



Ülke Milli Marşları ve İstiklal Marşı'nın Özgünlüğünün Veri Analizi¹
Data Analysis of the Authenticity of the National Anthems and Turkish National Anthem

Günay KILIÇ¹İbrahim BUDAK^{2*}Selçuk Burak HAŞILOĞLU³¹ Pamukkale Üniversitesi, gkilic@pau.edu.tr, ORCID: 0000-0003-2236-7535² Kastamonu Üniversitesi, ibrahimbudak04@gmail.com, ORCID: 0000-0001-7762-6114³ Pamukkale Üniversitesi, selcukburak@hasiloglu.com, ORCID: 0000-0003-4512-6531

* Yazışılan Yazar/Corresponding author

Makale Geliş/Received: 26.09.2022

Makale Kabul/Accepted: 27.10.2022

Araştırma Makalesi / Research Paper

DOI: 10.47097/piar.1180263

Öz

Milli marşlar ülkelerin bağımsızlık göstergelerinden biridir ve ülkelerin bağımsızlık duygusundan bahseder. Bu çalışmada Türkiye Cumhuriyeti milli marşı İstiklal Marşı'nın veri analizi yapılmıştır. İstiklal Marşı analiz edilirken ortak dil bağının bulunduğu Türk Keneşi üye ülke marşları, ortak dini değerleri bulunduğu düşünülen İslam İş Birliği Teşkilatı üye ülke marşları ve Türkiye'nin bir Avrupa ülkesi olmasından ötürü Avrupa Birliği Üye ülkelerinin marşları analize dahil edilmiştir. Analiz veri madenciliği açısından kümeleme analizi ve birliktelik analizini tekniklerini içermektedir. Metin madenciliği açısından ise kelime boyutunda, başlık boyutunda ve duygu analizi boyutunda İstiklal Marşı ve diğer milli marşların benzerlikleri incelenmiştir. Analiz sonucunda kümeleme analizi ile veriler iki küme olarak hesaplanmış bir kümede karşılaştırılan tüm ülke marşları yer alırken İstiklal Marşı diğer kümede tek başına yer almıştır. Bu İstiklal Marşının özgünlüğünü ortaya koymada önemli bir göstergedir. Yine metin madenciliği başlığı altındaki 3 boyuttaki analizler İstiklal Marşı ile karşılaştırılan milli marşlar arasında bir benzerlik ortaya koyamamıştır. Birliktelik analizi sonuçları ise milli marşlardaki ortak kelimelerin ve kelimelerin birlikte kullanımını göstermesi açısından önemlidir.

Anahtar Kelimeler: Milli marşlar, Gizli Ayırma Tahsisi, Duygu Analizi, Kümeleme Analizi, Birliktelik Analizi.

Jel Kodları: C88, Y10, Z00.

Abstract

National anthems are one of the indicators of independence of countries and talk about the sense of independence of countries. In this study, the National Anthem of the Republic of Turkey, Turkish National Anthem, is analyzed by data mining techniques. While analyzing Turkish National Anthem, the anthems of the member states of the Turkic Council with which they have a common language tie, the anthems of the member states of the Organization of Islamic Cooperation, which are thought to have common religious values, and the anthems of the European Union Member States, since Turkey is a European country, were included in the analysis. The analysis includes cluster analysis and association analysis techniques in terms of data mining. In terms of text mining, the similarities of Turkish National Anthem and other national anthems were examined in terms of word size, title size and sentiment analysis. As a result of the analysis, all the country anthems that were compared in one cluster, whose data were calculated as two clusters by cluster analysis, were included, while Turkish National Anthem took place alone in the other cluster. This is an important indicator in revealing the originality of Turkish National Anthem. Again, the analyzes in 3 dimensions under the title of text mining did not reveal a similarity between Turkish National Anthem and the national anthems compared. The results of the association analysis are important in terms of showing the use of common words and words in national anthems together.

Keywords: National anthems, Latent Dirichlet Allocation, Sentiment Analysis, Cluster Analysis, Association Rule Analysis.

Jel Codes: C88, Y10, Z00.

¹ Bu çalışmanın dar kapsamı, birliktelik ve kümeleme analizi dâhil edilmemiş sürümü, 04.11.2021 tarihinde Pamukkale Üniversitesi Rektörlüğü tarafından düzenlenen İstiklal Marşı ve Mehmet Akif Ersoy Sempozyumunda "Metin Madenciliği ile İstiklal Marşı ve Diğer Ülke Marşlarının Benzerliklerinin Değerlendirilmesi" başlığında sunulmuştur. Kıymetli hocamız Prof. Dr. Feyzullah EROĞLU anısına yayımlanan bu özel sayı için İstiklal Marşı'nın özgünlüğünün ve kullanılan ifadelerin birlikteliği analizleri yapılmıştır. Vatan sevgisini, adil olmayı, haksızlıklara karşı direnmeyi, güç zehirlenmesi yaşayanların karşısında dik durmayı benimseyen ve öğrencilerine aşılayan değerli hocamıza müteşekkirimiz.

1. GİRİŞ

Milli marşlar ülkelerin en önemli bağımsızlık sembollerinden biridir. İçerdiği duygular bir milletin ortak değerleridir. Milli marşlarda, ait olduğu ülkeye ait bağımsızlık duygusundan ve var olma mücadelesinden bahsedilmektedir. Vatan ve millet sevgisinin işlendiği milli marşlar, aynı milletin (ırk) kurduğu ülkelerde veya aynı dini benimsemiş farklı ülkelerde benzerlik gösterebilmektedir.

Çalışma kapsamında 12 Mart 1921 Tarihinde kabul edilen İstiklal Marşı'nın Türkiye'nin dâhil olduğu veya aday olduğu ülke örgütlerinin marşları ile kıyaslanması amaçlanmıştır. Bu amaç doğrultusunda belirlenen örgütler Avrupa Birliği, Tük Keneşi ve İslam İş Birliği Teşkilatı'dır. Çalışmada belirlenen ülke topluluklarındaki her bir ülkenin milli marşının ortak dilde değerlendirilmesi amacı ile İngilizceye çevrilerek analiz edilmiştir.

Uluslar, uluslararası siyasi toplantılarda ve spor müsabakalarında bazı sembolleri kimlik olarak kullanır. Çağdaş toplumlarda bu semboller milli marş, amblem ve bayraktır (Gilboa ve Bodner, 2009: 460). Bu sembollerden milli marşlar ulus devletlerin ortaya çıktığı ve milliyetçiliğin yaygınlaştığı on dokuzuncu yüzyılda yaygınlaşmıştır. Milli marşlar incelendiğinde marşların yapıldığı dönemdeki ülkelerin içinde bulunduğu şartlara göre farklılıklar olabilmektedir. Ülkeden ülkeye farklı konular bulunmakla birlikte, ülkelerin bayraklarından ve doğal güzelliklerinden bahsetmek Tanrı'yı yüceltmek ve ülkenin kurtuluşu için dua etmek genel kavramlardır (Aktaş, 2014: 73-74).

