



# Üniversite Bilgi Yönetim Sistemi Servis Destek Taleplerinin Konu Modelleme Tabanlı Analizi

Aytuğ Onan<sup>1\*</sup>, Adnan Yalçın<sup>2</sup>, Erdem Atik<sup>2</sup>

<sup>1</sup> İzmir Katip Çelebi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ORCID: 0000-0002-9434-5880)  
<sup>2</sup> İzmir Katip Çelebi Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilim Dalı, İzmir, Türkiye

(Bu yayın 26-27 Haziran 2020 tarihinde HORA-2020 kongresinde sözlü olarak sunulmuştur.)

(DOI: 10.31590/ejosat.780642)

**ATIF/REFERENCE:** Onan, A., Yalçın, A., & Atik, E. (2020). Üniversite Bilgi Yönetim Sistemi Servis Destek Taleplerinin Konu Modelleme Tabanlı Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 389-397.

## Öz

Kurumumuz (İzmir Kâtip Çelebi Üniversitesi) bünyesinde Üniversite Bilgi Yönetim Sistemi(ÜBYS) geliştirilmekte ve kendi kurumumuz dışında çeşitli kurumlar tarafından da ÜBYS altındaki farklı bileşenler kullanılmaktadır. Yazılım geliştirmenin doğasından kaynaklanan ve yazılımın büyüklüğü ile doğru orantılı olarak süreç içerisinde çeşitli hatalar oluşabilmekte, oluşan hatalardan bağımsız olarak çeşitli geliştirmelerin yapılması istenmektedir. Bu çalışmada, Üniversite Bilgi Yönetim Sistemi(ÜBYS) geliştirilirken farklı kurum ve bu kurumlardaki kişilerden gelen hata bildirimleri ve isteklerin konu modelleme yöntemlerine dayalı analizi gerçekleştirilmektedir. Servis destek taleplerinin konu modellemesi sonucunda ortaya çıkan konu modelleri ÜBYS servis destek veri koleksiyonundan bulunup çıkartılan anahtar kavramlar olarak adlandırılabilirler. Çıkartılan konular bir terim koleksiyonu olarak ifade edilmekle birlikte metin dokümanlarının (ÜBYS servis destek taleplerinin) büyük bir kısmını özetlemek için çok değerlidir. Dahası verilerdeki gizli kalıplar ve anlamsallık ortaya çıkarılmış olmaktadır. Büyük boyutlu dokümanlardan denetimsiz bir şekilde gizli yapıyı keşfeden konu modelleme güçlü bir yöntemdir. Konu modelleme yapılandırılmamış (unstructured) metin gövdelerinin büyük koleksiyonlarını anlamamıza, bilgileri düzenlememize ve sunmamıza yardımcı olur. Konu modellemesi için kullanılan gizli Dirichlet tahsisi (latent Dirichlet allocation), her belgenin bir konu koleksiyonu olarak kabul edildiği ve belgedeki her kelimenin konulardan birine karşılık geldiği bir konu modelleme yöntemidir. Dolayısıyla, bir belge(metin verisi) verildiğinde LDA, belgeyi temel alarak her konu grubunu o grubu en iyi açıklayan bir dizi kelimenin olduğu konu gruplarına kümeleştirir. Bu çalışmada ÜBYS bünyesindeki Öğrenci Bilgi Sistemi (AIS), Personel Bilgi Sistemi (HRM), Elektronik Belge Yönetim Sistemi(ERMS) ve Bilimsel Araştırma Projeleri (SRP) olmak üzere dört bileşen ele alınmıştır. Bu bileşenlere ait servis destek taleplerinin konu modellemesi için en temel yöntemlerden biri olan gizli Dirichlet tahsisi ile analizi gerçekleştirilmiştir. Bileşenlerden elde edilen metin belgeleri üzerinde temel konulara ve konulara ilişkin temel anahtar sözcüklere ilişkin analiz ve görseller sunulmaktadır.

**Anahtar Kelimeler:** Yazılım Geliştirme, Servis Destek Talepleri, Metin Madenciliği, Konu Modelleme, Gizli Dirichlet Tahsisi.

## Topic Modelling Based Analysis of Service Support Requests on University Information Management System

### Abstract

University Information Management System (ÜBYS) is being developed within our institution (İzmir Kâtip Çelebi University) and various institutions outside of our own institution use different components under ÜBYS. Various errors may arise in the process in line with the size of the software and arise from the nature of the software development, and various improvements are requested regardless of the errors that occur. In this study, we present a topic modelling based analysis on service support requests provided by different institutions and people on university information management system. Topic models that arise because of topic modelling of service support requests can be regarded as key concepts derived from the data collection. The issues raised are expressed as a collection of terms, but are of great value to summarize most of the text documents (ÜBYS service support requests). Moreover, hidden patterns and semantics in the data are revealed. Topic modeling helps us understand large collections of unstructured text

\* Sorumlu Yazar: İzmir Kâtip Çelebi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye, ORCID: 0000-0002-9434-5880, [aytug.onan@ikcu.edu.tr](mailto:aytug.onan@ikcu.edu.tr)

bodies, organize and present information. Latent Dirichlet allocation, which is used for topic modeling, is an important method of topic modeling in which each document is considered a collection of topics and each word in the document corresponds to one of the topics. Therefore, when a document (text data) is given, LDA clusters each subject group into subject groups with a set of words that best describe that group based on the document. In this study, four components, namely, Student Information System (AIS), Personnel Information System (HRM), Electronic Document Management System (ERMS) and Scientific Research Projects (SRP) within the body of information system are taken into account. Analysis of the service support requests of these components has been carried out by latent Dirichlet allocation, which is one of the most basic methods for topic modeling. The study presents the main results and visualizations obtained by the latent Dirichlet allocation method.

