

To cite this article: Koç, O., Aydın, Ö. (2021). Uluslararası Hakemli Yayınlar Bibliyometrik Bir Bakış ile Madenciliği Yöntemlerinin Uygulanması: Balıkesir Üniversitesi Örneği. International Journal of Social and Humanities Sciences (IJSHS), 5(3), 143-176

Submitted: October 07, 2021

Accepted: December 17, 2021

ULUSLARARASI HAKEMLİ YAYINLARA BİBLİYOMETRİK BİR BAKIŞ İLE MADENCİLİĞİ YÖNTEMLERİNİN UYGULANMASI: BALIKESİR ÜNİVERSİTESİ ÖRNEĞİ¹

Okan Koç²
Özlem Aydın³

ÖZET

Bu çalışmanın amacı, Balıkesir Üniversitesi'nin 1994- 2020 yılları arasında Web of Science' da yer alan yayınlarının, bibliyometrik analizini gerçekleştirmek ve bu yayınları metin madenciliği yöntemleri ile değerlendirmektir. Değerlendirme sonucunda metin madenciliği yöntemlerinin öncelikli alan tercihlerinde nasıl kullanılabileceği üzerinde durulacaktır. Bu doğrultuda Balıkesir Üniversitesi Web of Science bibliyometrik veri tabanında 1994-2020 yılları arasında indekslenen toplam 4176 yayın bibliyometrik analiz tekniği kullanılarak incelenmiştir. İlgili yayınlar; yayın diline, çalışmanın türüne, yayın yaptıkları ülkelere, çalıştıkları kurumlara, en çok atıf yapılan yayınlara, en çok yayın yapılan alanlara ve yazarlara göre değerlendirilmiştir. Araştırma sonucunda elde edilen bulgular, Balıkesir Üniversitesi'nin 1994-2020 yılları arasındaki uluslararası yayın performansını ortaya koymayı amaçlamaktadır.

Üniversiteye ilişkin yayınlar WoS veri tabanındaki Research Areas kategorisine göre dört ayrı alana ayrılmıştır. Bu alanlar temel bilimler, tıp ve sağlık bilimleri, mühendislik ve mimarlık son olarak sosyal bilimlerdir. Bu ayırımın amacı hem yayınların genel analizini hem de bu dört alt alanın ayrıntılı analizini yapmaktır. Araştırma sonucunda elde edilen bulgular ışığında; uluslararası bilimsel yayın sayısının artırılmasına yönelik yapılması gerekenler değerlendirilmiştir. Bununla birlikte metin madenciliği yönteminin sonuçlarına yer verilmiştir. Analiz sonuç-

¹ Bu makale 2021/090 no'lu, Balıkesir Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenen "Üniversitelerin Öncelikli Araştırma Alanlarının Belirlenmesinde Bir Öneri Modeli" isimli proje kapsamında üretilmiştir.

² Öğr. Gör. Dr., Balıkesir Üniversitesi, okan.koc@balikesir.edu.tr, ORCID ID: <https://orcid.org/0000-0002-5356-5940>

³ Doktora Öğrencisi. Balıkesir Üniversitesi, ozlem.aydin@balikesir.edu.tr, ORCID ID: 0000-0001-5861-2999

ları veri setinde bulunan WoS kategorisi bilgisi ile karşılaştırılmıştır. Karşılaştırma sonucu detaylı çalışma alanlarına karar verilmiştir. Buna göre Balıkesir Üniversitesi'nin WoS' da yer alan yayınlarını en iyi temsil eden anahtar kelimelerin "biyokimya ve moleküler biyoloji, kimya, organik kimya olduğu sonucuna ulaşılmıştır. Tıp ve sağlık bilimlerinde ise bu kategoriler "endokrinoloji ve metabolizma, kardiyak ve kardiyovasküler sistemler, biyopsiler, toksikoloji" olarak tespit edilmiştir. Sosyal bilimlerde ise "dilbilim, eğitim ve eğitim araştırmaları" kategorileri olarak gözlenmiştir.

Anahtar Kelimeler: Balıkesir Üniversitesi, Bibliyometrik Analiz, Bibliyometri, Web of Science, Metin Madenciliği, TFIDF

APPLICATION OF MINING METHODS WITH A BIBLIOMETRIC PERSPECTIVE TO INTERNATIONAL REVIEWED PUBLICATIONS: THE CASE OF BALIKESİR UNIVERSITY

ABSTRACT

The aim of this study is to perform bibliometric analysis of Balıkesir University's publications in Web of Science between 1994 and 2020 and to evaluate these publications with text mining methods. As a result of the evaluation, it will be emphasized how text mining methods can be used in priority field preferences. In this direction, a total of 4176 publications indexed between 1994-2020 in the Balıkesir University Web of Science bibliometric database were examined using the bibliometric analysis technique. Related publications; were evaluated according to the language of publication, the type of study, the countries with which they were published, the institutions they worked with, the most cited publications, the most published fields and authors. The findings obtained as a result of the research aim to reveal the international publication performance of Balıkesir University between 1994-2020.

University-related publications are divided into four separate areas according to the Research Areas category in the WoS database. These fields are basic sciences, medicine and health sciences, engineering and architecture, and finally social sciences. The purpose of this distinction is to make both a general analysis of publications and a detailed analysis of these four sub-areas. Text mining analysis was performed using the keywords of the publications for both the entire dataset and the dataset divided into four. In the light of the findings obtained as a result of the research, what should be done to increase the number of international scientific publications was evaluated. In addition, the results of the text

mining method are included. The analysis results were compared with the Wos categories information in the dataset. As a result of the comparison, detailed study areas were decided. Accordingly, it was concluded that the keywords that best represent the publications of Balıkesir University in WOS are included in the Wos Categories of 'biochemistry and molecular biology, chemistry, organic chemistry'. It has been observed that the words that best represent the field of basic sciences are in the Wos Categories of 'biochemistry and molecular biology, chemistry, organic chemistry'. In medicine and health sciences, these categories were determined as 'endocrinology and metabolism, cardiac and cardiovascular systems, biopsies, toxicology'. In social sciences, it has been observed as 'linguistics, education and educational research' categories.

Keywords: Balıkesir University, Bibliometric Analysis, Bibliometry, Web of Science, Text Mining, TFIDF

GİRİŞ

Bilimsel çalışmaların belirli aralıklarla incelenmesi, değerlendirilmesi, yoğunlaşan konuların açığa çıkarılması ve bununla birlikte o alanda faaliyet gösteren kurum, kuruluş, ve araştırmacıların faaliyetlerinin belirlenmesi mevcut durumun analizi açısından oldukça önemlidir. Bilimsel iletişim açısından bilginin kaydedilmiş olduğu ortam, bilgi kaynağının türü, uluslararası paydaşları, saygınlığı, geçerliliği ve dili her geçen gün değişmekte ve önemli bir kriter olarak karşımıza çıkmaktadır.

Bilimsel iletişimi derinden etkileyen en önemli unsurların başında bilginin kayıt edildiği ve erişime sunulduğu ortamların dijitalize edilmesi gelmektedir. Bugün için bilgi ve iletişim teknolojilerindeki gelişime paralel olarak bilimsel iletişim açısından bakıldığında elektronik kaynakların hız, maliyet, aynı anda çoklu kullanım ve zaman gibi avantajlarından ötürü daha çok tercih edildiği görülmektedir. Her geçen gün bilimsel literatürdeki elektronik yayınların atıf grafiği atarak devam etmekte, bu doğrultuda görünürlüğün de üst düzeye çıkması, bilginin küresel çapta yaygınlaşmasını mümkün kılmaktadır.

Bilimsel iletişim açısından uluslararası anlamda en önemli platformlardan birisi de Web of Science'dır. Akademik yükselmelerde burada indekslenen dergilerde yayın yapmak önemli kriter olması bir yana, ortaya çıkarılan yayının tüm dünya tarafından kabul görmesi, saygınlığının ve geçerliliğinin onaylanması anlamına da gelmektedir.

Web of Science, Science Citation Index, Social Science Citation Index ve Arts and Humanities Citation Index olmak üzere uluslararası bibliyografik veritabanlarından oluşmaktadır. Bilimsel materyalin türü, yazarı, almış olduğu atıf sayısı, kurum ve kuruluşları ve bilgi kaynağına ait bibliyografik veri içeren bir platformdur (Web of Science, 2020). Web of Science bibliyografik veri tabanı daha çok yayın türü olarak dergilere odaklanmakta olup, kitaplar, bildiriler ve bilimsel raporlar pek fazla değerlendirmeye alınmamaktadır. Doğa bilimleri, mühendislik ve sağlık bilimleri için makalenin bilimsel iletişim için açısından daha çok tercih edildiği bilinmekle birlikte sosyal bilimler açısından kitapların önemli bir yere sahip olması konunun bu yönden de ele alınmasını zorunlu kılmaktadır (Lariviere, Archambault, Gingras ve Vignola-Gagne, 2006). WoS üzerindeki bibliyometrik bir çalışmanın çıktıları doğa bilimleri, mühendislik, tıp ve sağlık bilimleri için etkinliği ve etkililiği ortaya koymakla birlikte, sosyal bilimler için aynı şeyi söylemek pek mümkün değildir (Nederhof, 2006; Hicks ve Wang, 2011). Ülkelere ait sosyal ve beşeri bilimler alanındaki çıktıların sayısal analizinde WoS verilerinin tek başına doğru bir sonuç sunamayacağı görülmektedir.

WoS üzerinde bibliyometrik bir analiz için önemli bir konu da dil unsurudur. Neredeyse indekslenen yayınların tamamına yakınının İngilizce olduğu görülmektedir. Dünya çapındaki en kapsamlı süreli yayınlar listesini sunan Ulrich's veritabanı kıyasladığımızda, WoS'da yer alan yayınların daha çok İngilizce olduğu görülmektedir (Archambault, Vignola-Gagné, Côté, Lariviere ve Gingras, 2006, s. 329). WoS üzerinde indekslenen yayınların toplamında baskın olarak 9 dilin öne çıktığı tespit edilmiş olup, hemen her alanda ilk sırada İngilizcenin yer aldığı görülmektedir (Mongeon ve Paul-Haus, 2016, s. 221).

Bibliyometrik yöntemler çeşitli alanlardaki araştırmaların analiz edilmesinde kullanılmaktadır. Bu doğrultuda ortak veri platformu olarak bibliyometrik veri sağlayan Web of Science ya da Scopus gibi veritabanları incelenmektedir. Ayrıca Ulrichs Süreli yayınlar veritabanı da konuyla ilgili bir başka çalışma alanı olarak değerlendirilmektedir (Mongeon ve Paul-Hus, 2016, s. 214). Çalışma kapsamında bibliyometrik analiz için Web of Science veritabanı kullanılmıştır.

Dijital teknolojilerdeki hızlı değişim ve ilerlemelerin, veritabanları, veri havuzları ve bulut tabanlı ortamlarda tutulan bilginin katlanarak artmasında doğrudan etkisi olduğu görülmektedir. Veritabanları üzerinde yer alan dijital bilgi belirli bir sistematığe göre yapılandırılmaktadır. Diğer yandan internet üzerinde yer alan her bilgi kaynağı için aynı şeyi söylemek mümkün değildir. Belirli bir sistematığe sahip olarak kümelendirilmiş veritabanlarından bilgiyi süzmek için me-

tin madenciliği yöntemi kullanılmaktadır. Metin madenciliği, bilgiye erişim, metin analiz, bilgi çıkarma, kümeleme, sınıflandırma, görselleştirme, veritabanı teknolojisi, makine öğrenimi ve veri madenciliğini içeren çok disiplinli alandır (Rajman ve Vesely, 2004).

