştırmacılar için tasarlanmış olsa da veri analizi için kısa Python komut dosyaları yazabilen, alana yeni girenler için de uygundur. Böylelikle programlama tecrübesi olmayan kullanıcılar da programı kolaylıkla kullanabilmektedir.

Orange veri madenciliği widget'larla çalışır. Bu widget'lar Orange'ın hesaplama birimleridir. Widget'lar verileri okur, işler ve görselleştirir. Bunları kullanarak kümeleme de yapabilir veya tahmine dayalı modeller oluşturulabilir (Nuriu, 2023).

Orange, kullanıcının widget'lar yerleştirdiği ve bir veri analizi iş akışı oluşturduğu bir Canvas arayüzünden oluşur. Widget'lar, veri analizi iş akışlarının yapı taşlarıdır. Bir widget, ekranın sol tarafındaki widget listesinden seçtiğiniz şeye bağlı olarak eylemler gerçekleştirir (Mala vd., 2024). Başka bir deyişle, widget'lar verilerin keşfedilmesine yardımcı olur.

Orange Canvas, Orange için görsel bir programlama ortamıdır. Burada tüm widget'lar entegre edilir ve widget'lar arasındaki tüm ilişkiler belirginleşir.



Şekil 3.3. Orange model

3.4 Orange3 Text Mining: Özellikler ve Uygulama Alanları

Orange, ücretsiz olarak erişilebilen açık kaynaklı bir veri madenciliği aracıdır ve Python ile yazılmış bilgisayar tabanlı zeka ve veri madenciliği programlama dilidir (Kaur

ve Singh, 2024). Orange, veri madenciliği, bilgisayar tabanlı zeka ve veri analizi için bünyesinde bir den çok eklenti ve araç barındıran görsel bir programlamadır (Demsar vd., 2004). Bu eklenti ve araçlar, temel, alt küme kararı ve ön işlemeden, öğrenme hesaplamalarının gözlemsel değerlendirmesine ve algısal gösterime kadar birçok işlem yapmaya olanak tanır (Mohi., 2020). Orange, yenilik, güvenilirlik veya kalite kavramı söz konusu olduğunda çok etkilidir (Ratra ve Guila, 2020).

Orange programının işlevleri, farklı widget'lar (örneğin, bag of words, corpus, stopwords, text mining vb.) ile görsel olarak temsil edilir. Arayüzde her widget'ın kısa bir açıklaması mevcuttur. Programlama, widget'ları canvasa yerleştirerek ve giriş ve çıkışlarını bağlayarak gerçekleştirilir (Jovic vd., 2014). Arayüz oldukça özenli ve görsel olarak çekici olup, keyifli bir kullanıcı deneyimi sunmaktadır.

Orange yazılımı, isteğe bağlı eklentiler yüklenerek genişletilebilir. Bünyesinde bulundurduğu eklentilerden text mining, metinsel verileri kategorize etmek için Kümeleme, kayıtları, gözlemleri veya vakaları benzerliklere dayalı olarak sınıflara ayırma tekniğidir (Maulana ve Wibovo, 2025). Metin kümeleme, benzer içeriğe sahip metinlerin bir araya getirilmesi için bir metin grubunu kümelere ayırma işlemidir. Bu yöntem, elde edilen verilerden son derece ilgili bilgileri belirleyerek bir belge içindeki benzerlikleri değerlendirmeyi kolaylaştırabilir.

Programda bulunan eklenti, metin ön işleme sürecine olanak tanır. Ham veriler, harfleri küçük harfe dönüştürme, noktalama işaretlerini, sembolleri ve sayıları kaldırma, tokenizasyon, durdurma(stopwords) kelimelerini kaldırma ve kök bulma(stemming) gibi metin ön işleme aşamalarından geçer (Hibatullah vd., 2025). Bu ön işleme sürecinin sonunda, ham veriler analiz için uygun hale gelmiş olur (Muñoz, 2019).

Sonuç olarak Orange, keşifsel veri analizi ve görselleştirme için kullanılan açık kaynaklı bir veri madenciliği araç setidir. Python üzerine kuruludur, deney seçimi platformu görevi görür ve yenilik, güvenilirlik veya kalite gerektiğinde oldukça kullanışlıdır (Arabiat, 2025).

Bu süreç, verilerin kalitesini ve sınıflandırmanın doğruluğunu artırmayı amaçlar.

4. ANALİZLER VE BULGULAR

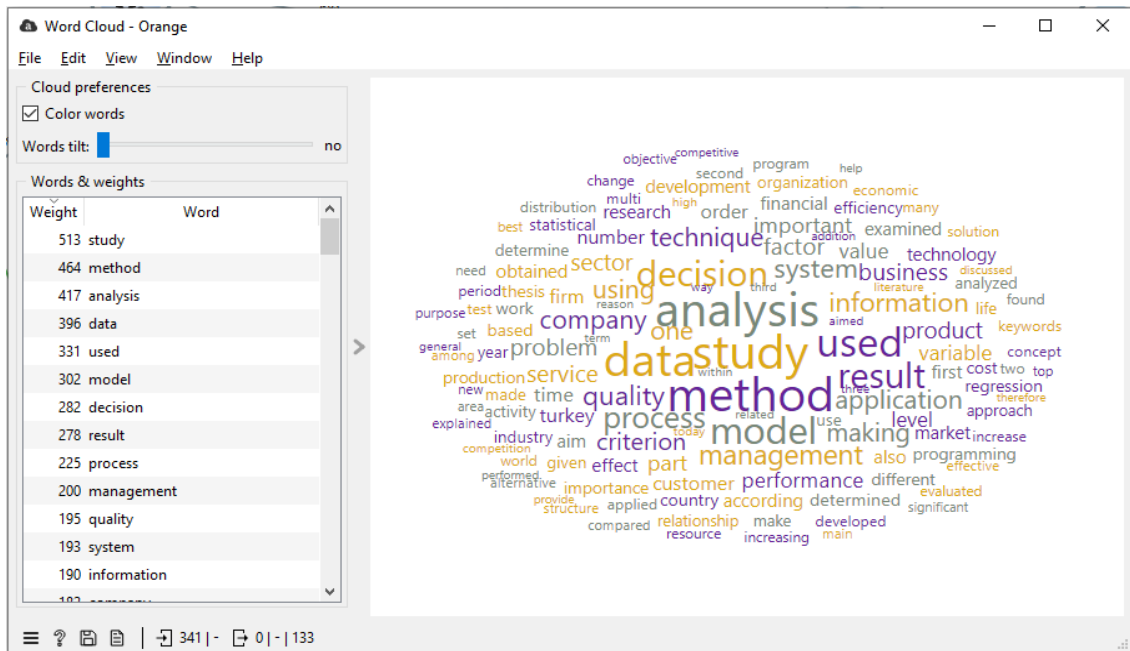
Çalışmanın bu bölümünde Orange Data Mining programı aracılığıyla YÖK Ulusal Tez Merkezi web sitesi üzerinden erişilmiş olan verilerin analizi ve bulgularına yer verilmiştir.

4.1. Kelime Sıklığı

Orange Data Mining programında Metin ön işleme eklentisi kullanılarak metin belgesi farklı türde belirteçlere dönüştürülmüş, Kelime bulutu seçeneği kullanılarak tüm belirteçler görsel olarak görüntülenmiştir(Şekil 4).

Kelime bulutları, kelimelerin geçme sıklığını görüntülemek için kullanılabilir (Hasan vd., 2021). Bir kelimenin geçme sıklığı ne kadar yüksekse, kelime bulutundaki harflerinin boyutu da o kadar büyük olur (Thange vd., 2021). Kelime bulutları, metin ön işleme aşamasının bir varyasyonudur. Kelimelerin rengi, görüntüyü daha çekici ve anlaşılması daha kolay hale getirir (Anggraini ve Suroyo, 2019). Metin ön işleme sürecinden geçtikten sonra, kelime bulutu aşağıdaki resimdeki gibi bir görüntü sunar. Aşağıdaki sonuç görüntüsü, en yüksek kelime sayısına sahip kelime bulutunu göstermektedir.

Kelime bulutu işleminin sonuçları, bu çalışmada en sık geçen kelimenin ‘Study’ olduğunu (513 kez) göstermektedir (Şekil 6’ya bakınız). Ayrıca, ‘method’ kelimesi 464 kez geçme yoğunluğuyla ikinci en sık geçen kelimedir. Ardından, en sık geçen kelimeler sırasıyla ‘analysis’, ‘data’, ‘used’, ‘model’, ‘decision’, ‘result’



4.1.1 Kelime Bulutu Bulgularının Tartışılması ve Sonraki Analizlerle Entegrasyonu

Kelime bulutu analizi, çalışmada incelenen metin korpusunda en sık tekrar eden ve literatürün kavramsal ağırlık merkezini oluşturan terimleri görselleştirmek amacıyla kullanılmıştır. Bu analiz, konu modellemesi, hiyerarşik kümeleme ve çok boyutlu ölçekleme sonuçlarını destekleyici nitelikte betimleyici bir araç olarak değerlendirilmiştir.

Kelime bulutunda öne çıkan terimler arasında analysis, decision, data, model, result, method, system, information, management, quality, performance, customer ve service kavramlarının belirgin biçimde baskın olduğu görülmektedir. Bu bulgu, literatürün güçlü biçimde analitik, veri temelli ve karar verme odaklı bir karakter taşıdığını ortaya koymaktadır. Özellikle analysis ve data kavramlarının yüksek frekansla yer alması, incelenen çalışmaların büyük ölçüde nicel yöntemlere ve ampirik analizlere dayandığını göstermektedir.

Kelime bulutu sonuçları, konu modellemesi analizinde ortaya çıkan temalarla yüksek düzeyde örtüşmektedir. LDA analizinde belirlenen modelleme, performans değerlendirme, karar verme ve bilgi sistemleri temalarının, kelime bulutunda da baskın kavramlar olarak yer alması, yöntemler arası tutarlılığı güçlendirmektedir. Bu durum, literatürde analitik modellerin ve sistem temelli yaklaşımların merkezî bir konumda olduğunu teyit etmektedir.

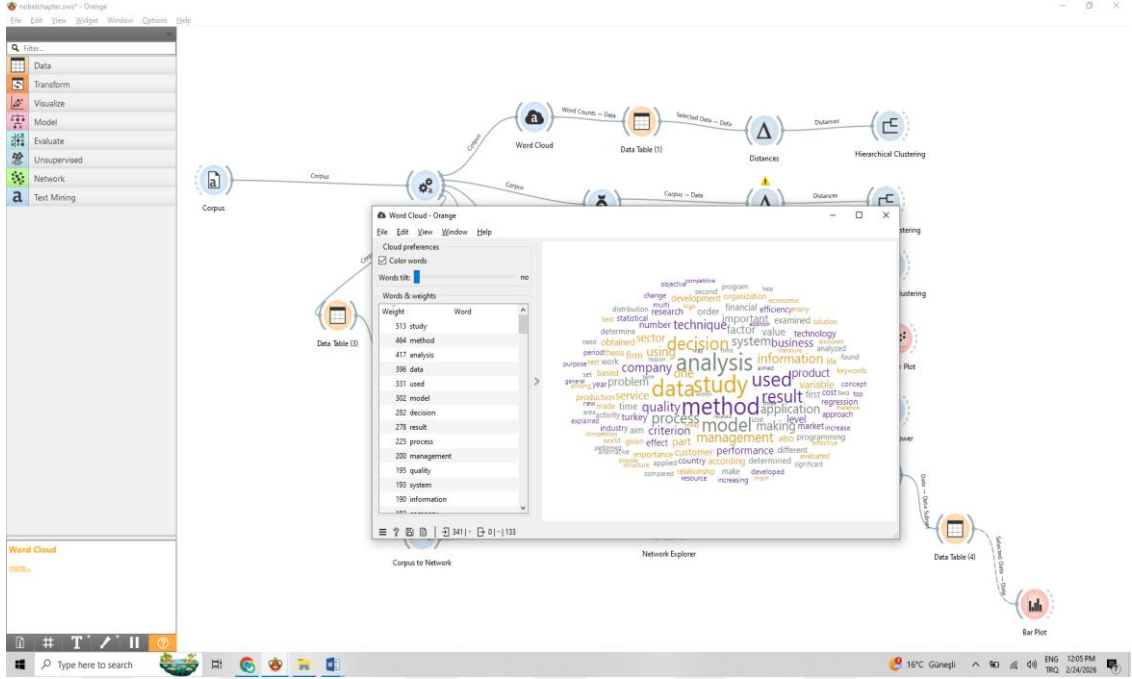
Ayrıca kelime bulutunda customer, quality, service ve value gibi kavramların öne çıkması, müşteri odaklılık ve hizmet kalitesine yönelik çalışmaların literatürde önemli bir yer tuttuğunu göstermektedir. Bu bulgu, hiyerarşik kümeleme analizinde belirlenen müşteri deneyimi ve algılanan değer kümesiyle doğrudan uyumludur. MDS analizinde bu temaların diğer temalardan görece ayrılmış biçimde konumlanması, kelime bulutu bulgularıyla birlikte değerlendirildiğinde, müşteri merkezli yaklaşımların literatürde bağımsız ve derinleşen bir araştırma alanı haline geldiğini ortaya koymaktadır.

