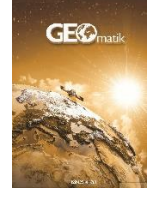




## Geomatik

<https://dergipark.org.tr/tr/pub/geomatik>

e-ISSN 2564-6761



### Metin ve konum bazlı sosyal medya analizleriyle Antarktika'yı haritalama yaklaşımı

Ayşe Giz Gülnerman<sup>\*1</sup>, Fevzican Karakuş<sup>2</sup>, Necip Enes Gengeç<sup>3</sup>, Himmət Karaman<sup>3</sup>, Hasan Hakan Yavaşoğlu<sup>3,4</sup>, Burcu Özsoy<sup>4,5</sup>

<sup>1</sup> Ankara Hacı Bayram Veli Üniversitesi, Tapu Kadastro Yüksekokulu, Tapu Kadastro Bölümü, Ankara, Türkiye, [ayse.gulnerman@hbv.edu.tr](mailto:ayse.gulnerman@hbv.edu.tr)

<sup>2</sup> Hacettepe Üniversitesi, Mühendislik Fakültesi, Geomatik Mühendisliği Bölümü, Ankara, Türkiye, [fevzican.karakuş@hacettepe.edu.tr](mailto:fevzican.karakuş@hacettepe.edu.tr)

<sup>3</sup> İstanbul Teknik Üniversitesi, İnşaat Fakültesi, Geomatik Mühendisliği Bölümü, İstanbul, Türkiye, [gengec@itu.edu.tr](mailto:gengec@itu.edu.tr), [karamanhi@itu.edu.tr](mailto:karamanhi@itu.edu.tr)

<sup>4</sup> TÜBİTAK Marmara Araştırma Merkezi, Kutup araştırmaları Enstitüsü, Türkiye, [yavasoglu@itu.edu.tr](mailto:yavasoglu@itu.edu.tr)

<sup>5</sup> İstanbul Teknik Üniversitesi, Denizcilik Fakültesi, Tuzla, İstanbul, Türkiye, [ozsoybu@gmail.com](mailto:ozsoybu@gmail.com)

Kaynak Göster: Gülnerman, A. G., Karakuş, F., Gengeç, N. E., Karaman, H., Yavaşoğlu, H. H., & Özsoy, B. (2024). Metin ve konum bazlı sosyal medya analizleriyle Antarktika'yı haritalama yaklaşımı. *Geomatik*, 9 (2), 175-184

<https://doi.org/10.29128/geomatik.1417673>

#### Anahtar Kelimeler

Kitle kaynaklı haritalama  
Mekansal bilgi çıkarımı  
Metin bazlı analizler  
Kutup araştırmaları  
Sosyal medya

#### Araştırma Makalesi

Geliş: 10.01.2024  
Revize: 08.02.2024  
Kabul: 13.02.2024  
Yayınlanma: 15.08.2024



#### Öz

Antarktika bulunduğu konum ve iklim özellikleri sebebiyle insanların aktif olarak yaşamadığı bir kıtadır. Bu nedenle Antarktika kıtasındaki coğrafi bilgi üretimi farklı açılardan kısıtlıdır. Geleneksel haritalama teknolojileri olarak kabul edilen uzaktan algılama, fotogrametri ve yersel ölçme yöntemleri ile Antarktika dijital veri tabanına katkı sunulmaktadır. Bu tekniklerin yanı sıra, son 10 yılda yeni haritalama teknolojileri ortaya çıkmıştır. İnsanı bulunduğu çevreyi algılayan bir sensör olarak tanımlayan bu teknolojiler "kitle kaynaklı haritalama" olarak adlandırılmaktadır. Sosyal medya platformları, sunduğu kitle kaynaklı veriler ve üretilen bilgi çıkarım algoritmaları sayesinde bu yeni haritalama teknolojisine katkı vermektedir. Ancak sosyal medya kaynağında üretilen veriler veri miktarı, tamlığı, yanlılığı gibi konularda düzenli veri üretilen teknolojilere göre belirsizlikler içermektedir. Sosyal medyanın düzensiz veri yapısı ve veri kalitesindeki belirsizlikleri, üretilen bilgi çıkarım algoritmalarında değişikliklere neden olmakta ve üretilen sonuçların anlamlılığı da tartışılmaktadır. Bu çalışmada Antarktika kıtası için toplanan sosyal medya verileri üretilen metin bazlı bilgi çıkarım algoritması ve sonuçları tartışılmakta ve ilk defa Antarktika kıtası için metin ve konum bazlı bilgi çıkarım algoritması sunulmaktadır. Aynı zamanda bu çalışmanın önerdiği sosyal medya veri analiz yöntemleri veri azlığı olan, karma doğal dil kullanılan bölgelerde coğrafi yer adlarının çıkarılması ve konuşulan güncel konuların haritalanması için tekrarlanabilir.

### Mapping Antarctica through text and location-based social media analysis

#### Keywords

Crowdsourced mapping  
Geo-information extraction  
Text based analysis  
Polar science  
Social media

#### Research Article

Received: 10.01.2024  
Revised: 08.02.2024  
Accepted: 13.02.2024  
Published: 15.08.2024

#### Abstract

Antarctica is a continent where people do not actively live due to its location and climate characteristics. Consequently, geographic information production in Antarctica is limited from various perspectives. Conventional mapping technologies, such as remote sensing, photogrammetry, and ground surveying methods, contribute to the digital database of Antarctica. In addition to these techniques, new mapping technologies have emerged in the last decade. One of these technologies, defined as "crowdsourced mapping," considers humans as sensors perceiving their environment. Social media platforms contribute to this new mapping technology through the data they provide, and the information inference algorithms generated. However, data produced in social media sources contain uncertainties in terms of data quantity, completeness, and bias compared to regularly generated data in traditional technologies. The unstructured data and uncertainties in data quality of social media lead to changes in information extraction algorithms, and the significance of the generated results is also debated. In this study, social media data collected for the Antarctica continent are discussed in terms of the text-based information extraction algorithm and its results. In this study, a text and location-based information extraction algorithm is presented for the Antarctica continent for the first time. In addition, the social media data analysis methods suggested in this study are repeatable for extracting geographic names and mapping current topics in regions with data scarcity and mixed natural language use.

## 1. Giriş

Günümüzde yeryüzünde gelişen olaylarla ilgili bilgi edinmek için çeşitli veri kaynakları kullanılmaktadır. Bu veri kaynaklarının coğrafi kodla buluşması ile ortaya çıkan verilere, coğrafi veri adı verilmektedir. Son yıllarda geleneksel coğrafi veri elde etme kaynaklarına ek olarak yeni yöntemler ortaya çıkmıştır (Psyllidis, 2020; Anbaroglu, 2017). Bu veri elde etme yöntemleri sensörlerin küresel konumlama sistemleriyle entegrasyonu sonucunda yaygınlaşmıştır (Heipke, 2010). Özellikle insan kaynaklı sistemler olan giyilebilir cihazlar ve akıllı telefonlar aracılığı ile coğrafi veri sıradan insanlar tarafından sağlanabilir hale gelmiştir. Gönüllü coğrafi veri olarak adlandırılan bu kaynaklardan sağlanan verilerin kalitesi birçok açıdan eleştirilmekte olsa da bu kaynaklardan sağlanan verilerin işlenerek coğrafi bilgiye dönüşmesi birçok alanda oldukça önemli katkılar sunmaktadır (Anbaroglu ve ark., 2021; Flanagan ve Metzger, 2008; Hacı ve Gökgöz, 2021; Senaratne ve ark., 2017). Özellikle hızlı, basit ve düşük maliyetli olarak edinilebilecek bu veriler afet acil durumlarının yönetiminde (Gulnerman ve ark., 2021), kentsel mekân algısının ölçülmesinde (Liu ve ark., 2020), insan hareketlerinin değişimi (Calafiore ve ark. 2021) üzerine araştırmalarda ve kentsel sorunların yönetiminde (Taşkanat ve ark., 2018) kullanılmıştır. Sosyal medya platformları milyarlarca kayıtlı kullanıcısı ile dünya çapında çeşitli yapısal olmayan veri sunmaktadır. Bu yapısal olmayan veriler daha çok kentsel alanlar için değerlendirilmiştir (Bilgi ve ark., 2019).