Metin madenciliği, bilgisayar bilimi, dilbilim, tıp, kütüphane ve bilgi bilim gibi birçok alanda kullanılmaktadır. Kütüphane ve bilgi bilimi alanında metin madenciliği kullanılmakla birlikte bibliyometrik analizlere de yer verilmektedir.

Dijital ortamdaki bu veri artışı mevcut dataların çeşitli bilgisayar yöntemleri ile analiz edilerek içerdikleri saklı bilgilerin keşfedilmesinin de önemini artırmaktadır. Veri yığınlarından bilgiye geçiş her alanda artı değerler yaratmaktadır. Bu çalışma kapsamında bibliyometrik analiz yöntemleri metin madenciliği yöntemleri ile birlikte kullanılmış ve Balıkesir Üniversitesine ait WoS' da yer alan yayınların anahtar kelimelerinden hangi alanlarda hangi spesifik konuların daha çok çalışıldığı bilgisine ulaşılmaya çalışılmıştır. Bu bilgi aynı zamanda Balıkesir Üniversitesi açısından hangi anahtar kelimeleri içeren yayınların WoS' da daha fazla yer aldığı bilgisini kullanıcıya sağlamaktadır. Bu çalışmanın sonuçları uluslararası hakemli yayınlardaki mevcut durum analizini yaparken kullanılabileceği gibi gelecek planlamasında da referans olarak kullanılabilir.

LİTERATÜRE BAKIŞ

Son yıllarda bibliyometrik analizlerle, ülkelere, kurumlara göre, konu bazlı, atıf temelli, yazar ya da makale sayılarına göre farklı türde araştırmaların yapıldığı görülmektedir. Daha çok belirli bir disipline ya da konuya göre genel eğilimleri ortaya çıkarmaya yönelik sürdürülen çalışmalarda durum saptamaları yapılmaktadır (Koç, 2021, s.23).

Bununla birlikte ülkelere ve kurumlara yönelik atıf ve yayın sayılarının genel seyrinin gösterildiği bibliyometrik analizlerin gerçek resmi görmede tek başına yeterli olmadığı, içerisinde kullanılan anahtar kelimelerin daha da ayrıntılı analizlerde kullanılması gerektiği belirtilmektedir (Suk ve diğerleri, 2011). Diğer yandan üzerinde durulan bir başka konu çalışmalarda ortak olarak kullanılan kelimelerin belirlenerek sağlıklı bir analizin mümkün olabileceği dile getirilmiştir (Ohniwa, Hibino ve Takeyasu, 2010).

Konuyla ilgili olarak literatüre bakıldığında Web of Science verilerinin kullanıldığı birçok araştırma söz konusudur. Bibliyometrik bir analizle Çin'in makine öğrenmesi konusunda destek vektör makineleri için gerçekleştirilen bilimsel çalışmaları temel alan, diğer ülkelerle genel bir kıyaslama yapan, konuyla ilgili

kurumsal iş birliği yapan kurumları ortaya çıkaran Yu ve Wang'ın çalışması ülke ve konu temelli bibliyometrik çalışmalara örnek oluşturmaktadır (2020).

Ülkelerin WoS üzerindeki bilimsel çalışmalarını temel alan; hangi alanlarda daha çok araştırma yapıldığını ortaya koyan, kronolojik olarak araştırma alanlarındaki değişiklikleri belirleyen incelemelerin yapıldığı da görülmektedir. 2008 yılında İspanya'nın 1990-2005 yılları arasındaki WoS üzerindeki yayın performansı ele alan çalışmada, hangi disiplinlerde daha çok yoğunlaştığı ve disiplinler arası çalışmaların genel seyrinin belirlendiği görülmektedir (Vargas-Quesada ve diğerleri, 2008).

Dergi temelli bir çalışmada Computer Methods and Programs in Biomedicine dergisindeki 1970- 2017 yılları arasındaki biyomedikal araştırmalarındaki makalelerin genel seyri ortaya çıkarılmıştır (Shukla ve diğerleri, 2019).

Web of Science üzerinde insan güvenirliliği konusunda genel eğilimleri ortaya çıkarmak adına 1984- 2018 yılları arasındaki toplamda 1463 yayın incelenmiştir. Yayınlar üzerinde en sık kullanılan anahtar kelimelerin neler olduğu belirlenmiştir (Tao ve diğerleri, 2020).

Hindistan'daki biyoteknoloji fakültelerinin araştırma performansını ölçmeye yönelik gerçekleştirilen bir diğer çalışmada 1997-2006 yılları arasındaki Web of Science verileri incelenmiştir (Sevukan ve Sharma, 2008).

Bir başka çalışma da, Leiden Üniversitesi Tıp Merkezi'nin bibliyometrik yöntemle atıf ve araştırma çıktılarının analizi gerçekleştirilmiş, tıp merkezinin yayın performansı değerlendirilmiştir (Van Raan, 2014).

Üniversitelerin yayın performansını ölçmeye yönelik gerçekleştirilen bir diğer çalışma ise Orissa Üniversitesi Tarım Teknolojileri Enstitüsü'nün 2008- 2012 yılları arasındaki Scopus üzerindeki bilimsel çıktılarını ve performansı ele alan araştırmadır (Maharana, 2013).

Sambalpur Üniversitesi'nin 2007- 2011 yılları arasında Web of Science'da indekslenen yayınlarını ele alan bir diğer çalışma da üniversitenin bilimsel performansı değerlendirilmiştir (Maharana ve Sethi, 2013).

Burdwan Üniversitesi'nin kimya alanındaki bilimsel çıktılarını ele alan bir diğer çalışmada alana özgü genel seyir, kurum ve yazar performansları değerlendirilmiştir (Nandi ve Bandyopadhyay, 2009).

Ülkemiz açısından bakıldığında ise uluslararası literatüre benzer bir gelişimin olduğu görülmektedir. Literatürdeki çalışmaların belli alanlar özelinde gerçekleştirildiği, yayın türüne göre daha çok dergi ve tez içeriklerinin ele alındığı bibliyometrik çalışmalar yer almaktadır. Bununla birlikte daha çok konu bazlı bib-

zarin ilgili yıllar arasında almış oldukları atıf sayıları, h-indeksleri ile en çok yayın yapılan dergilerin etki değerleri belirlenmiştir.

Bibliyometrik analiz tekniği ile elde edilen Balıkesir Üniversitesi'nin 1994-2020 yılları arasındaki Web of Science veritabanında indekslenen toplam 4176 yayına ait anahtar kelimeler metin madenciliği yöntemleri kullanılarak ayrıca bir analize tabi tutulmuştur.

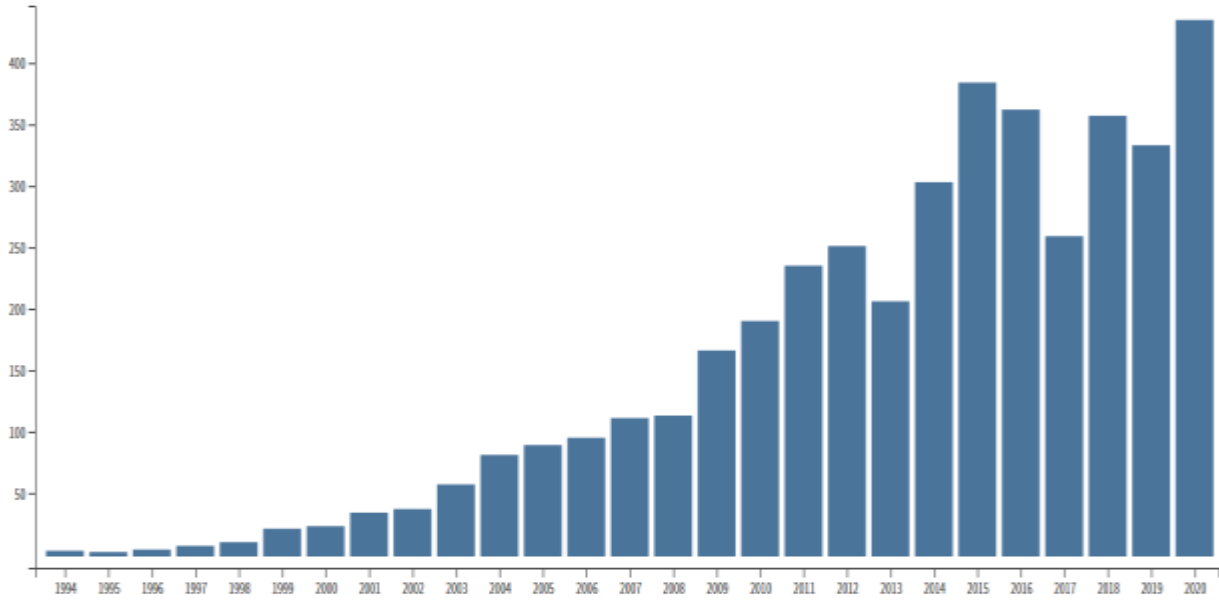
Bilimsel Üretim Performansı

Tablo 1. Yıllara Göre Yayın Sayıları

Yıllar	Yayın Sayısı	%
2020	440	10.536
2019	334	7.998
2018	358	8.573
2017	260	6.226
2016	362	8.669
2015	384	9.195
2014	303	7.256
2013	206	4.933
2012	251	6.011
2011	235	5.627
2010	190	4.550
2009	166	3.975
2008	113	2.706
2007	111	2.658
2006	95	2.275
2005	89	2.131
2004	81	1.940
2003	57	1.365
2002	37	0.886
2001	34	0.814
2000	23	0.551
1999	21	0.503
1998	10	0.239
1997	7	0.168
1996	4	0.096
1995	2	0.048

1994	3	0.072
Toplam	4176	100

Tablo 1’de Web of Science bibliyografik veritabanı tarafından indekslenen Balıkesir Üniversitesi yayınlarının yıllara göre dağılımı görülmektedir. Buna göre en fazla yayın 2020 (440 yayın) yılında yapılmıştır.



Grafik 1 Yıllara göre Yayın sayıları

Grafik 1’de yıllara göre yayınların genel seyri görülmektedir. Yayınların yaklaşık %79.53’ünün son on yılda yapıldığı görülmektedir.

Tablo 2. Web of Science’da yayınların türe göre dağılımı

Yayınların Türe Göre Dağılımı	Yayın Sayısı	Yüzde %
Makale	3613	86.518
Bildiriler	355	8.501
Toplantı Özeti	197	4.717
Mektup	46	1.102
Görüşler	39	0.934
Kitap Bölümü	29	0.694
Diğer	62	1,486
Toplam	4341	100

Araştırma sonuçları birçok biçimde yayınlanmaktadır. Web of Science’da yer alan Balıkesir Üniversitesi araştırmalarının yayın türüne göre dağılımına bakıl-

dığında ilk sırada 3613 adet (%87) ile makalelerin geldiği görülmektedir. Bildirilerin sayısı ise 355 (% 8) olarak tespit edilirken, toplantı özeti 197 (%5), mektup 46 (%1), görüşler 39 (&1), kitap bölümlerinin sayısı 29 adet (%1) olarak belirlenmiştir. Toplamda 261 yayının türünün bildiri, görüş, mektup olmasının yanı sıra makale olarak da sisteme yüklendiği tespit edilmiştir. Bilimsel yayının sayısı bakımından makalelerin sıralamada en önde yer alması oldukça olağan bir durum olarak görülmektedir.