Kelime bulutunda management, process, system ve decision gibi kavramların birlikte görünmesi, stratejik ve operasyonel düzeylerin literatürde iç içe ele alındığını göstermektedir. Bu durum, hiyerarşik kümeleme analizinde ortaya çıkan stratejik–operasyonel bütünleşme yapısını desteklemekte; karar verme süreçlerinin hem yönetsel hem de teknik boyutlarıyla ele alındığını ortaya koymaktadır.

Genel olarak değerlendirildiğinde, kelime bulutu analizi literatürde baskın olan kavramsal eksenlerin veri analitiği, karar verme, performans ve kalite yönetimi ile müşteri odaklı değer yaratma etrafında yoğunlaştığını göstermektedir. Bu sonuçlar, konu modellemesi ile belirlenen tematik yapıların frekans temelli bir doğrulamasını sunmakta ve hiyerarşik kümeleme ile MDS analizlerinde gözlenen yapısal ayrılmaları betimleyici düzeyde desteklemektedir.

4.1.2 Yöntemsel Katkı Açısından Değerlendirme

Kelime bulutu analizi, tek başına çıkarımsal bir yöntem olmaktan ziyade, konu modellemesi ve kümeleme analizleriyle birlikte kullanıldığında literatürün kavramsal yoğunluk noktalarını görünür kılmaktadır. Bu çalışmada kelime bulutu, ileri analitik yöntemlerle elde edilen bulguların yorumlanabilirliğini ve şeffaflığını artıran tamamlayıcı bir araç olarak işlev görmüştür.



Şekil 4.2. Kelime bulutu- Abstractlarda kelime sıklığı

4.2. Hiyerarşik Kümeleme Analizi Bulguları

Orange Data Mining kullanılarak yapılan kelime kümeleme analizi, veri setindeki kelimelerin gizli temalarını bulmayı amaçlamaktadır. Kümeler bir dendrogram görüntüsünde gösterilmiştir.

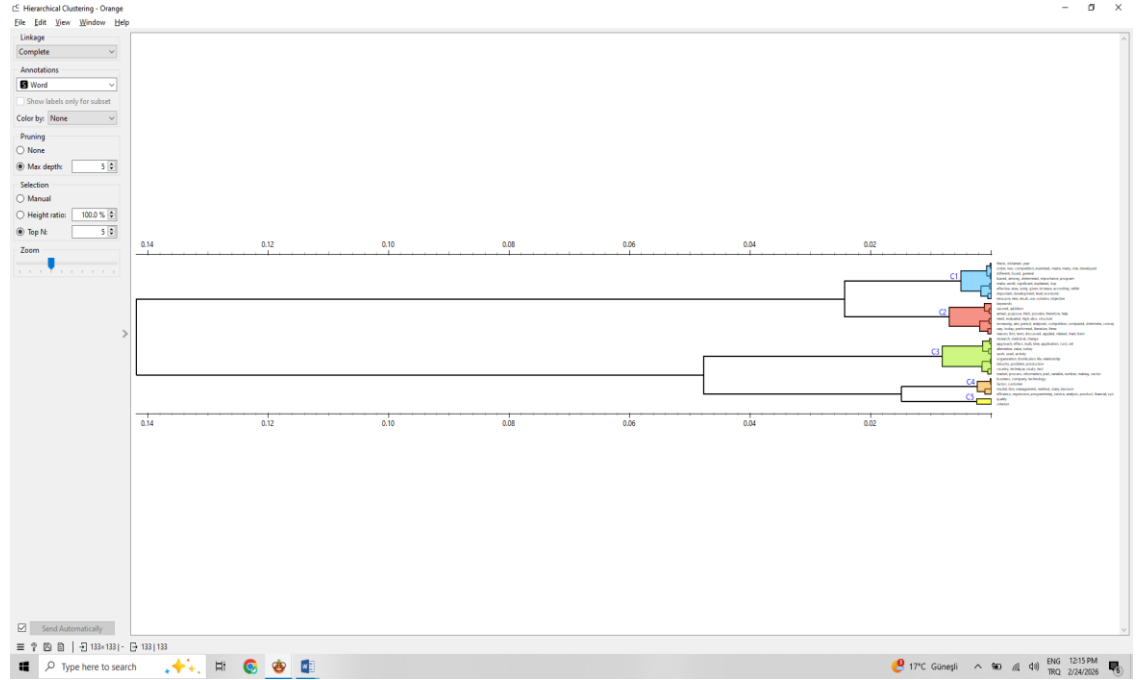
Kosinüs ölçüsü, seyrek vektörler için yorumlanması kolay ve hesaplanması basit olduğundan, metin madenciliği ve bilgi erişiminde yaygın olarak kullanılmaktadır (Dhillon ve Modha, 2001). Çalışmada, metinlerdeki benzerlikleri belirlemek için terim sıklığına (term frequencies) dayalı kosinüs uzaklık metriği kullanılmıştır. Kosinüs benzerliği, iki vektör arasındaki açı veya açının kosinüsü ile de tanımlanabilir (Al-Anzi ve AbuZeina, 2017). Bu, aynı bileşime sahip ancak farklı toplamaları olan belgelerin aynı şekilde ele alınmasına olanak tanır ve bu da onu metin belgeleri için en popüler ölçü haline getirir.

Kosinüs Benzerliği Hesaplaması:

Sorgu vektörü 'x' ile belge vektörü 'y' arasındaki benzerlik, aşağıdaki kullanılarak hesaplanır:

$$\text{Sim}(x, y) = (x \cdot y) / (|x| |y|)$$

Bu çalışmada literatürde öne çıkan kavramsal yapıların belirlenmesi amacıyla, metin madenciliği yaklaşımı kapsamında kelime temelli hiyerarşik kümeleme analizi gerçekleştirilmiştir. Analizler, Orange veri madenciliği yazılımı kullanılarak yürütülmüş; kümeler arası benzerliğin hesaplanmasında Complete Linkage (tam bağlantı) yöntemi tercih edilmiştir. Bu yöntemin seçilme gerekçesi, kümeler içindeki kavramsal tutarlılığı maksimize etmesi ve tematik olarak daha homojen gruplar üretmesidir.



Şekil 4.3. Kelime kümeleme görselleştirmesi

Dendrogram incelendiğinde, yaklaşık 0.05 mesafe eşiği dikkate alındığında kavramların beş ana küme (C1–C5) altında anlamlı biçimde ayrıştığı görülmektedir. Kümeler arası birleşmelerin daha yüksek mesafe değerlerinde gerçekleşmesi, kümeler arasında belirgin tematik farklılıkların bulunduğu işaret etmektedir

Tablo 4.1. Analiz için kullanılan bileşenlerin detayları Kaynak:orangedatamining.com Ziyaret tarihi: 11.11.2025

Widget Adı	İşlevi / Kullanım Amacı
File	Veri setinin Orange programına aktarılmasını sağlar. CSV, Excel veya metin dosyaları içe aktarılabilir.
Corpus	Kullanıcı giriş alanına veri sağladığında, bu veri metin kümesine dönüştürülür ve analiz için uygun biçime dönüştürür.
Preprocess Text	metni daha küçük birimlere (token'lara) ayırır, filtreler, normalleştirme (kök bulma, lemmatizasyon) işlemlerini gerçekleştirir.
Word Cloud	Metinden kelime bulutu oluşturur.Kelimeler sıklıklarına göre listelenir
Corpus Viewer	metin dosyalarını (metin kümesinin örneklerini) görüntülemek için tasarlanmıştır.
Data Table	Veri setinin tablo şeklinde gösterimini sağlar.
Distributions	ayrık veya sürekli özniteliklerin değer dağılımını görüntüler . Veriler bir sınıf değişkeni içeriyorsa, dağılımlar sınıfa göre bölünebilir.
Box Plot	Herhangi bir yeni veriyi kontrol etmek, yinelenen değerler, aykırı değerler ve benzeri anormallikleri hızlı bir şekilde keşfetmek için kullanılır
Scatter Plot	Değişkenler arasındaki ilişkileri görselleştirir
Hierarchical Clustering	mesafe matrisinden rastgele nesne türlerinin hiyerarşik kümelenmesini hesaplar ve karşılık gelen bir dendrogram gösterir.
Heat Map	iki boyutlu bir matristeki öznitelik değerlerini görselleştirmek için kullanılan grafiksel bir yöntemdir.
Multidimensional Scaling (MDS)	noktalar arasındaki mesafeleri mümkün olduğunca iyi eşleştirmeye çalışarak, noktaların düşük boyutlu (iki boyutlu) bir izdüşümünü bulan bir tekniktir.
Silhouette Plot	veri kümeleri içindeki tutarlılığın grafiksel bir temsilini sunar ve kullanıcıya küme kalitesini görsel olarak değerlendirme olanağı sağlar
K-Means	Küme sayısına göre veriler için grup oluşturur
Save Data	Analiz sonucunda ulaşılan verileri dışa aktarır
Concordance	sorgulanan kelimeyi metinde bulur ve bu kelimenin kullanıldığı bağlamı görüntüler
t-sne	t-SNE , MDS'ye benzer bir boyut indirgeme tekniğidir; burada noktalar olasılık dağılımlarına göre 2 boyutlu hale gelir.

4.3 Küme Bazlı Tematik Yapı

Küme 1 (C1): Stratejik Yönetim ve Rekabet Odaklı Kavramlar:

Birinci küme; brand, value, competition, market, firm, strategy gibi kavramlardan oluşmaktadır. Bu küme, işletmelerin rekabetçi çevrede konumlanması, pazar dinamikleri ve stratejik değer yaratma süreçlerine odaklanan literatürü temsil etmektedir. Kavramların birlikte kümelenmesi, çalışmalarda pazar temelli stratejilerin ve rekabet avantajının merkezi bir rol oynadığını göstermektedir.

Küme 2 (C2): Ürün, Yenilik ve Değer Yaratma:

İkinci küme; product, innovation, customer, value, service gibi kavramları içermektedir. Bu küme, ürün geliştirme, yenilikçilik ve müşteri için değer yaratma süreçlerini konu alan çalışmaları yansıtmaktadır. C1 ile olan görece yakınlık, stratejik yönelimlerin ürün ve yenilik çıktılarıyla doğrudan ilişkili olduğunu ortaya koymaktadır.

Küme 3 (C3): Uygulama, Süreç ve Yetkinlik Temelli Yaklaşım:

Üçüncü küme; application, skills, process, practice, interaction gibi kavramlardan oluşmaktadır. Bu yapı, bireysel ve örgütsel düzeyde uygulama süreçlerine, yetkinlik geliştirmeye ve teknoloji-insan etkileşimine odaklanan çalışmaları temsil etmektedir. Küme, stratejik yaklaşımların operasyonel düzeyde nasıl hayata geçirildiğini açıklayan literatürle örtüşmektedir.

Küme 4 (C4): İş Süreçleri ve Karar Verme Mekanizmaları:

Dördüncü küme; business, decision, analysis, performance, management gibi kavramları içermektedir. Bu küme, yönetim bilişim sistemleri, karar destek sistemleri ve performans analitiği literatürünü yansıtmaktadır. C3 ile olan yakın ilişki, uygulamaya dayalı süreçlerin yönetsel karar alma mekanizmalarıyla bütünleştiğini göstermektedir.