**Keywords:** Software Development, Service Support Requests, Text Mining, Topic Modelling, Latent Dirichlet Allocation.

## 1. Giriş

Günümüzdeki gelişmeler veri birikiminin artmasına neden olmaktadır. Bunun sonucu olarak, depolanan verinin boyutu da her geçen gün artmaktadır. Depolanan verinin çok büyük oranda metin verisi olduğu göz önüne alındığında, metin verilerinin otomatik analizi oldukça önemli bir araştırma problemi haline dönüşmektedir. Bu doğrultuda geniş ölçekli metin verilerinde arama, anlama ve işleme görevlerini yerine getirecek otomatik araçlara ihtiyaç ortaya çıkmıştır. Bu boşluğu doldurmak için konu modelleme yöntemleri, makine öğrenmesi, doğal dil işleme ve bilgi çıkarımı süreçlerinde yaygın şekilde uygulanmaya başlanmıştır. Konu modelleme, geniş ölçekli metin koleksiyonlarından anlamsal bilgiye erişimde uygulanan denetimsiz bir makine öğrenmesi yöntemidir. Literatürde, araştırmacılar tarafından geliştirilen mevcut pek çok konu modeli bulunmaktadır. Bu konu modelleri arasında en yaygın kullanılan yöntemlerden birisi gizli Dirichlet tahsisi (latent Dirichlet allocation, LDA) yöntemidir [1-5]. Metin belgeleri gibi ayrık verileri modellemek için geliştirilen temel bir grafiksel model olan LDA, metin belgesi içerisinde yer alan gizli konuları ortaya çıkarmayı hedefler [6]. LDA yönteminin dayandığı temel fikir, konuların sabit bir sözlük üzerinden olasılık dağılımına sahip olması ve dokümanların gizli konuların rastgele bileşiminden oluşmasıdır. Bu temel fikre göre LDA, doküman koleksiyonundaki konuları, konuları oluşturan kelimelerin konular altındaki olasılıklarını, dokümanlar için o dokümanı oluşturan kelimelerin hangi konulara atandığını ve her doküman için bu dokümandaki konuların dağılımını öğrenmektedir [7].

Bu çalışma kapsamında, içerisinde dört farklı kategoride 17,831 adet kayıttan oluşan bir veri seti kullanılmıştır. Yapısal olmayan bu veriler üzerinde modellemenin uygulayabilmesi için öncelikle metinlerin işlenmesi sağlanmıştır. İşlenen metinler daha sonra yapısal bir forma çevrilmiştir. Gizli Dirichlet tahsisi (LDA) kullanılarak bu dokümanlara (kayıtlara) ait konu modelleri oluşturulmuştur [8, 9]. Konu modeli sayıları, karşılaştırma yapabilmek için 4 ve 8 olarak belirlenmiş olup, tüm kayıtlar için önce 4 konu modeli daha sonra da 8 konu modeli kümeleme işlemi gerçekleştirilmiştir. Her bir konu modellemesi için, konuların birbiri ile olan ve konulardaki kelimelerin konularla olan ilişkileri görsel olarak gösterilmiştir. Konuların birbirleri ile ne oranda iç içe geçtikleri ve ayrıştıkları belirtilmiştir.

Çalışmanın geri kalanı şu şekildedir: İkinci bölümde ilgili çalışmalar, üçüncü bölümde çalışmanın metodolojisi, dördüncü bölümde konu modelleme tabanlı analiz sonuçları ve tartışma sunulmaktadır.

## 2. İlgili Çalışmalar

Bilindiği kadarıyla, yazılım geliştirme sürecinde servis destek taleplerinin konu modelleme tabanlı analizine yönelik daha önce gerçekleştirilmiş benzer bir çalışma bulunmamaktadır. Bunun ile birlikte, konu modelleme yöntemlerinin son yıllarda metin madenciliği alanında sıklıkla kullanıldığı görülmektedir. Bu bölümde, metin madenciliğinde konu modellemeye yönelik gerçekleştirilen çalışmalara kısaca değinilmektedir.

Ekinci vd. [10] yapmış oldukları çalışmada, Türkiye’de tıp alanında çalışmakta olan araştırmacıların hangi konularda daha aktif çalıştığını saptamanın yararlı olacağını düşünerek, tıp alanında son yıllarda güncel olan konu başlıklarının otomatik olarak belirlenmesine yönelik olarak gizli Dirichlet tahsisi yöntemi tabanlı bir analiz sunmaktadır.

Onan vd. [11] tarafından gerçekleştirilen çalışmada, metin madenciliğinde kullanılan temel veri kümeleri üzerinde duygu analizi gerçekleştirilmesinde konu modelleme tabanlı öznelik seçiminin etkinliği değerlendirilmiştir. Gizli Dirichlet tahsisi yöntemi sonucu elde edilen konular, öznelik olarak alınarak, temel makine öğrenmesi ve sınıflandırıcı topluluğu yöntemlerinin etkinlikleri değerlendirilmiştir. Benzer şekilde, Onan [12] tarafından gerçekleştirilen çalışmada, Türkçe twitter mesajları üzerinde duygu analizi gerçekleştirmek için gizli Dirichlet tahsisine dayalı öznelik çıkarımının etkinliği incelenmiştir.

Karami vd. [13] yapmış oldukları çalışmada, bulanık kümeleme yöntemi kullanarak bulanık mantık tabanlı bir konu modelleme yöntemi önerisinde bulunmaktadır. Geliştirilen bulanık mantık tabanlı konu modelleme yöntemi, tıbbi metin belgelerinin modellenmesi ve sınıflandırılması için kullanılmıştır. Tıbbi metin belgeleri üzerinde yapılan deneysel analizler, bulanık küme teorisinin, tıbbi belgelerin konu modellemesinde etkin olduğunu göstermektedir.

Bagheri vd. [14] dokümanın cümle yapısını esas alan ve LDA’nın bir uzantısı olarak tasarladığı ADM-LDA (Aspect Detection Model-LDA) yöntemini belirgin özellikleri kullanıcı yorumlarından çıkartmak için önermişlerdir. Bu yaklaşımda sözcük torbası yaklaşımı yok sayılmış ve özelliklerin bir Markov Zincirinden geldiği ve koşullu bağımsız oldukları varsayılmıştır. Wang vd. [15] FL-LDA (Finegrained Labeled LDA) ve UFL-LDA (Unified Finegrained Labeled-LDA) olmak üzere iki tane yarı denetimli konu modelleme yöntemi ile kullanıcı yorumlarından ürün özelliklerini çıkartmışlardır. FLLDA yönteminde ürün özelliklerinin çekirdek kümesi modele dâhil edilerek bu çekirdek küme ile ilişkili kelimelerin ürün özellikleri olarak etiketlenmesi amaçlanmıştır. UFL-LDA ise FL-LDA’nın iyileştirilmiş bir biçimi olarak sunulmuştur.