Bu çalışma, mekânsal veri üretiminin zor olduğu Antarktika kıtasına ait alternatif bir veri kaynağı olan sosyal medya kaynağını değerlendirmektedir. Bu değerlendirme özellikle üretilen sosyal medya verilerini metin bazlı analiz yöntemleri ile ele almakta, üretilen metin bazlı analiz adımlarının sonuçları ile Antarktika'yı coğrafi etiketler ile haritalamayı hedeflemektedir. Bu sayede kıtada geçici sürelerde yaşayan veya kıtayı kısa sürelerde ziyaret eden araştırmacıların, turistlerin ve çalışanların kullandıkları ortak ve farklı diller, coğrafi isimler ortaya konulmaktadır. Coğrafi etiketlerin ortaya çıkarılması kıtada gerçekleşen mekânsal aktiviteler hakkında önemli çıkarımlar yapılmasını sağlamaktadır. Özellikle Antarktika kıtasında farklı dillerde üretilen coğrafi isimlendirmelerin keşfedilmesi, zaman içerisinde kıtada gerçekleşen mekânsal aktivite bilgilerinin ortaya koyması bakımından önemlidir. Bu anlamda çalışma, Antarktika kıtasının mekânsal kullanımının uzun soluklu keşi için bir başlangıçtır.

## 2. Yöntem

Sosyal medya verileri üzerine yapılan araştırmalarda metin tabanlı, konum tabanlı, görüntü tabanlı ve ağ tabanlı olmak üzere dört temel analiz yöntemi kullanılmaktadır. Bu yöntemlerden bir ya da birkaçının seçimi bilgi çıkarımı için hedeflenen konuya, sosyal medya platformundan sağlanan verinin içeriğine ve yapısına bağlıdır. Bu çalışmada Antarktika kıtası sınırları içerisinde üretilmiş veriler üzerinden coğrafi yer adlarının çıkarılması ve popüler konuların haritalanması amaçlanmıştır. Bu amaca bağlı olarak 2006 yılından

itibaren hizmet veren Twitter (yeni adıyla X) verileri toplanmış, veri içeriğine bağlı olarak geliştirilen metin ve konum bazlı algoritma araştırması yapılmıştır. Veri elde etme ve verinin düzenlenmesi için belirlenen yöntem adımları 2.1 alt başlığında, verinin metin ve konum bazlı bilgi çıkarım algoritması için izlenen adımlar da 2.2. alt başlığı altında sunulmaktadır.

### 2.1. Veri elde etme ve filtreleme

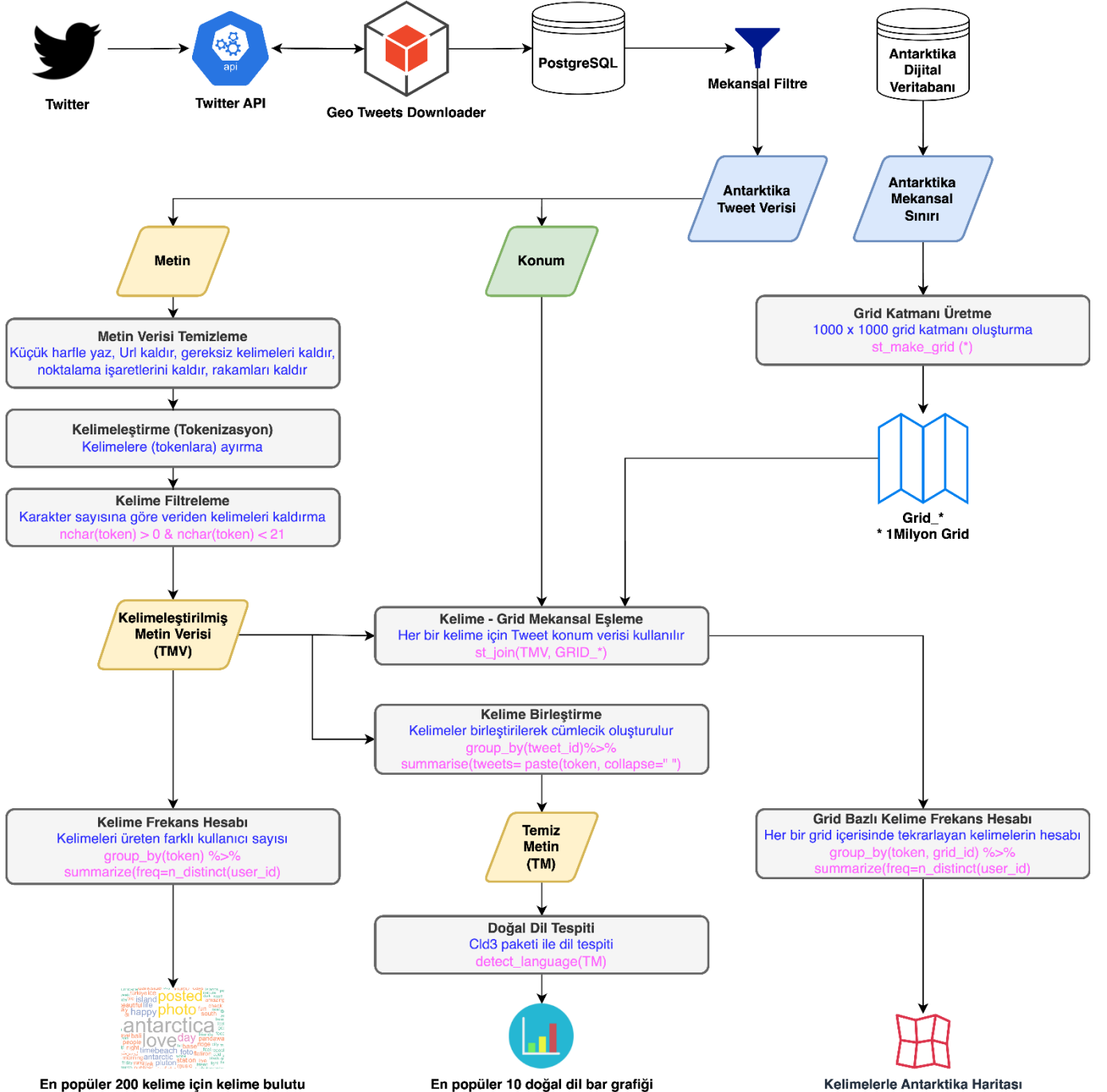
Sosyal medya platformları, kullanıcıları tarafından üretilen verileri çeşitli yöntemlerle sunmaktadır. Bu yöntemlerin en başında Uygulama Programlama Arabirimi (Application Programming Interface- API) gelmektedir. API ile veri sunan popüler sosyal medya platformları arasında Flickr, Twitter, Meta, Instagram yer almaktadır (Lomborg ve Bechmann, 2014; URL-1). Sosyal medya platformlarının sunduğu bu yöntemi kullanabilmek için verisi elde edilecek sosyal medya platformundan API izinlerinin alınmış ve anahtar kodlarının edinilmiş olması gerekmektedir. Bu anahtar kodlar sosyal medya platformları tarafından kullanıcıya (veri istemcisine) özgü olarak tanımlanmakta ve veri elde etme talebine göre sınıflanabilmektedir. Örneğin, Twitter Şubat 2023 tarihinden önce birden çok API türü tesis etmiş (akademik, ticari, sıradan, standart vb.) ve kullanıcıların amaçlarına yönelik API anahtarlarını sağlamıştır. Firma Şubat 2023 tarihinde ücretsiz erişim sağladığı API türlerini kaldıracağını duyurmuş ve şirket politikasındaki değişikliğe gitmiştir (URL-2). Twitter halihazırda daha önceden akademik çalışmalar için sunduğu ücretsiz veri kaynağını kapatmış ve eski döneme ait sağlamış olduğu API anahtarlarının kullanımını durdurmuştur.