Küme 5 (C5): Müşteri Deneyimi ve Algılanan Değer:

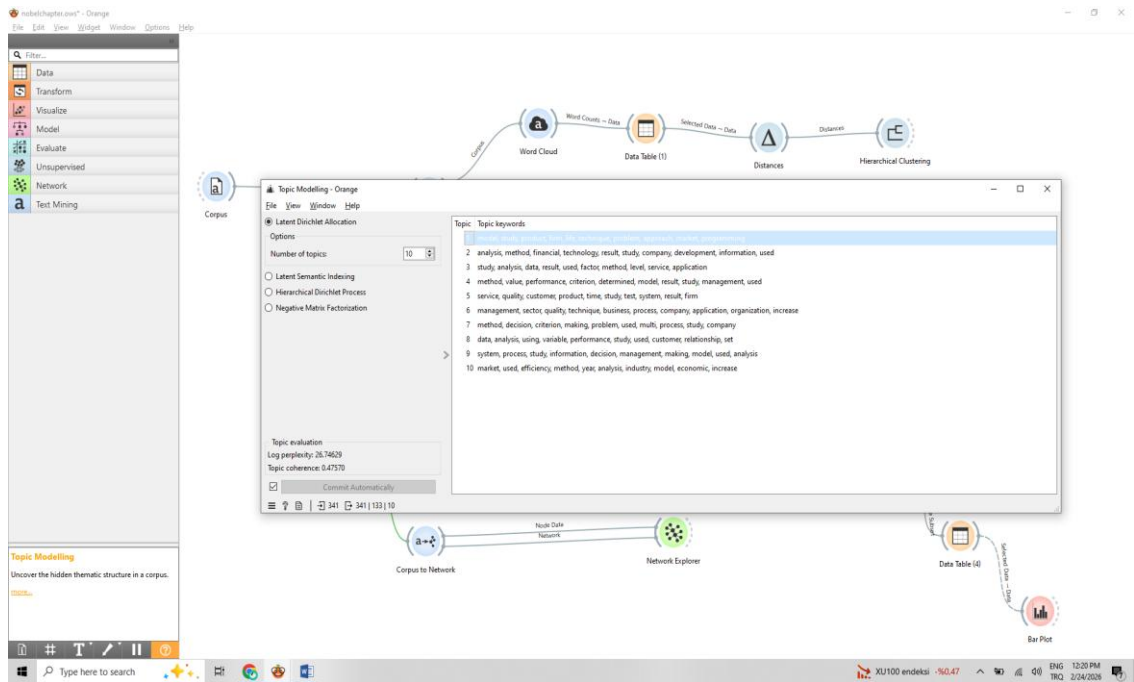
Beşinci küme; customer, experience, service, satisfaction, value gibi kavramlardan oluşmaktadır. Bu küme, müşteri odaklılık, hizmet kalitesi ve deneyim temelli değer yaratma yaklaşımlarını temsil etmektedir. Diğer kümelerden görece ayrışması, müşteri deneyiminin literatürde bağımsız ve özgün bir tematik alan olarak ele alındığını ortaya koymaktadır.

Genel olarak, hiyerarşik kümeleme analizi sonuçları, literatürün stratejik (C1–C2), operasyonel ve yönetsel (C3–C4) ve kullanıcı/müşteri odaklı (C5) olmak üzere üç ana eksen etrafında yapılandığını göstermektedir. Bu çok katmanlı yapı, stratejik kararların operasyonel süreçler ve müşteri deneyimi ile bütünleşik bir biçimde ele alındığını ortaya koymaktadır. Elde edilen bulgular, çalışmanın kavramsal çerçevesinin temellendirilmesi ve araştırma sorularının gerekçelendirilmesi açısından önemli katkılar sunmaktadır.

4.4. Konu Modellemesi (Topic Modeling) Analizi Bulguları

Bu çalışmada metin korpusunun altında yatan tematik yapının ortaya çıkarılması amacıyla Konu Modellemesi (Topic Modeling) yöntemi uygulanmıştır. Analizler Orange veri madenciliği yazılımı kullanılarak gerçekleştirilmiş ve olasılıksal bir modelleme yaklaşımı olan Latent Dirichlet Allocation (LDA) tercih edilmiştir. LDA yöntemi, belgelerin birden fazla konudan oluşabileceği varsayımına dayanmakta ve literatürde tematik örüntülerin keşfi için yaygın olarak kullanılmaktadır.

Model parametreleri kapsamında konu sayısı 10 olarak belirlenmiştir. Modelin uygunluğu log perplexity (26.75) ve topic coherence (0.476) değerleri üzerinden değerlendirilmiş olup, elde edilen değerler konuların anlamlı ve yorumlanabilir bir bütünlük sunduğunu göstermektedir.



Şekil 4.4. Orange yazılımında oluşturulan LDA konu modelleme çıktısı

4.4.1 Konu Bazlı Tematik Yapı

Konu 1: Modelleme ve ürün odaklı yaklaşımlar

Bu konu; model, study, product, firm, approach, market gibi anahtar kelimeler etrafında şekillenmektedir. Tema, ürün temelli analizler, firma düzeyinde modelleme çalışmaları ve pazar odaklı metodolojik yaklaşımları temsil etmektedir. Literatürde özellikle nicel modelleme ve ürün performansı analizleriyle örtüşmektedir.

Konu 2: Finansal ve teknolojik analizler

analysis, method, financial, technology, result, company gibi kavramları içeren bu konu, finansal performans değerlendirmeleri ve teknolojik gelişmelerin firma sonuçları üzerindeki etkisini inceleyen çalışmaları yansıtmaktadır.

Konu 3: Veri analizi ve uygulama çalışmaları

Bu konu; data, analysis, result, method, service, application gibi kelimelerden oluşmaktadır. Ampirik veri analizleri ve uygulamaya dönük araştırmaların baskın olduğu bir temayı temsil etmektedir.

Konu 4: Performans ve değer yönetimi

Value, performance, criterion, management, result gibi kavramlar bu konunun merkezindedir. Tema, performans ölçümü, değer temelli yönetim ve kriterlere dayalı değerlendirme yaklaşımlarını içermektedir.

Konu 5: Hizmet kalitesi ve müşteri odaklılık

Service, quality, customer, product, system anahtar kelimeleriyle tanımlanan bu konu, hizmet kalitesi, müşteri memnuniyeti ve sistem temelli hizmet süreçlerine odaklanan çalışmaları yansıtmaktadır.

Konu 6: Yönetim ve iş süreçleri

Bu konu; management, business, process, organization, technique gibi kavramları içermektedir. İş süreçleri, örgütsel yapı ve yönetim teknikleri ekseninde şekillenen literatürü temsil etmektedir.

Konu 7: Karar verme ve problem çözme

Decision, problem, criteria, process, company gibi anahtar kelimelerden oluşan bu tema, çok kriterli karar verme, problem çözme yaklaşımları ve yönetsel karar süreçlerine odaklanmaktadır.

Konu 8: Değişkenler ve müşteri ilişkileri

Variable, performance, relationship, customer kavramları etrafında şekillenen bu konu, değişkenler arası ilişkiler ve müşteriyle ilgili performans göstergelerini ele alan çalışmaları kapsamaktadır.

Konu 9: Bilgi sistemleri ve analitik

System, information, decision, model, analysis gibi anahtar kelimeler bu temayı tanımlamaktadır. Yönetim bilişim sistemleri, bilgi temelli karar alma ve analitik modeller ön plandadır.

Konu 10: Pazar, verimlilik ve ekonomik analiz

Market, efficiency, industry, economic, growth kavramlarını içeren bu konu, sektör analizi, verimlilik ve ekonomik performans değerlendirmelerine odaklanan çalışmaları temsil etmektedir.

4.5. Konu Modellemesi ve Hiyerarşik Kümeleme Bulgularının Entegrasyonu

Bu çalışmada uygulanan konu modellemesi (Latent Dirichlet Allocation – LDA) ve kelime temelli hiyerarşik kümeleme analizleri, literatürdeki kavramsal yapının çok katmanlı ve bütünsel bir biçimde ele alındığını ortaya koymaktadır. İki tamamlayıcı yöntem birlikte değerlendirildiğinde, literatürün yalnızca baskın temalar üzerinden değil, aynı zamanda bu temalar arasındaki yapısal ilişkiler üzerinden de anlamlandırılmasına olanak sağladığı görülmektedir.

Hiyerarşik kümeleme analizi sonuçları, kavramların stratejik, operasyonel/yönetsel ve müşteri odaklı olmak üzere üç ana üst düzey eksen etrafında gruplaştığını göstermektedir. Bu yapı, konu modellemesi sonuçlarıyla büyük ölçüde örtüşmektedir. Özellikle LDA analizinde ortaya çıkan modelleme, performans, karar

verme ve müşteri deneyimi temalarının, hiyerarşik kümeleme ile belirlenen kümelerle tutarlı bir biçimde ayrıştığı gözlemlenmiştir.

Stratejik düzeyi temsil eden kümeler (C1 ve C2), LDA analizinde ağırlıklı olarak modelleme, pazar, ürün, yenilik ve ekonomik performans temalarıyla örtüşmektedir. Bu durum, literatürde stratejik kararların çoğunlukla nicel modelleme yaklaşımları ve pazar temelli analizler üzerinden ele alındığını göstermektedir. Özellikle ürün ve yenilik odaklı konuların stratejik yönetim literatürüyle bütünleşik biçimde ele alınması, değer yaratma süreçlerinin üst düzey yönetsel bakış açısıyla ele alındığını ortaya koymaktadır.

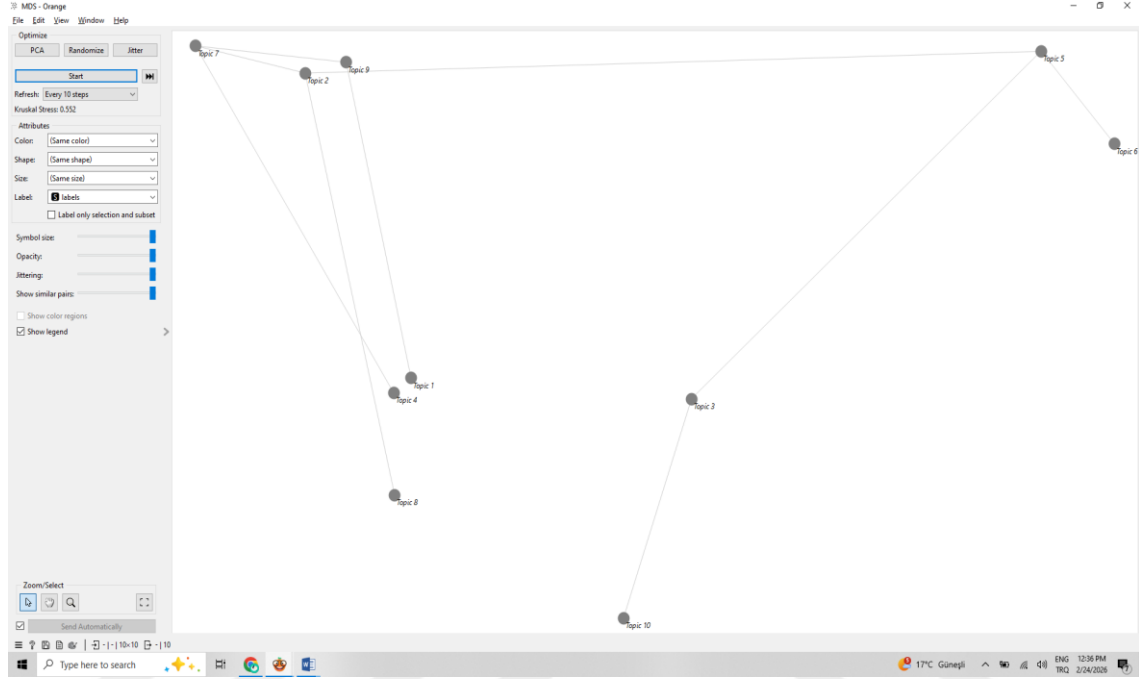
Operasyonel ve yönetsel eksenini temsil eden kümeler (C3 ve C4) ile LDA sonuçlarında belirlenen iş süreçleri, karar verme, bilgi sistemleri ve performans analitiği temaları arasında güçlü bir uyum bulunmaktadır. Bu bulgu, stratejik yönelimlerin operasyonel süreçler ve karar destek mekanizmaları aracılığıyla somutlaştığını göstermektedir. Özellikle karar verme ve süreç yönetimi temalarının hem konu modellemesi hem de kümeleme analizinde merkezi bir konumda yer alması, literatürde analitik ve sistem temelli yaklaşımların artan önemine işaret etmektedir.

Müşteri odaklı eksenini temsil eden küme (C5) ise, LDA analizinde ortaya çıkan hizmet kalitesi, müşteri deneyimi ve algılanan değer temalarıyla doğrudan örtüşmektedir. Bu kümenin diğer kümelerden görece daha ayrılmış olması, müşteri deneyimi ve hizmet odaklı çalışmaların literatürde bağımsız ve derinleşen bir araştırma alanı haline geldiğini göstermektedir. Bu durum, geleneksel firma ve süreç merkezli yaklaşımlardan, kullanıcı ve deneyim merkezli yaklaşımlara doğru bir paradigmatik kaymayı da yansıtmaktadır.

Her iki yöntemin birlikte sunduğu bulgular, literatürdeki tematik yapının hiyerarşik ve bütünleşik bir karakter taşıdığını ortaya koymaktadır. Konu modellemesi, baskın araştırma temalarını ortaya çıkarırken; hiyerarşik kümeleme bu temaların birbirleriyle olan kavramsal yakınlıklarını ve ayrışma düzeylerini görünür kılmaktadır. Bu bütünleşik yaklaşım, literatürün yalnızca “hangi konuların çalışıldığı” sorusuna değil, aynı zamanda “bu konuların birbirleriyle nasıl ilişkilendiği” sorusuna da yanıt vermektedir.