Devletler işbirliği ve dayanışmayı güçlendirmek amacı ile bazı örgütler ve teşkilatlar kurumaktadır. Bu çalışmada Avrupa Birliği (AB), İslam İş Birliği Teşkilatı (İİT) ve Türk Dili Konuşan Ülkeler İşbirliği Konseyi (Türk Keneşi) üye ülkeleri araştırmanın kapsamı olarak belirlenmiştir.

AB, ikinci dünya savaşı sonrasında yıkıma uğrayan Avrupa'da barışın yeniden sağlanması ve ekonomik alanda kuvvetli bir iş birliği kurulması amacı ile kurulmuştur. Bugün itibari ile üye ülkeler Almanya, Avusturya, Belçika, Bulgaristan, Çek Cumhuriyeti, Danimarka, Estonya, Finlandiya, Fransa, Güney Kıbrıs Rum Yönetimi, Hırvatistan, Hollanda, İrlanda, İspanya, İsveç, İtalya, Letonya, Litvanya, Lüksemburg, Macaristan, Malta, Polonya, Portekiz, Romanya, Slovakya, Slovenya ve Yunanistan'dır. AB'ye aday konumda aralarında Türkiye'nin bulunduğu 5 aday ülke vardır. Türkiye dışındaki aday ülkeler şu Kuzey Makedonya, Karadağ, Sırbistan ve Arnavutluk'tur (<https://www.ab.gov.tr/3.html>; Erişim Tarihi: 16.04.2021).

İİT, üye devletlerin arasında iş birliği ve dayanışmayı güçlendirmek ve İslam Dünyasının hak ve çıkarlarını korumak amacı ile 25 Eylül 1969 da kurulmuştur. Suriye'nin üyeliği askıya alınmıştır. Teşkilatın diğer 56 üyesi Afganistan, Arnavutluk, Azerbaycan, Bahreyn, Bangladeş, Benin, Birleşik Arap Emirlikleri (BAE), Brunei, Darüsselam, Burkina Faso, Cezayir, Cibuti, Çad, Endonezya, Fas, Fildişi Sahili, Filistin, Gabon, Gambiya, Gine, Gine Bissau, Guyana, Irak, İran, Kamerun, Katar, Kazakistan, Kırgızistan, Komorlar, Kuveyt, Libya, Lübnan, Maldivler, Malezya, Mali, Mısır, Moritanya, Mozambik, Nijer, Nijerya, Özbekistan, Pakistan, Senegal, Sierra Leone, Somali, Sudan, Surinam, Suudi Arabistan, Tacikistan, Togo, Tunus, Türkiye, Türkmenistan, Uganda, Umman, Ürdün, Yemen'dir (<https://www.mfa.gov.tr/islam-isbirligi-teskilati.tr.mfa>; Erişim Tarihi: 16.04.2021).

Türk Keneşi, üye ülkelerin, haklarının arasındaki tarihi bağları, ortak dil, kültür ve gelenekleri temel alarak, kapsamlı iş birliğinin daha da derinleştirilmesi amacı ile 13.10.2009 tarihinde Nahcivan antlaşması ile kurulmuştur. Üye ülkeler Azerbaycan, Kazakistan, Kırgızistan, Özbekistan ve Türkiye'dir (https://www.turkkon.org/assets/pdf/temel_belgeler/Nahcivan_Anlasmasi_Turkce_20140417_193951.pdf Erişim Tarihi: 16.04.2021).

Araştırma kapsamındaki ülkelerin marşları farklı dillerde olduğundan, değerlendirme yapabilmek adına öncelikle marşların ortak dilde (İngilizce) karşılıkları alınmıştır. Milli marşlarla ilgili yapılan çalışmaların öncülleri Tablo 1'de verilmiştir.

Tablo 1. Milli Marşlar ile İlgili Çalışmalar

Yazar	Konu
Csepeli ve Örkény (1998: 37-56).	Avrupa'daki milli marşların görüntüleri
Souza (2008: 1-188).	İngilizce yazılmış milli marşların sistematik bir işlevsel incelemesi
Essam (2015: 1-11).	Mısır'ın ulusal marşının ve onu çevirisi üzerine bir vaka çalışması
Lauenstein vd. (2015: 309-329).	Milli marşlarda milliyetçilik, cinsiyet ve hayali bir ailenin temsili üzerine bir araştırma
Abril (2016: 77-94).	Bir milli marş: Vatansızlık sembolü mü yoksa demokratik eylem mi?
Dlodlo (2016: 1-181).	Zimbabve'de milli marş aracılığıyla ulus inşasına yönelik bir araştırma
Van Ginderachter (2020: 105-124).	Belçika bayrağı ve ulusal marşının karşılaştırılması
Arslan vd. (2021: 21-29).	Pakistan milli marşının üslup ve metinsel analizi
Akarsu (2022: 1-35).	İstiklal marşı ve icrasına yönelik bir analiz çalışması

Tablo 1'de ulusal marşlara yönelik literatürdeki örnek çalışmalar verilmiştir. Çalışmada İstiklal Marşı'na diğer milli marşların benzerlikleri farklı analizlerle değerlendirilmiştir. Bunlar kelime frekansı, marşlardaki başlıklar (GDA), duygu analizi, birliktelik analizi ve kümeleme analizidir. GDA ile milli marşlardaki konular tanımlanmış ve en sık kullanılan 10 kelime verilmiştir. Duygu analizi ile marşlar olumlu, olumsuz ve nötr olarak sınıflandırılarak milli marşımıza olan en fazla benzer olan 10 ülke gösterilmiştir. Birliktelik analizi ile marşlarda ortaya çıkan kelimelerden en güçlü kelimenin hangisi olduğunu ve hangi kelimelerle güçlü bir ilişki kurulduğu vurgulanmıştır. Son olarak kümeleme analizi ile ülke marşları 6 kümede sınıflandırılmıştır.