Zheng vd. [16] restoran, otel, MP3 çalar ve kamera yorumları üzerinde gerçekleştirdikleri, konu modelleme tabanlı analiz ile belirgin ürün özelliklerini kullanıcı yorumlarından çıkartmışlardır. Önerdikleri yöntem, cümle seviyesinde olup bir cümleyi oluşturan tüm kelimelerin aynı konuya ait olduğu fikrini esas almaktadır. Yine burada da özellikler ve duygu ifadeleri eşzamanlı olarak çıkartılmışlardır.

Jo ve Oh [17] aynı başlık altındaki ürün özelliklerinin yorum içerisinde birbirine yakın oldukları fikrinden yola çıkarak bir cümledeki tüm kelimelerin tek bir ürün özelliği ile ilişkili olduğu yaklaşımını varsayan S-LDA (Sentence-LDA) yöntemini önermişlerdir. Sonra ise bu yöntemin gelişmiş bir hali olan ASUM'u (Aspect Sentiment Unification) geliştirmişlerdir. Bu yöntemde ürün özellikleri ve duygu ifadeleri birlikte modellenerek (özellik, duygu ifadesi) çiftlerinin çıkartılması sağlanmıştır. Bu amaçla duygu ifadelerinin küçük bir kümesinden de yararlanılmıştır. Yöntemler, elektronik ve restoran veri kümelerine uygulanarak ürün özellikleri ve (özellik, duygu ifadesi) çiftleri başarılı bir şekilde elde edilmiştir.

Li vd. [18] dokümandaki belirgin özellikleri çıkarmak amacıyla Sentiment-LDA ve DependencySentiment-LDA olmak üzere iki yöntem önermişlerdir. Önerdikleri yöntemler ile ürün özelliklerini çıkartırken, eşzamanlı olarak duygu ifadelerini de çıkartmışlardır. Bu yöntemleri; dokümandaki duygu ifadelerinin temel varlık ile ilişkili olması fikrinden yola çıkarak geliştirmişler ve duygu ifadeleri ve belirgin özellikleri bütün olarak ele alınmıştır. Dependency Sentiment-LDA'da ayrıca duygu yönlerinin belirlenmesini amaçlamışlardır.

Xianghua vd. [19] Çince yazılmış yorumlar üzerinden LDA ile ilk adımda genel konuları çıkartmışlardır. Lokal konuların ve ilişkili duygu ifadelerinin çıkartılmasında ise LDA'yı tüm dokümanda denemek yerine kayan pencere tabanlı bir yöntem önermişlerdir.

Onan vd. [20] tarafından gerçekleştirilen çalışmada, üniversite bilgi sisteminden elde edilen servis destek taleplerine ilişkin metin belgelerinde, makine öğrenmesi ve temel metin temsil yöntemlerinin etkinlikleri değerlendirilmiştir. Bu çalışmada, TF-IDF tabanlı temsil edilen metin belgeleri, Naive Bayes, k-en yakın komşu, C4.5 karar ağacı, rastgele orman algoritması ve destek vektör makineleri sınıflandırıcılarıyla değerlendirilmiştir. Deneysel analizlerde, destek vektör makineleri ile %92.26 doğru sınıflandırma oranı elde edildiği görülmüştür.

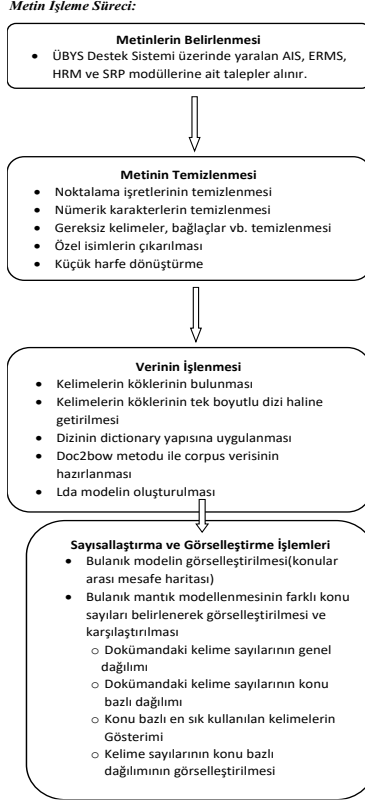
### 3. Metodoloji

Bu bölümde, üniversite bilgi sisteminden elde edilen servis destek taleplerinin gizli Dirichlet tahsisi yöntemine dayalı analizinde izlenen yöntem tanıtılmaktadır.

Metin madenciliğinde temel olarak, yapısal olmayan verinin yapılandırılmış veriye dönüştürülmesi gerekir. Öncelikle veritabanından alınan veri setinin istenilen özellikte olması gerekmektedir. İstenilen özelliklerde toplanan veri seti üzerinde dizgeciklere ayırma (tokenization), durdurma sözcükleri filtreleme (stopword filtering), kök bulma (stemming) ön işleme adımları yapılır. Daha sonra veri seti doğal dil işleme yöntemleri ile metin madenciliği algoritmalarıyla işlenmeye hazır hale getirilir. Metin verisindeki kelimeler token olarak ifade edilir ve her token için tekil bir indeks numarası olacak şekilde sözlük yapısı (dictionary) oluşturulur. Bu yapı kelimeler ve onlara ait indeks (id) numaralarını barındıran bir sözlük yapısıdır. Sözlük yapısı derleme dönüştürülerek ilgili metinlerin düzenli ve yapısal olarak bir arada bulunması sağlanır. Derlem, kelimelerin indekslerinden ve metin verisi içindeki frekanslarından oluşur. Oluşturduğumuz derlem bir doküman terim matrisidir ve LDA konu modeli için girdi matrisidir. LDA konu modelleri derlem kullanılarak 4 ve 8 toplam konu sayılı olarak oluşturulur. Oluşturulan konu modelleri daha sonra karşılaştırılarak ve görselleştirilerek analiz edilmiştir. Şekil 1'de konu modelleme sürecinde izlenen adımlar ve temel etkinlikler özetlenmektedir. Konu modellemede izlenen süreç şu şekilde özetlenebilir:

- Toplanacak metinlerin belirlenmesi: ÜBYS destek sistemi üzerinden iletilen talepler önceden belirlenmiş 4 kategori üzerinden elde edilmiştir.
- Metinleri organize edilip düzenlenme işlemi: Metin belgeleri üzerinde verilerin temizlenmesi gerçekleştirilmiştir. Öncelikli olarak ham veriler üzerinde sınıflandırma aşamasında hiçbir anlam ifade etmeyecek olan noktalama karakterleri temizlenmiştir. Ardından sayısal karakterlerin temizlenmesi işlemi uygulanmıştır. Böylelikle KVKK kapsamında yer alabilecek kimlik bilgileri vb. sayısal karakterlerden arındırılması sağlanmıştır. Devamında metin üzerinde yer alan kelimeler normalizasyon işlemine tabi tutulmuştur. Normalizasyon yöntemi ile yazım hataları bulunan kelimeler saptanarak, karşılık olarak uygun kelimelere dönüştürülmüş, böylelikle metinler en uygun hale getirilmiştir. Böylelikle özniteliklerin belirlenmesi öncesinde anlam içermeyen, hatalı olabilecek kelimelerin normalize edilmesi ile sınıflandırma aşamasında özniteliklerin en aza indirgenmesi sağlanmıştır. Ardından durdurma sözcükleri filtreleme (stopword filtering) işlemi uygulanmıştır. Böylelikle değerlendirmeye katılmaması gereken sözcükler kapsam dışında bırakılmıştır. Stopwords filtering için bazı veri setleri belirlenmiş, bu setlerin birleşiminden ortak bir havuz oluşturulmuştur. Bu havuz içinde yeralan verilerin metinlerden çıkartılması ile gereksiz kelimeler ile bağlaç, edat vb kelimeler egale edilmiş hale gelmiştir. Bu işlemin ardından Türkçe isim listesi (Ali, Ahmet, Mehmet, vb...) için yine aynı yöntem uygulanmış ve metinlerde yeralan isimler ortadan kaldırılmıştır.
- Temizleme işlemi sonrasında metinlerin tamamı küçük harfe dönüştürülerek, programsal anlamda ortak bir küme üzerinde rahat şekilde işlem yapılmasına hazır hale getirilmiştir.
- Kelimenin köklere indirgenme işlemi: Normalize edilmiş metinler üzerinde yeralan kelimelerin köklere ayrılma işlemi uygulanmıştır. Bu işlemin amacı aynı köke sahip fakat farklı ekleri bulunan kelimelerin tekilleştirme işleminin sağlanmasıdır. Bağlaçlar, imleçler, kalıplaşmış kısaltmalar vb durumlar gözönünde bulundurularak, kelimelerin en yalın hale getirilmesi sağlanmıştır. Kelimelerin kök halinin saptanması amacıyla türkçenin kural tabanlı yapısını gözönünde bulunduran Zemberek Kütüphanesi tercih edilmiştir.

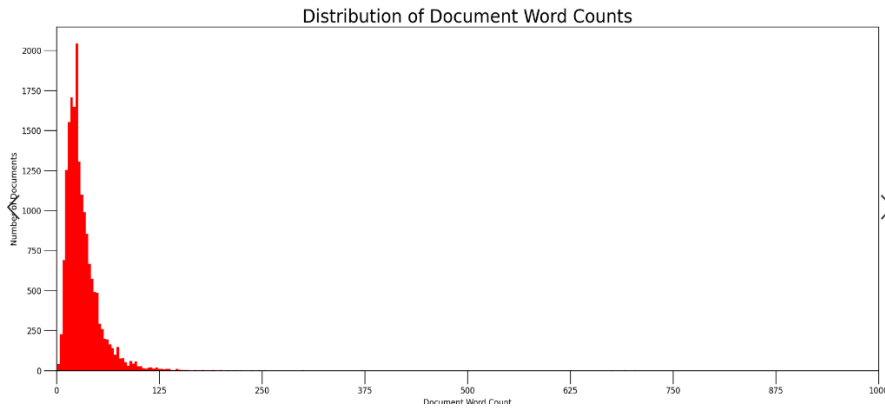
- Modelin oluşturulması: Kök haline getirilmiş kelimeler, doküman tabanlı olacak şekilde dizi olarak edinilir. Kelime dizileri gensim kütüphanesi kullanılarak token olarak ifade edilir ve her token için indeks numarası oluşturulur. İlgili diziler sözlük yapısına dönüştürülür. Sözlük haline getirilen nesne ardından terim frekanslarını içerek yapıya yani corpusa dönüştürülür. Kelime-frekans eşleşiminin(corpus) ilk indexindeki (0,1)'de ifade edilmek istenen, kelime indeksi 0 olan kelimenin ilk metin verisinde(dokümanda) kaç kere tekrarlanmış olduğudur. Frekans ağırlıkları belirlenmiş kelimeler, belirtilen konu sayısı parametresi ile birlikte LDA modelinin oluşturulması sağlanır.



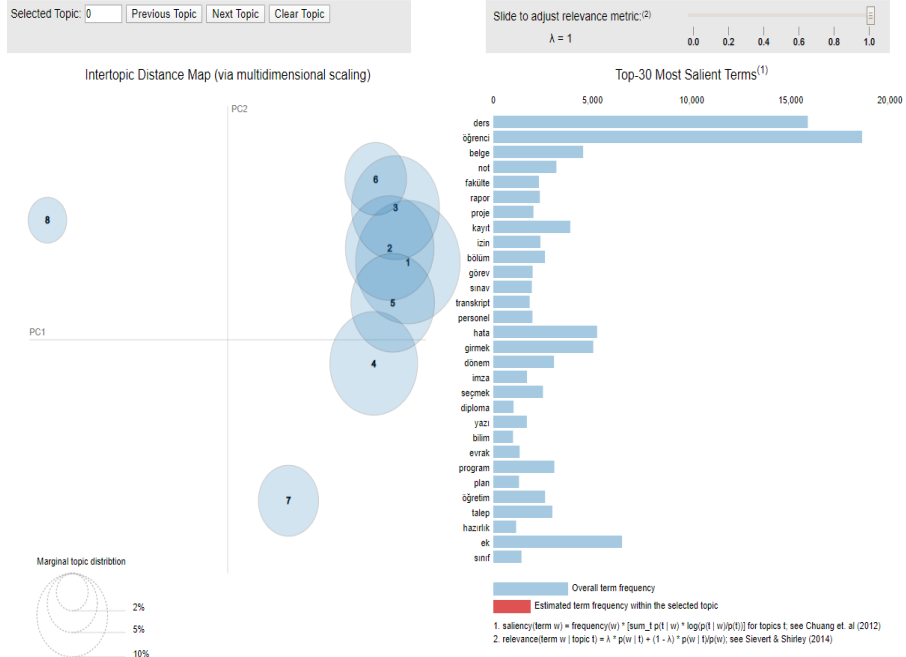
Şekil. 1. Metin işleme sürecinde izlenen temel adımlar