Sonuç olarak, konu modellemesi ve hiyerarşik kümeleme analizlerinin entegrasyonu, literatürde stratejik yönetim, operasyonel süreçler ve müşteri deneyimi

arasında çok yönlü ve dinamik bir etkileşim bulunduğunu ortaya koymaktadır. Bu bulgular, gelecekte yapılacak çalışmalar için disiplinlerarası, çok düzeyli ve bütünlük araştırma tasarımlarının önemini vurgulamakta; aynı zamanda literatürde hâlen yeterince bütüncül biçimde ele alınmamış olan strateji–operasyon–müşteri deneyimi etkileşimlerine yönelik yeni araştırma fırsatlarına işaret etmektedir.



Şekil 4.5. LDA ile oluşturulan 10 konunun MDS görselleştirmesi

4.5.1 Konu Modellemesi, Hiyerarşik Kümeleme ve MDS Bulgularının Bütünlük Yorumu

Bu çalışmada literatürün kavramsal yapısını çok boyutlu biçimde ortaya koyabilmek amacıyla konu modellemesi (LDA), kelime temelli hiyerarşik kümeleme ve çok boyutlu ölçekleme (Multidimensional Scaling – MDS) yöntemleri birlikte kullanılmıştır. Bu üç yöntemin bütünlük değerlendirilmesi, literatürdeki temaların yalnızca içerik açısından değil, aynı zamanda kavramsal yakınlık, ayrışma ve konumsal ilişkiler bağlamında da analiz edilmesine olanak sağlamıştır.

Hiyerarşik kümeleme analizi, literatürdeki kavramların stratejik, operasyonel/yönetimsel ve müşteri odaklı olmak üzere üç ana üst düzey eksen etrafında yapıldığını göstermiştir. Konu modellemesi sonuçları, bu üst düzey ayrışmayı destekler biçimde, modelleme, performans, karar verme, hizmet kalitesi ve müşteri deneyimi gibi temaların baskın olduğunu ortaya koymuştur. MDS analizi ise, bu temaların iki boyutlu

uzaydaki görelî konumlarını görselleştirerek, kümeler arası ilişkilerin daha açık biçimde yorumlanmasını sağlamıştır.

MDS haritası incelendiğinde, Topic 1 ve Topic 4'ün birbirine oldukça yakın konumlandığı görülmektedir. Bu durum, konu modellemesinde bu temaların ortak biçimde modelleme, performans ve metodolojik yaklaşımlar etrafında şekillendiğini doğrulamaktadır. Aynı şekilde Topic 2, Topic 7 ve Topic 9'un görelî olarak birbirine yakın bir küme oluşturması, finansal analiz, karar verme ve bilgi temelli yaklaşımların literatürde birlikte ele alındığını göstermektedir. Bu bulgu, hiyerarşik kümeleme analizinde ortaya çıkan operasyonel ve yönetsel kümelerle yüksek düzeyde tutarlılık sergilemektedir.

Öte yandan, Topic 5 ve Topic 6'nın MDS uzayında daha uzak ve görece izole bir konumda yer alması, bu temaların literatürde hizmet kalitesi, müşteri deneyimi ve yönetim uygulamaları ekseninde daha özgün bir araştırma alanı oluşturduğunu göstermektedir. Bu durum, hiyerarşik kümeleme analizinde müşteri odaklı kümenin diğer kümelerden daha belirgin biçimde ayrışmasıyla örtüşmektedir. Bu ayrışma, literatürde kullanıcı ve müşteri deneyimine yönelik çalışmaların giderek bağımsız bir paradigma haline geldiğini düşündürmektedir.

MDS analizinde Topic 3 ve Topic 10'un daha uç konumlarda yer alması ise, bu temaların pazar, verimlilik ve makro düzey ekonomik analizlerle ilişkili olduğunu göstermektedir. Bu temaların stratejik kümelerle olan görece mesafesi, literatürde makro ekonomik değerlendirmelerin mikro düzey uygulama ve müşteri odaklı çalışmalardan kavramsal olarak ayrıştığını ortaya koymaktadır. Bu bulgu, hiyerarşik kümeleme analizinde stratejik kümelerin daha üst düzeyde ve daha geniş mesafe değerlerinde birleşmesiyle tutarlıdır.

Üç yöntemin birlikte sunduğu bulgular, literatürdeki tematik yapının hiyerarşik, çok katmanlı ve konumsal olarak anlamlı bir yapı sergilediğini ortaya koymaktadır. Konu modellemesi baskın araştırma temalarını belirlerken, hiyerarşik kümeleme bu temaların kavramsal bloklar halinde nasıl organize olduğunu göstermekte; MDS ise bu bloklar arasındaki görelî yakınlık ve uzaklıkları görsel ve analitik olarak desteklemektedir. Bu bağlamda elde edilen sonuçlar, literatürün yalnızca tematik çeşitliliğini değil, aynı zamanda bu temalar arasındaki yapısal ilişkileri de ortaya koymaktadır.

Sonuç olarak, konu modellemesi, hiyerarşik kümeleme ve MDS analizlerinin bütünlük kullanımı, stratejik yönetim, operasyonel süreçler, karar verme mekanizmaları ve müşteri deneyimi arasında çok yönlü ve dinamik bir etkileşim bulunduğunu göstermektedir. Bu bütüncül yaklaşım, literatürde sıklıkla ayrı ayrı ele alınan bu temaların aslında birbirini besleyen ve tamamlayan yapılar olduğunu ortaya koymakta; gelecekteki çalışmalar için disiplinlerarası ve çok düzeyli araştırma tasarımlarının önemini güçlü biçimde vurgulamaktadır.

4.6. Concordance (Bağlam Analizi) Bulgularının Tartışılması

Bu çalışmada uygulanan concordance analizi, kelime bulutu ve konu modellemesi ile nicel olarak belirlenen kavramların, metin bağlamı içerisinde nasıl kullanıldığını ve hangi anlamsal çerçeveler içinde konumlandığını ortaya koymak amacıyla gerçekleştirilmiştir. Bu analiz, özellikle literatürde yüksek frekansa sahip kavramların yüzeysel tekrarın ötesinde, hangi tematik ve metodolojik bağlamlarda kullanıldığını anlamaya olanak sağlamaktadır.

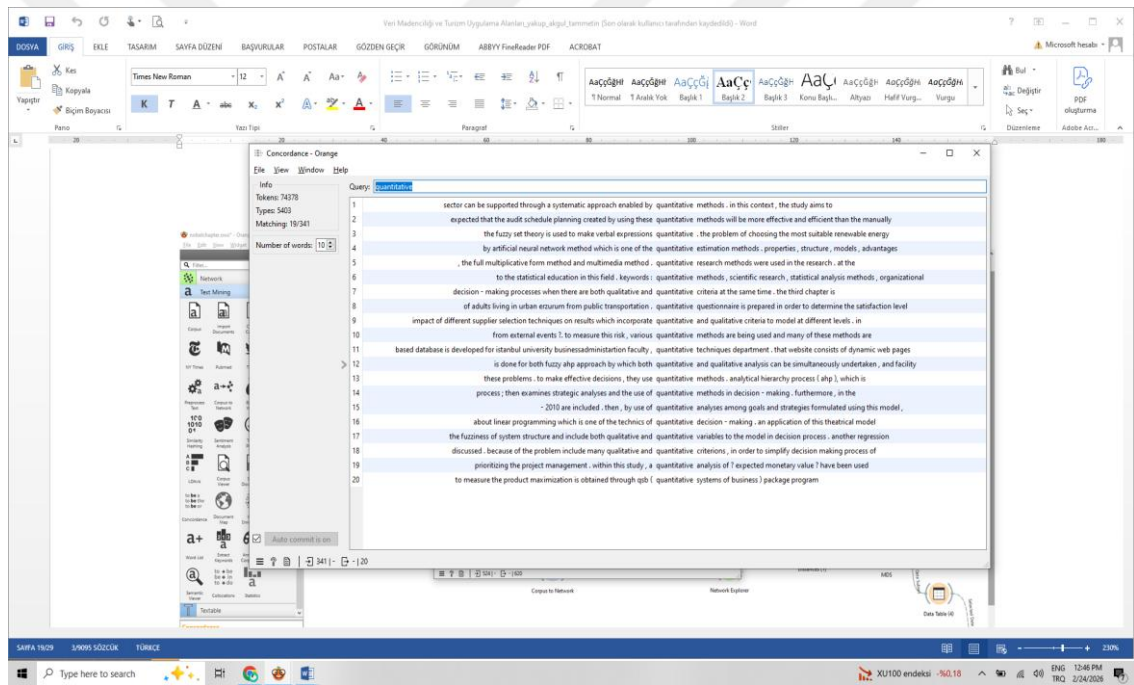
Concordance çıktıları incelendiğinde, *quantitative* kavramının metinlerde sıklıkla method, analysis, decision, model, system, evaluation, estimation ve optimization gibi terimlerle birlikte kullanıldığı görülmektedir. Bu bulgu, literatürde nicel yaklaşımların ağırlıklı olarak analitik karar verme, modelleme ve performans değerlendirme bağlamında ele alındığını göstermektedir. Özellikle nicel yöntemlerin, karmaşık problemlerin çözümünde sistematik ve rasyonel araçlar olarak konumlandırıldığı dikkat çekmektedir.

Bağlam analizinde nicel yaklaşımların; scheduling, forecasting, risk analysis, supplier selection, energy planning ve project management gibi uygulama alanlarıyla ilişkilendirildiği görülmektedir. Bu durum, konu modellemesi analizinde ortaya çıkan uygulama, süreç ve karar verme temalarıyla doğrudan örtüşmektedir. Dolayısıyla, nicel yöntemlerin yalnızca teorik bir araç olarak değil, aynı zamanda pratik ve sektörel problemlerin çözümünde yaygın biçimde kullanıldığı anlaşılmaktadır.

Concordance analizi ayrıca, quantitative decision making ve quantitative methods ifadelerinin sıklıkla information systems ve management kavramlarıyla birlikte geçtiğini göstermektedir. Bu bulgu, hiyerarşik kümeleme analizinde belirlenen

operasyonel/yönetmel küme ile uyumludur ve literatürde karar destek sistemleri ile nicel analizlerin bütünleşik biçimde ele alındığını ortaya koymaktadır. MDS analizinde bu temaların birbirine yakın konumlanması da, kavramsal yakınlığın görsel bir doğrulamasını sunmaktadır.

Bağlam analizi, müşteri ve hizmet odaklı kavramların (service quality, customer satisfaction, performance improvement) nicel yöntemlerle birlikte ele alındığını da göstermektedir. Bu bulgu, müşteri deneyimi ve kalite yönetimi temalarının literatürde sezgisel yaklaşımlardan ziyade ölçülebilir ve analiz edilebilir değişkenler üzerinden incelendiğini ortaya koymaktadır. Bu durum, kelime bulutu analizinde öne çıkan quality ve performance kavramlarının anlamsal bağlamını derinleştirmektedir.



Şekil 4.6. “Quantitative” teriminin konkordans satırlarının analizine örnek olarak sunulmaktadır.

4.6.1 Yöntemsel Bütünlük ve Katkı

Bu çalışmada concordance analizi, kelime frekansına dayalı betimleyici bulgular ile ileri düzey çıkarımsal yöntemler (LDA, hiyerarşik kümeleme ve MDS) arasında mikro düzeyde anlamsal bir köprü işlevi görmektedir. Böylece, literatürde hangi kavramların öne çıktığı sorusunun yanı sıra, bu kavramların hangi bağlamlarda ve hangi amaçlarla kullanıldığı da sistematik biçimde açıklanabilmektedir.

Bu bütünleşik yaklaşım, literatür incelemelerinde sıklıkla eleştirilen “frekans var ama anlam yok” sorununu aşarak, tematik yapıların bağlamsal geçerliliğini güçlendirmektedir

4.7. t-SNE Tabanlı Konu Modellemesi ve Hiyerarşik Kümeleme Bulgularının Tartışılması

Bu çalışmada literatürdeki tematik yapıyı ortaya koymak amacıyla t-Distributed Stochastic Neighbor Embedding (t-SNE) tekniği kullanılmıştır. t-SNE, yüksek boyutlu metin verilerinin düşük boyutlu bir uzayda görselleştirilmesini sağlayan, özellikle yerel komşuluk ilişkilerini korumada güçlü bir yöntemdir. Bu nedenle elde edilen görselleştirme, çalışmalar arasındaki tematik benzerlik ve ayrışmaları ortaya koymada uygun bir analitik çerçeve sunmaktadır.

t-SNE çıktısında her bir nokta tek bir çalışmayı temsil etmekte; noktalar arasındaki yakınlık, metinlerin konu dağılımlarının birbirine benzerliğini yansıtmaktadır. Renkler ise konu modellemesi sonucunda belirlenen baskın temaları göstermektedir. Görselleştirme genel olarak incelendiğinde, çalışmaların tek bir homojen yapıdan ziyade, birden fazla yerel yoğunlaşma (local clusters) oluşturduğu görülmektedir. Bu durum, literatürün tematik açıdan parçalı ve çok boyutlu bir yapıya sahip olduğunu göstermektedir.