2. GİZLİ AYIRMA TAHSİSİ (LATENT DİRİCHLET ALLOCATION) (GAT/LDA)

Milli marşların başlık boyutunda analizi için Gizli Ayırma Tahsisi (GAT) kullanılmıştır. GAT'ın genel fikri, bir belge yazan bir kişinin aklında belirli konular olduğu hipotezine dayanmaktadır. İlk olarak Blei ve arkadaşları (2003) tarafından konu tanımlama, varlık çözümleme ve kelime kökü sınıflandırılması olarak ortaya atılmıştır. GAT, bir konu hakkında yazmak, o konudaki kelime havuzundan belirli bir olasılıkla bir kelime seçmek anlamına gelmektedir. Bütün bir belge daha sonra farklı konuların bir karışımı olarak temsil edilebilir. Bir belgenin yazarı tek bir kişi olduğunda, bu konular kişinin bir belgeye bakış açısını ve özel kelime dağarcığını yansıtmaktadır. Birden çok kullanıcının kaynaklara açıklama eklediği durumlarda, ortaya çıkan konular belgenin ortaklaşa paylaşılan bir görünümünü meydana

getirmektedir. GAT, konuların etiketleri belgeyi tanımlamak için ortak bir sözlük oluşturmaktadır. Daha genel olarak, GAT, bu verilerin özelliklerini gözlemlenmeyen kümeler halinde gruplayarak verilerin benzerliğini açıklamaya yardımcı olmaktadır. Bu kümelerin bir karışımı daha sonra gözlemlenebilir verileri oluşturmaktadır (Krestel vd. 2009: 62). GAT, birlikte ortaya çıkan kelimeleri konulara göre kümeleyen bir dil modelidir. Son yıllarda, GAT, bilgisayarla görme problemlerini çözmek için yaygın olarak kullanılmaktadır. Örneğin, GAT bir görüntü koleksiyonundan nesnelere keşfetmek ve görüntüleri farklı sahne kategorilerine ve bireylerin eylemlerini sınıflandırmak için kullanılmıştır. Ayrıca görsel gözetimde, GAT, kalabalık ve yoğun bir sahnede meydana gelen atomik aktiviteler ve etkileşimler modellenmiştir. Son olarak, düşük seviyeli görsel kelimeleri (görüntü yamaları, uzamsal ve zamansal ilgi noktaları veya hareketli pikseller) eş anlamlarını kullanarak semantik anlamları olan (nesnelere, nesnelere bölümlerine, insan eylemleri) konulara göre kümelendirilmiştir (Wang ve Grimson, 2007: 1577).

3. DUYGU ANALİZİ

Duygu analizi, insanların görüşlerini yazılı bir metinden analiz etmekle ilgilidir. Doğal dil işlemede en popüler araştırma alanlarından biridir. Aynı zamanda metin madenciliği, veri madenciliği ve web madenciliği alanlarında yoğun olarak karşımıza çıkmaktadır. Duygu analizine işletme ve toplum için önemi nedeniyle bilgisayar bilimleri dışında da yönetim bilimleri ve sosyal bilimlerde sıklıkla karşılaşılmaktadır. Duygu analizine Twitter, sosyal ağlar, bloglar, forumlar, yorumlar vb. sosyal medyanın farklı alanlarında rastlanmaktadır (Liu, 2012: 1).

Duygu analizi, olumlu veya olumsuz görüşlerin ortaya çıkarılmasına yardımcı olur. Çoklu bakış açısı sorularına cevap verme ve özetleme, fikir odaklı bilgi çıkarma ve metin analizi gibi görevlerle cümle düzeyinde analiz yapar. Örneğin, insanların görüşlerini içeren soru sorma ve cevabı merak etme sistemine ihtiyaç duyulursa, duygu analizinin olumlu ve olumsuz duyguların ifadelerini tam olarak tanımlayabilmesi gerekir (Wilson vd., 2005: 347).

Verilen metnin polaritesi önce cümle düzeyinde hesaplanmaktadır. Cümledeki kelimeler cümledeki sıraya göre kodlanmıştır. İlişkilendirme sözlüğü yardımıyla, sözlükte tanımlanmayan pozitif terim için +1, negatif için -1 ve nötr kelimeler için 0 olarak tanımlanır. Sözcük değerleri pozitif, negatif ve nötr olarak ağırlıklandırılır. Bu şekilde kelimelerin kutupsallığı vurgulanarak daha iyi hesaplanması sağlanmaktadır. Amaç, farklı kelimeler kullanarak polariteyi daha doğru hesaplamaktır (Balbi vd., 2018: 674).

4. KÜMELEME ANALİZİ

Bilgi işlemede veri, hareket veya işleme için verimli bir forma çevrilmiş bilgidir. Günümüz bilgisayarlarına ve iletişim ortamlarına göre veri, ikili dijital forma dönüştürülen bilgidir. Verilerin tekil özne veya çoğul özne olarak kullanılması kabul edilebilir. Ham veriler, verileri en temel dijital biçiminde tanımlamak için kullanılan bir terimdir. İnsanlar her tür ölçüm ve gözlemden gelen farklı veri yapıları ile ilgilenmektedir. Veriler farklı tekniklerle analiz edilerek karar alma noktasında temel oluşturmaktadır. Veriyi tanımlamada en önemli işlevlerden biri, verileri kategoriye, gruplara ya da kümelere ayırmaktır (Xu ve Wunsch, 2008: 1; Haşiloğlu ve Budak, 2019: 65). Verileri ayırmak için k-ortalamlar kümeleme algoritması sık kullanılmaktadır.

Veri madenciliğinde x-ortalamlar kümeleme algoritması, k-ortalamlar kümeleme algoritmasının geliştirilmiş şeklidir (Pelleg ve Moore, 2000: 727). k-ortalamlar algoritması, ham veri seti verildiğinde, orijinal veri kümesinin verilerin "k" alt kümelerine bölünmesini sağlamaya çalışmaktadır. Bahsedilen "k", benzersiz alt kümelerin sayısını ya da uygun veri madenciliği terminolojisinde grupları veya kümeleri temsil etmektedir. Her ne kadar k-ortalamlar algoritması için birçok geliştirmeler yapılmış olsa da temel matematiksel formülasyon Eşitlik (1)'deki gibi kullanılmalıdır (Tucker vd., 2010: 600):

$$f = \sum_{t=1}^L \sum_{x_i \in S_t} \|x_i - y_t\|^2 \quad (1)$$

Burada;

S_t = Bir veri noktaları kümesidir.

y_t = Bir S_t kümesinin merkezidir.

x_i = Bir küme içinde mevcut bir veri noktasıdır.

L = Toplam küme sayısı (kullanıcı tarafından önceden belirtilen)'dir.

Pelleg ve Moore (2010) Bayesian Bilgi Ölçütü (BIC) gibi bir kriteri optimum hale getirerek, kümeleme sayısını otomatik olarak hesaplayabilen X-ortalamlar algoritmasını önermişlerdir. X-ortalamlar kümeleme algoritması, k-ortalamlar algoritmasının üç temel alanını geliştirmeyi hedeflemektedir. Bu üç kademe aşağıdaki gibi ifade edilebilir:

- Yöntem, kullanıcı tarafından belirlenen küme sayısını ortadan kaldırır.
- Yöntem, hesaplanabilir bir ölçek geliştirir.
- Yöntem, küme merkezlerini güncellemek için arama kriterlerini geliştirir.