Tablo 3. Birlikte en çok yayın yapılan kurumlar

Birlikte yayın yapılan kurum	Yayın Sayısı	Yüzde %
Uludağ Üniversitesi	234	5.603
Selçuk Üniversitesi	112	2.682
Ankara Üniversitesi	110	2.634
Atatürk Üniversitesi	105	2.514
Dokuz Eylül Üniversitesi	90	2.155
Ege Üniversitesi	90	2.155
İstanbul Üniversitesi	90	2.155
Gazi Üniversitesi	88	2.107
Sakarya Üniversitesi	85	2.035
Hacettepe Üniversitesi	82	1.964
Çanakkale Onsekiz Mart Üniversitesi	80	1.916
Erciyes Üniversitesi	77	1.844
Anadolu Üniversitesi	76	1.820
Bezmialem Vakıf Üniversitesi	74	1.772
Muğla Sıtkı Kocaman Üniversitesi	73	1.748
İnönü Üniversitesi	70	1.676
İstanbul Teknik Üniversitesi	69	1.652
Orta Doğu Teknik Üniversitesi	67	1.604
Eskişehir Osmangazi Üniversitesi	65	1.557
Adnan Menderes Üniversitesi	59	1.413
Akdeniz Üniversitesi	58	1.389
Kocaeli Üniversitesi	58	1.389
Karadeniz Teknik Üniversitesi	57	1.365
Necmettin Erbakan Üniversitesi	57	1.365
Dicle Üniversitesi	56	1.341
Fırat Üniversitesi	55	1.317

Dumlupınar Üniversitesi	50	1.197
Celal Bayar Üniversitesi	49	1.173
Marmara Üniversitesi	48	1.149
Ondokuz Mayıs Üniversitesi	47	1.125
Uşak Üniversitesi	46	1.102
Afyon Kocatepe Üniversitesi	45	1.078
Pamukkale Üniversitesi	45	1.078
TÜBİTAK	45	1.078
Süleyman Demirel Üniversitesi	44	1.054
Sağlık Bilimleri Üniversitesi	43	1.030
Yüzüncü Yıl Üniversitesi	42	1.006
Abant İzzet Baysal Üniversitesi	41	0.982
Cumhuriyet Üniversitesi	41	0.982
Aksaray Üniversitesi	40	0.958
Recep Tayyip Erdoğan Üniversitesi	40	0.958
Mustafa Kemal Üniversitesi	37	0.886
Floransa Üniversitesi	36	0.862
Gaziosmapaşa Üniversitesi	34	0.814
Diğer Kurumlar	1266	29,214
Toplam	4176	100

Balıkesir Üniversitesi akademisyenleri toplamda 1367 kurumla ortak yazarlığı paylaşmış olup, Tablo 3'e göre 1994- 2020 yılları arasında Web of Science 'da en çok birlikte yayın yapılan kurum Uludağ Üniversitesi (234 yayın) olarak tespit edilmiştir. Sırayla bakıldığında Selçuk Üniversitesi ile 112 yayın, Ankara Üniversitesi 110 yayın ve Atatürk Üniversitesi ile 105 yayın yapıldığı görülmektedir.

Tablo 4. Yayın desteği alınan kurumlar

Yayın desteği alınan kurumlar	Yayın Sayısı	Yüzde %
TÜBİTAK	424	10.129 %
Balıkesir Üniversitesi	347	8.290 %
TC Kalkınma Bankası	72	1.720 %
Avrupa Komisyonu	51	1.218 %
TC Milli Eğitim Bakanlığı	25	0.597 %
Amerika- Ulusal Bilim Vakfı	23	0.549 %

Sakarya Üniversitesi	23	0.549 %
Uludağ Üniversitesi	21	0.502 %
TÜBA	19	0.454 %
İstanbul Üniversitesi	15	0.358 %
Uşak Üniversitesi	14	0.334 %
Atatürk Üniversitesi	13	0.311 %
Balıkesir Üniversitesi BAP	12	0.287 %
Dokuz Eylül Üniversitesi	12	0.287 %
Londra Kraliyet Topluluğu	12	0.287 %
Selçuk Üniversitesi	12	0.287 %
Çin Doğa Bilimleri Vakfı	11	0.263 %
Gazi Üniversitesi	10	0.239 %
Diğer	1918	73,339
Toplam	1334	100 %

Web of Science verileri incelendiğinde toplamda 4176 yayından 3034'ü için çeşitli kurumlardan destek alındığı saptanmıştır. Yayınlar destek sağlayan toplam kurum sayısı ise 463'tür. Oransal olarak bakıldığında bu durum toplam yayın sayısının %72'sine denk gelmektedir.

Bilimsel üretimin her geçen artıyor olması, üretilen yayın sayısının daha da yukarıya taşınması ve toplumsal olarak karşılığının alınabilmesi amacıyla ülkemizde yayın teşvikleri verilmektedir. Ancak son yıllarda bu durumun kötüye kullanımının önüne geçilebilmesi ve kaliteli yayınların üretilmesi daha da önemli hale gelmiştir. Bu noktada atıf sayıları temel alınırken, uluslararası indekslerce taranan dergilerdeki yayınların çeşitli kurum ve vakıflarca fonlanması da söz konusudur. Özellikle parasal desteği sağlanan projelerin bilimsel yayına dönüştürüldüğü görülmektedir.

Tablo 5. Yayınların dergilere göre dağılımı

Dergi Adı	Yayın Sayısı	Atıf Sayısı	Etki Değeri
Procedia Social And Behavioral Sciences	90	297	0
Febs Journal	82	5	5.542
Fresenius Environmental Bulletin	60	345	0.489
Journal of Enzyme Inhibition And Medicinal Chemistry	42	377	5.051

Turkish Journal of Mathematics	28	110	0.803
Aip Conference Proceedings	27	33	0
Journal of Magnetism and Magnetic Materials	27	348	2.993
Kafkas Üniversitesi Veteriner Fakültesi Dergisi	27	52	0.685
Turkish Journal of Botany	25	271	1.489
Journal of Materials Science Materials in Electronics	24	115	2.478
Journal of The Faculty of Engineering and Architecture of Gazi University	23	31	0.968
Milli Folklor	23	13	0
Turkish Journal of Veterinary Animal Sciences	21	65	0.581
5th World Conference on Educational Sciences	19	32	0
Hacettepe Journal of Mathematics and Statistics	18	30	0.929
Journal of Superconductivity And Novel Magnetism	18	200	1.506
Sensors and Actuators B Chemical	18	383	7.46
Acta Crystallographica a Foundation and Advances	17	0	2.29
Food Chemistry	17	723	7.514
Journal of Optoelectronics And Advanced Materials	17	61	0.587
Journal of Molecular Structure	16	160	3.196
Journal of Hazardous Materials	15	117	10.588
Zeitschrift Fur Naturforschung Section a Journal of Physical Sciences	15	54	0

Bilimsel dergilerinin kalitesini ve saygınlığı ölçmeye odaklı çalışmaların başında etki değeri (impact faktör) araştırmaları önemli bir yere sahiptir. Bir derginin etki değeri ne kadar yüksek ise, bilim dünyasında saygınlığını, güvenilirliği ve bilinirliği o denli yüksektir. Bilimsel yayınların etki faktörünün ölçülmesi bir takım eleştirilere rağmen günümüzde araştırmacıların yayın yaparken dikkate aldığı bir yapıdır. Web of Science üzerinde yer alan dergilerin etki değerinin yüksek olması ve etki değeri yüksek dergilerde yapılan yayınların uluslararası tanınırlığı artırması konuyu önemli kılmaktadır.

Etki değeri hesaplamaları bir dergide önceki iki yılda yayınlanan makalelere yapılan atıfların ilgili iki yıldaki makale sayısına bölünmesi şeklinde yapılmaktadır (Sanni S. A. ve Zainab, 2010).

Tablo 5 incelendiğinde, 2015- 2020 yılları arasındaki Balıkesir Üniversitesi araştırmalarının dergilere göre dağılımında ilk sırada 90 yayın ile Procedia Social And Behavioral Sciences yer almaktadır. İlgili yayındaki tüm araştırmaların, 2015 yılında Antalya’da gerçekleştirilen, “Proceedings of the 1st GlobELT Conference on Teaching and Learning English as an Additional Language” adlı kongrede sunulan bildirilerden oluştuğu görülmektedir. İlgili eserin etki değerini ölçülemediği. İlgili yayında yer alan araştırmalara toplamda 297 atıf alınmıştır.

İkinci sırada FEBS Journal (FEBS Congress) yer almaktadır. İlgili derginin yayınları incelendiğinde biyokimya ve moleküler biyoloji alanında düzenlenen FEBS Kongresinde sunulan bildirilerin daha sonradan makaleye dönüştürülerek yayımlandığı görülmektedir. Aynı yazarların özellikle Almanya ve Türkiye’de düzenlenen kongrelere katılarak birden fazla bildiri sunduğu ve bu bildirileri sonrasında makaleye çevrildiği tespit edilmiştir. İlgili derginin etki değeri (impact factor) 5.542 olarak belirlenmiştir. Dergi alana özgü 298 dergi içerisinde 72. sırada yer almaktadır. İlgili dergide yer alan yayınlara toplamda 5 atıf alınmıştır.

Üçüncü sırada 60 yayın ile Fresenius Environmental Bulletin yer almaktadır. İlgili derginin etki değeri 0.489 olarak belirlenmiştir. Dergi alana özgü 274 dergi içerisinde 270. sırada yer almaktadır. İlgili dergide yer alan yayınlara toplamda 345 atıf alınmıştır.

Dördüncü sırada yine 42 yayın ile Journal of Enzyme Inhibition And Medicinal Chemistry yer almaktadır. İlgili derginin etki değeri 5.051 olarak belirlenmiştir. Dergi alana özgü 298 dergi içerisinde 89. sırada yer almaktadır. İlgili dergide yer alan yayınlara toplamda 377 atıf alınmıştır.

Beşinci sırada Turkish Journal of Mathematics yer almaktadır. İlgili derginin etki değeri 0.803 olarak belirlenmiştir. Dergi alana özgü 330 dergi içerisinde 213. sırada yer almaktadır. İlgili dergide yer alan yayınlara toplamda 110 atıf alınmıştır.

Tablo 6. Yayınların yazarlara göre dağılımı

Yazar Adı	Yayın Sa- yısı	Atıf Sayısı	h- index
Hakan Köçkar	126	687	19
Oktay Arslan	122	1 684	22
Muzaffer Alkan	111	473	12
Hülya Kara	99	638	20
Feray Köçkar	92	428	14
Mehmet Doğan	90	8 173	54
Rıfat Çapan	85	521	16
Serap Doğan	82	593	16
Tuncay Dirmenci	81	509	16
Mursel Alper	79	610	17

Yayınların yazarlara göre dağılımı, alınan atıf sayıları ve h-index verileri Tablo 6'de verilmiştir. Buna göre 2015-2020 yılları arasında Web of Science'da en çok yayını indekslenen yazar 126 yayın ile Hakan Köçkar olmuştur. Sırasıyla diğer yazarlar; Oktay Arslan 122 yayın, Muzaffer Alkan 111, Hülya Kara 99 ve Feray Köçkar 92'şer yayın üretmiş olduğu görülmektedir.

Yazarların atıf sayısına bakıldığında ise 8173 atıf ile Mehmet Doğan en çok atıf alan yazar olmuştur. Sırasıyla diğer en çok atıf alan yazarlar; Oktay Arslan 1684 atıf, Hakan Köçkar 687 atıf, Hülya Kara 638 atıf ve Mursel Alper 610 atıf aldığı görülmektedir.