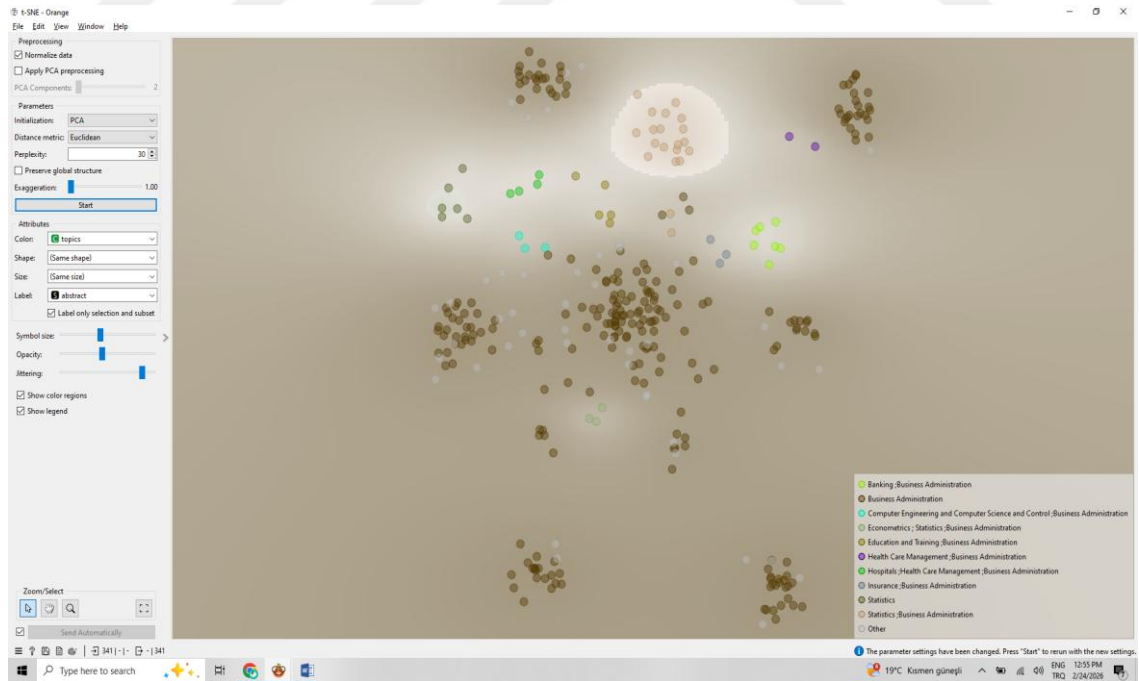
Merkezi bölgede yer alan ve görece daha yoğun olan küme, nicel yöntemlerin genel yönetim, karar verme ve analiz odaklı çalışmalarda yaygın biçimde kullanıldığını düşündürmektedir. Bu çekirdek küme, konu modellemesi çıktılarında ortaya çıkan “nicel analiz”, “karar destek sistemleri” ve “yönetimsel uygulamalar” temalarıyla yüksek düzeyde örtüşmektedir. t-SNE’nin yerel yapıları ön plana çıkaran doğası dikkate alındığında, bu merkezî yoğunlaşma, literatürde yönetsel açıdan benzer çalışmaların güçlü bir çekirdek oluşturduğunu göstermektedir.

Buna karşılık, t-SNE haritasının farklı bölgelerinde konumlanan alt kümeler, daha özgül uygulama alanlarını temsil etmektedir. Sağlık yönetimi, eğitim, sigortacılık ve bankacılık gibi alanlara odaklanan çalışmaların, kendi içinde daha sıkı kümeler oluşturduğu görülmektedir. Bu bulgu, nicel yöntemlerin farklı sektörlerde benzer analitik araçlarla ele alınmasına rağmen, bağlamsal farklılıklar nedeniyle tematik olarak ayrıştığını ortaya koymaktadır.

Hiyerarşik kümeleme sonuçları ile t-SNE görselleştirmesi birlikte değerlendirildiğinde, bazı temaların diğerlerine kıyasla daha belirgin sınırlarla ayrıldığı anlaşılmaktadır. Özellikle istatistik ve ekonometri odaklı çalışmaların, uygulama ağırlıklı yönetim çalışmalarından daha net biçimde ayrıştığı görülmektedir. Bu durum, literatürde yöntem-merkezli ve uygulama-merkezli iki ana araştırma ekseninin varlığına işaret etmektedir.

t-SNE haritasında birbirine görece yakın konumlanan ancak farklı renklere sahip kümeler, temalar arasında kavramsal bir geçişkenliğin bulunduğunu göstermektedir. Bu geçişkenlik, karar verme, performans değerlendirme ve verimlilik analizi gibi konuların farklı sektörlerde benzer nicel tekniklerle ele alınmasından kaynaklanmaktadır. Dolayısıyla literatür, keskin sınırlarla ayrılmış temalardan ziyade, örtüşen ve etkileşimli tematik alanlar içermektedir.

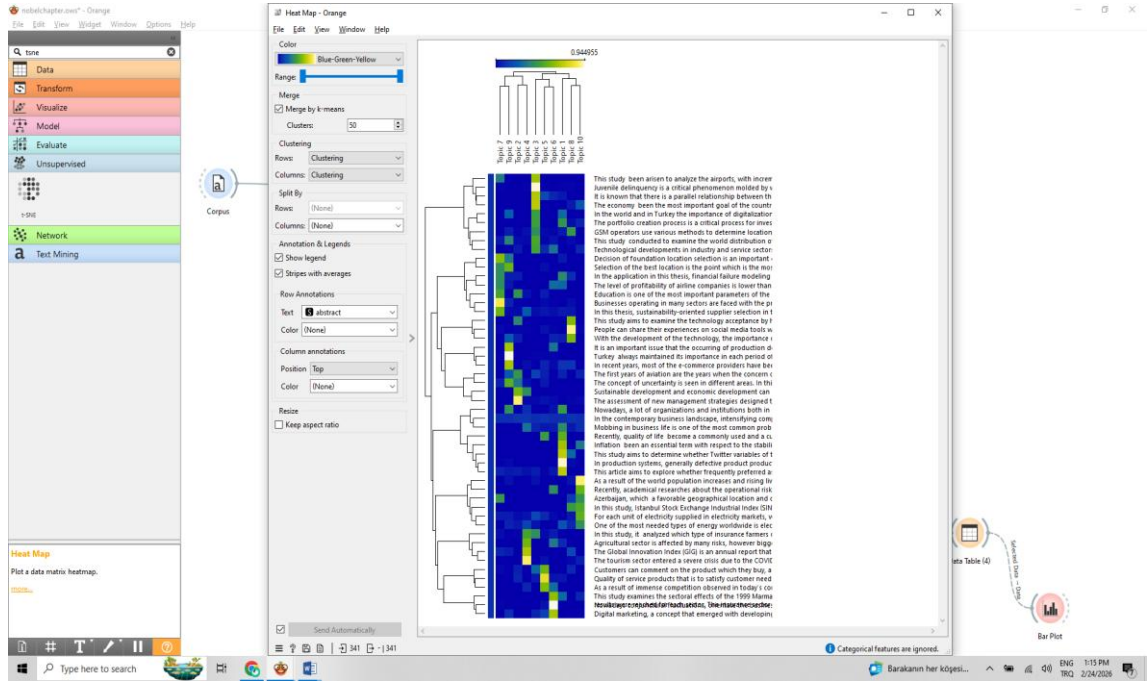
Son olarak, t-SNE uzayının periferisinde yer alan küçük ve yoğun kümeler, literatürde henüz sınırlı sayıda çalışma ile temsil edilen ancak içerik açısından tutarlı niş araştırma alanlarını işaret etmektedir. Bu alanlar, gelecekte nicel yöntemlerin farklı disiplinlerde nasıl evrilebileceğine dair önemli araştırma fırsatları sunmaktadır.



Şekil 4.7. t-SNE sonuçları (topic)

4.8. Heat Map Bulgularının Yorumu (Hiyerarşik Kümeleme Entegrasyonu)

Isı haritası analizi, dokümanlar (satırlar) ile konu temsilleri (sütunlar) arasındaki yoğunluk ilişkilerini görselleştirmekte ve metinsel benzerlik örüntülerini hiyerarşik kümeleme ile birlikte ortaya koymaktadır. Renk skalasında mavi tonlar düşük katkıyı, yeşil ve sarı tonlar ise ilgili konunun doküman içerisindeki görece yüksek ağırlığını temsil etmektedir.



Şekil 4.8. Heat map

1. Konu Temelli Dikey Kümelenme (Sütunlar)

Isı haritasının üst kısmında yer alan dendrogram, belirli konuların birlikte kümelendiğini göstermektedir. Bu durum, konu modellemesi sonucunda elde edilen bazı temaların birlikte ortaya çıkma eğilimi sergilediğini ve metinlerde eş zamanlı olarak kullanıldığını ortaya koymaktadır.

Özellikle:

- Yöntem, analiz, model, karar verme ve performans odaklı konuların aynı alt kümelerde toplanması,
- Yönetim, kalite, süreç ve organizasyon temalarının birbirine yakın konumlanması,

Literatürde yöntem-merkezli ampirik çalışmalar ile uygulama ve yönetim odaklı çalışmaların ayrışan fakat kendi içinde tutarlı tematik bloklar oluşturduğunu göstermektedir.

2. Doküman Temelli Yatay Kümelenme (Satırlar)

Satır dendrogramı incelendiğinde, özet metinlerin belirli kümelerde yoğunlaştığı görülmektedir. Bu kümelenme, dokümanların:

- benzer problem tanımlarını,
- benzer metodolojik yaklaşımları,
- ya da ortak uygulama alanlarını

paylaştığını göstermektedir.

Bu bulgu, veri setinin rastgele dağılmış metinlerden oluşmadığını, aksine tematik olarak yapılandığını ortaya koymaktadır.

3. Yoğunluk Desenleri ve Tematik Baskınlık

Bazı doküman kümelerinde belirli konuların sarı/yeşil renklerle öne çıkması, bu dokümanların belirli temalar etrafında yoğunlaştığını göstermektedir. Buna karşılık, mavi tonların baskın olduğu dokümanlar, daha genel veya çok temalı bir içerik yapısına işaret etmektedir.

Bu durum şu şekilde yorumlanabilir:

- Yüksek yoğunluklu hücreler → tematik uzmanlaşma
- Düşük ve dağınık yoğunluk → tematik çeşitlilik

4. Konu Modelleme ve t-SNE ile Tutarlılık

Heat map çıktısı, daha önce elde edilen konu modellemesi ve t-SNE görselleştirmeleri ile yüksek düzeyde tutarlılık göstermektedir:

- Konu modellemesinde tanımlanan baskın temalar, ısı haritasında belirgin yoğunluk desenleriyle desteklenmektedir.

- t-SNE analizinde birbirine yakın konumlanan dokümanların, ısı haritasında da benzer satır kümeleri içerisinde yer aldığı gözlemlenmektedir.

Bu bütüncül yapı, kullanılan metin madenciliği tekniklerinin aynı tematik yapıyı farklı analitik perspektiflerden doğruladığını göstermektedir.



5. TARTIŞMA

Bu çalışmanın amacı, 2015-2025 yılları arasında Türkiye’de işletme anabilim dalı alanında sayısal yöntemler üzerine yazılan yüksek lisans tezlerindeki baskın araştırma temalarını metin madenciliği teknikleri kullanarak incelemektir. Kelime sıklığı analizi, konu modelleme ve kümeleme teknikleri ile elde edilen bulgular, bu alandaki lisansüstü araştırmaların kavramsal yönelimine ilişkin birkaç önemli kalıbı ortaya koymaktadır.

İlk olarak, kelime bulutu analizi, “çalışma”, “yöntem”, “analiz”, “veri”, “model”, “karar” ve “sonuç” gibi terimlerin çok yüksek sıklıkta ortaya çıktığını göstermiştir. Bu bulgu, incelenen tezlerin analitik ve veri odaklı araştırma yaklaşımlarına güçlü bir şekilde yöneldiğini göstermektedir. Bu terimlerin öne çıkması, istatistiksel analiz, modelleme ve ampirik değerlendirmeyi vurgulayan nicel araştırmanın temel özellikleriyle tutarlıdır. Nicel ve hesaplamalı araştırma geleneklerinin genellikle veri analizine ve metodolojik titizliğe güçlü bir vurgu ile karakterize edildiği önceki metin madenciliği ve bibliyometrik çalışmalarda da benzer kalıplar bildirilmiştir.