API kullanımı için programlama dillerini bilmek, istenilen veriye ulaşım sağlamaktadır; ancak bu verinin istenilen şekilde filtrelenmesi, çekilmesi için programlama bilgisi ve yeteneği gerektirmektedir. Buna ek olarak, hazır masaüstü veya bulut tabanlı uygulamalar (örn: Gephi, Crowdtangle, Netlytic) amaca yönelik veri elde etmeyi kolaylaştırmakta ve bilinen analiz yöntemlerini de barındırarak sosyal medya araştırmacılarına programlama bilgisi olmadan veri üzerinde araştırma yapma imkânı sağlamaktadır (Batinca ve Treleavan, 2015).

Bu çalışmada Şubat 2023 öncesi ücretsiz olarak akademik çalışmalar için sorumlu yazara sunulan Twitter API anahtarı ile veriler toplanmıştır. Twitter API anahtarı ile 2017-2023 yıllarına ait veriler toplanmıştır. Toplanan veriler için Geo Tweet Downloader (Gulnerman ve ark., 2016) masaüstü uygulaması kullanılmıştır. Bu uygulama ile yalnızca konum verisi içeren tweetler toplanmaktadır. Twitter API akademik uygulamalar için sağladığı anahtarlara yalnızca verinin 1%'ine erişim imkânı vermektedir (Cvetojevic ve ark., 2016). Bu 1% veri Twitter'ın rastgelelik algoritmasına göre çalışmakta ve normalde üretilen verilerin çok az bir yüzdesi konum verisi eklenerek üretildiği için, toplanan verilerin çoğunluğu da konum verisi içermemektedir. GeoTweetsDownloader veri toplarken sadece konum bilgisi içeren veri istemi gönderdiği için, araştırılan bölge için üretilen konum bazlı veri elde etme potansiyelini artırmaktadır.

Bu çalışma için elde edilen veriler bir tablo yapısında PostgreSQL veritabanında depolanmıştır. Tablo yapısında yer alan kolonlar; "id", "tweet", "kullanıcı\_adi", "tarih", "enlem", "boylam" olarak tanımlanmıştır. Tweet içerikleri için dil sınırlaması bulunmamaktadır. Tarih kolonu "Gün-Ay-Yıl Saat:Dakika:Saniye" olarak tutulmaktadır. Antarktika Tweet Verisi katmanı oluşturmak için toplanan verilere mekânsal filtre

uygulanmıştır. Mekânsal filtreleme işlemi için Quantarctica (Matsuoka ve ark., 2021) uygulamasından elde edilen Antarktika coğrafi sınır katmanı kullanılmıştır. Mekânsal filtreleme işlemi için gerekli olan mekânsal işlem kabiliyeti PostgreSQL veritabanının PostGIS eklentisi ile sağlanmıştır. Bu sayede mekânsal filtreleme için st\_contain fonksiyonu kullanılarak Antarktika Tweet Verisi katmanı hazırlanmıştır (Şekil 1).



Şekil 1. Veri akışı ve analiz şeması.

## 2.2. Metin ve konum bazlı analiz yöntemi

Metin bazlı analizler trend konuların belirlenmesinde, duygu analizlerinde, seçim tahminleri gibi geniş yelpazeli alanlarda kullanılmaktadır (Wang ve ark., 2017; Chauhan ve ark., 2021). Bu çalışmada ise coğrafi olarak tekrarlanan metinler üzerinden Antarktika coğrafyasının anlaşılması hedeflenmiştir.

Metin verisinden özellik çıkarımı üzerine farklı yaklaşımlar bulunmaktadır. Irfan ve ark. (2015), bu yaklaşımları sözcüksel, söz dizimsel ve anlamsal olarak sınıflandırmıştır. Bu çalışmada kelime frekansına bağlı izlenen yöntem sözcüksel yaklaşım sınıfında değerlendirilmektedir.

Stock (2018), coğrafi alanların sosyal medya analizleri ile tanımlanma yöntemlerini detaylı bir

literatür taraması ile sunmuştur. [Gülnerman ve Karaman \(2020\)](#), coğrafi tekrarlık temelli metin analizini kullanarak İstanbul ve Londra için coğrafi veri sözlüğü üretimi üzerine bir vaka analizi gerçekleştirmiştir. İstanbul için Türkçe, Londra genelinde ise baskın olarak İngilizce kullanılması ve yer adlarının çıkarımı için takip edilen metin desenleri sayesinde anlamlı sonuçların üretilmesini başarmışlardır. Ancak kullanılan metin bazlı desenin Antarktika verilerinde yer almaması ve kentsel alanda referans alınabilen yol ağının Antarktika kıtasında bulunmaması hem metin hem de mekân bazlı analiz yöntemlerinin farklılaşmasını gerektirmektedir. Ayrıca Antarktika’da farklı doğal dilleri kullanan insanların varlığı sebebiyle metin bazlı analizlerde bu durumun da göz önünde bulundurulması gerekliliği ortaya çıkmaktadır.

**Şekil 1**’de veri analiz adımları sunulmaktadır. Veri analizinden önce 2.1.’de anlatıldığı gibi Antarktika tweet veri setinin elde edilmesi için izlenen adımlara şekilde yer verilmiştir. Veri analizi metin verilerinin temizlenmesi ile başlamaktadır. Metin temizleme için izlenen yöntemler kadar bu yöntemlerin uygulanma sıralaması da önemlidir. Bu nedenle özellikle metin bazlı analiz adımlarında sıranın değiştirilmesi ortaya çıkan sonucu da etkilemektedir. Temizleme adımlarının tasarlanması deneysel çalışmalarla gerçekleşmiştir. Bu deneysel çalışma sonuçlarına göre metin verisinin en temiz hali için takip edilen adımlar:

1. Küçük harfle yaz
2. URL kaldır
3. Gereksiz kelimeleri kaldır
4. Noktalama işaretlerini kaldır
5. Rakamları kaldır

Üçüncü adımda bahsi geçen gereksiz kelimeler (stopwords), tek başına anlam içermeyen kelimeler (ör. acaba, sanki, belki) olarak tanımlanmaktadır. Her dil için bu kelimelerin listesi tanımlanmıştır. R programlama dilinde “stopwords” kütüphanesi ([Benoit ve ark., 2021](#)) farklı kaynakları kullanarak 50’den fazla doğal dil için gereksiz kelimeler listesini sunmaktadır. Bu sayede farklı dilleri kapsayıcı gereksiz kelime temizliği yapılabilmesini mümkün kılmaktadır. Antarktika veri setinden birçok doğal dil için (İngilizce, Çince, Arapça, Türkçe, Rusça, İspanyolca vb.) stopwords kütüphanesi kullanılarak gereksiz kelimeler çıkarılmıştır.

Veri temizliğinin ardından, metinlerin kelimelere (tokenlara) ayrılma işlemi gerçekleştirilmiştir. Bu işleme tokenizasyon (parçalara ayırma) denilmektedir ([Çınar, 2023](#)). Bir metin verisini parçalara ayırma işleminde parçalar içerisinde yer alacak kelime sayısı önceden belirlenir. Bu kelime sayısı  $n\_gram$  ile ifade edilir. Eğer bir metin verisi tek kelime ( $n=1$ ) içerecek parçalara bölünüyorsa  $uni\_gram$ , iki kelime ( $n=2$ ) içerecek parçalara bölünüyorsa  $bi\_gram$  tokenizasyonu yapılmıştır. Bu çalışmada her bir tweet metni tek kelime parçalara ayrılmıştır. Bu parçalara ayırma işlemi çalışma boyunca “kelimeleştirme” olarak adlandırılmıştır. Bu işlem sayesinde her bir kelime ayrı olarak ele alınabilmektedir. Ayrıca temizleme sonrası kelimeler arasında yer alacak boşlukların temizlenmesi amaçlanmaktadır.