Diğer taraftan 2005 yılından itibaren bilimsel üretkenliğin ve etkinin ölçülebilmesi amacıyla oluşturulan h-indeks (Egghe, L. ve Rousseau, 2019) sıralamasına bakıldığında ilk sırayı 54 puan ile Mehmet Doğan almıştır. Sırasıyla diğer h-indeks sıralaması; Oktay Arslan 22, Hülya Kara 20, Hakan Köçkar 19, ve Mursel Alper 17 h-indeks puanına sahiptir.

Tablo 7. Birlikte en çok yayın yapılan ülkeler

Yazar Adı	Yayın Sa- yısı	Atıf Sayısı	h- index
Amerika	185	13 802	32
İngiltere	176	3 706	31
İtalya	63	1 310	21
Hindistan	50	517	14

Çin Halk Cumhuriyeti	34	1 641	14
Almanya	33	411	11
Kanada	32	1 629	16
Suudi Arabistan	28	191	6
Mısır	27	219	8

Yayınların ülkelere göre dağılımı, alınan atıf sayıları ve h-index verileri Tablo 7’de verilmiştir. Buna göre 2015-2020 yılları arasında Web of Science’da indekslenen yayınların birlikte yapıldığı ülkelere göre dağılımına bakıldığında en çok yayın yapılan ülke 185 yayın ile Amerika olmuştur. Sırasıyla diğer ülkeler; İngiltere 176 yayın, İtalya 63 yayın, Hindistan 50 yayın, Çin Halk Cumhuriyeti 34 yayın, Almanya 33 yayın, Kanada 32 yayın, Suudi Arabistan 28 yayın ve Mısır ile 27 yayın üretmiş olduğu görülmektedir.

Yayınların atıf sayılarına bakıldığında ise; en çok atıfı Amerika almış (13 802 atıf), onu sırasıyla İngiltere (3 706 atıf) ve Çin (1 641 atıf) izlemiştir. H-indeks sıralamasına göre ise, Amerika ile yapılan yayınların h-indeksleri ilk sırada yer alırken (h- indeks 32), onu İngiltere (h- indeks 31) ve İtalya (h- indeks 21) izlemektedir.

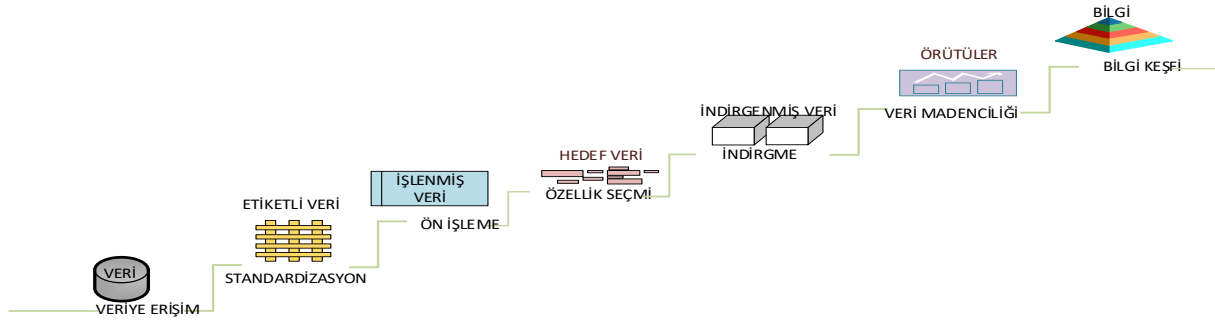
Tablo 8. Yayınların dile göre dağılımı

Yayın Dili	Yayın Sayısı	%
İngilizce	3974	95.104%
Türkçe	202	4.896%
Toplam	4176	100

Web of Science verilerine bu doğrultuda bakıldığında 2015-2020 yılları arasındaki yayınların % 95’ i İngilizce iken % 5’i Türkçedir.

Metin Madenciliği

Veri madenciliğinin bir alt dalı olarak gelişen metin madenciliği, bilgisayar yöntemlerinin kullanılarak, yazılı verilerin analizinin yapıldığı bir bilim dalıdır (Feldman ve Sanger, 2009). Metin madenciliği uygulamaları gerçekleştirilirken bazı temel işlemler yapılır. Şekil 1’ de temsili gösterilen bu işlemler sıra ve içerik olarak veri setine göre değişiklik gösterebilmektedir. Çalışma kapsamında yapılan analizlerde Python programlama dilinde yazılan kodlar ve bu dildeki kütüphaneler kullanılmıştır.



Şekil 1. Metin madenciliği adımları

Veriye erişim aşaması; analize dahil edilecek verinin elde edilmesini kapsamaktadır (Feldman ve Sanger, 2009). Bu çalışmada veriler bibliyometrik analiz tekniği kullanılarak Web of Science veritabanından elde edilmiştir. İlgili veri tabanında arama alanına belirli kriterler girilerek bu kriterlere uygun yayınlar bir excel dosyası formatında saklanmıştır. Veri seti Balıkesir Üniversitesi'nin 1994-2020 yılları arasında Web of Science veritabanında indekslenen İngilizce dilinde yazılmış toplam 4169 yayınından oluşturmaktadır. Analize kullanılan veri seti yazarların yayınlarını tanımlarken kullandıkları anahtar kelimelerdir. Veri setinde bu bilgiyi 'Author Keywords' kolonu içermektedir. Data setindeki toplam 4169 yayın yer almaktadır fakat yapılan ön incelemede 292 adet yayında 'Author Keywords' bilgisi olmadığı gözlenmiştir. Bu yayınlar veri setinden çıkarılmıştır. Bu işlem sonucunda veri seti 3877 adet yayından oluşturulmuştur. Veri setinde bulunan anahtar kelimeler herhangi bir metin madenciliği ön işleme yöntemi uygulanmadan önce değerlendirilmiş ve Tablo 9'de yer alan istatistiksel verilere ulaşılmıştır.

Tablo 9.Ön işleme öncesi anahtar kelimelere ilişkin istatistiksel veriler

Gösterge	Değer
Toplam anahtar kelime sayısı	18127
Anahtar kelime sayısı ortalaması	2,61
Standart sapma	0,98
İlk %25'lik dilim	2
Medyan	2
Son %75'lik dilim	3

3877 yayında toplam 18.127 adet anahtar kelime kullanılmıştır. Her bir anahtar kelime ortalama 2,61 kelimedenden oluşmakta ve verini %25'inde ve %50' sinde anahtar kelimeler iki kelimedenden son %25'inde ise üç kelimedenden oluşmuştur. De-

polanan veri setinin WoS veri tabanındaki ‘WoS Categories’ sınıfına göre dört ayrı alana bölünmüş bir kopyası oluşturulmuştur. Bu dört alan ‘temel bilimler, mühendislik ve mimarlık, tıp ve sağlık bilimleri ile sosyal bilimler’ dir. Tablo 10’ da yayınların bu dört ana alana göre dağılımı ve veri setindeki yüzde ağırlığı verilmiştir. Çalışmada kapsamında analizler hem bütün veri seti üzerinde hem de dört ayrı alan üzerinde ayrı ayrı yapılmış ve sonuçları paylaşılmıştır. Metin ön işleme aşamasından itibaren veri analizinde ve sonuçların özetlenmesinde Python programlama dilinde yazılmış kodlar kullanılmıştır.

Tablo 10. Web of Science’da yayınların dört ana alana göre dağılımı

Yayımların Konuya Göre Dağılımı	Yayın Sayısı	Yüzde %
Temel Bilimler	1.366	35
Tıp ve Sağlık Bilimleri	1.109	29
Mühendislik ve Mimarlık	1.048	27
Sosyal Bilimler	354	9
Toplam	3877	100

Standardizasyon aşaması; veri kümelerinin elemanlarının Word, basit metin, resim vs. farklı formatlarda kaydedilmiş olması durumunda veri madenciliği yöntemlerini uygulayabilmek için metinlerin; CSV, XML, ARFF gibi formatlara dönüştürüldüğü aşamadır (Mannila, 1996). WoS, veri setinin excel formatında indirilmesini sağladığı için bu aşama analizlerde kullanılmamıştır.

Metin ön işleme aşaması; verileri analize dahil etmeden veriler üzerinde; veri tipi dönüştürme, gruplandırma, eksik veya gürültülü verilerin veri setinden çıkarılması gibi işlemlerin gerçekleştirilerek verilerin analize hazır hale getirildiği aşamadır (Ergün, 2017). Metin ön işleme aşamalarından burada çalışma kapsamında kullanılanlara yer verilecektir. Bunların dışında farklı aşamalar da mevcuttur. Bu çalışmada öncelikle Author Keywords sütununda yer alan anahtar kelimeler Pandas Data Frame’e dönüştürülerek python dilinde yazılmış bir fonksiyon ile anahtar kelimeler sırasıyla boşluklar yardımıyla kelimelere ayrılmış, bütün harfler küçük harfe dönüştürülmüş, noktalama işaretleri temizlenmiş, sayılar veri setinden çıkarılmıştır. Ardından metin madenciliğinde stopwords olarak tanımlanan ve Türkçeye durak kelimeleri olarak çevrilen ‘a, and, the, also, either vs.’ gibi kelime ve metinlerin anlamını etkilemeyen, bununla birlikte metinlerde sıkça kullanılan sözcükler ile ekler Pythonda bulunan NLTK kütüphanesindeki listeler kullanılarak veri setinden çıkarılmıştır. Son olarak yine NLTK kütüphanesinde bulunan wordnet kullanılmış İngilizcede kelimelerin sonlarına eklenen

's, ies, ed, ing' gibi çekim ekleri temizlenmiş ve kelime köklerine (lemmatization) ulaşılmıştır. Bu işlem veri setinde aynı anlama gelen bir kelimenin birden fazla kez kullanılmasını engellemektedir.

Özellik seçimi aşaması; Burada metinler için belirleyici olan önemli kelimeler tespit edilir ve analiz edilecek veri kümesine dahil edilir. Bu basamakta bilgisayarlar için anlamsız karakterlerden oluşan kelimeler sayısallaştırılarak anlamlı işlenebilir hale getirilir (Berry & Kogan, 2010). Metin madenciliğinde özellik seçimi için farklı yöntemler kullanılmakla birlikte burada çalışma kapsamında kullanılan TF-IDF (term frequency–inverse document frequency) yöntemine değinilecektir. TF- IDF yöntemi iki ölçüt kullanır. Bunlardan biri TF diğeri ise IDF dir.

TF (terim sıklığı): kelimenin dokümanda geçme sayısının dokümandaki toplam kelime sayısına oranıdır.

DF (doküman sıklığı): toplam doküman sayısının ilgili kelimeyi içeren doküman sayısına oranıdır.

IDF (ters doküman sıklığı): DF değerinin logaritmasının alınması ile bulunur. Böylece dokümanlarda daha sık geçen kelimelerin ağırlık değeri azalarak ağırlık değerleri normalize edilir.