Konu modelleme sonuçları, literatürün tematik yapısı hakkında daha derin bir anlayış sağlamaktadır. LDA analizi, modelleme ve ürün odaklı yaklaşımlar, finansal ve teknolojik analizler, veri analizi ve uygulama çalışmaları, performans ve değer yönetimi, hizmet kalitesi ve müşteri odaklılık, yönetim ve iş süreçleri, karar verme ve problem çözme, değişken tabanlı ilişki çalışmaları, bilgi sistemleri ve analitik, pazar ve ekonomik analiz dahil olmak üzere on baskın tema ortaya koymuştur. Bu temalar, nicel yöntemler programları kapsamında yürütülen araştırmaların yalnızca metodolojik çalışmalarla sınırlı olmadığını, aynı zamanda işletme yönetiminin çeşitli fonksiyonel alanlarına da uzandığını göstermektedir.

Bu temalar literatürle ilişkilendirildiğinde, çeşitli benzerlikler gözlemlenebilir. Metin madenciliği ve konu modelleme tekniklerini kullanan önceki çalışmalar da performans analizi, karar verme modelleri, teknolojik yenilik ve müşteriyle ilgili değişkenler gibi temaları yönetim ve işletmeyle ilgili alanlarda baskın araştırma konuları olarak belirlemiştir. Bu tutarlılık, Türk yüksek lisans tezlerinde gözlemlenen araştırma gündeminin büyük ölçüde uluslararası akademik trendleri takip ettiğini göstermektedir.

Çalışmanın bir diğer önemli bulgusu ise kavramların üç ana eksen etrafında kümelenmesidir: stratejik, operasyonel/yönetimsel ve müşteri odaklı bakış açıları. Bu

yapı, işletme yönetimi araştırmalarının çok boyutlu doğasını yansıtmaktadır. Stratejik konular genellikle pazar analizi, verimlilik ve ekonomik performansa odaklanırken; operasyonel konular yönetim süreçleri, analitik modeller ve performans ölçümünü vurgular; müşteri odaklı konular ise hizmet kalitesi, müşteri memnuniyeti ve ilişki yönetimini öne çıkarır. Bu kümelerin konu modellemesinden elde edilen sonuçlarla uyumu, analitik bulguların güvenilirliğini güçlendirir.

Karar verme ve performansla ilgili temaların öne çıkması, modern organizasyonlarda veri odaklı yönetim uygulamalarının artan önemiyle de açıklanabilir. Büyük veri kümelerinin ve gelişmiş analitik araçların artan kullanılabilirliğiyle birlikte, organizasyonlar giderek daha fazla nicel modellere ve karar destek sistemlerine güvenmektedir. Sonuç olarak, lisansüstü araştırmalar, organizasyonel karar verme süreçlerini iyileştirebilecek analitik çerçevelere odaklanma eğilimindedir.

Ayrıca, bilgi sistemleri ve analitik ile ilgili konuların varlığı, iş ortamlarının devam eden dijital dönüşümünü yansıtmaktadır. Organizasyonlar bilgi teknolojilerini operasyonlarına entegre ettikçe, bilgiye dayalı karar verme ve analitik sistemler üzerine yapılan araştırmalar giderek daha önemli hale gelmektedir. Bu bulgu, dijital teknolojilerin ve veri analitiğinin çağdaş yönetim uygulamalarını şekillendirmedeki rolünü vurgulayan son literatürle tutarlıdır.

Analitik modelleme, karar destek sistemleri, performans değerlendirmesi ve müşteri odaklı araştırmaların entegrasyonu, alanın daha disiplinlerarası ve veri odaklı bir araştırma paradigmasına doğru evrildiğini göstermektedir.

6. SONUÇ

Çalışmada, 2015 ile 2025 yılları arasında YÖK Ulusal Tez Merkezi web sitesinde yayınlanmış olan İşletme Anabilim dalı altında sayısal yöntemler bilim dalı alanında hazırlanmış olan tezler metin madenciliği yöntemleri kullanılarak analiz edilmiştir. Çalışmada öncelikle metin madenciliği kavramı, uygulama alanları, süreçleri ve metin madenciliği ile ilgili genel bilgiler açıklanmıştır. Sonraki bölümde metin madenciliği yöntemleri kullanılarak yapılan çalışmalara yer verilmiştir. YÖK Ulusal Tez Merkezi web sitesi üzerinden erişilen tezlerin verileri (başlıklar anahtar kelimeler, özetler gibi) excel tablosuna aktarılıp Orange data mining programı aracılığıyla LDA algoritması kullanılarak konu modellemesi gerçekleştirilmiştir. Tez özetleri ve anahtar kelimelerin Orange veri madenciliği platformu ile analiz edilmesi ve Latent Dirichlet Allocation algoritması ile konu modellemesi uygulanması yoluyla, çalışma literatürün kavramsal yapısını ve tematik dağılımını belirlemiştir.

LDA konu modelleme aşamasında öncelikle performans ölçüleri elde edilip bunlara göre optimum konu sayıları belirlenmiştir. Bu işlemlere göre konu modellemesi yapılarak konular kümeler biçiminde görselleştirilmiştir. Konu modellerinde bulunan her konuya ait en çok tekrar eden 10 kelime sıklık histogramı şeklinde görselleştirilmiştir. Bunlara ek olarak, konu modellerinde bulunan konuların bir biryle olan ilişkisi ağ topolojisi oluşturularak gösterilmiştir. Ayrıca konu modellerinde bulunan her konu için kelime bulutları oluşturulmuş, ısı haritaları üzerinden konu-belge dağılımları incelenerek çok boyutlu olan konu modelleri iki boyuta indirgenmiş ve konu modeli yoğunluk haritası ortaya çıkarılmıştır. Orange Data Mining programı kullanılarak yapılan kelime sıklığı, kelime kümeleme, konu modelleme (LDA), çok boyutlu ölçekleme (MDS), ve konkordans analizleri, araştırma alanının hangi temalar etrafında yoğunlaştığını ortaya koymuştur.

Çalışma bulguları, yıllar içinde sayısal yöntemler alanındaki tez sayısında önemli artışlar olduğunu ve akademik verimliliğin belirli dönemlerde hızlandığını ortaya koymaktadır. Bu artış, dijital dönüşüme ve bilgi sistemlerinin entegrasyonuna yönelik artan akademik ilgiyle ilişkilendirilebilir. Araştırmaların teknoloji odaklı konulara yoğunlaştığı ve bu alanda önemli katkılar sağladığı söylenebilir.

Sonuçlar, tezlerde en sık kullanılan terimlerin analiz, veri, yöntem, model, karar ve sonuç gibi nicel araştırma uygulamalarıyla yakından ilişkili olduğunu ortaya koymuştur. Bu bulgular, bu alandaki lisansüstü araştırmaların ampirik analiz ve analitik modellemeye güçlü bir şekilde vurgu yaptığını göstermektedir. Konu modelleme analizi ayrıca, modelleme yaklaşımları, finansal ve teknolojik analiz, ampirik veri uygulamaları, performans ve değer yönetimi, hizmet kalitesi ve müşteri ilişkileri, iş süreç yönetimi, karar verme modelleri, değişken tabanlı ilişki analizi, bilgi sistemleri ve pazar odaklı ekonomik çalışmalar dahil olmak üzere on ana tematik alanı ortaya koymuştur.

Bu bulgular, Türkiye’deki işletme yönetimi alanındaki nicel yöntemlerin, çeşitli ancak birbirine bağlı bir araştırma yapısıyla karakterize edildiğini göstermektedir. Stratejik, operasyonel ve müşteri odaklı temaların entegrasyonu, lisansüstü araştırmaların hem teorik bilgiye hem de pratik yönetsel uygulamalara katkıda bulunduğunu göstermektedir.

Metodolojik açıdan bakıldığında, bu çalışma aynı zamanda büyük akademik belge kümelerinin analizinde metin madenciliği tekniklerinin yararlılığını da göstermektedir. Kelime sıklığı analizi ve konu modelleme gibi otomatik analiz araçları uygulanarak, araştırmacılar büyük metin külliyatları içindeki gizli tematik yapıları sistematik olarak belirleyebilirler. Bu nedenle, literatür analizinde metin madenciliği yöntemlerinin kullanımı, araştırma eğilimlerini anlamak için verimli ve objektif bir yaklaşım sağlamaktadır.

Bu çalışmanın bulguları, gelecekteki araştırmalar için çeşitli çıkarımlar sunmaktadır. Birincisi, araştırmacılar araştırma eğilimlerini daha kapsamlı bir şekilde anlamak için doktora tezleri, dergi makaleleri veya uluslararası veri tabanlarını dahil ederek veri setini genişletebilirler. İkincisi, gelecekteki çalışmalar, araştırma konularının zaman içindeki evrimini incelemek için ağ analizi, duygu analizi veya dinamik konu modelleme gibi ek analitik teknikler uygulayabilir. Son olarak, farklı ülkeler veya akademik disiplinler genelinde araştırma eğilimlerini inceleyen karşılaştırmalı çalışmalar, nicel yöntemler araştırmasının küresel gelişimine dair değerli bilgiler sağlayabilir.

Sonuç olarak, bu çalışma, Türkiye’de işletme alanında sayısal yöntemler üzerine yapılan lisansüstü araştırmaların tematik yapısına ilişkin sistematik bir genel bakış

sunarak literatüre katkıda bulunmaktadır. Sonuçlar, veri odaklı analitik yaklaşımların artan önemini vurgulamakta ve metin madenciliği yöntemlerinin meta düzey akademik analizleri nasıl destekleyebileceğini göstermektedir.



7. KISITLAR VE ÖNERİLER

Çalışmada, 2015 ile 2025 yılları arasında Türkiye’de İşletme Anabilim Dalı alanında yayınlanan yüksek lisans tezleri incelenmiş olup doktora tezleri hariç tutulmuştur. Bu tezler belirli parametrelere göre detaylı olarak analiz edilirken, çalışmanın bazı sınırlamaları bulunmaktadır. Üniversitelerde İşletme Anabilim Dalı haricinde nicel yöntemler kullanılarak yazılan ve diğer bölümlerde (örneğin, bilgisayar mühendisliği, hukuk, edebiyat, felsefe) hazırlanan tezler çalışmaya dahil edilmemiştir. İşletme alanıyla ilgili olmadığı düşünülen konuları ele alan bu tezlerin dışlanması bir sınırlamadır. Olası yanlış sınıflandırmaları önlemek için metin madenciliği yöntemleri kullanılarak yapılacak çalışmalarda sayısal yöntemler kullanılarak yazılmış başka alanlardaki tezlerin de dahil edilmesi önerilmektedir. Bu çalışmada tezlerin analizi metin madenciliği, konu modellemesi ve kelime sıklığı yöntemleri kullanılarak gerçekleştirilmiştir. Ancak, bu tür yöntemler, belirli kelimeler bağlamları dikkate alınmadan sık sık kullanılıyorsa yanlış yorumlamalara yol açabilir. Tezlerdeki bazı terimler, sıklıklarına bağlı olarak sayısal yöntemler konularına ait gibi görünebilir, ancak tezlerin daha geniş bağlamı bu sonucu desteklemeyebilir. Bu sınırlamayı hafifletmek için, gelecekteki çalışmalar, bu tezlerin bir alt kümesini doğrulamak üzere alan uzmanlarını içerebilir. Daha geniş bir uzman değerlendirmesinin, çalışmanın güvenilirliğini artırmada ve yanlış sınıflandırmaları azaltmada önemli bir rol oynayacağına inanılmaktadır. Bu çalışmada metin madenciliği yöntemi kullanıldığı için, analiz tezlerin özetleri ve temel bilgileriyle sınırlı kalmıştır. Zaman ve kaynak kısıtlamaları nedeniyle, tezlerin tam metinlerini analiz etmek mümkün olmamıştır. Gelecekteki çalışmalar, konuların daha doğru sınıflandırılmasını sağlayabilecek ve daha güvenilir sonuçlar verebilecek tezlerin tam metin analizini içerecek şekilde bu araştırmanın kapsamını genişletebilir. Mevcut çalışma yalnızca Türkiye’de yapılan ve YÖK Ulusal Tez Merkezi web sitesi üzerinden erişilebilen yüksek lisans(doktora tezleri hariç tutulmuştur) tezlerini incelemiştir. Gelecekteki araştırmalar ayrıca farklı ülkelerden tezleri de içerebilir ve küresel ölçekte sayısal yöntemler alanına daha derin bakış açıları sağlayan karşılaştırmalı çalışmalara olanak tanıyabilir.