#### 4. Verilerin Görselleştirilmesi ve Analiz Sonuçları

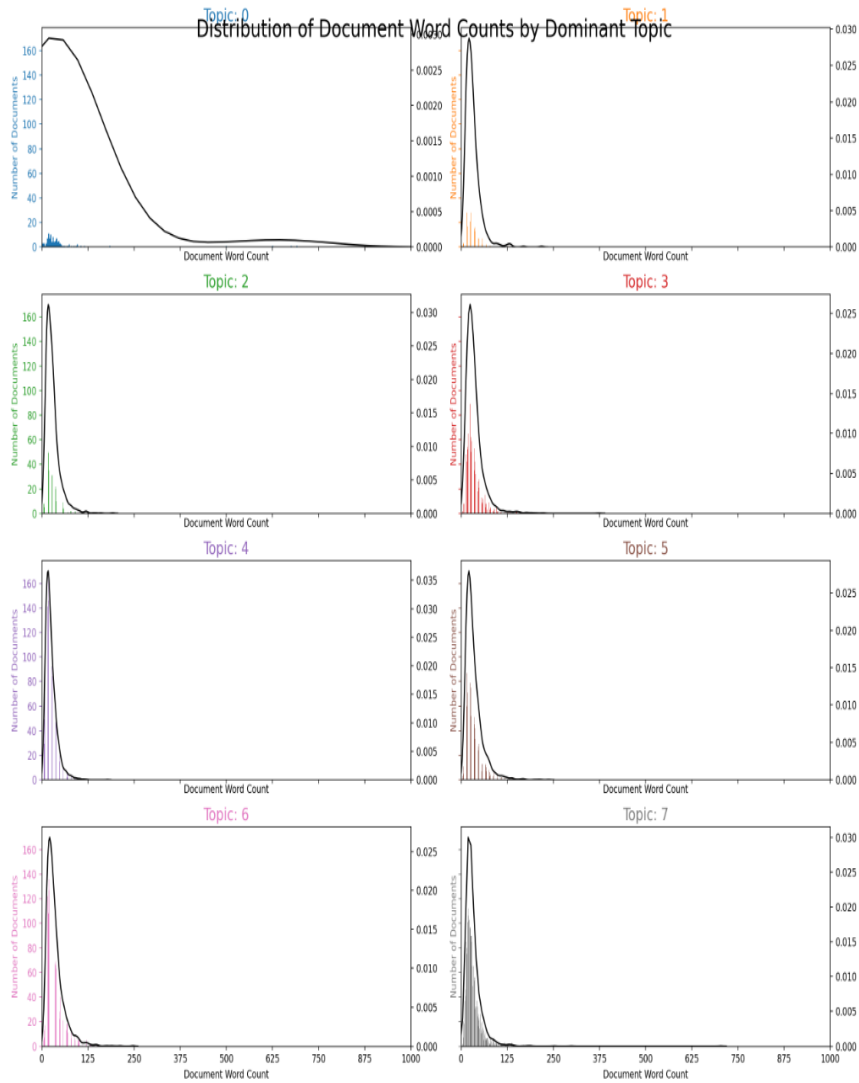
Belirtilen işlemlerin sonrasında toplam konu sayısı (num\_topics) 8 olarak belirlenerek konu modeli oluşturulmuştur. Konu modelleme işleminin gerçekleştirilmesinde Gensim kütüphanesinden yararlanılmıştır [21]. Oluşturulan modelin pyLDAvis gensim kütüphanesi ile “Konular Arası Mesafe Haritası” hazırlanmıştır ( num\_topics = 8). Şekil 2’de metin belgesindeki sözcüklerin genel dağılımı sunulmaktadır. Şekil 3’te ise toplam konu sayısı 8 alınarak konular arası mesafe haritasına ilişkin görsel sunulmaktadır.



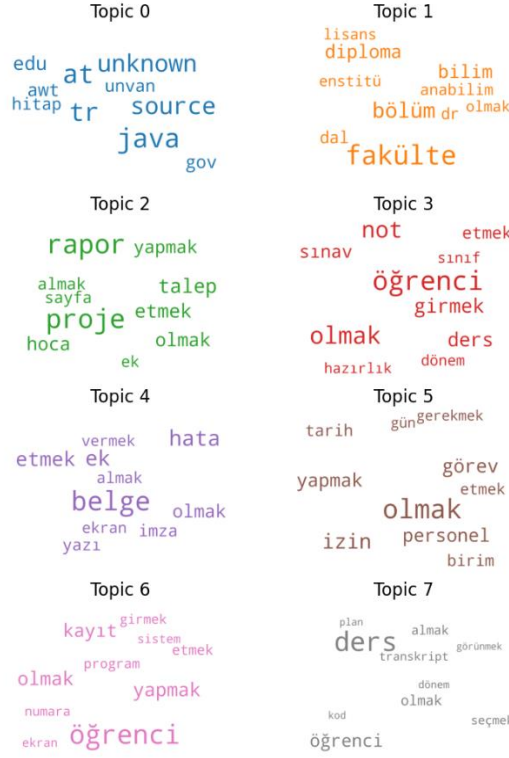
Şekil. 2. Metin belgesindeki kelime sayılarının genel dağılımı (Toplam konu sayısı 8 için)