TF-IDF (terim sıklığı - ters doküman sıklığı): TF ve IDF değerlerinin çarpımıdır. Bu çarpım terimin dokümandaki ağırlığını verir. Dokümanın ağırlıklı olarak hangi konu ile ilgili olduğunu araştırdığımız analizlerde ilgili terimin dokümanda ne kadar ağırlıkta geçtiği bilgisini verir (HUANG, YIN, & HOU , 2021). Aşağıdaki formülle hesaplanır.

$$W_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

$tf_{x,y}$; x' in y dokümanındaki kullanılma sıklığı

df_x ; x' i içeren doküman sayısı

N ; Toplam doküman sayısı

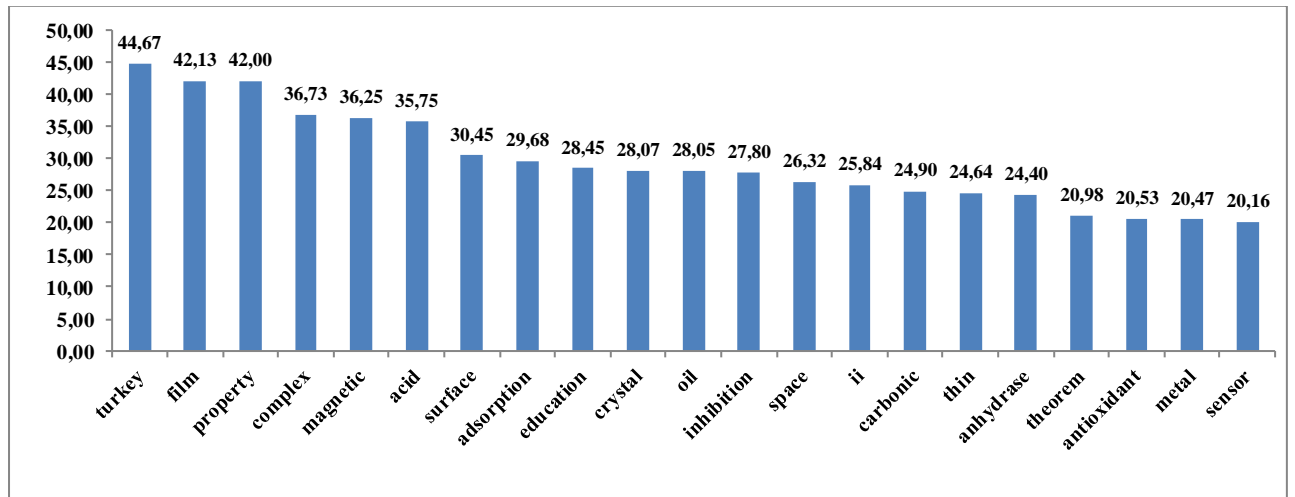
TF-IDF ağırlıkları n-gram yöntemi kullanılarak iki veya daha fazla kelimenin yan yana gelme ağırlıklarını yani birlikte kullanılma ağırlıklarını da hesaplayabilmemize olanak verir. Çalışmada n-gram parametresi bir, iki ve üç olan sonuçlara ilişkin örnekler paylaşılacaktır. Bu sonuçlar paylaşılırken öncelikle veri setinin bir bütün olarak değerlendirildiği analiz sonuçlarına yer verilecektir. Bu kapsamda elde edilen sonuçlar tablo, grafik ve kelime bulutları ile görselleştirilecektir.

Kelime bulutları metinlerde sık kullanılan anahtar kelimelerin görselleştirilmesi için kullanılan bir yöntemdir (Lohmann, Heimerl, Bopp, Burch, ve Ertl, 2015). Bu yöntem metinlerin daha hızlı analiz edilmesini ve daha sade bir şekilde görselleştirilmesini sağlar

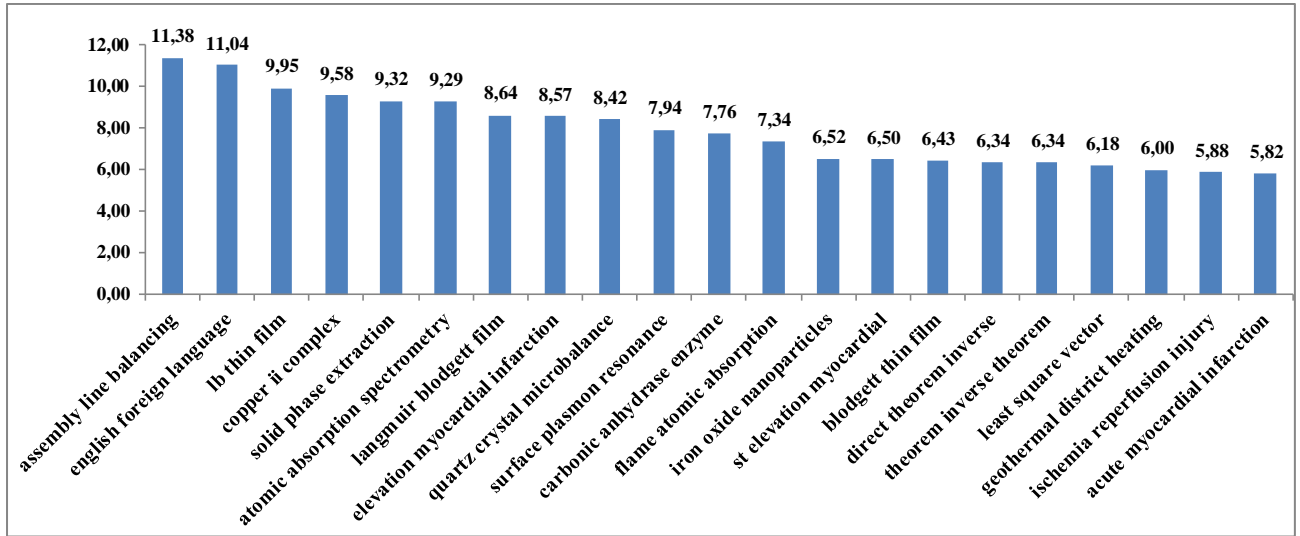
(Lohmann, Heimerl, Bopp, Burch, ve Ertl, 2015).

Ardından temel bilimler, mühendislik ve mimarlık, sağlık ve veteriner bilimleri ile sosyal bilimler alanlarının ayrı ayrı sonuçlarına yer verilecektir

N-gram değeri bir olarak verildiğinde tüm veri setini temsil edebilecek 3584 özellik için bir ağırlık hesaplanmıştır. Bu özelliklerden 20 tanesine ve bunların TF-IDF ile hesaplanmış ağırlık değerlerine Grafik 2’ de yer verilmiştir. Grafik incelendiğinde en yüksek ağırlığa sahip özelliği '44,67' ile 'turkey' olduğu ve bu özelliği 'film', 'property', 'complex' vs. özelliklerinin takip ettiği görülmüştür. Şekil 2’ de tüm veri setine ait kelime bulutu paylaşılmıştır. Kelime bulutlarının metinlerde sık kullanılan anahtar kelimelerin görselleştirilmesinde kullanıldığına değinmiştik. Kelime bulutları oluşturulurken kelime sayısı parametresi 100 olarak belirlenmiş böylece doküman setinde sık kullanılan 100 anahtar kelime görselleştirilmiştir. Şekil incelendiğinde grafikte yer alan kelimelerin kelime bulutunda da (turkey, property, complex, magnetic, carbonic, anhydrase, acid, thin, film vs.) öne çıktığı görülmektedir.



Grafik 2 Tüm veri seti n-gram bir TF-IDF ağırlıkları

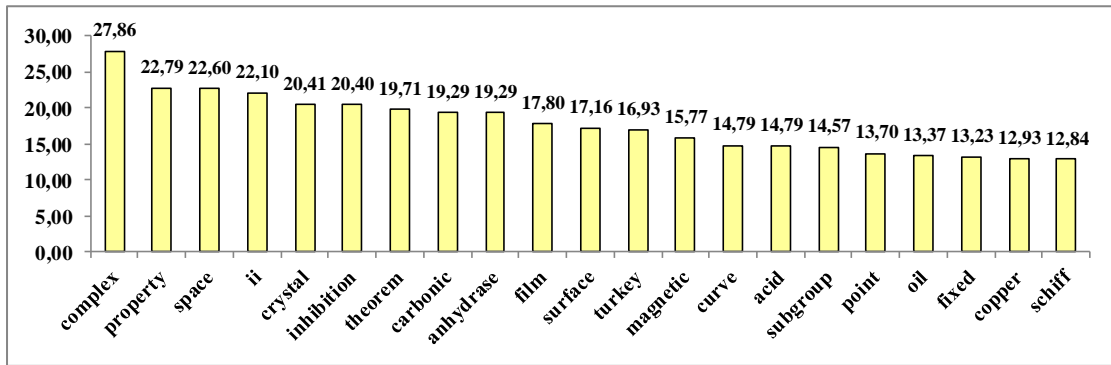


Grafik 4 Tüm veri seti n-gram üç TF-IDF ağırlıkları

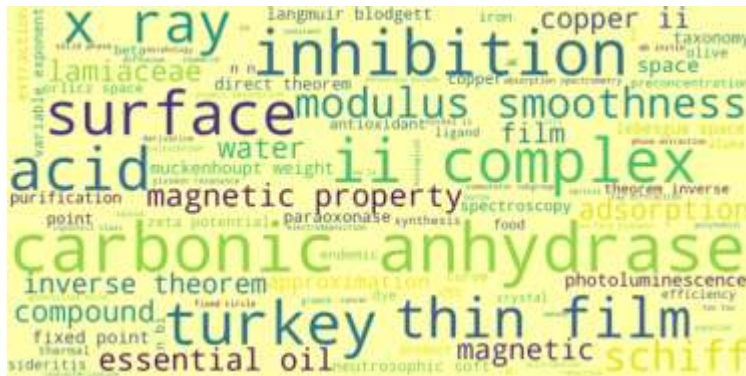
Tüm veri seti için uygulanan analiz yöntemleri WoS kategorileri temel alınarak oluşturulan veri gruplarına da ayrı ayrı uygulanmış ve sonuçları aşağıda sunulmuştur.

Temel Bilimler Veri Seti

Veri setinde yer alan 1366 adet yayın WoS kategorisine göre temel bilimler alanında yapılmıştır. TF-IDF yöntemiyle yapılan analiz sonucunda bu veri setini en iyi temsil eden 20 kelime ve bunların ağırlık değerleri Grafik 5’de gösterilmiştir. Temel bilimler veri setine ait kelime bulutuna Şekil 3’de yer verilmiştir.