Kelimeleştirme (Tokenizasyon) işleminin ardından her bir kelime karakter sayısına bakılarak filtreleme işlemi yapılmıştır. Karakter sayısı 0’dan büyük ve 20’den küçük olan kelimeler (tokenlar) veri seti içerisinde bırakılmıştır. Bazı dillerde tek karakter bir anlam ifade etmezken ve bu karakter sayısı filtreleme adımında filtrelenirken, Çince ve Japonca gibi farklı dillerde tek bir karakter bir kelimeyi ifade edebilmektedir. Buna ek olarak kelimelerin karakter sayılarının çok olması da bazı metinlerde yazım hatası, birleşik yazım ve/veya etiket (hashtag) gibi kelimeler olduğu gözlemlenmektedir. Birden çok doğal dilden metinleri içeren veri seti göz önüne alınarak bir kelime için karakter sayısı üst sınırı 20 kabul edilip filtreleme yapılmıştır.

Veri temizleme, kelimeleştirme ve filtreleme aşamalarından sonra, kelime frekansı hesaplanmıştır. Kelime frekansı için kaç farklı Twitter kullanıcısı tarafından paylaşıldığı dikkate alınmıştır. Bu sayede benzer içeriği tekrarlayan bot ya da tekil kullanıcının bilgi çıkarımında yanlılık (bias) yaratmasının önüne geçilmesi hedeflenmiştir. Yanlılık birçok çeşidi (mekân, zaman, katılım vb.) olan bir araştırma konusudur ([Basiri ve ark. 2019](#)). Veriden çıkarılacak analiz sonuçlarının bozulmasına, hedeflenen çıktılara ulaşılmamasına yol açabilir ([Gengeç, 2023](#)). Bu çalışmada bahsedilen yanlılık veriyi üreten tekil kullanıcıların yapabileceği sürekli ve çok sayıdaki paylaşımların bir konumda kullanılan yaygın bir etiket varmışçasına bir çıkarım yapılmasına neden olmaktadır. Buna terminolojide kullanıcı yanlılığı (user bias) denilmektedir ([Tsou ve ark. 2017](#)) ve bu çalışmada kelime frekansı hesabında kullanıcı sayısının dikkate alınması ile bu yanlılık problemi ortadan kalkmaktadır.

Mekânsal kelime frekansı hesabında ilk olarak her bir kelimenin konum verisi eşlenmiştir. Hücre bazlı tekrarlılık hesabı için 1000 x 1000’lik grid katmanı üretilmiştir. Grid üretimi için R programlama dilinde mekânsal analiz fonksiyonlarının yer aldığı sf kütüphanesi ([Pebesma, 2018](#)) kullanılmıştır. Bu kütüphanenin içerisinde yer alan `st_make_grid` fonksiyonu grid üretimi için kullanılacak bir alan tanımı, en ve boy için grid sayısı ile koordinat referans sistemi tanımı istemektedir. Antarktika Dijital Veritabanı (ADD) ([URL-3](#)) ve QGIS Quantarctica eklentisi ([Matsuoka ve ark., 2021](#)) içerisinde yer alan Antarktika veri sınır katmanı alan tanımında, Antarctic Polar Stereographic (EPSG:3031) koordinat referans sistemi kullanılarak, satır ve sütun sayısı 1000 x 1000 olmak üzere 1 milyon hücreye sahip grid katmanı üretilmiştir.

Bu çalışmada deneysel olarak farklı grid boyutları araştırılmıştır. Ancak daha büyük grid boyutları mekânsal tanımlama kabiliyetini düşürmektedir. Daha küçük grid boyutlarında ise anlamlı bir detay çıktısına rastlanmamıştır. Bu nedenle 1000 x 1000’lik grid katmanının üretilmesine karar verilmiştir. Üretilen grid katmanı hücresel kelime frekansının hesaplanması için analiz imkânı sunmaktadır. Frekans hesabı için kelime konum bilgileri ile hücreler mekânsal eşlemeye tabi tutulmaktadır. Hücre bazlı frekans hesabında her bir kelimeyi üreten farklı kullanıcı sayısı dikkate alınmaktadır.



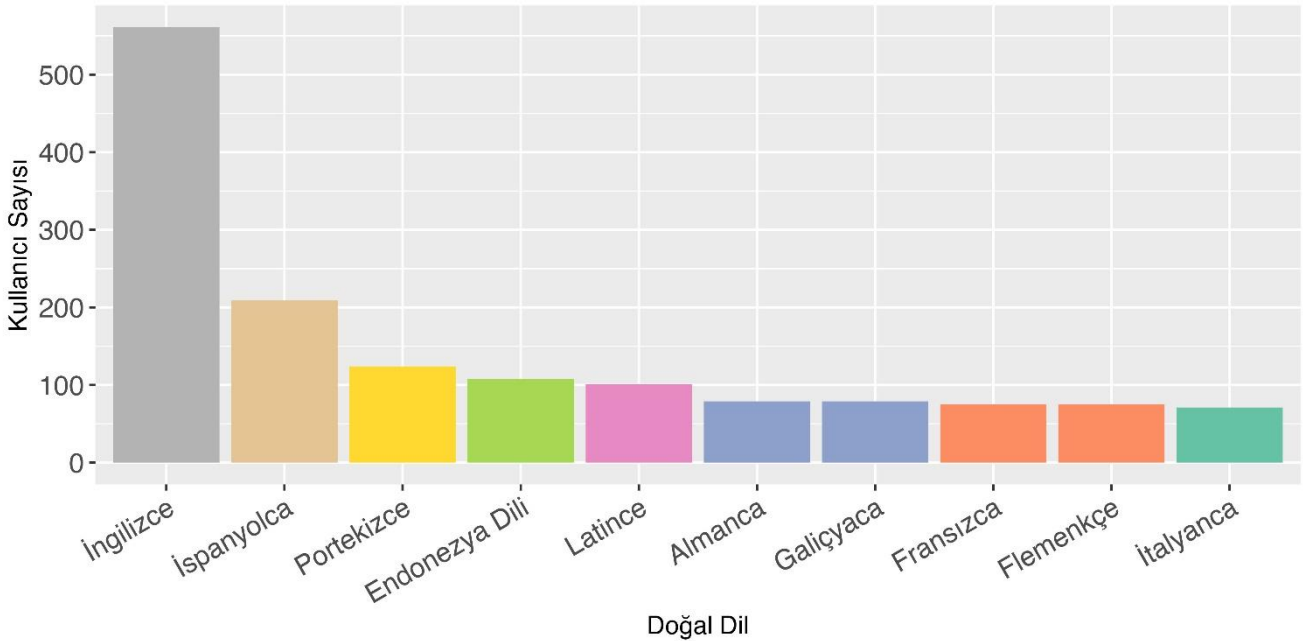
Frekans hesaplarına ek olarak, temizlenmiş kelime verisi tweet id'lerine bağlı kalınıp birleştirilerek cümlecik oluşturulmuştur. Bu cümlecikler üzerinden doğal dil tespit algoritması çalıştırılarak Antarktika sosyal medyasında popüler olarak kullanılan dillerin tespiti hedeflenmiştir. Dil tespiti için R cld3 paketi (Ooms, 2023) kullanılmıştır. Bu paket Google'ın Compact Language Detector kütüphanesini kullanarak geniş dil tanıma desteği sunması ve hızlı olması özellikleri nedeniyle tercih edilmiştir.