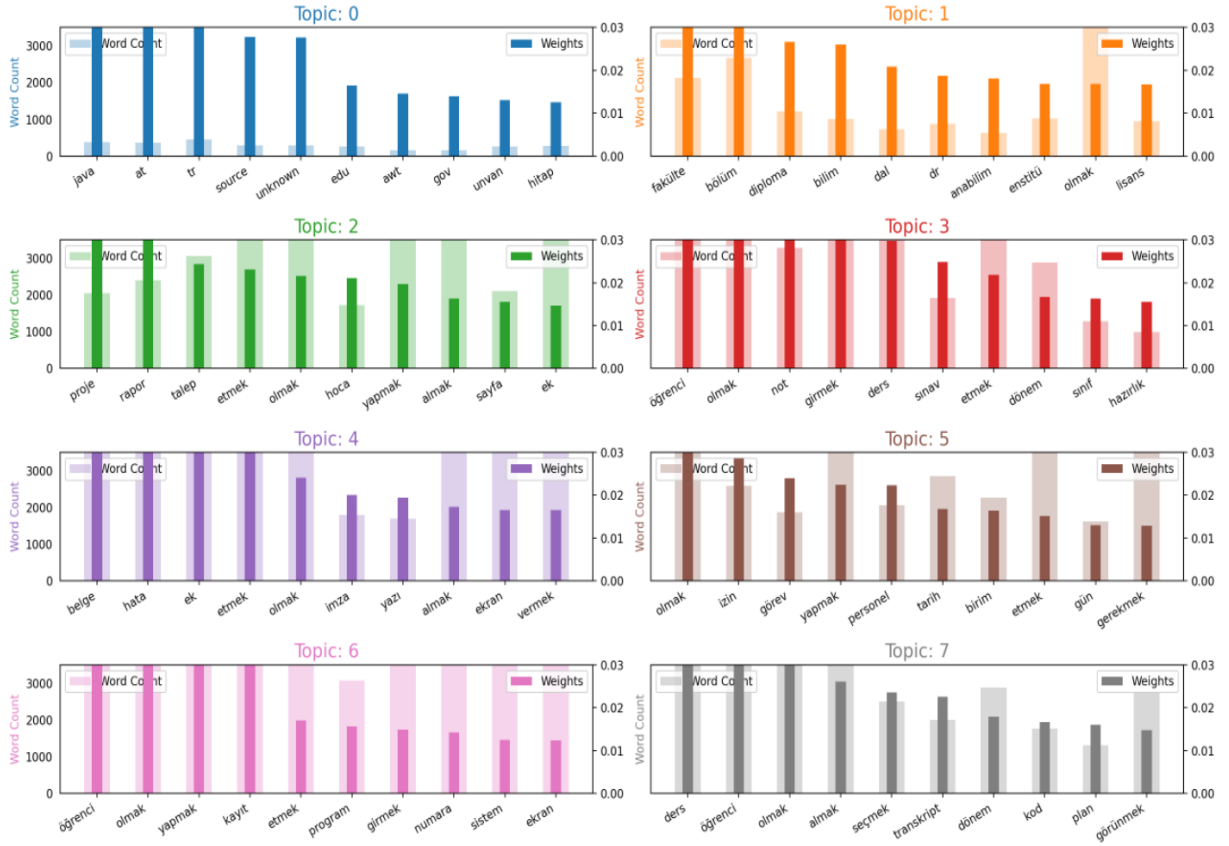
Şekil. 3. Toplam konu sayısı 8 için konular arası mesafe haritası



Şekil. 4. Metin belgesindeki kelime sayılarının konu tabanlı dağılımı (Toplam konu sayısı 8 için)



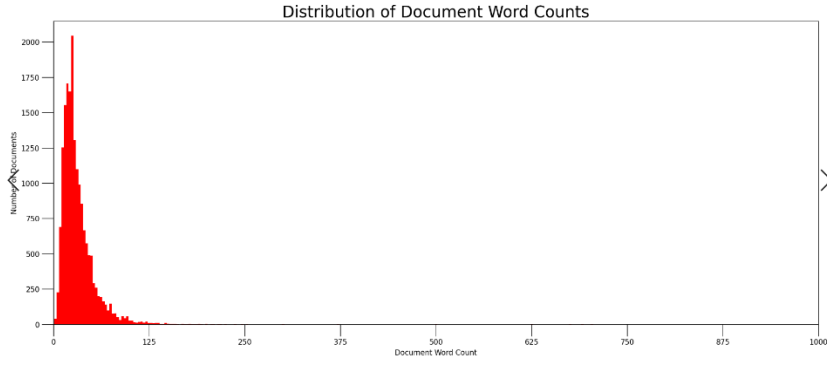
Şekil. 5. Konu tabanlı en sık kullanılan kelimeler (Toplam konu sayısı 8 için)



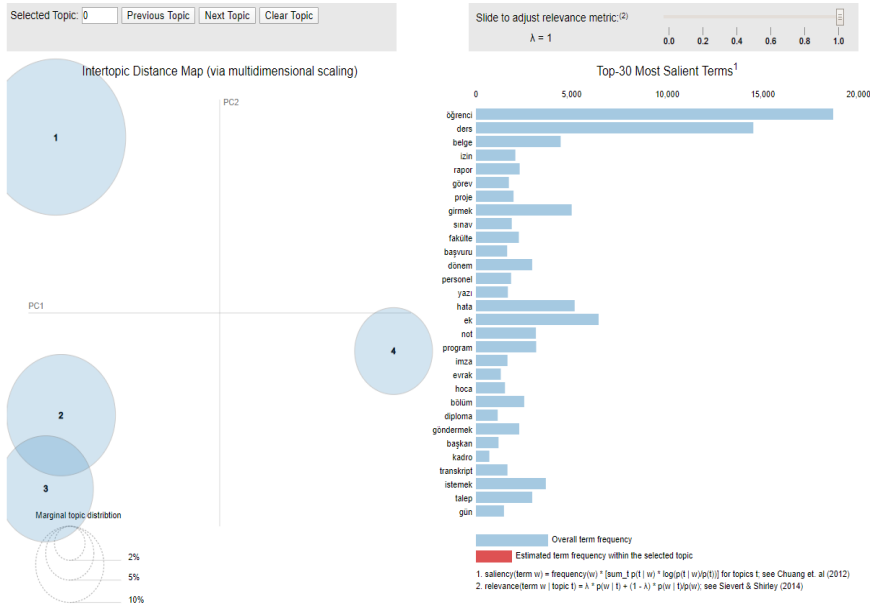
Şekil. 6. Kelime sayılarının konu tabanlı dağılımı (Toplam konu sayısı 8 için)