Matematiksel olarak BIC formülasyonu Eşitlik (2)'deki gibi gösterilebilir (Pelleg ve Moore, 2000: 730; Tucker vd., 2010: 601):

$$BIC(M_j) = t_j(K) + \frac{R_j}{2} \log S \quad (2)$$

Burada;

$t_j(K)$ = Maksimum olabilirlik noktasında alınan verilerin logaritmik olma olasılığıdır.

K = Verilen veri kümesini temsil eder.

$R_j = M_j$ cinsinden parametre sayısını gösterir.

S = Kümeye girecek aday merkezlerin toplam veri noktasının sayısıdır.

5. BİRLİKTELİK ANALİZİ

Birliktelik analizi, veri madenciliğinin en önemli işlevlerinden biridir ve araştırmacılar tarafından üzerinde çalışılan en popüler tekniktir. Birliktelik analizi, işlem veri tabanında sık öge kümelerini keşfeden ilişkilendirme veya birliktelik madenciliğinde önemli bir aşamadır (AL-Zawaidah vd. 2011: 312). Birliktelik kuralları, gruplar, sınıflandırma, kümeleme ve korelasyon gibi veri kümelerinden ilginç desenler bulmaya çalışan birçok veri madenciliği görevinin özünü oluşturmaktadır (Han ve Kamber, 2006: 3). Sık görülen öge kümelerini bulmak için birçok algoritma önerilmiştir. Bunların tümü iki sınıfa ayrılabilir: aday oluşturma veya örüntü büyümesi (Agrawal vd., 1993: 2007).

Apriori, aday oluşturma yaklaşımının bir temsilcisidir. Uzunluk (k) sık öge kümelerine dayalı olarak uzunluk (k+1) aday öge kümeleri üretmektedir. Öge kümelerinin sıklığı, işlemlerdeki oluşumlar sayılarak tanımlanmaktadır. Apriori algoritmasının yürütülmesi kolaydır ve çok basittir, veri tabanındaki tüm sık kullanılan öge kümelerini araştırmak için kullanılmaktadır. Algoritma, k bir öge kümelerini oluşturmak için k öge setinin kümesinin kullanıldığı sık öge değerlerini bulmak için veri tabanında arama yapmaktadır. Her k-öge setinin frekansı için minimum destek eşliğinden büyük veya buna eşit olması gerekmektedir. Aksi takdirde, aday öge kümeleri olarak adlandırılır. İlkinde, algoritma, veri tabanındaki her bir ögeyi sayarak yalnızca bir öge içeren bir öge setlerinin sıklığını bulmak için veri tabanını taramaktadır. Bir öge setinin sıklığı, iki öge setindeki öge setlerini bulmak için kullanılır ve bu da 3 öge setini bulmak için kullanılmaktadır. Daha fazla k öge seti kalmayana kadar bu şekilde devam etmektedir. Bir öge kümesi sık değilse, ondan gelen herhangi bir büyük alt küme de sık değildir; bu koşul, veri tabanındaki arama alanından budama yapmaktadır (Wu vd., 2008: 12).

FP-büyümesi (FP-growth), 2000 yılında Han tarafından önerilmiştir ve örüntü büyüme yaklaşımını temsil etmektedir. FP büyümesi, Spesifik veri yapısını (FP-tree) kullanmaktadır. FP-büyümesi, sık görülen öge kümelerini, bir öge veri kümelerindeki tüm sıklıkları koşul örüntüsü bazında, koşul örüntüsünde bularak keşfetmektedir. FP-tree ile ilişkilendirilen düğüm yapısının bağlantısına dayalı olarak verimli bir şekilde oluşturulmaktadır. FP büyümesi, aday öge kümelerini açıkça oluşturmamaktadır (Al-Maolegi ve Arkok, 2014: 22).

Birliktelik analizi veri seti arasındaki ilişkileri analiz etmede destek ve güven değerlerini kullanır.

Birliktelik analizinde kullanılan destek (Support), Güven(confidence) ve kaldıraç(lift) değerleri sırası ile Eşitlik 3,4 ve 5 yardımı ile hesaplanır (Sabah ve Bayraktar, 2020:73).

Destek: Bir kuralın veri setinde tekrarını gösterir.

$$Destek (A \rightarrow B) = \frac{Sayı(A, B)}{N} \quad (3)$$

Güven: Bir A değişkenin B değişkeni ile birlikte bulunma olasılığıdır.

$$Güven (A \rightarrow B) = \frac{Sayı(A, B)}{Sayı(A)} \quad (4)$$

Kaldıraç: Bu oran verilen A ve B değerlerinin bağımsız olup olmadıklarını gösteren bir değerdir. Oran 1'den büyük ise değişkenler birbirine bağlıdır.

$$Kaldıraç(A \rightarrow B) = \frac{Destek(A, B)}{Destek(A) * Destek(B)} \quad (5)$$