8. KAYNAKÇA

- Abbasiantaeb, Z., Verberne, S., & Wang, J. (2025). *Tracing science-technology-linkages: A machine learning pipeline for extracting and matching patent in-text references to scientific publications*. *Information Processing & Management*, 62(6), 104264.
- Ahadi, A., Singh, A., Bower, M., & Garrett, M. (2022). *Text mining in education—A bibliometrics-based systematic review*. *Education Sciences*, 12(3), 210.
- Akilan, A. (2015, February). Text mining: Challenges and future directions. In 2015 2nd International Conference on Electronics and Communication Systems (ICECS) (pp. 1679-1684). IEEE.
- Al-Anzi, F. S., & AbuZeina, D. (2017). *Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing*. *Journal of King Saud University-Computer and Information Sciences*, 29(2), 189-195.
- Alawadh, H. M., Meraj, T., Aldosari, L., & Tayyab Rauf, H. (2024). *An Efficient Text-Mining Framework of Automatic Essay Grading Using Discourse Macrostructural and Statistical Lexical Features*. *SAGE Open*, 14(4), 21582440241300548.
- Aleqbqie, H. J., Sfoq, M. S., & Albeer, R. A. (2024). *A review of text mining techniques: Trends and applications in various domains*. *Iraqi Journal for Computer Science and Mathematics*, 5(1), 45–63.
- Alkamli, S., Alabduljabbar, R., 2024. Understanding privacy concerns in ChatGPT: a data-driven approach with LDA topic modeling. *Heliyon* 10 (20), e39087. <https://doi.org/10.1016/j.heliyon.2024.e39087>.
- Alksher, M. A., Azman, A., Yaakob, R., Kadir, R. A., Mohamed, A., & Alshari, E. M. (2016, August). A review of methods for mining idea from text. In 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP) (pp. 88-93). IEEE.
- Anggraini, N., & Suroyo, H. (2019). *Comparison of sentiment analysis against digital payment t-cash and go-pay in social media using orange data mining*. *Journal of Information Systems and Informatics*, 1(2), 152-163.
- Arabiat, A. M. (2025). *Intelligent Model for Detecting GAN-Generated Images Based on Multi-Classifer and Advanced Data Mining Techniques*. *International Journal of Electrical and Electronic Engineering and Telecommunications*, 14(3), 147-157.
- Aubakirov, S., Akhmetov, I., Gelbukh, A., & Mussabayev, R. (2025). *Dynamic optimization of min-df in the GreedSum algorithm for enhanced extractive summarization*. *Artificial Intelligence Review*, 58(9), 270.
- Belhekar, V., & Bhargava, R. (2023). *Development of word count data corpus for Hindi and Marathi literature*. *Applied Corpus Linguistics*, 3(3), 100070.
- Bemke-Switilnik, M., & Drabek, A. (2016). *A citation analysis of the research reports of the Central Mining Institute. Mining and Environment using the Web of Science, Scopus*,

- BazTech, and Google Scholar: A case study. Journal of Scientometric Research*, 4(3), 153.
- Blanchard, A. (2007). *Understanding and customizing stopword lists for enhanced patent mapping. World Patent Information*, 29(4), 308-316.
- Blei, D. M., & Jordan, M. I. (2003). *Modeling annotated data. SIGIR Forum*, 127–134.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet allocation. Journal of Machine Learning Research*, 3, 993–1022.
- Bryman, A. (2016). *Social research methods. Oxford university press*.
- Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). *Making words work: Using financial text as a predictor of financial events. Decision Support Systems*, 50(1), 164–175.
- Cogburn, D., Ochieng, T., Shikako, K., Woods, J., & Aydin, M. (2025). *Uncovering policy priorities for disability inclusion: NLP and LLM approaches to analyzing CRPD state reports. Data & Policy*, 7, e61.
- Cotrina-Teatino, M. A., Marquina-Araujo, J. J., Mamani-Quispe, J. N., Arango-Retamozo, S. M., & Gonzalez-Vasquez, J. A. (2025). *Bibliometric and systematic analysis of the literature on the Hartwick rule in non-renewable resources. Resources Policy*, 107, 105654.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications*.
- Çelikbağ, M. A. (2025). *PREDICTING STUDENT PERFORMANCE IN ONLINE ENGLISH LANGUAGE LEARNING DURING CHALLENGING TIMES THROUGH LEARNING ANALYTICS*.
- Dang, S., & Ahmad, P. H. (2015). *A review of text mining techniques associated with various application areas. International Journal of Science and Research (IJSR)*, 4(2), 2461-2466.
- Das, A. K., Nandi, S., & Dutta, B. (2020, December 24). *Text Mining: Research Trend Analysis, Forecast and Citation Approach. Library Philosophy and Practice (e-journal)*.
- Demsar, J., & Zupan, B. (2021). *From experimental machine learning to interactive data mining. Orange: Data Mining Fruitful & Fun*.
- Demšar, J., Zupan, B., Leban, G., & Curk, T. (2004, September). *Orange: From experimental machine learning to interactive data mining. In European conference on principles of data mining and knowledge discovery (pp. 537-539). Berlin, Heidelberg: Springer Berlin Heidelberg*.
- Dhillon, I. S., & Modha, D. S. (2001). *Concept decompositions for large sparse text data using clustering. Machine Learning*, 42(1–2), 143–175.
<https://doi.org/10.1023/A:1007612920971>

- Dvořák, J., Tripes, S., Sokolova, M., & Musilova, I. (2022). *Trends in business strategy research, bibliometric analysis and text mining*. *Journal of Business Economics and Management (JBEM)*, 23(6), 1377-1397.
- Dzhunushalieva, G., & Teuber, R. (2024). *Roles of innovation in achieving the Sustainable Development Goals: A bibliometric analysis*. *Journal of Innovation & Knowledge*, 9(2), 100472.
- Elitaş, T., Kır Elitaş, S., & Ayhan, N. (2025). *Identify Future Trending Topics by Thematic Mapping of the Cinema Phenomenon Using Machine Learning and LDA*. *Sage Open*, 15(4), 21582440251390678.
- Eltaiba, N., Hosseini, S., & Hosseini, K. (2025). *Benefits and impact of technology-enhanced learning applications in higher education in Middle East and North Africa: A systematic review*. *Global Transitions*, 7.
- Farrell, M. J., Le Guillarme, N., Brierley, L., Hunter, B., Scheepens, D., Willoughby, A., Yates, A., & Mideo, N. (2024). *The changing landscape of text mining: A review of approaches for ecology and evolution*. *Proceedings of the Royal Society B*, 291(2024), 1–12.
- Galluccio, C., Beccherle, P., & Petrucci, A. (2025). *The narrative on tourism sustainability in Italian news: a text mining approach*. *Big Data Research*, 100541.
- García-León, R. A., & Teichert, T. (2024). *Food and social media: a research stream analysis*. *Management Review Quarterly*, 74(2), 1145-1183.
- Gatzert, N., & Reichel, P. (2024). *Sustainable investing in the US and European insurance industry: a text mining analysis*. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 49(1), 26-62.
- Gulati, S., Sinhababu, A., & Chakravarty, R. (2023). *Understanding the research landscape of smart libraries using text mining and data visualization: A use case of Voyant Tool*. *COLLNET Journal of Scientometrics and Information Management*, 17(1), 159-174.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). *Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation*. *Tourism management*, 59, 467-483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and. Techniques*, Waltham: Morgan Kaufmann Publishers.
- Hasan, M. R., Abdunurova, A., Wang, W., Zheng, J., & Shams, S. R. (2021). *Using deep learning to investigate digital behavior in culinary tourism*. *Journal of Place Management and Development*, 14(1), 43-65.
- Hevey, D. (2018). *Network analysis: a brief overview and tutorial*. *Health psychology and behavioral medicine*, 6(1), 301-328.
- Hibatullah, I., Rahman, R. I. N., Maulana, H., Utomo, D., & Kuswanto, A. D. (2025). *Sentiment Analysis of Twitter Data Related to the Ratification of the TNI Bill Using Orange Data Mining*. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(3), 2087-2094.

- Hicks, D. J., & Lobato, E. J. (2025). *Race science in mainstream psychology, 1960–2010*. *Quantitative Science Studies*, 6, 879-900.
- Huang, J. Y., & Liu, J. H. (2020). *Using social media mining technology to improve stock price forecast accuracy*. *Journal of Forecasting*, 39(1), 104-116. <https://doi.org/https://doi.org/10.1002/for.2616>
- Huang, Z., & Jiang, Z. (2024). *Text mining of syntactic complexity in L2 writing: an LDA topic modeling approach*. *International Review of Applied Linguistics in Language Teaching*.
- Hudaefi, F. A., Caraka, R. E., & Wahid, H. (2022). *Zakat administration in times of COVID-19 pandemic in Indonesia: a knowledge discovery via text mining*. *International Journal of Islamic and Middle Eastern Finance and Management*, 15(2), 271-286.
- Jha, V., Manjunath, N., Shenoy, P. D., & Venugopal, K. R. (2016, January). Hsra: Hindi stopword removal algorithm. In 2016 international conference on microelectronics, computing and communications (MicroCom) (pp. 1-5). IEEE. <https://doi.org/10.1109/MicroCom.2016.7522593>
- Jovic, A., Brkic, K., & Bogunovic, N. (2014, May). An overview of free software tools for general data mining. In 2014 37th International convention on information and communication technology, electronics and microelectronics (MIPRO) (pp. 1112-1117). IEEE
- Kao, A., & Poteet, S. R. (2007). *Natural language processing and text mining*. Springer.
- Kaur, B., & Singh, W. (2014). *Review on heart disease prediction system using data mining techniques*. *International journal on recent and innovation trends in computing and communication*, 2(10), 3003-3008.
- Kaur, J., & Buttar, P. K. (2018). *Stopwords removal and its algorithms based on different methods*. *International Journal of Advanced Research in Computer Science*, 9(5), 81-88.
- Keijzer, C., & van Lange, M. (2025). *Dutch War Letters (1935–1950) to Digital Data*. *Journal of Open Humanities Data*, 11(1).
- Kilinc, M., Aydin, C., Erdogan Aydin, G., & Balci, D. (2025). *Exploring the Urban Heat Island Effect: A Bibliometric and Topic Modeling Analysis*. *Sustainability*, 17(17), 8072.
- Kim, S. Y., Park, Y., & Shin, Y. W. (2024). *An Analysis on Research Trends in Rural Spaces using Text Mining*. *인간식물환경학회지 (JPPE)*, 27(6), 499-508.
- Kondo, R., Yoshida, T., & Hisano, R. (2023). *Masked prediction and interdependence network of the law using data from large-scale Japanese court judgments*. *Artificial Intelligence and Law*, 31(4), 739-771.
- Kurday, M., & Vladova, G. (2025). *Learning analytics and educational data mining applications: bibliometric and ChatGPT-based analysis of research publications from 2014 to 2023*. *Journal of Computers in Education*, 1-36.
- Ladani, D. J., & Desai, N. P. (2020, March). Stopword identification and removal techniques on tc and ir applications: A survey. In 2020 6th International Conference on Advanced

- Computing and Communication Systems (ICACCS) (pp. 466-472). IEEE. <https://doi.org/10.1109/ICACCS48705.2020.9074166>
- Larsen, V. H., & Thorsrud, L. A. (2019). *The value of news for economic developments. Journal of econometrics*, 210(1), 203-218. <https://doi.org/10.1016/j.jeconom.2018.11.013>
- Lee, J. (2024). *Bridging Bytes and Behaviors: Unraveling the Multifaceted Interplay of Technology and Employee Dynamics Over Time Through Text-Mining and Systematic Literature Review. Journal of Information Science Theory & Practice (JISaP)*, 12(4).
- Li, J., Feng, Y., Li, G., & Sun, X. (2020). *Tourism companies' risk exposures on text disclosure. Annals of tourism research*, 84, 102986.
- Liu, B., & Yao, G. (2025). *Media Representation of Autism in Chinese Official News: A Topic-Modeling-Assisted Discourse Study. SAGE Open*, 15(3), 21582440251382553.
- Macanović, A., & Przepiorka, W. (2024). *A systematic evaluation of text mining methods for short texts: Mapping individuals' internal states from online posts. Behavior Research Methods. https://doi.org/10.3758/s13428-024-02291-5*
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., & Adam, S. (2021). *Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In Computational methods for communication science (pp. 13-38). Routledge.*
- Mala, Y. A., Selivyorstova, T. V., & Huda, A. I. (2024). *Використання технології Orange для інтелектуального аналізу даних в освітній галузі [Use of Orange technology for data mining in the field of education]. https://doi.org/10.34185/1562-9945-3-152-2024-12*
- Marsden, P. (2005). *Network analysis. In Encyclopedia of Social Measurement. Elsevier.*
- Maulana, I. A., & Wibowo, A. P. W. (2025). *Analysis of public sentiment text clustering on tax increases using Orange data mining on Twitter. Brilliance: Research of Artificial Intelligence*, 5(1), 93-99.
- Mirzyńska, A., Kosch, O., Schieg, M., Šuhajda, K., & Szarucki, M. (2021). *Exploring concomitant concepts in the discussion on the circular economy: A bibliometric analysis of Web of Science, Scopus and Twitter. Technological and Economic Development of Economy*, 27(6), 1539-1562.
- Mnasri, M. (2019). *Recent advances in conversational NLP: Towards the standardization of Chatbot building. arXiv preprint arXiv:1903.09025.*
- Mohi, Z. R. (2020). *Orange Data Mining as a tool to compare Classification Algorithms. Dijlah Journal of Sciences and Engineering*, 3(3), 13-23
- Mostafa, M. M., Alhur, M., & Moustafa, A. M. (2025). *A structural topic modeling of communication research: insights from over a century of journals' abstracts. International Journal of Information Management Data Insights*, 5(2), 100364.