### 3. Bulgular

Antarktika Twitter verisi 2.1'de bahsedilen yöntemlerle elde edilmiş, veri tabanına kaydedilmiş ve Antarktika kıta sınırı baz alınarak filtrelenmiştir. Elde edilen veri seti yaklaşık 150 bin adet tweet verisi içermektedir. Metin verisi üzerinde gerçekleştirilen veri temizliği aşamalarından sonra geriye yaklaşık 50 bin tweet içeriği kalmıştır. Tweet içeriklerinin birçok farklı doğal dili barındırdığı öngörülerek doğal dil tespit analizi verinin açıklanması ve anlaşılması için gerçekleştirilmiştir. R cld3 paketi ile yapılan bu tespit

işleminde 97 farklı dilin Antarktika sahasında kullanıldığı belirlenmiştir. Ancak 2079 adet tweet içeriği bu paket aracılığı ile tespit edilememiştir. Buna ek olarak belirlenme yapılan cld3 paketinin hata payı da bu çıkarımda göz önünde bulundurulmalıdır. Bu tespit işlemine göre Antarktika sahasında en yaygın kullanılan 10 dil Şekil 2'de kullanıcı sayıları ile birlikte verilmiştir.

Bu sonuçlar kıtada en çok araştırma faaliyeti gösteren ülkelerin veya kıtayı ziyaret eden turistlerin kullandığı dilleri göstermektedir (URL-4; URL-5). İngilizce kıtada aktif olarak faaliyet gösteren Amerika Birleşik Devletleri, İngiltere, Yeni Zelanda, Avustralya gibi ülkeler tarafından, İspanyolca Arjantin, İspanya, Şili, Uruguay gibi ülkeler tarafından yaygın olarak kullanılmaktadır. Kıtaya komşu ülkelerden Brezilya'nın kullandığı dil Portekizce ilk 10'da yer almaktadır. Çin kıtada aktif faaliyet göstermesine rağmen Çince en çok kullanılan diller arasında görülmemektedir. Bu durum Çin halkının Twitter uygulaması yerine milli sosyal medya platformları olan WeChat uygulamasını kullanmalarıyla açıklanabilir. Bu bulgu Twitter verisinde katılım yanlılığı olduğunun bir göstergesidir.



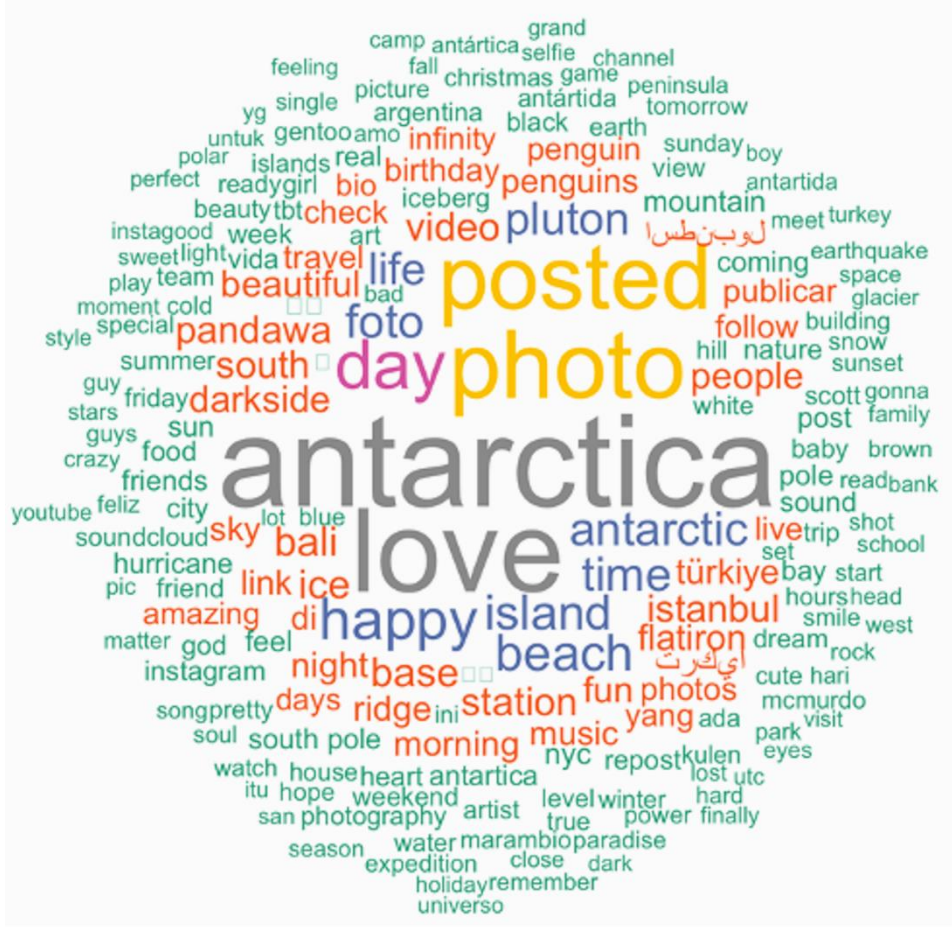
Şekil 2. Antarktika'da en çok kullanılan 10 doğal dil.

Antarktika sahasında en sık kullanılan, tekrarlanan kelimeler Şekil 3'de kelime bulutu aracılığı ile sunulmuştur. Bu kelime bulutuna göre en sık kullanılan dilin yine İngilizce olduğu, bazı kelimelerin farklı dillerde tekrarlandığı (antarctica, antartida, antartica vb.) görülmektedir. Buna ek olarak bazı yer isimlerinin ("antarctica", "argentina", "türkiye") tekrarlandığı ve sık kullanılan kelimeler arasında girdiği söylenebilir. Kelime anlamlarına bakarak tweet içeriklerinde fotoğraf paylaşımının yaygın olduğunu, canlılara (penguin, gentoo), araştırma istasyonlarına (base, station) ve popüler bazı yer adlarına (mcmurdo, peninsula, scott) yer verildiği anlaşılmaktadır.

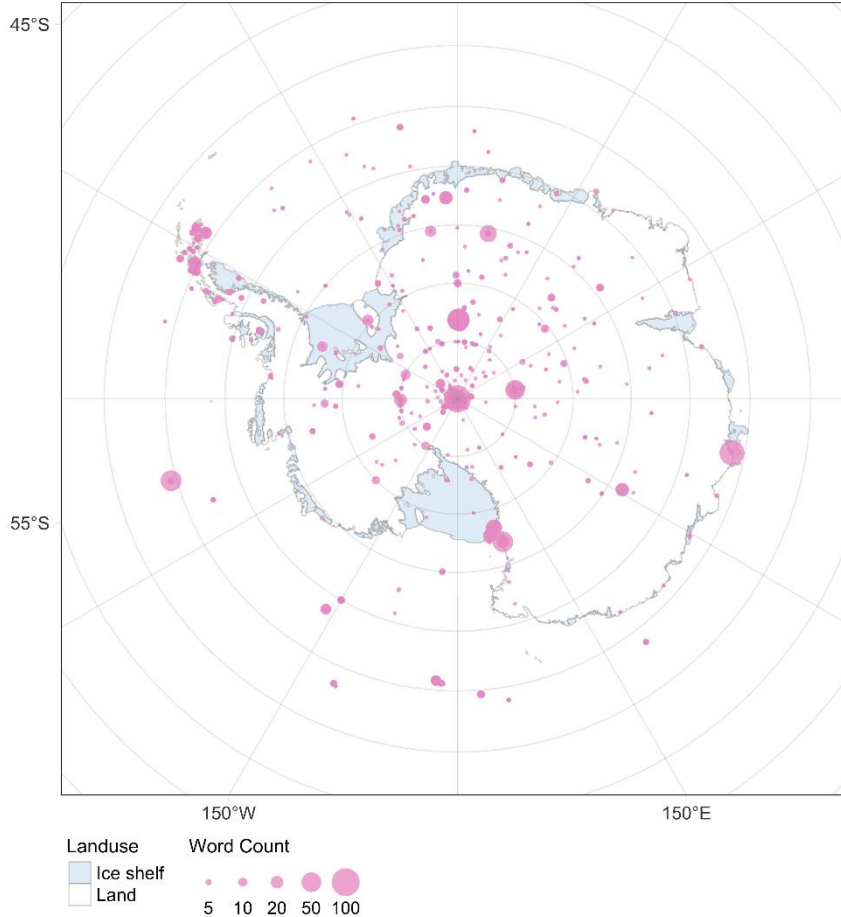
Kelimelerin mekânsal tekrarının tespiti için Antarktika'yı kaplayan ve 1 milyon hücreyi içeren grid