Şekil 4'te metin belgesindeki kelime sayılarının konu tabanlı dağılımı, Şekil 5'te konu tabanlı en sık kullanılan kelimeler, Şekil 6'da ise kelime sayılarının konu tabanlı dağılımı sunulmaktadır. Konu modelleme tabanlı metin analizinde, farklı toplam konu sayı

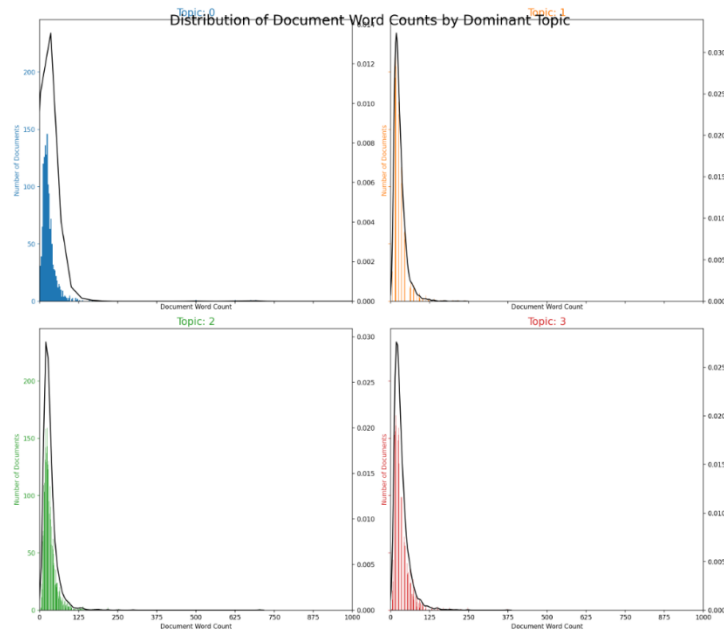
parametreleri değerlendirilerek, elde edilen konuların belirlenmesi ve görselleştirilmesinde değerlendirilmiştir. Şekil 7’de toplam konu sayısı 4 için, metin belgesindeki sözcüklerin genel dağılımı sunulmaktadır.



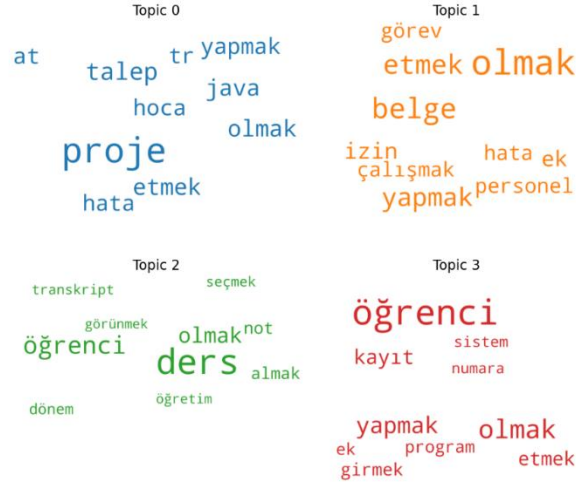
Şekil. 7. Metin belgesindeki kelime sayılarının genel dağılımı (Toplam konu sayısı 4 için)



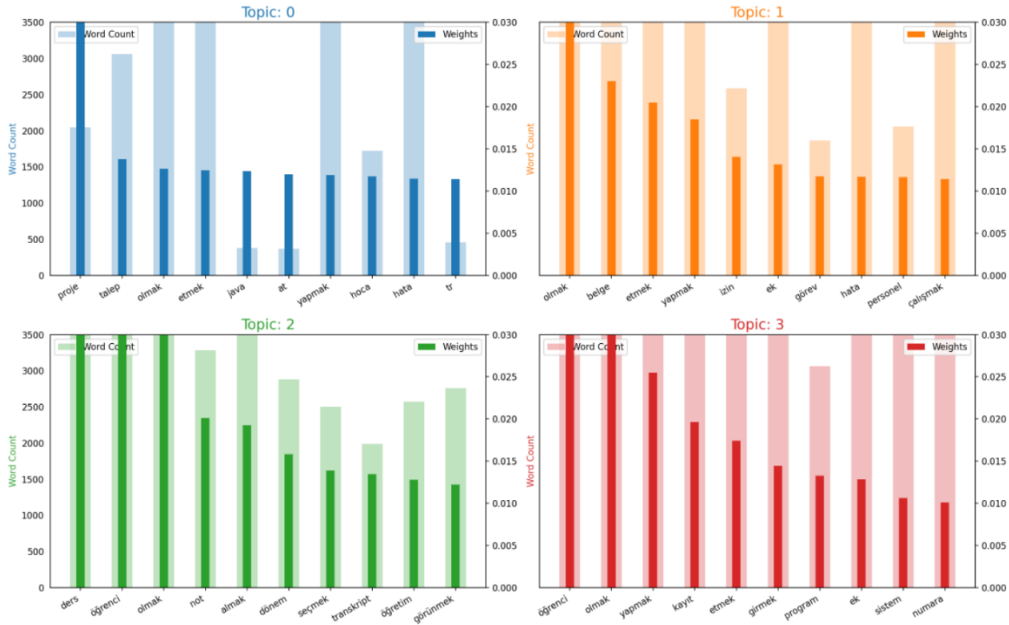
Şekil. 8. Toplam konu sayısı 4 için konular arası mesafe haritası



Şekil. 9. Metin belgesindeki kelime sayılarının konu tabanlı dağılımı (Toplam konu sayısı 4 için)



Şekil. 10. Konu tabanlı en sık kullanılan kelimeler (Toplam konu sayısı 4 için)



Şekil. 11. Kelime sayılarının konu tabanlı dağılımı (Toplam konu sayısı 4 için)

Karşılaştırma için sabit bir doküman ele alınmış ve gizli Dirichlet tahsisi tabanlı konu modelleme ile üretilmiş 2 modelin (4 ve 8 toplam konu parametresine göre oluşturulmuş modeller) çıktısına göre hangi konuya ne kadar ilişkili olduğunun detayı gösterilmektedir. Örnek Doküman: “fen fakülte moleküler biyoloji genetik bölüm ad öğrenci öğrenci dosya aramak tıklamak ek hata karşılaşmak saygı”