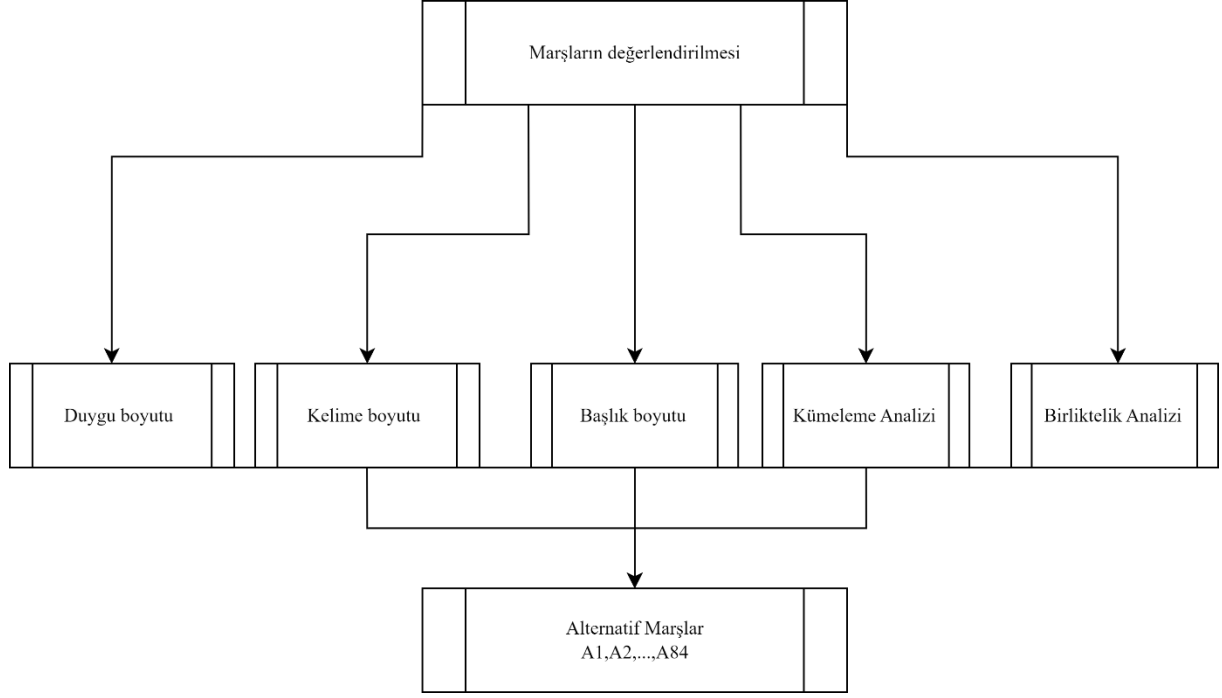
6. MARŞLARIN DEĞERLENDİRİLMESİ

Bütün çalışmada benzerlik için referans alınan milli marş Türkiye milli marşı *İstiklal Marşı'* dir. Belirtilen bütün boyutlarda referans marşa olan uzaklık, x İstiklal Marşı'ndaki değer, y hedef ülke marşındaki değer olmak üzere, farkların kareleri toplamı Eşitlik 6 ile hesaplanmıştır.

$$Benzerlik = \sum(x - y)^2 \quad (6)$$

Hesaplanan değer küçük olması farkların hesaplandığı milli marşın, hesap yapılan boyutta İstiklal Marşı'na daha benzer olduğunun göstergesidir. Çalışmada belirlenen ülkelere ait 85

milli marş Metin Madenciliği açısından 3 boyutta değerlendirilmiştir. Veri madenciliği tekniklerinden kümeleme analizi ve birliktelik analizi kullanılarak toplam 5 boyutlu bir çalışma yapılmıştır. Bu boyutların gösterimi Şekil 1'de verilmiştir.

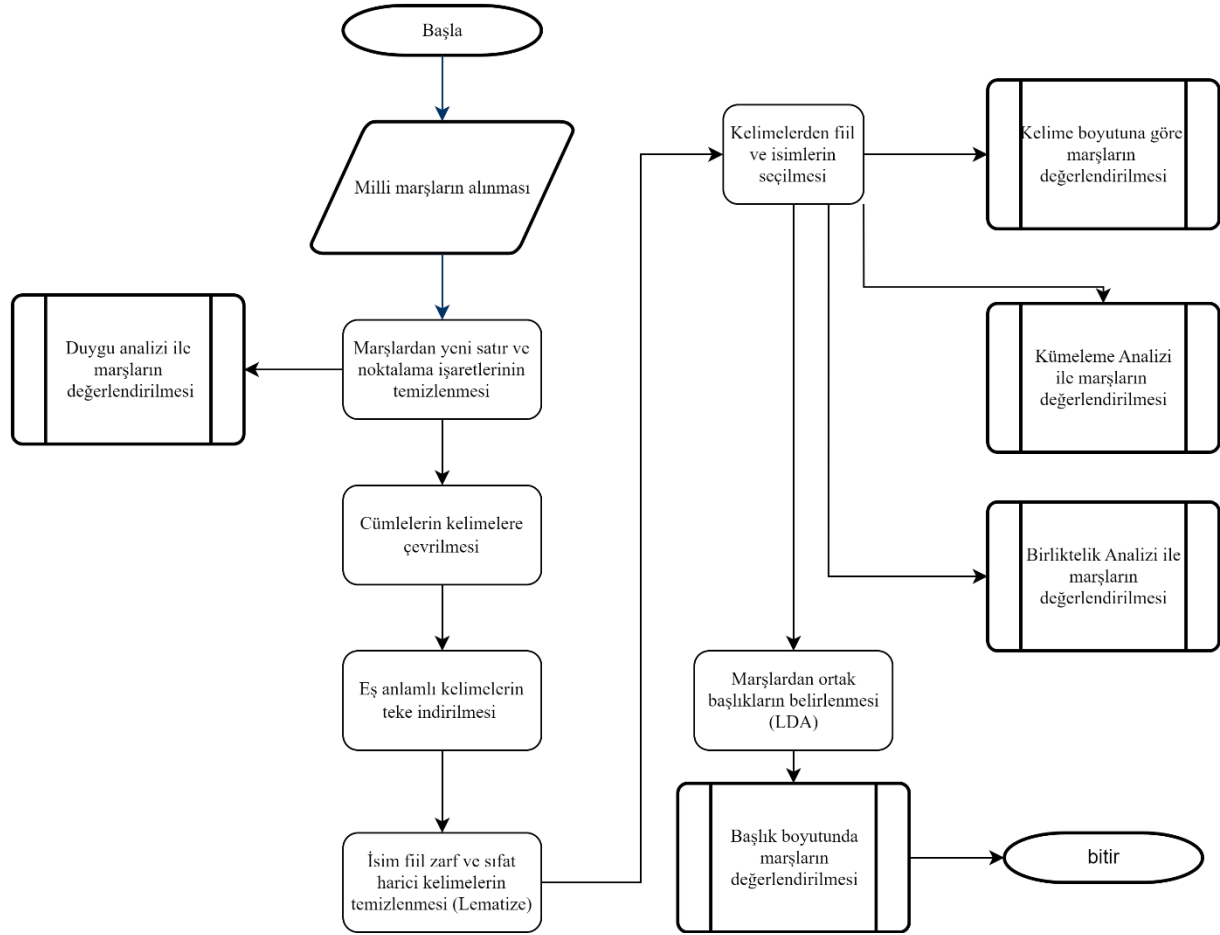


Şekil 1. Milli Marşların Değerlendirilmesi

İstiklal Marşı'nın diğer ülke marşları ile benzerliğinin ortaya konması adına önce marşların bir dizi işlemlerden geçirilmesi ve bu işlemlerin ardından belirlenen boyutlara göre benzerliklerinin ortaya konması amaçlanmıştır. Uygulamaya ait akış şeması Şekil 2'de verilmiştir.

Şekil 2'deki akış şemasında belirtilen işlem adımları Python programlama dilinde kodlanmıştır. Kodlama için PyCharm geliştirme ortamı ve matris işlemleri için numpy ve panda, doğal dil işleme için nltk, spacy ve gensim makine öğrenmesi için sklearn kütüphanesi ve altındaki modüller kullanılmıştır. Duygu analizi ise nltk kütüphanesindeki SentimentIntensityAnalyzer modülü ile gerçekleştirilmiştir.