katmanı üretilmiştir. Böylece, mekânsal olarak farklı kullanıcılar tarafından üretilen kelimeler ortaya çıkarılmıştır. Bu grid bazlı mekânsal tekrarlılık araştırması sonucunda 2100 kelime tespit edilmiştir. Şekil 4 'de tekrarlanan kelimeler temsili kabarcık haritası ile gösterilmiştir. Her bir balon bir kelimeyi, balonun büyüklüğü ise kelimenin kaç farklı kullanıcı tarafından paylaşıldığını temsil etmektedir. Kullanıcı sayısı arttıkça balon büyüklüğü artmaktadır.

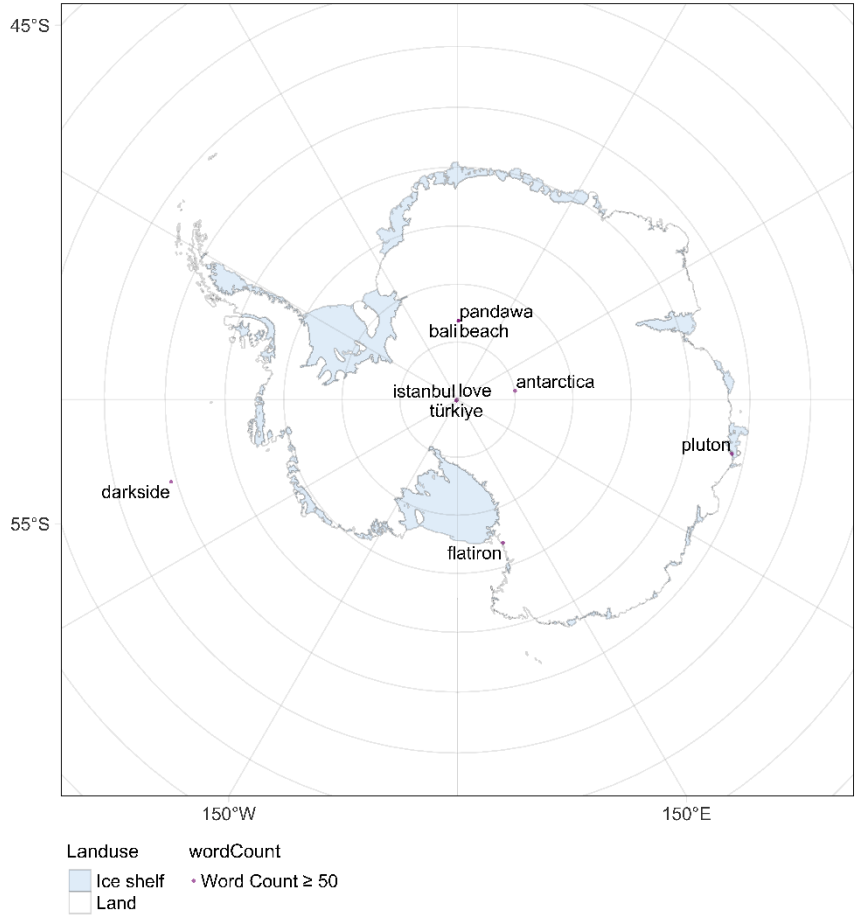
Her bir kelime etiketlenerek haritalandırıldığında Antarktika'daki mekânları tanımlamaktadır (Şekil 5). Bu kelimelere harita üzerinde bakarak Antarktika'daki mekanların kullanımlarıyla ilgili bilgi çıkarmak mümkündür.



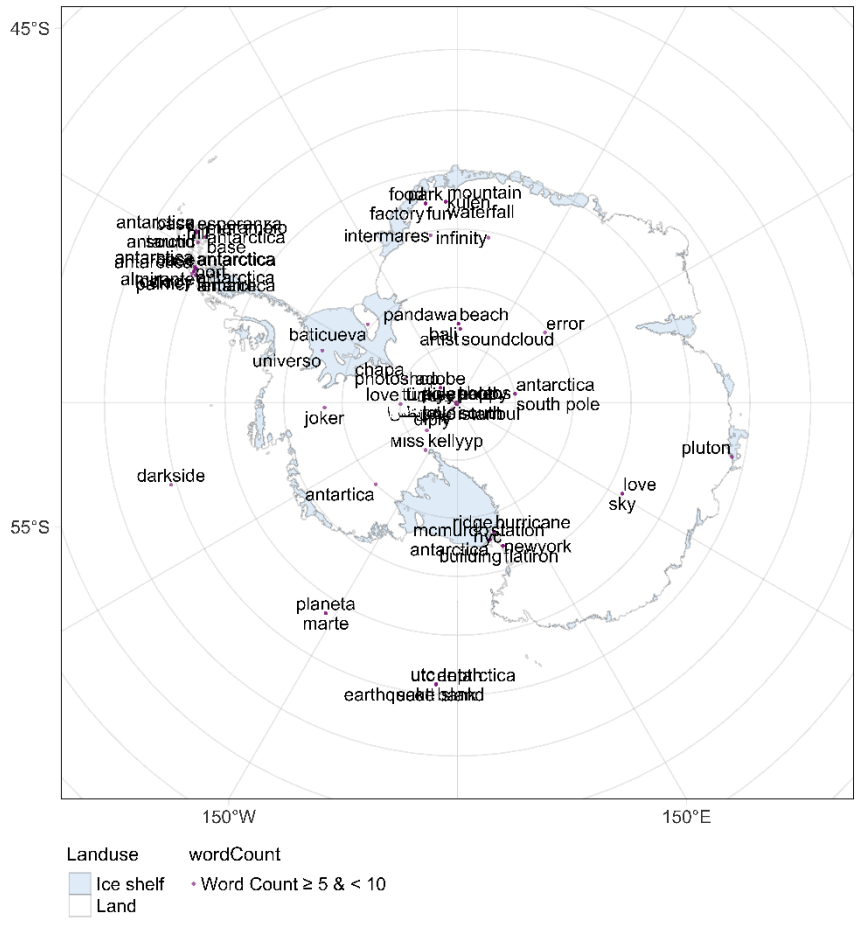
Şekil 3. Antarktika sınırı içerisinde en çok paylaşılan 200 kelime.