Örnek Dokümanın Konu Modelleri ile İlişkileri:

- Toplam Konu Sayısı = 4:
  - Konu => %0.595
  - Konu => %0.015
  - Konu => %0.372
  - Konu => %0.016
- Toplam Konu Sayısı = 8
  - Konu => %0.086
  - Konu => %0.477
  - Konu => %0.399



Örnek dokümanın farklı konu sayıları için gizli Dirichlet tahsisi tabanlı konu modelleme ile elde edilmiş konu modellerine dahil olma olasılıkları görülmektedir. Benzer durumlar her bir doküman için oluşmaktadır. Bir doküman için metin sınıflandırmadaki gibi belli bir kümeye ait olmaktan ziyade farklı kümelere ne kadar dâhil olduğu durumu görülmektedir.

## 5. Sonuç

Bu çalışmada, Üniversite Bilgi Yönetim Sistemi (ÜBYS) geliştirilirken farklı kurum ve bu kurumlardaki kişilerden gelen hata bildirim ve isteklerin gizli Dirichlet tahsisi kullanılarak konu modellemesi yapılmıştır. Yapısal olmayan veriler modelin uygulanabilmesi için öncelikle yapısal bir forma çevrilmiştir. Latent Dirichlet Allocation (LDA) kullanılarak 4 ve 8 konulu olmak üzere 2 adet konu modeli örneği oluşturulmuştur. Her bir konu modellemesi için, konuların birbiri ile olan ilişkileri ve konulardaki kelimelerin konularla olan ilişkileri görsel olarak gösterilmiştir. Konuların birbirleri ile ne oranda iç içe geçtikleri ve ayrıştıkları belirtilmiştir. Sonuç olarak, ÜBYS sistemi idari ve akademik süreçlerin tamamını barındıran ve birbiri ile tam entegre çalışan bir sistem olduğu göz önünde bulundurulduğunda, doküman olarak baz alınan taleplerin içerikleri de bahsi geçen durumla ilgili olarak konu modelleme analizine tam uyumlu bir veri seti sunmaktadır. Bu durumla ilişkili olarak birbiri içine geçmiş dokümanların, konu modeli sayısı arttığında bununla doğru orantılı olarak ilişkili oldukları konular da artmaktadır. Yine aynı şekilde konu modeli sayısı arttığında konuların birbirleri ile olan etkileşimlerinin de arttığı görülmüştür. İlerleyen çalışmalarda, gizli Dirichlet tahsisine dayalı elde edilen konular üniversite bilgi sistemi üzerinde elde edilen servis destek talepleri üzerinde öznitelik olarak kullanılarak, makine öğrenmesine dayalı bir otomatik sınıflandırma çerçevesi geliştirilmesi amaçlanmaktadır.

## Kaynakça

- Schwarz, C. (2018). Idagibbs: A command for topic modeling in Stata using latent Dirichlet allocation. *The Stata Journal*, 18(1), 101-117.
- Sun, M., & Zheng, H. (2018, September). Topic Detection for Post Bar Based on LDA Model. In *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 136-149). Springer, Singapore.
- Shah, A. H. (2019). How episodic frames gave way to thematic frames over time: A topic modeling study of the Indian media's reporting of rape post the 2012 Delhi gang-rape. *Poetics*, 72, 54-69.
- Karami, A., Ghasemi, M., Sen, S., Moraes, M. F., & Shah, V. (2019). Exploring diseases and syndromes in neurology case reports from 1955 to 2017 with text mining. *Computers in biology and medicine*, 109, 322-332.
- Onan, A., Bulut, H., & Korukoglu, S. (2017). An improved ant algorithm with LDA-based representation for text document clustering. *Journal of Information Science*, 43(2), 275-292.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modeling? and how to fix it using search-based software engineering. *Information and Software Technology*, 98, 74-88.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2002). Latent dirichlet allocation. In *Advances in neural information processing systems* (pp. 601-608).
- D'Urso, P., & Leski, J. M. (2019). Fuzzy clustering of fuzzy data based on robust loss functions and ordered weighted averaging. *Fuzzy Sets and Systems*, 389, 1-28.
- Ekinci, E., Omurca, S. İ., KIRIK, E., & TAŞÇI, Ş. Tıp Veri Kümesi için Gizli Dirichlet Ayrımı. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 22(64), 67-80.
- Onan, A., Korukoglu, S., & Bulut, H. (2016). LDA-based Topic Modelling in Text Sentiment Classification: An Empirical Analysis. *Int. J. Comput. Linguistics Appl.*, 7(1), 101-119.
- Onan, A. (2017). Türkçe Twitter Mesajlarında Gizli Dirichlet Tahsisine Dayalı Duygu Analizi. *Akademik Bilişim*, 8-10.
- Karami, A., Gangopadhyay, A., Zhou, B., & Karrazi, H. (2015, August). Flatm: A fuzzy logic approach topic model for medical documents. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)* (pp. 1-6). IEEE.
- Bagheri, A., Saraee, M., & De Jong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems*, 52, 201-213.
- Wang, T., Cai, Y., Leung, H. F., Lau, R. Y., Li, Q., & Min, H. (2014). Product aspect extraction supervised with online domain knowledge. *Knowledge-Based Systems*, 71, 86-100.
- Zheng, X., Lin, Z., Wang, X., Lin, K. J., & Song, M. (2014). Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowledge-Based Systems*, 61, 29-47.
- Jo, Y., & Oh, A. H. (2011, February). Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 815-824).
- Li, F., Huang, M., & Zhu, X. (2010, July). Sentiment analysis with global topics and local dependency. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.
- Xianghua, F., Guo, L., Yanyan, G., & Zhiqiang, W. (2013). Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems*, 37, 186-195.
- Onan, A., Atik, E., & Yalçın, A. (2020). Machine learning approach for automatic categorization of service support requests on university information management system. In *Proceedings of the Second International Conference on Intelligent and Fuzzy Systems* (pp. 1-7).
- Rehůrek, R., & Sojka, P. (2011). Gensim—statistical semantics in python. *Retrieved from genism.org*.