Geliştirilen algoritma ile her bir boyut altında ilgili modül yardımı ile işlemler gerçekleştirilmiş ve bulgular alt başlıklarda verilmiştir.



Şekil 2. Uygulama Akış Şeması

6.1. Kelime Boyutuna Göre Marşların Benzerlikleri

Milli marşların eş anlamlı kelimeleri benzersiz tek kelime haline dönüştürüldükten sonra milli marşların 950 farklı kelime ile oluşturulduğu tespit edilmiştir. 85 ülke marşı toplam 4170 kelimedenden oluşmaktadır. Kelimeleri işlenmiş marşlardan en az kelime sayısına sahip marş 9 kelime ile Brunei milli marşı iken en fazla kelime sayısına sahip marş ise 206 kelime ile Hollanda milli marşıdır. Milli marşlardaki kelime sayılarındaki farklılıklar her bir kelimenin milli marşlardaki ağırlığını değiştirmektedir. Örnek verilir ise "unity" (birlik) kelimesi Avusturya ve Türkiye milli marşlarında 1'er kez geçmektedir. Unity kelimesi 46 kelimelik Avusturya marşının % 2,173913 'ünü oluştururken 174 kelimelik İstiklal Marşının % 0,5747126'sını oluşturmaktadır. Kelime Avusturya milli marşı açısından İstiklal Marşına oranla 3,782608696 kat daha fazla ağırlığa sahiptir denilebilir. Bütün kelimelerin ağırlıklarına göre İstiklal Marşına uzaklıkları belirlenmiş en yakın 10 marşa sahip ülke Hollanda, Macaristan, Fransa, Kamerun, Estonya, Libya, Tunus, Yemen, Sırbistan ve Romanya olarak sıralanmıştır.

6.2. Başlık Boyutuna Göre Marşların Benzerlikleri

Gizli Dirichlet Tahsisi metin gibi ayrık verilerin her bir ögesinin belli olasılık ile belli konulara atandığı üç seviyeli bir Bayes modelidir (Blei, vd. 2003: 993). Bu konu modellenmesi sayesinde doğal dil işleme alanında metinler etkin bir şekilde konulara ayrıştırılır.

Çalışmada milli marşlar 4 konuya ayrılmış ve bu konularda her bir marşın konuya ait katsayılar belirlenmiştir. Bir marşın 4 konuya ait katsayıları toplamı 1 etmektedir. Marşların benzerlikleri bu katsayılar üzerinden hesaplanmıştır. Her bir konuya ait 10 anahtar kelime Tablo 2'de verilmiştir.

Tablo 2. Başlıklara Ait Anahtar Kelimeler

	1	2	3	4	5	6	7	8	9	10
Umut	let	Arise	happiness	follow	flag	earth	freedom	want	unity	people
Yaşam	shall	Want	live	let	glory	follow	Life	die	blood	day
Zafer	glory	Time	look	man	blessing	root	Field	die	earth	maintain
Dünya	earth	people	blood	freedom	look	flag	struggle	piece	live	enjoy

Tablo 2'de Başlık 1, umut; Başlık 2 yaşam; Başlık 3 zafer ; Başlık 4 dünya temaları ile ilişkilendirilebilir. Kelime boyutuna benzer şekilde marşların her bir başlığa dâhil olma katsayıları ile İstiklal Marşı'nın katsayıları ile arasındaki yine Eşitlik 1 ile farkların kareleri toplamı hesaplanmıştır. Hesaplamalar ile benzerlikler ortaya konulmuştur. Başlık boyutunda İstiklal Marşı'na en benzer 10 marşın benzerlik oranları ve katsayıları Tablo 3'te verilmiştir.

Tablo 3. Başlıklara Göre Marşların Katsayıları

Başlık 1	Başlık 2	Başlık 3	Başlık 4	Benzerlik	Ülke
0,0057	0,1421	0,0056	0,8466	0	Türkiye (Referans Değer)
0,0511	0,0525	0,0532	0,8432	0,01236664	Guyana
0,0521	0,0511	0,0513	0,8455	0,01252366	Çad
0,0432	0,0438	0,0448	0,8681	0,01306803	Özbekistan
0,0439	0,0429	0,0442	0,869	0,0132916	Fas
0,0424	0,0425	0,0425	0,8726	0,01330466	Hırvatistan
0,0635	0,0663	0,0634	0,8067	0,01401933	Brunei
0,0643	0,065	0,0658	0,8049	0,0147413	Bahreyn
0,0365	0,0363	0,0362	0,8911	0,01505889	Kuveyt
0,033	0,0356	0,0319	0,8995	0,01557764	Moritanya
0,0343	0,0346	0,033	0,8981	0,01577722	Mozambik

Tablo 3 incelendiğinde İstiklal Marşı'nın baskın konusu Konu 4'ün benzerlikleri ortaya koymada etkin rol oynadığı görülmektedir. Kelimelerin başlık içinde ağırlıkları farklı olduğundan başlık ve kelime boyutuna göre sıralamada farklılıklar olduğu söylenebilir.

6.3. Duygu Analizi ile Marşların Benzerliklerinin Ortaya Konulması

Daha önceki iki kelime boyut frekansları ve kelimelerin başlıkları oluşturması üzerine yoğunlaşmaktaydı. Bu boyutta ise milli marşların bir bütün olarak içerdiği duygulara odaklanılmıştır. Doğal dil işleme ile marşlar, olumsuz tarafsız ve olumlu olmak üzere 3 farklı duygu ile tanımlanmıştır. Bir marşta bu duygular belirli katsayılarla sahiptir. İstiklal marşına bu üç farklı duyguya ait katsayılar ve bu katsayıların diğer marşlardaki duygu katsayılarına göre uzaklıkları yine Eşitlik 1 yardımı ile hesaplanmış ve Duygu analizi ile benzerlikler ortaya

konulmuştur. Duygu analizi ile elde edilen değerler ve İstiklal Marşı'na benzerlikleri Tablo 4'te verilmiştir.