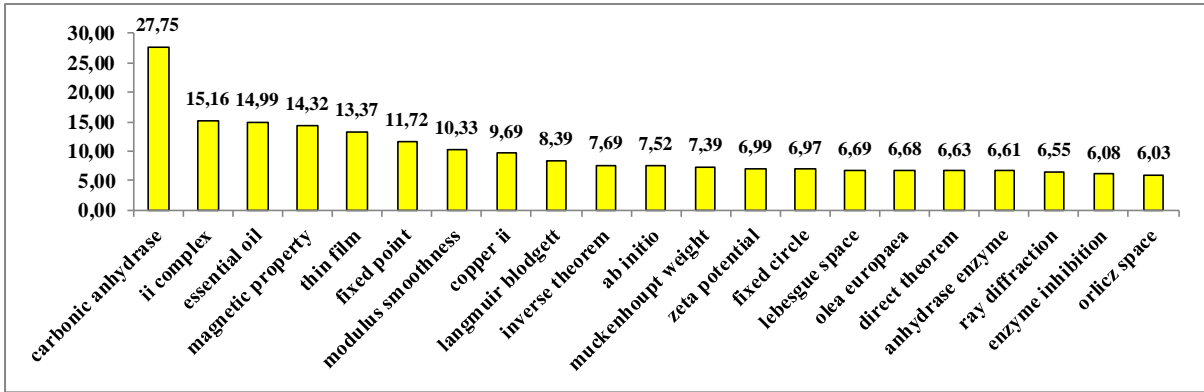


Grafik 5 Temel bilimler veri seti n-gram bir terim ağırlıkları



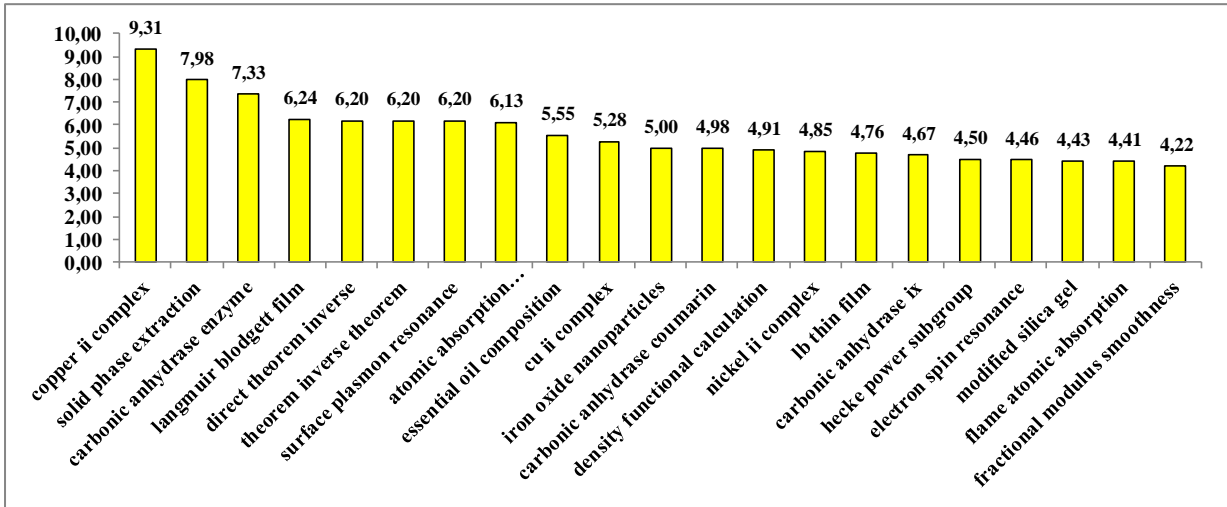
Şekil 3 Temel bilimler veri seti kelime bulutu

N-gram parametresinin iki seçilmesi durumunda veri setini en iyi temsil eden 20 kelime grubu ve bunların ağırlık değerleri Grafik 6'de gösterilmiştir.



Grafik 6 Temel bilimler veri seti n-gram iki terim ağırlıkları

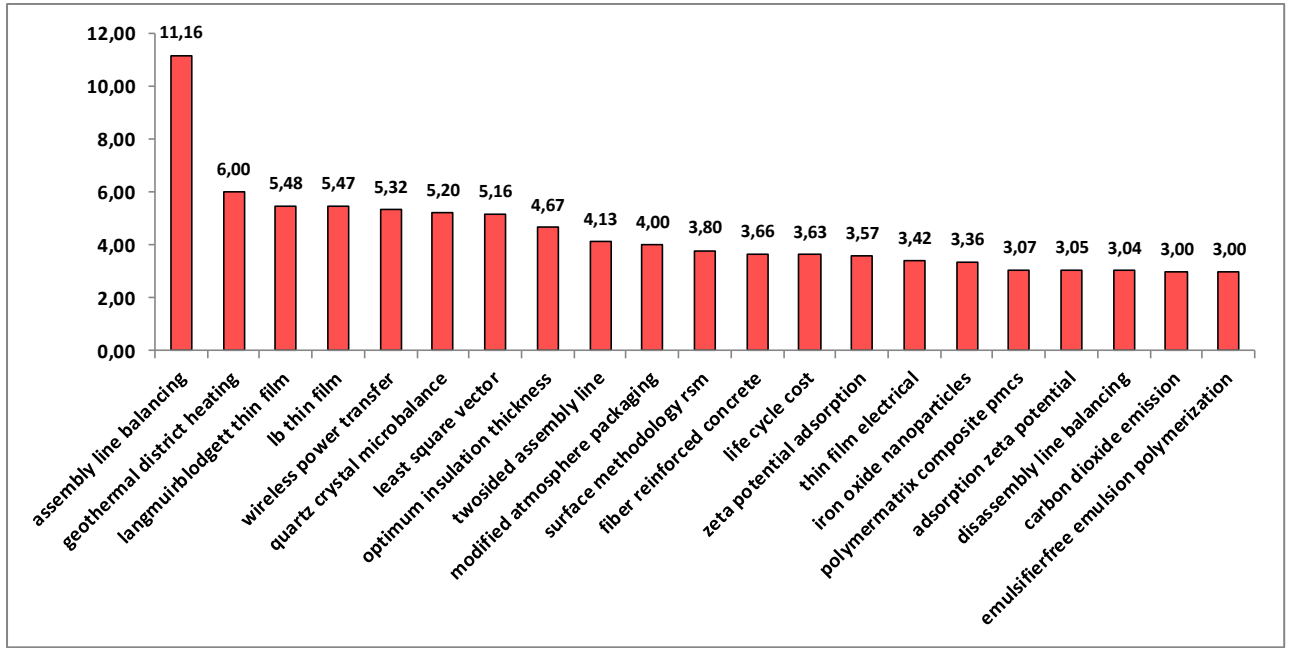
N-gram parametresinin iki seçilmesi durumunda veri setini en iyi temsil eden 20 kelime grubu ve bunların ağırlık değerleri Grafik 7'da gösterilmiştir.



Grafik 7 Temel bilimler veri seti n-gram üç terim ağırlıkları

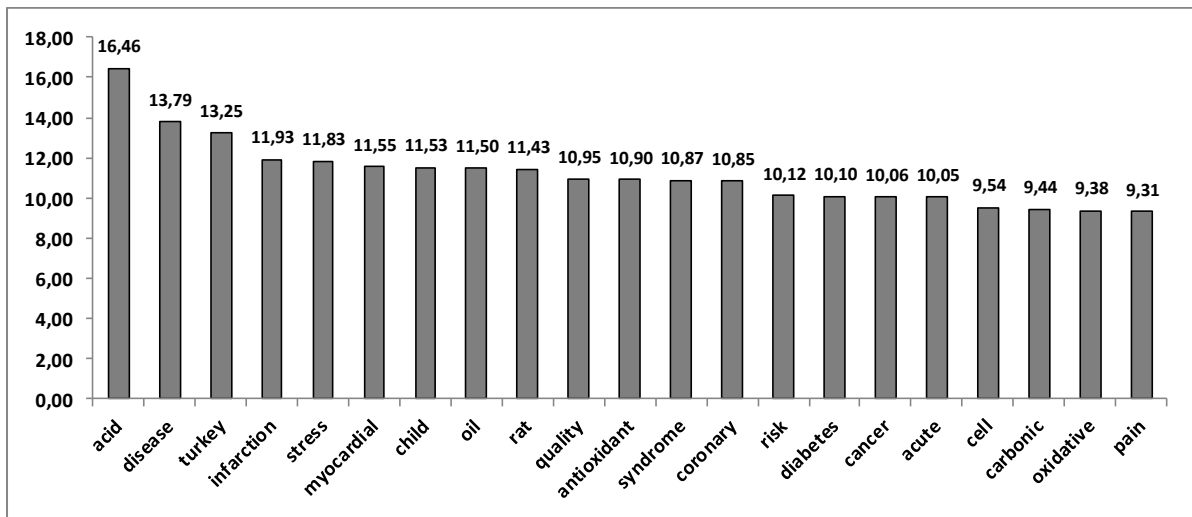
Mühendislik ve Mimarlık Veri Seti

Veri setinde yer alan 1048 adet yayın Wos kategorisine göre mühendislik ve mimarlık bilimleri alanında yapılmıştır. TF-IDF yöntemiyle yapılan analiz sonucunda bu veri setini en iyi temsil eden 20 kelime ve bunların ağırlık değerleri Grafik 8'de gösterilmiştir. Mühendislik ve mimarlık veri setine ait kelime bulutuna Şekil 4' de yer verilmiştir. N-gram parametresinin iki seçilmesi durumunda veri setini en iyi temsil eden 20 kelime grubu ve bunların ağırlık değerleri Grafik 9'da üçlü seçilme durumundaki bilgilere ise Grafik 10' de yer verilmiştir.



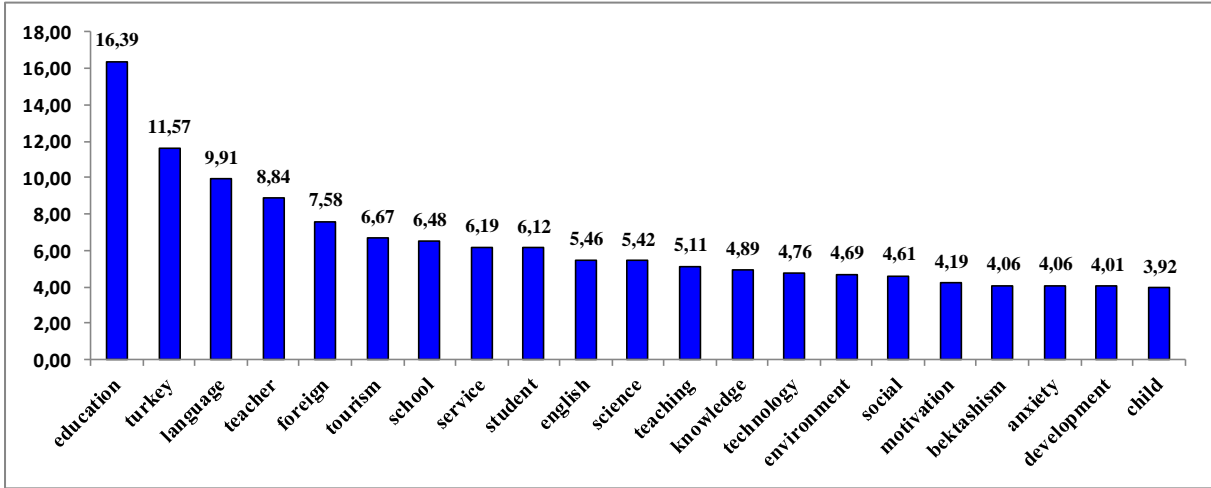
Grafik 10 Mühendislik ve mimarlık veri seti n-gram üç terim ağırlıkları
Tıp ve Sağlık Bilimleri Veri Seti

Veri setinde yer alan 1109 adet yayın Wos kategorisine göre tıp ve sağlık bilimleri alanında yapılmıştır. TF-IDF yöntemiyle yapılan analiz sonucunda bu veri setini en iyi temsil eden 20 kelime ve bunların ağırlık değerleri Grafik 11’de gösterilmiştir. Tıp ve sağlık bilimleri veri setine ait kelime bulutuna Şekil 5’ de yer verilmiştir. N-gram parametresinin iki seçilmesi durumunda veri setini en iyi temsil eden 20 kelime grubu ve bunların ağırlık değerleri Grafik 12’da üçlü seçilme durumundaki bilgilere ise Grafik 13’ de yer verilmiştir.