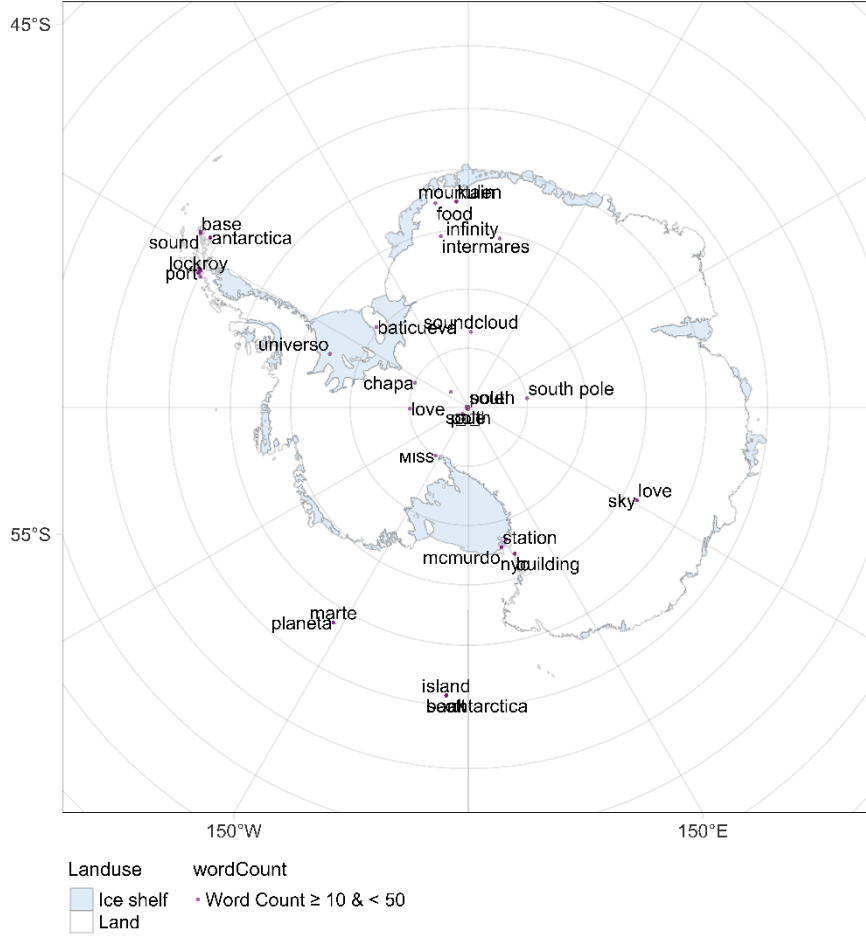
Şekil 4. Antarktika sınırı içerisinde kullanıcı sayısına göre kelimelerin mekânsal dağılımları.



(a)



(b)



(c)

Şekil 5. Antarktika'nın en popüler kelimeleri (a) grup 1, (b) grup 2, (c) grup 3.

Şekil 5 (a) 'da en çok tekrarlanan 10 kelime yer almaktadır. Bu kelimeler arasında Antarktika'da yer alan Flatiron vadisi ve Pluton buzulunun adı görülmektedir. Ancak Antarktika'da yer almayan popüler olarak ziyaret edilen yer adlarını (örn: İstanbul, Türkiye, Bali, Pandawa) da görüyoruz. Okyanus üzerinde 'darkside' (karanlık taraf) kelimesinin 56 farklı kullanıcı tarafından tekrarlandığı görülmektedir. Bu kelime ADD yer isimleri listesinde yer almamaktadır. Bu kelime Antarktika dışında kalan alanlar için de popüler bir yer adı değildir. Bu nedenle, bu coğrafyada bulunan kişilerin haritadaki konumu 'darkside' olarak tanımladığı düşünülmektedir. Tüm harita etiketlerinin haritalanması mümkün olmadığı için Şekil 5 (b) ve (c)'de harita görselleştirme için kullanılan R tmap kütüphanesinin (Tennekes, 2018) üst üste gelen etiketleri sil özelliği kullanılmıştır. Görsellerdeki okunabilirliği arttırmak için tüm kelimelere yer verilememiştir. Şekil 5 (b)'de en az 10 ve en çok 50 kere tekrarlanan kelimeler yer almaktadır. Bu gruba 62 kelime girmektedir. Bu grupta south ve pole kelimelerinin Antarktika merkezine yakın noktalarda farklı gridlerde tekrarlandığı gözlenmektedir. Yine bu grupta sıklıkla anılan bilinen araştırma üsleri (port locroy (Rusya), mcmurdo (ABD)) ve yer isimlerinin (sound canal, kulen dağları) ortaya çıkarılabildiği görülmektedir. Şekil 5 (c)'de ise 10'dan az 5'e eşit veya çok sayıda bir grid içerisinde tekrarlanan kelimelerin mekânsal dağılımı verilmektedir. Özellikle Antarktik Yarımadası olarak tabir edilen, Antarktika'ya giriş

kapılarından biri olan Şili'ye en yakın kısmında Antarktika kelimesinin sıklıkla ancak başka gridlerde tekrarlandığı görülmektedir. Bu grupta Şekil 5 (b)'de de görülen bazı kelimelerin (plenata marte, baticueva, universo, intermares vb.) yakın konumlarda tekrarlandığını görmek mümkündür. Bu kelimelerin ADD yer adları listesinde yer almamaktadır. Bu kelimelerin farklı dillerde olması ama benzer bölgelerde tekrarının sıklıkla görülmesi nedeniyle farklı dillerde bölgelerin farklı isimlerle tanımlanabildiği çıkarımı yapılmaktadır.

#### 4. Sonuçlar

Bu çalışmada Twitter verileri metin ve konum tabanlı olarak analiz edilmiş ve bu analizler yardımıyla verilerin Antarktika'daki araştırmalar için kullanılabilirliği incelenmiştir. Elde edilen analiz sonuçlarına göre Antarktika'da yapılan paylaşımlar için kullanılan diller arasından İngilizce baskın dil olsa da yapılan paylaşımlarda kullanılan dilin diğer coğrafi bölgelerde beklenilene göre daha fazla çeşitlilik içerdiği gözlemlenmiştir. Bu sonuç, kıtadaki araştırmacıların ve diğer ziyaretçilerin çeşitliliği hakkında bilgi vermektedir. Bu çeşitliliğin bir sonucu olarak yapılan analizlerin farklı dillerin karakterleri de dikkate alınarak özelleştirilmesi gerekmektedir. Örneğin kelime filtreleme için belirlenen kelime uzunluk sınırları için belirlenen alt sınırın Çince



ve Japonca gibi dillerdeki tek harften oluşan kelimeler için uygun olmadığı gözlemlenmiştir.

Coğrafi etiketlerin sosyal medya metinlerden çıkarılması için yapılan literatürdeki Antarktika kıtası dışında yapılan araştırmalarda kullanılan yol ağından türetilen Voronoi gibi referans katmanlar bu çalışmada bahsedilen verilerin Antarktika için bulunmaması nedeniyle kullanılamamıştır. Bunun yerine benzeri bir coğrafi referans oluşturmak ve verileri konumsal olarak özetleyebilmek için gridlerden yararlanılmıştır.

Analizlerde, sıkça kullanılan kelimelerin coğrafi dağılımları incelenmiştir. Buna göre paylaşım yapılan konumların sıklıkları dikkate alındığında paylaşımların kıta üzerindeki yerleşim ve insan aktivitesinin yoğunlaştığı yerleri işaret ettiği gözlemlenmektedir. Bu ilişkiye dayanarak ilgili verinin bilinen insan aktivitesini tutarlı bir şekilde temsil ettiği ifade edilebilir. Paylaşılan kelimeler ve konuma göre paylaşım sıklıkları irdelendiğinde de benzer sonuçlar elde edilmiştir. Buna göre sık paylaşılan yer adı, tür adı gibi kelimelerin kıta üzerinde buldukları/rastlandıkları yerler ile ilişkili sonuçlar elde edilmiştir. Bu tutarlılığa ek olarak Bölüm 3'te değinilen 'darkside' örneğinde olduğu gibi semantik tutarlılığa sahip ancak daha önce coğrafi ya da metin kaynaklarda belgelenmemiş genel kabul gördüğü değerlendirilen yer adları ile karşılaşmıştır. Bu ve benzeri örneklerin artırılması Twitter ve benzeri şekilde analiz edilebilecek diğer sosyal medya kaynaklarının konum verisi kaynaklarındaki eksiklikleri tamamlayıcı bir kaynak olarak kullanılabilirliğini arttıracaktır.