Tablo 4. Milli Marşların Duygu Değerleri

Olumsuz	Nötr	Olumlu	Benzerlik	Ülke
0,100	0,683	0,217	0	Türkiye(Referans Değer)
0,092	0,69	0,218	0,000114	Portekiz
0,072	0,697	0,231	0,001176	Letonya
0,072	0,700	0,228	0,001194	Mozambik
0,077	0,717	0,206	0,001806	Katar
0,080	0,665	0,255	0,002168	Sırbistan
0,091	0,655	0,254	0,002234	Finlandiya
0,059	0,710	0,230	0,002579	Bangladeş
0,122	0,641	0,238	0,002689	Hollanda
0,061	0,723	0,217	0,003121	Gine-Bissau
0,068	0,670	0,261	0,003129	Kuzey Makedonya

6.4. Marşların Kelime, Duygu ve Başlık Boyutu Açısından Bir Arada Değerlendirilmesi

Bölüm 6.1, 6.2 ve 6.3 te bulunan marşlar arasındaki benzerlik değerleri eşit ağırlıklar verilerek bir arada değerlendirildiğinde oluşan benzerlik Tablo 5'te verilmiştir.

Tablo 5. Bütün Boyutlar ile Genel Benzerlik

Benzerlik	Ülke
0	Türkiye (Referans Ülke)
0,047377	Mozambik
0,059204	Libya
0,061565	Danimarka
0,063238	Gine-Bissau
0,063657	Guyana
0,064351	Filistin
0,067951	Burkina-Faso
0,069840	Hırvatistan
0,070030	Macaristan
0,071424	Sierra Leone

6.5 Milli Marşların Kümeleme Analizi

Milli marşlar içerisinde geçen kelimelere (eş anlamlı sözcükler dahil) göre ülkelerin benzerlikleri ve buldukları küme analizi için k-ortalamlar ve x-ortalamlar algoritmaları kullanılmıştır. k-ortalamlar kümeleme analizi Bölüm 4'te açıklandığı üzere veri madenciliğinde veri kümesinin kümelerine ayrılmasında etkin bir biçimde kullanılmaktadır. k-ortalamlar kümeleme analizi ayrılacak küme sayısına ihtiyaç duyar. Diğer bir kullanılan x-ortalamlar kümeleme analizi de temelde k-ortalamlar kümeleme analizini kullanır. X-ortalamlar kümeleme analizi k-ortalamlar kümeleme analizinin aksine küme sayısına ihtiyaç duymaz. Her iki kümeleme analizi için RapidMiner 9.1 (akademik sürüm) yazılımından yararlanılmıştır. RapidMiner 2001 yılında Dortmund Üniversitesinde geliştirmeye başlanılan geliştirilmesi devam eden veri madenciliği paket programıdır. İçinde

birçok makine öğrenme algoritmasını bulundurmaktadır ve akademik çalışmalarda kullanımına ücretsiz destek vermektedir.

RapidMiner Programında X-Means operatörü ile Öklid uzaklığına göre BregmannDivergences ölçüsü kullanılarak yapılan analiz sonucunda 85 ülke iki kümeye ayrılmıştır. Kümelerin birinde 84, diğerinde ise bir ülke yer almaktadır. Üye sayısı tek olan kümedeki ülke Türkiye'dir. Yani İstiklal Marşı, diğer 84 ülkenin mili marşından farklı olduğu tespit edilmiştir. Küme sayısı altıya zorlandığında (K-Means operatörü) küme üye dağılımları Tablo 6'da sunulmuştur. Tablodan görüleceği üzere, Türkiye, Lübnan, Irak ve Hollanda tek başına birer kümede yer almaktadır.

Tablo 6. Ülkelerin Kümlere Göre Dağılımı

Ülke	Küme	Ülke	Küme
Almanya	Küme 1	Cibuti	Küme 1
Avusturya	Küme 1	Suudi Arabistan Krallığı	Küme 1
Belçika	Küme 1	Senegal	Küme 1
Bulgaristan	Küme 1	Sudan	Küme 1
Çekya	Küme 1	Suriye Arap Emirlikleri	Küme 1
Danimarka	Küme 1	Surinam	Küme 1
Estonya	Küme 1	Sierra Leone	Küme 1
Finlandiya	Küme 1	Somali	Küme 1
Fransa	Küme 1	Umman	Küme 1
GKRY	Küme 1	Gabon	Küme 1
Hırvatistan	Küme 1	Gambiya	Küme 1
İrlanda	Küme 1	Guyana	Küme 1
İsveç	Küme 1	Gine	Küme 1
Letonya	Küme 1	Gine-Bissau	Küme 1
Litvanya	Küme 1	Filistin	Küme 1
Lüksemburg	Küme 1	Komorlar	Küme 1
Macaristan	Küme 1	Katar	Küme 1
Malta	Küme 1	Kazakistan	Küme 1
Polonya	Küme 1	Kamerun	Küme 1
Portekiz	Küme 1	Fildişi Sahili	Küme 1
Romanya	Küme 1	Kuveyt	Küme 1
Slovakya	Küme 1	Libya	Küme 1
Slovenya	Küme 1	Maldivler	Küme 1
Yunanistan	Küme 1	Mali Cumhuriyeti	Küme 1
Arnavutluk	Küme 1	Malezya	Küme 1
Karadağ	Küme 1	Mısır	Küme 1
Kuzey Makedonya	Küme 1	Fas	Küme 1
Sırbistan	Küme 1	Moritanya	Küme 1
Azerbaycan	Küme 1	Mozambik	Küme 1
Ürdün	Küme 1	Nijer	Küme 1
Afganistan	Küme 1	Nijerya	Küme 1
Endonezya	Küme 1	Yemen	Küme 1
Uganda	Küme 1	Kırgızistan	Küme 1
İran	Küme 1	Özbekistan	Küme 1

Pakistan	Küme 1		
Bahreyn	Küme 1	İtalya	Küme 2
Brunei	Küme 1	Cezayir	Küme 2
Bangladeş	Küme 1		
Benin	Küme 1	Hollanda	Küme 3
Burkina-Faso	Küme 1		
Tacikistan	Küme 1	Irak	Küme 4
Türkmenistan	Küme 1		
Çad	Küme 1	Lübnan	Küme 5
Togo	Küme 1		
Tunus	Küme 1	Türkiye	Küme 6

6.6 Milli Marşlarda Geçen Sözcüklerin Birliktelik Analizi

Birliktelik analizi milli marşlar içerisinde geçen kelimelere (eş anlamlı sözcükler dahil) göre yapılmıştır. Analizlerde Bölüm 5'te açıklanan ve birliktelik analizinde kullanılan destek (support), kaldıraç (lift) ve güven (confidence) eşitliklerinden yararlanılarak yapılmıştır.

Birliktelik analizleri için yine RapidMiner 9.1 (akademik sürüm) yazılımının FP-Growth ve Create Association Rules operatörlerinden yararlanılmıştır. Yapılan analizlere göre en yoğun kullanılan sözcük "toprak"tır (Support:0,859). Ayrıca vatan-toprak, yaşam-toprak, son dilek-toprak ve kan-toprak sözcüklerinin birliktelikleri güçlü ve anlamlı olduğu bulunmuştur (Tablo 7).