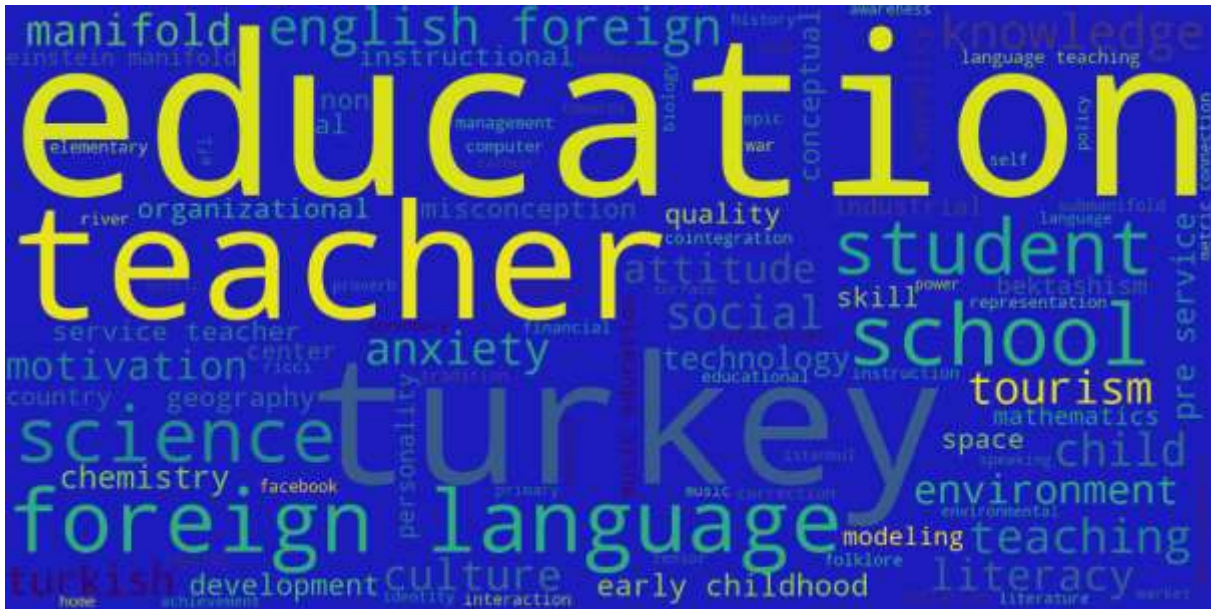


Grafik 11 Tıp ve sağlık bilimleri veri seti n-gram bir terim ağırlıkları

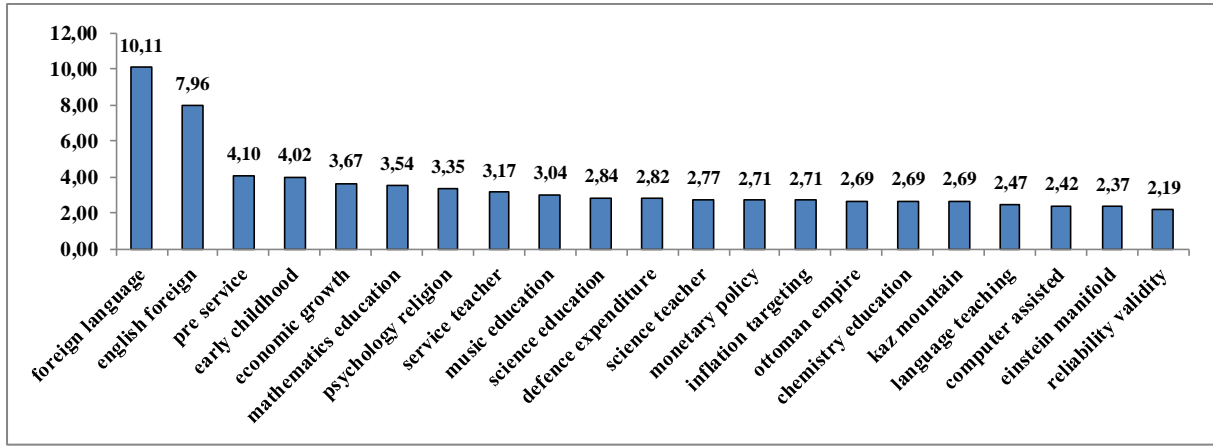
Veri setinde yer alan 354 adet yayın Wos kategorisine sosyal bilimler alanında yapılmıştır. TF-IDF yöntemiyle yapılan analiz sonucunda bu veri setini en iyi temsil eden 20 kelime ve bunların ağırlık değerleri Grafik 14’ de gösterilmiştir. Sosyal bilimleri veri setine ait kelime bulutuna Şekil 6’ da yer verilmiştir. N-gram parametresinin iki seçilmesi durumunda veri setini en iyi temsil eden 20 kelime grubu ve bunların ağırlık değerleri Grafik 15’da üçlü seçilme durumundaki bilgilere ise Grafik 16’ da yer verilmiştir.



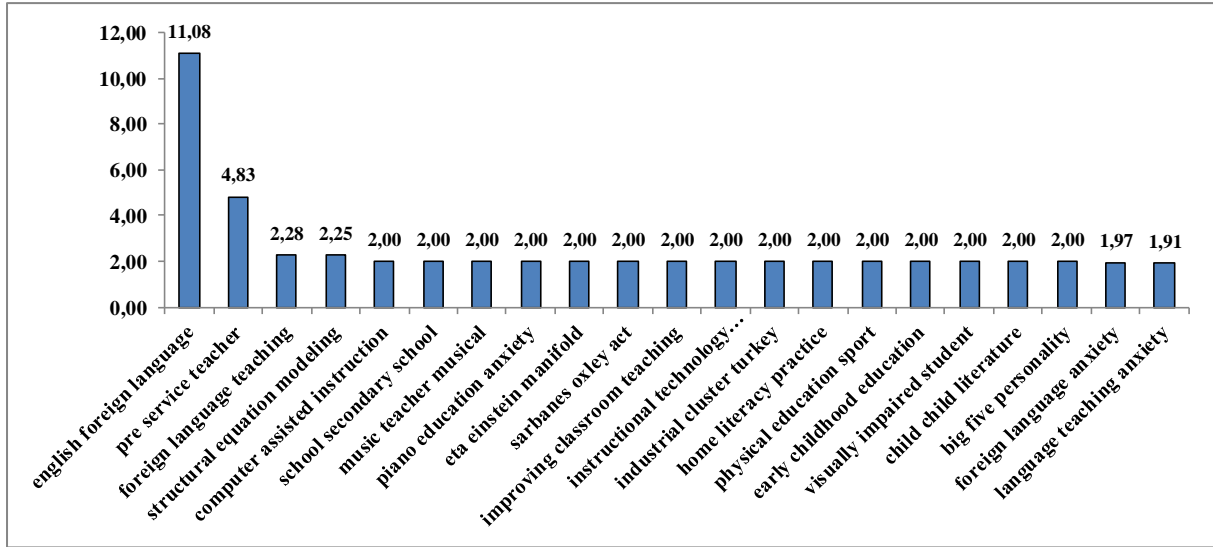
Grafik 14 Sosyal bilimler veri seti n-gram bir için 20 terim ve ağırlıkları



Şekil 6 Sosyal bilimler veri seti kelime bulutu



Grafik 15 Sosyal bilimler veri seti n-gram iki için 20 terim ve ağırlıkları



Grafik 16 Sosyal bilimler veri seti n-gram üç için 20 terim ve ağırlıkları

SONUÇ VE ÖNERİLER

Araştırmada 2015- 2020 yılları arasında Web of Science bibliyometrik veritabanında indekslenen Balıkesir Üniversitesi'nin üretmiş olduğu yayınların bibliyometrik analizi gerçekleştirilmiştir. Bu çalışmayla kurumun hangi alanlarda daha aktif bir bilimsel iletişimi olduğu, yayınların gelişim düzeyleri, etkinlikleri ortaya çıkarılmaya çalışılmıştır. Elde edilen bibliyometrik veriler Balıkesir Üniversitesi'nin bilimsel bilginin gelişimine katkısını ortaya koymakta, araştırmacıların verimliliğini sınamakta, güçlü ve zayıf yönlerini belirlemektedir.

2015- 2020 yılları arasındaki Balıkesir Üniversitesi yayın sayısına bakıldığında ve kendi ölçeğindeki diğer üniversitelerle kıyaslandığında istenilen seviyede olmadığı görülmektedir. 1992 yılında kurulan 32 üniversitenin WoS üzerinde indekslenen yayın sıralamasında Balıkesir Üniversitesi son sıralarda yer almaktadır. İlgili yıllar arasında Kocaeli Üniversitesi'nin 19.247, Isparta Süleyman De-

mirel Üniversitesi'nin 10,841, Pamukkale Üniversitesi'nin 8,689, Aydın Adnan Menderes Üniversitesi'nin 7386, Çanakkale On Sekiz Mart Üniversitesi'nin 7,097, Manisa Celal Bayar Üniversitesi'nin 6,945 yayınının WoS üzerinde yer aldığı tespit edilmiştir. 2017 yılında yayın sayısının düşüş sebebi Balıkesir Üniversitesi'nden bazı akademisyenlerin suç unsurlarına karışması doğrultusunda ihraç edilmesi olduğu düşünülmektedir. 2018 yılına gelindiğinde yayın sayısında yukarı doğru bir ivmenin tekrar yakalandığı görülmektedir.

Web of Science'da yer alan Balıkesir Üniversitesi araştırmalarının büyük çoğunluğunun 1366 yayın (%35) ile temel bilimler ve 1109 yayın (%29) ile tıp ve sağlık bilimlerinde olduğu görülmektedir. Tıp ve sağlık bilimlerinde yayın sayısının diğer alanlara göre daha fazla olmasının en önemli nedeni olarak, bu alanda görevli akademisyen sayısının diğer alanlara göre fazla olması ve Web of Science tarafından indekslenen dergi sayısının bu alanda fazla olmasından kaynaklandığı düşünülmektedir. Toplam yayın sayısı içerisinde sosyal ve beşeri bilimler alanında yayınların az olduğu görülmektedir. Bu durumun en önemli nedeni ise Web of Science tarafından bu alanda indekslenen dergi sayısı fen ve sağlık bilimlerine göre daha az olmasından kaynaklanmakta olup, sosyal ve beşeri bilimlerde yayın yapan akademisyenlerin daha çok ulusal indekslerde yayınlanan dergilere yöneldiğini düşündürmektedir.

Web of Science verileri incelendiğinde toplamda 4176 yayından 3034'ü için çeşitli kurumlardan destek alındığı saptanmıştır. En fazla yayın desteği alınan kurumun TÜBİTAK olduğu tespit edilmiştir. Bilimsel üretimin her geçen artıyor olması, üretilen yayın sayısının daha da yukarıya taşınması ve toplumsal olarak karşılığının alınabilmesi amacıyla ülkemizde yayın teşvikleri verilmektedir. Ancak son yıllarda bu durumun kötüye kullanımının önüne geçilebilmesi ve kaliteli yayınların üretilmesi daha da önemli hale gelmiştir. Bu noktada atıf sayıları temel alınırken, uluslararası indekslerce taranan dergilerdeki yayınların çeşitli kurum ve vakıflarca fonlanması da söz konusudur. Özellikle parasal desteği sağlanan projelerin bilimsel yayına dönüştürüldüğü görülmektedir.

Makale başına düşen yazar sayısı 1,44 olarak tespit edilirken, en çok katkıda bulunan kurumlar arasında Uludağ Üniversitesi, Selçuk Üniversitesi, Ankara Üniversitesi ve Atatürk Üniversitesi yer almıştır. Birlikte yayın yapılırken kurumlar arası işbirliğinde bölgesel yakınlık ve yayın yapan akademisyenlerin daha önceki görev yaptığı kurumlardaki akademisyenleri tercih ettiği düşünülmektedir.

Balıkesir Üniversitesi arařtırmalarının dađılımlında bakıldıđında ilk sırada dergilerin yer aldıđı görölmektedir. Kongrelerde sunulan bildirilerin daha sonradan makaleye dönüřtürölerek yayımlandıđı görölmektedir.

İlgili yıllar arasında indekslenen yayınların toplam atıf sayısı 53 883 olarak tespit edilmiřtir. En çok yayın yapılan ilk altı dergideki toplam atıf sayısı 1134 (% 2) iken, ilk sırada yer alan FEBS Journal dergisindeki yayınların yalnızca 5 atıf aldıđı görölmüřtür.