- Mujahid, M., Rustam, F., Shafique, R., Chunduri, V., Villar, M.G., Ballester, J.B., Diez, I.d. I.T., Ashraf, I., 2023. Analyzing sentiments regarding ChatGPT using novel BERT: a machine learning approach. *Information* 14 (9), 474. <https://doi.org/10.3390/info14090474>.
- Muñoz, K. (2019). *Orange3 Text Mining documentation*. Biolab.
- Musilová, I., Dvořák, J., Jánský, J., & Bolek, V. (2023). *Trends in Performance Research in Relation to Business Strategy: Bibliometric Analysis and Text Mining*. *Central European Business Review*, 12(3), 143.
- Musunuru, K. (2025). *Emerging Technologies for Socioenvironmental Auditing: Identification of Factors, Challenges and Technologies Using Text Mining and Analysis*. *Revista Kawsaypacha: Sociedad y Medio Ambiente*, (15).
- Naz, F., Kumar, A., Agrawal, R., Garza-Reyes, J. A., Majumdar, A., & Chokshi, H. (2024). *Artificial intelligence as an enabler of quick and effective production repurposing: an exploratory review and future research propositions*. *Production Planning & Control*, 35(16), 2154-2177.
- Nie, B., & Sun, S. (2017). *Using text mining techniques to identify research trends: A case study of design research*. *Applied Sciences*, 7(4), 401. <https://doi.org/10.3390/app7040401>
- Nuriu, R. (yıl). *Sentiment analysis in Twitter using Orange* (Master's thesis). Epoka University, Faculty of Architecture and Engineering.
- Okoye, K., Campos, E., Das, A., Chakraborty, V., Ghosh, M., Chakrabarti, A., & Hosseini, S. (2025). *Impact of digitalized-education upon sustainable education and practice: A systematic review and meta-analysis of literature based on pre-intra-and-post pandemic and rural education development*. *Sustainable Futures*, 100851.
- Oner, B., Hakli, O., & Zengul, F. D. (2024). *A text mining and network analysis of topics and trends in major nursing research journals*. *Nursing open*, 11(1), e2050.
- Özmantar, M. F., Gökdağ, K., Hangül, T., & Agaç, G. (2024). *Research themes and trends in the field of teacher educators: A topic modelling study*. *Teaching and Teacher Education*, 148, 104696.
- Piwowar-Sulej, K., Wawak, S., Tyrańska, M., Zakrzewska, M., Jarosz, S., & Sołtysik, M. (2023). *Research trends in human resource management. A text-mining-based literature review*. *International Journal of Manpower*, 44(1), 176-196.
- Queiroz, M. M., Roque, C., Moura, F., & Marôco, J. (2024). *Understanding the expectations of parents regarding their children's school commuting by public transport using latent Dirichlet Allocation*. *Transportation research part A: policy and practice*, 181, 103986.
- Raaymann, S., & Spinler, S. (2024). *Measuring supply chain resilience along the automotive value chain—A comparative research on literature and industry*. *Transportation Research Part E: Logistics and Transportation Review*, 192, 103792.

- Rani, R., & Lobiyal, D. K. (2022). *Performance evaluation of text-mining models with Hindi stopwords lists. Journal of King Saud University-Computer and Information Sciences*, 34(6), 2771-2786.
- Ratra, R., & Gulia, P. (2020). *Experimental evaluation of open source data mining tools (WEKA and Orange). International Journal of Engineering Trends and Technology*, 68(8), 30-35.
- Sandu, A., Cotfas, L.-A., Stănescu, A., & Delcea, C. (2024). *A bibliometric analysis of text mining: Exploring the use of natural language processing in social media research. Applied Sciences*, 14(8), 3144.
- Shi, X., & Chen, Z. (2021). *Listening to your employees: analyzing opinions from online reviews of hotel companies. International Journal of Contemporary Hospitality Management*, 33(6), 2091-2116.
- Shrestha, S. (2024). *Using Text Mining and Sentiment Analysis to Enhance Customer Experience and Service Quality in Hotel Operations: A Case Study of a Finnish Hotel (Master's thesis). Novia University of Applied Sciences.*
- Smith, G. (2020). *Data mining fool's gold. Journal of Information Technology*, 35(3), 182-194.
- Song, C., Guo, J., Gholizadeh, F., & Zhuang, J. (2022). *Quantitative analysis of food safety policy—Based on text mining methods. Foods*, 11(21), 3421. <https://doi.org/10.3390/foods11213421>
- Song, G., Wu, J., & Wang, S. (2021). *[Retracted] Text Mining in Management Research: A Bibliometric Analysis. Security and Communication Networks*, 2021(1), 2270276.
- Sonk, M., Tunger, D. Trend mining with Orange(2024) – using topic modeling in futures research with the example of urban mobility. *Eur J Futures Res* 12, 6 (2024). <https://doi.org/10.1186/s40309-024-00229->
- Sormanen, N., Hietajärvi, L., & Wilska, T. A. (2025). *Young people's online political participation in an apolitical digital sphere: identifying political topics and expression patterns on a youth discussion forum. Journal of Youth Studies*, 1-18.
- Stansfield, C., O'Mara-Eves, A., & Thomas, J. (2017). *Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Research synthesis methods*, 8(3), 355-365.
- Sumathy, K. L., & Chidambaram, M. (2013). *Text mining: concepts, applications, tools and issues-an overview. International Journal of Computer Applications*, 80(4). <https://doi.org/10.5120/13851-1685>
- Sundaram, G., & Berleant, D. (2023). *Automating systematic literature reviews with natural language processing and text mining: A systematic literature review. In International Congress on Information and Communication Technology (pp. 73-92). Springer, Singapore.*

- Taecharungroj, V., 2023. What can ChatGPT do? analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data Cogn. Comput.* 7 (1), 35. <https://doi.org/10.3390/bdcc7010035>.
- Thakur, K., & Kumar, V. (2022). *Application of text mining techniques on scholarly research articles: Methods and tools. New Review of Academic Librarianship*, 28(3), 279-302.
- Thange, U., Shukla, V. K., Punhani, R., & Grobbelaar, W. (2021, January). Analyzing COVID-19 dataset through data mining tool orange. In 2021 2nd international conference on computation, automation and knowledge management (ICCAKM) (pp. 198-203). IEEE.
- Tian, Y., & Chen, K. H. (2024). *Emerging landscapes of alternative-academic careers in library and information science: Evolutionary patterns and prospects in the Chinese context. Humanities and Social Sciences Communications*, 11(1), 1-10.
- Tran, P. P. (2024). *High-Quality Human Resources in the Labor Market: A First Assessment of Potential and Obstacles. SAGE Open*, 14(3), 21582440241267259.
- Tyagi, A. (2021). *A review study of natural language processing techniques for text mining. International Journal of Engineering Research & Technology*, 10(9), 586-589.
- Urru, S., Sciannameo, V., Lanera, C., Salaris, S., Gregori, D., & Berchialla, P. (2022). *A topic trend analysis on COVID-19 literature. Digital health*, 8, 20552076221133696.
- Valverde-Berrocoso, J., González-Fernández, A., & Acevedo-Borreg, J. (2022). *Disinformation and multiliteracy: A systematic review of the literature. Comunicar: Media Education Research Journal*, 30(70), 93-105.
- Vestrelli, R., Colladon, A. F., Fabiani, C., & Pisello, A. L. (2025). *Energy narratives in Europe: Exploring the link between online news and renewable energy production. Energy Research & Social Science*, 128, 104345.
- Vinayavekhin, S., Li, F., Banerjee, A., & Caputo, A. (2023). *The academic landscape of sustainability in management literature: Towards a more interdisciplinary research agenda. Business Strategy and the Environment*, 32(8), 5748-5784.
- Wang, Y., Wang, L., Wang, X., & Xie, F. (2024). *Decoding online literature: A Quantitative Analysis of Language Features in Chinese Online Literature. Glottometrics*, 57.
- Weissenfels, S., Nissen, A., & Smolnik, S. (2025). *Advancing digital health in information systems research: Insights from a text mining analysis. Electronic Markets*, 35(1), 1-25. <https://doi.org/10.1007/s12525-025-00768-w>
- Wilbur, W. J., & Sirotkin, K. (1992). *The automatic identification of stop words. Journal of information science*, 18(1), 45-55.
- Witten, I. H., Don, K. J., Dewsnip, M., & Tablan, V. (2004). *Text mining in a digital library. International Journal on Digital Libraries*, 4(1), 56–59. <https://researchcommons.waikato.ac.nz/server/api/core/bitstreams/1eafdaf4-519e-4ab3-ade6-23d114884315/content>

- Wu, H., Zhang, L., Huang, Z., Chen, Z., Yang, W., Tong, X., & Li, S. (2024). *Unraveling minnan imagery: a comprehensive analysis of traditional and modern minnan nursery rhymes through complex networks*. *Heritage Science*, 12(1), 180.
- Wu, X., Deng, K., & Zhao, T. (2024). *Mapping Computational Communication Research: A Methodological Breakthrough and Thematic Exploration*. *Journal of Social Computing*, 5(3), 242-260.
- Xia, H., Muskat, B., Karl, M., Li, G., & Law, R. (2024). *Destination competitiveness research over the past three decades: a computational literature review using topic modelling*. *Journal of Travel & Tourism Marketing*, 41(5), 726-742.
- Xu, X. (2018). *Does traveler satisfaction differ in various travel group compositions? Evidence from online reviews*. *International Journal of Contemporary Hospitality Management*, 30(3), 1663-1685. <https://doi.org/10.1108/IJCHM-03-2017-0171>
- Ye, J., & Xue, M. (2021). *Influences of sentiment from news articles on EU carbon prices*. *Energy Economics*, 101, 105393. <https://doi.org/10.1016/j.eneco.2021.105393>.
- Zeng, J., et al. (2019). *Operationalization of next-generation sequencing and decision support for precision oncology*. *JCO Clinical Cancer Informatics*, (3), 1–12. <https://doi.org/10.1200/CCI.19.00089>
- Zhang, L., Wang, Y., Xu, Z., & Chen, L. (2025). *Quantitative Study on the Synergistic Effect of China's Plastic Restriction Policy from 2008 to 2025*. *Sustainability*, 17(16), 7355.
- Zhang, Y., Lu, X., & Wang, Y. (2025). *Research on evaluation of implementation effect of science and technology innovation policy*. *Sustainable Futures*, 10, 101228.
- Zhao, R., Yang, M., Long, D., Fan, X., Zhang, Z., Duan, Y., & Chen, J. (2026). *The evolution of transformative resilience research topics: A data-driven analysis*. *Cities*, 168, 106425.
- Zhao, Y., Xu, X., & Wang, M. (2019). *Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews*. *International journal of hospitality management*, 76, 111-121. <https://doi.org/10.1016/j.ijhm.2018.03.017>
- Zheng, H., & Pee, L. G. (2024). *Research knowledge utilisation for societal impact: Information practices based on abductive topic modelling*. *Journal of Information Science*, 50(1), 129-144.
- Zheng, Y., Zeng, M., Deng, B., Bin, H., Liu, F., & Lin, X. (2024). *Topic Mining and Sentiment Analysis based on the Online Review from MOOC Platform's Logistics Course*. *SAGE Open*, 14(4), 21582440241303486.
- Zhu, H., & Liu, X. (2024). *Do linguistic features of research article titles affect received online attention? A corpus-based analysis*. *Library Hi Tech*, 42(6), 2000-2016.
- Zyoud, M. M., Bsharat, T. R. K., & Dweikat, K. A. (2024). *Quantitative research methods: Maximizing benefits, addressing limitations, and advancing methodological frontiers*. *ISRG Journal of Multidisciplinary Studies*, 2(4), 11-14.

ÖZGEÇMİŞ

Adı-Soyadı: Ömer YILMAZ

Eğitim ve Mesleki Geçmişi:

- 2020-2026 Emniyet Genel Müdürlüğü – Memur
- 2012-2017 Afyon Kocatepe Üniversitesi – İngilizce İşletme Bölümü

Yabancı Dil Bilgisi:

- İngilizce