Tablo 7. Anlamlı Birliktelik Kural Değerleri

Öncül	Sonuç	Destek	Kaldıraç	Güven
Vatan	Toprak	0,271	1,116	0,958
Yaşam	Toprak	0,200	1,106	0,950
Son dilek	Toprak	0,224	1,053	0,905
Kan	Toprak	0,224	1,006	0,864

7. SONUÇ

Metin madenciliği ve veri madenciliği kullanıldığı alanların genişlemesi itibari ile her geçen gün popülerliği artan yöntemlerdir. Bu çalışmada bir milletin ortak duygularına tercüman olan milli marşların değerlendirilmesi hedeflenmiştir. Milli marşların metinsel veri türü olmasından ötürü metin madenciliği ve veri madenciliği teknikleri ile analiz edilmesi mümkündür. Bu çalışmada incelenecek ve Referans marş olarak seçilecek marş Türkiye Cumhuriyeti milli marşı İstiklal Marşı'dır. İstiklal Marşının yanı sıra Türkiye Cumhuriyeti'ne ilişkisi bulunan 84 ülkenin marşı değerlendirmeye alınmıştır. Çalışma 5 farklı boyut içermektedir. Bu boyutların üçü metin madenciliği ikisi ise veri madenciliği teknikleri olarak değerlendirilebilir. Kelime, Duygu ve Başlık boyutunu içeren Metin Madenciliği teknikleri ile elde edilen bulgulardan oluşan sıralamalarda farklılıklar gözlemlenmiştir. Başlıklardan başlıklara kelimelerin ağırlıkları farklı olduğu için başlıklarda sıralamalar değişmiştir. Duygu analizi metinlerin orijinal halleri ile gerçekleştirildiğinden anlamların farklı olması beklenebilir. Bütün boyutlarda en benzer ülkeler içinde farklı ülke topluluklarından ülkeler görülebilir. Metin madenciliği teknikleri sonucunda İstiklal Marşı'na benzer bir marş tespit edilememiştir. Bu bulgu, kümeleme analizi ile test edilmiştir. Kümeleme analizi sonucunda İstiklal Marşı'nın tek başına bir kümede yer alması diğer marşlarda benzerliğinin

bulunmadığını göstermektedir. Başka bir ifadeyle Türkiye Cumhuriyeti İstiklal Marşı, kendisi ile kıyaslanan 84 ülkenin milli maşından farklı ve özgündür. Yine tüm milli marşların ortak sözcüğü "toprak" olup, bu kavramın öncülleri (tetikleyicileri) vatan, yaşam, son dilek ve kan sözcükleridir.

KAYNAKÇA

- Abril, C. R. (2016). A national anthem: Patriotic Symbol or Democratic Action? In *Patriotism and Nationalism in Music Education* (pp. 77-94). Routledge.
- Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining Association Rules Between Sets of Items in Large Databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207-216).
- Akarsu, S. İstiklal Marşı ve İcrasına Yönelik Bir Analiz Çalışması. *Turkish Studies*, 16(5), 1-35.
- Aktaş, H. E. (2014). Milli Marşların Siyaset Biliminin Bazı Kavramları Açısından Değerlendirilmesi. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(2), 71-92.
- Al-Maolegi, M., & Arkok, B. (2014). An Improved Apriori Algorithm for Association Rules. arXiv preprint arXiv:1403.3948.
- AL-Zawaidah, F. H., Jbara, Y. H., & Marwan, A. L. (2011). An Improved Algorithm for Mining Association Rules in Large Databases. *World of Computer Science and Information Technology Journal*, 1(7), 311-316.
- Arslan, M. F., Mehmood, M. A., & Haroon, H. (2021). Stylistic and Textual Analysis of Pakistani National Anthem. *Journal of Social Sciences and Humanities*, 1(2), 21-29.
- Balbi, S., Misuraca, M., & Scepti, G. (2018). Combining Different Evaluation Systems on Social Media for Measuring User Satisfaction. *Information Processing & Management*, 54(4), 674-685.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Csepeli, G., & Örkény, A. (1998). The Imagery of National Anthems in Europe. *Nation, Ethnicity, Minority And Border. Contributions to an International Sociology*, 37-56.
- Dlodlo, S. (2019). *An Investigation into Nation Building through the National Anthem in Zimbabwe: a Sociolinguistic Approach* (Doctoral dissertation).
- Essam, B. A. (2015). Melopoetics of the Contemporary "National Anthem" of Egypt and its Translations: A case Study. *Higher Education of Social Science*, 8(1), 1-11.
- Gilboa, A., & Bodner, E. (2009). What are Your Thoughts when the National Anthem is Playing? An Empirical Exploration. *Psychology of Music*, 37(4), 459-484.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*, 2nd. University of Illinois at Urbana Champaign: Morgan Kaufmann.
- Haşiloğlu, M., & Budak, İ. (2019). Sanal mağaza Drone Depo Yer ve Önceliklerinin Tespitine Yönelik Bir Araştırma Süreci Modeli. *Journal of Internet Applications and Management*, 10(2), 63-79.
- Krestel, R., Fankhauser, P., & Nejdl, W. (2009, October). Latent Dirichlet Allocation for Tag Recommendation. In *Proceedings of the Third ACM Conference on Recommender Systems* (pp. 61-68).

- Lauenstein, O., Murer, J. S., Boos, M., & Reicher, S. (2015). 'Oh motherland I pledge to thee...': a Study into Nationalism, Gender and The Representation of an Imagined Family Within National Anthems. *Nations and Nationalism*, 21(2), 309-329.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Pelleg, D., & Moore, A. W. (2000, June). X-means: Extending k-means with Efficient Estimation of the Number of Clusters. In *Icml* (Vol. 1, pp. 727-734).
- Sabah, L., & Bayraktar, H. (2020). Veri Madenciliği Birliktelik Kuralları ile Binaların Risk Durumlarının Analizi: Kaynaşlı, Düzce Örneği. *Gazi Mühendislik Bilimleri Dergisi*, 6(1), 70-78.
- Souza, A. A. D. (2008). Do the righth, be firm, be fair...: a Systemic Funcional Investigation of National Anthems Written in English.
- Van Ginderachter, M. (2020). *Encounters With the Belgian Flag and the National Anthem*. In the Everyday Nationalism of Workers (pp. 105-124). Stanford University Press.
- Wang, X., & Grimson, E. (2007). Spatial Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 20.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 347-354).
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 Algorithms in Data Mining. *Knowledge and Information Systems*, 14(1), 1-37.
- Xu, R. Ve Wunsch, D. (2008). *Clustering* (Vol. 10). John Wiley & Sons.