Birlikte yayın yapılan ölkelerin dađılımlına bakıldıđında ise, en çok yayının Amerikalı arařtırmacılarla ile yapıldıđı (185 yayın) görölmüřtür. Yine atıf sayıları açasından da Amerika ile ortaklařa yapılan yayınların daha çok atıf aldıđı belirlenmiřtir.

Yayınlarda dile göre dađılımlına bakıldıđında büyük bir çođunluđunun İngilizce olduđu görölmektedir. Günümüzde bilimsel iletiřim açasından İngilizcenin artan önemi arařtırmacıları, yayınlarının daha çok kiřiye ulařabilmesi için kendi dilleri yerine İngilizce yayın yapmaya itmektedir. Bu durum da dođal olarak ana dilde bilimsel yayın üretimini yavařlatmaktadır.

Metin madenciliđi yönteminde öncelikle veri seti oluřturulmuř, veri ön iřleme ařamaları tamamlanmıř, TF-IDF (özelliđ çıkarımı yöntemi olarak) yöntemi kullanılarak veri setindeki dokümanları en iyi temsil eden anahtar kelime ve kelime gruplarına ulařılmıřtır. Böylece Balıkesir üniversitesinin WoS veri tabanında yer alan yayınlarının hangi ana ve alt alanlarda toplandıđı bilgisine ulařılmıřtır. Ulařılan bilgiler üniversitenin SCI ve SCI- Expanded indeksli dergilerde kendisine yer bulabilen güçlü alanları ve bunların hangi alt çalıřma alanlarına yođunlařtıđını göstermesi bakımından önemlidir ve literatüre katkı sađlamaktadır. Bu çalıřma sonuçları; Üniversitelerin uluslararası yayınları ile ilgili yapacakları SWOT analizi, durum analizi gibi analizlerde yol gösterici niteliktedir. Bu konularda verecekleri stratejik kararlarda önemli bir bilgi kaynađı oluřturabilecek sonuçlar içermektedir.

Metin madenciliđi yöntemleri kullanılarak bařlangıçta tüm veri seti analiz edilmiřtir. Analiz sonucunda tüm veri setini en iyi temsil eden kelimelerin bazı örnekleri Grafik 2 ile özetlenmiřtir. Bu grafik temel bilimler veri setine ait Grafik 5, mühendislik mimarlık veri setine ait Grafik 8, tıp ve sađlık bilimleri veri setine ait Grafik 11, sosyal bilimler veri setine ait Grafik 14 ile karřılařtırılmıřtır. Yapılan karřılařtırma sonucu genel veri setini temsil eden kelimelerin temel bilimler veri seti ile daha fazla keřiřtiđi gözlenmiřtir ('complex, property, crystal, inhibition, turkey, film, magnetic, acid vs.). Buradan Balıkesir üniversitesinde

yapılan ve Wos veri tabanından erişilebilen yayınların temel bilimler alanında yoğunlaştığı sonucuna varılmıştır. Ayrıca tüm alanlarda ‘turkey’ kelimesinin dokümanları en iyi temsil eden kelimeler arasında yer alması araştırmaların Türkiye ekseninde dağıldığı sonucunu doğrulamaktadır.

N gram parametresi iki seçilerek dokümanları en iyi temsil eden ikili kelime grupları incelenmiştir. Tüm veri setini en iyi temsil eden ikili kelime grubu ‘carbonic anhydrase’ olduğu ve bu grubu ‘thin film, essential oil, magnetic property vs.’ gruplarının izlediği gözlemlenmiştir. Veri setinde yer alan WoS kategorileri bu kelime gruplarını içermeleri durumuna göre incelendiğinde çalışmaların ‘biyokimya ve moleküler biyoloji, kimya, organik kimya’ gibi alanlarda yoğunlaştığı görülmüştür. Bu veriler ışığında üniversitenin yayınlarının temel bilimler ana başlığında ‘biyokimya ve moleküler biyoloji, kimya, organik kimya’ alt kategorilerinde yoğunlaştığı söylenebilir.

Temel bilimler, mühendislik ve mimarlık bilimleri, tıp ve sağlık bilimleri, sosyal bilimler olarak ayrılan veri setlerine ilişkin inceleme sonuçları ayrı ayrı paylaşılmıştır. Analiz sonucu elde edilen veriler veri setinde yer alan WoS kategorileri ile ilişkilendirilmiştir. Böylece yayınların hangi alt alanlarda yoğunlaştıkları kontrol edilmiştir. Buna göre temel bilimler veri setini (tüm veri setinde elde edilen sonuçlara paralel biçimde) ‘biyokimya ve moleküler biyoloji, kimya, organik kimya’ alanlarındaki anahtar kelimelerin daha iyi temsil ettiği gözlenmiştir.

Mühendislik ve mimarlık veri seti özelinde çalışmaların ‘elektrik elektronik mühendisliği, bilgisayar bilimleri, endüstri mühendisliği’ alanlarında kullanılan kelime grupları ile temsil edildiği görülmüştür.

Sağlık ve veteriner bilimleri veri setinde ise ‘endokrinoloji ve metabolizma, kardiyak ve kardiyovasküler sistemler, biyopsiler, toksikoloji’ alanların kelime ve kelime grupları ile daha iyi temsil edildiği gözlenmiştir.

Soysal bilimler veri seti ‘dilbilim, eğitim ve eğitim araştırmaları’ alanlarının kelime ve kelime grupları ile daha iyi temsil edildiği gözlenmiştir.

KAYNAKÇA

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329–342.

Aydın, Ö. (2020). Mobbing İçerikli Yargı Kararlarının Makine Öğrenmesi Algoritmaları ile Sınıflandırılması. Balıkesir: Balıkesir Üniversitesi Fen Bilimleri Enstitüsü.

Berry, M. W., ve Kogan, J. (2010). *Text Mining: Applications and Theory*.

Ergün, K. (2017). (Pusula Yayıncılık) 09 08, 2021 tarihinde http://kergun.baun.edu.tr/veri_madenciligi_hafta11.pdf adresinden alındı

Feldman, R. ve Sanger, J. (2009). *The Text Mining Handbook*. Cambridge University Press.

Gözen, E. (2020). Rekreasyon Bilim Alanının Geçmişten Günümüze Bibliyometrik Analizi. *Gaziantep University Journal of Social Sciences*, 19(2). <http://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=891e4cca-2114-4e0f-b8e4-ca8fd2db6607%40pdc-v-sessmgr01>

Hicks, D., & Wang, J. (2011). Coverage and overlap of the new social sciences and humanities journal lists. *Journal of the American Society for Information Science and Technology*, 62(2), 284–294. doi:10.1002/asi.21458.

Huanh, C.-H., Yin, J., ve Hou , F. (2021). A Text Similarity Measurement Combining Word Semantic Information with TF-IDF Method. *Chinese journal of computers*, 5(34), 856—864.

Koç, O. (2021). Web of Science Üzerinde İndekslenen Kadına Yönelik Şiddet Araştırmalarına Bibliyometrik Bir Bakış. *İstanbul Üniversitesi Kadın Araştırmaları Dergisi*, 21, 19-35. DOI: 10.26650/iukad.2020.22.002

Kürklü, S. Biyoetik Konusunda Yapılan Araştırmaların Bibliyometrik Analizi; WOS Örneği. *Türkiye Biyoetik Dergisi*, 6 (3), 87-99. https://www.journalagent.com/tjob/pdfs/TJOB_6_3_87_99.pdf

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997–1004. doi:10.1002/asi.20349.

Lohmann, S., Heimerl, F., Bopp, F., Burch, M., ve Ertl, T. (2015). Con-centri Cloud: Word Cloud Visualization for Multiple Text Documents. 2015 19th International Conference on Information Visualisation. Barcelona, Spain.

Maharana, R. K. (2013). Bibliometric analysis of Orissa University of Agricultural Technology's research output as indexed in Scopus in 2008-2012. *Chinese Librarianship: An International Electronic Journal*, 36, 25-34. <http://www.white-clouds.com/iclc/cliej/cl36maharana.pdf>

Maharana, R. K.; Sethi, B. B. (2013). A bibliometric analysis of the research output of Sambalpur University's publication in ISI Web of Science during 2007-11. *Library Philosophy and Practice* (e-journal). Accessed online at: <http://digitalcommons.unl.edu/libphilprac/926>

Mannila, H. (1996). Data mining: Machine learning, statistics, and databases. *Proceedings of 8th International Conference on Scientific and Statistical Data Base Management*. Stockholm, Sweden.

Mongeon, P., Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics* 106, 213–228. <https://doi.org/10.1007/s11192-015-1765-5>

Nandi, A.; & Bandyopadhyay, A. K. (2009). Research contributions in chemistry at the University of Burdwan: An analytical study. *Annals of Library and Information Studies*, 56(3), 141-149.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100. doi:10.1007/s11192-006-0007-2.

Ohniwa, R. L., Hibino, A. ve Takeyasu, K. (2010). Trends in research foci in life science fields over the last 30 years monitored by emerging topics. *Scientometrics*, 85 (1), 111-127.

Porter, A. (2004). Technology futures analysis methods working group, technology futures analysis: Toward integration of the field and new methods. *Technology Forecasting and Social Change*, 71(3), 287–303.

Rajman, M. and Vesely, M. (2004), " From text to knowledge: document processing and visualization: a text mining approach ", in Sirmakessis, D.S. (Ed.), *Text Mining and its Applications* , Studies in Fuzziness and Soft Computing, Springer Berlin Heidelberg, pp. 7 – 24. https://doi.org/10.1007/978-3-540-45219-5_2

Sevukan, R., & Sharma, J. (2008). Bibliometric analysis of research output of biotechnology faculties in some Indian central universities. *DESIDOC Journal of Library & Information Technology*, 28(6), 11. <https://pdfs.semanticscholar.org/25ee/d2f920c76809e6f28a7a84c2a034e8b20fc2.pdf>

Shukla, N.; Merigo, J. M.; Lammers, T. ve Miranda, L. (2019). Half a century of computer methods and programs in biomedicine: A bibliometric analysis from 1970 to 2017. *Computer Methods and Programs in Biomedicine*. 183. <https://doi.org/10.1016/j.cmpb.2019.105075>

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595- 610.

Suk, F. M., Lien, G. S., Yu, T. C. ve Ho, Y. S. (2011). Global trends in *Helicobacter pylori* research from 1991 to 2008 analyzed with the Science Citation Index Expanded. *European Journal of Gastroenterology and Hepatology*, 23 (4), 295-301.

Tao, J., Qiu, D., Yang, F., & Duan, Z. (2020). A bibliometric analysis of human reliability research. *Journal of Cleaner Production*, 121041. <https://doi.org/10.1016/j.jclepro.2020.121041>

Van Raan, A. F. (2014). Advances in bibliometric analysis: research performance assessment and science mapping. *Bibliometrics. Use and abuse in the review of research performance*, 3, 17-28. https://portlandpress.com/DocumentLibrary/Umbrella/Wenner%20Gren/Vol%2087/WG_87_chapter%203.pdf

Vargas-Quesada, B., De-Moya-Anegón, F., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Guerrero-Bote, V. (2008). Evolución de la estructura científica española: ISI web of science 1990-2005. *El Profesional De La Información*, 17(1), 22-37. doi:<http://dx.doi.org/10.3145/epi.2008.ene.03>

Yu, D., Xu, Z. & Wang, X. (2020). Bibliometric analysis of support vector machines research trend: a case study in China. *Int. J. Mach. Learn. & Cyber.* 11, 715–728. <https://doi.org/10.1007/s13042-019-01028-y>