Antarktika kıtasında gerçekleşen aktivitelerin haritalanmasının hedeflendiği bu çalışmada elde edilen sonuçlar sınırlıdır. Bunun temel iki sebebi bulunmaktadır. İlki Antarktika kıtasındaki internet altyapısının kapsayıcı olmamasıdır. İkincisi ise kıtada seyir halinde olan araştırmacı ve turistlerin arazi keşifleri sırasında kullandıkları ve veri topladıkları cihazların bataryalarının soğuk sebebiyle hızla tükenmesi veya kısıtlı olmasıdır. Bu sebepler Antarktika kıtasında sosyal medya üzerindeki veri üretiminin anlık olması önündeki engellerdir. Bu sebeple Antarktika kıtasına ait sosyal medya araştırmalarında uzun süreli takip ve incelemeye ihtiyaç duyulmaktadır.

Bu çalışma, coğrafi etiketleme sonrasında coğrafi kodlama ve metin tabanlı konum belirleme algoritmaları açısından geliştirilebilir potansiyele sahiptir. Coğrafi kodlama, bölgenin otomasyon sistemleri ve gönüllü coğrafi verilerle analizi ile haritalanmasında önemli bir araç olabilir. Özellikle, az veriye sahip ve izole bir bölge olarak Antarktika'nın coğrafi kodlama ile haritalanması, doğal afetler, bilimsel araştırmalar ve çevre izleme gibi alanlarda kritik öneme sahip olabilir. Bu bağlamda, genel anlamda makine öğrenmesi algoritmalarının ve büyük dil modellerinin (Large Language Model-LLM) kullanımı, coğrafi kodlama alanında performansı artırabilir. Ayrıca, Antarktika bölgesi için özel olarak bölgeye uygun mekânsal veri tabanlarının entegrasyonu sayesinde daha yüksek performanslı coğrafi kodlama ve bilgi çıkarımları hedeflenebilir.

## Bilgilendirme/Teşekkür

Bu çalışma, TÜBİTAK KUTUP 1001 Programı kapsamlarında yürütülen, "Kutup Bölgelerinde Veri ve Bilgi Arttırımı için Kitle Kaynaklı Coğrafi Veri Elde Etme, Bilgi Çıkarma Algoritmaları ve Uygulaması" başlıklı projesi (Proje No: 122Y193) ile desteklenmiştir. 6 Şubat Depremi sonrası TÜBİTAK BİÇABA Burs Programı katkısı ve destekleri ile bu yayına katkı sunan Fevzican Karakuş projede yer almış ve desteklenmiştir.

## Araştırmacıların katkı oranı

**Ayşe Giz Gülnerman:** Literatür taraması, Uygulama, Makale yazımı; **Fevzican Karakuş:** Literatür taraması, Uygulama; **Necip Enes Gengeç;** Makale yazımı, Düzenleme; **Himmet Karaman:** Makale yazımı, Düzenleme; **Hasan Hakan Yavaşoğlu:** Düzenleme; **Burcu Özsoy:** Düzenleme

## Çatışma Beyanı

Herhangi bir çıkar çatışması bulunmamaktadır.

## Kaynakça

- Anbaroğlu, B. (2017). Gönüllü coğrafi bilgi: Mekansal bilişim çalışmalarına web 2.0 devrinde yeni bir yaklaşım. *Harita Dergisi*, 158, 1-9.
- Anbaroğlu, B., Güllüoğlu, N., Bilgin, G., & Aydınöğlu, A. Ç. (2021). Ulusal Coğrafi Bilgi Sistemi Ulaşım Veri Temasının Katılımcı Coğrafi Bilgi Yaklaşımıyla Güncellenmesi. *Geomatik*, 6(2), 115-123. <https://doi.org/10.29128/geomatik.714493>
- Basiri, A., Haklay, M., Foody, G., & Mooney, P. (2019). Crowdsourced geospatial data quality: Challenges and future directions. *International Journal of Geographical Information Science*, 33(8), 1588-1593. <https://doi.org/10.1080/13658816.2019.1593422>
- Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, 30, 89-116. <https://doi.org/10.1007/s00146-014-0549-4>
- Benoit, K., Muhr, D., & Watanabe, K. (2021). Stopwords: Multilingual stopword lists. R package version, 2.
- Bilgi, S., Gülnerman, A. G., Arslanoğlu, B., Karaman, H., & Öztürk, Ö. (2019). Complexity measures of sports facilities allocation in urban area by metric entropy and public demand compatibility. *International Journal of Engineering and Geosciences*, 4(3), 141-148. <https://doi.org/10.26833/ijeg.540180>
- Calafiore, A., Palmer, G., Comber, S., Arribas-Bel, D., & Singleton, A. (2021). A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Computers, Environment and Urban Systems*, 85, 101539. <https://doi.org/10.1016/j.compenvurbsys.2020.101539>
- Chauhan, P., Sharma, N., & Sikka, G. (2021). The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, 12, 2601-2627. <https://doi.org/10.1007/s12652-020-02423-y>

- Cvetojevic, S., Juhasz, L., & Hochmair, H. (2016). Positional accuracy of twitter and instagram images in urban environments. *GI Forum*, 1, 191-203. [https://doi.org/10.1553/giscience2016\\_01\\_s191](https://doi.org/10.1553/giscience2016_01_s191)
- Çınar, N. (2023). R ile Twitter verisi analizi: Veri toplama, sosyal ağ analizi ve metin analizi aşamaları. *Turkish Online Journal of Design Art and Communication*, 13(1), 193-224. <https://doi.org/10.7456//11301100/014>
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72, 137-148. <https://doi.org/10.1007/s10708-008-9188-y>
- Gengeç, A. G. G. (2023). Exploring crowdsourcing accountability for mapping Antarctica: a case study using 5 years of social media data. *Turkish Journal of Earth Sciences*, 32(8), 1041-1051. <https://doi.org/10.55730/1300-0985.1892>
- Gulnerman, A. G., Gengeç, N. E., & Karaman, H. (2016). Review of public tweets over Turkey within a pre-determined time. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 153-159. <https://doi.org/10.5194/isprs-annals-IV-4-W1-153-2016>
- Gulnerman, A. G., Karaman, H., & Basiri, A. (2021). New age of crisis management with social media. *Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data*, 131-160. [https://doi.org/10.1007/978-3-030-58232-6\\_8](https://doi.org/10.1007/978-3-030-58232-6_8)
- Gülnerman, A. G., & Karaman, H. (2020). Sosyal medyanın gönüllü coğrafi veri olarak kullanımı ve sosyal medya verilerinden coğrafya sözlüğü üretimi. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 20(2), 276-286. <https://doi.org/10.35414/akufemubid.667397>
- Hacar, M., & Gökğöz, T. (2021). A new approach for matching road lines using efficiency rates of similarity measures. *International Journal of Engineering and Geosciences*, 6(3), 146-156. <https://doi.org/10.26833/ijeg.791324>
- Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 550-557. <https://doi.org/10.1016/j.isprsjprs.2010.06.005>
- Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., ... & Li, H. (2015). A survey on text mining in social networks. *The Knowledge Engineering Review*, 30(2), 157-170. <https://doi.org/10.1017/S0269888914000277>
- Liu, Y., Yuan, Y., & Zhang, F. (2020). Mining urban perceptions from social media data. *Journal of Spatial Information Science*, (20), 51-55. <https://doi.org/10.5311/JOSIS.2020.20.665>
- Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society*, 30(4), 256-265. <https://doi.org/10.1080/01972243.2014.915276>
- Matsuoka, K., Skoglund, A., Roth, G., de Pomereu, J., Griffiths, H., Headland, R., ... & Melvær, Y. (2021). Quantarctica, an integrated mapping environment for Antarctica, the Southern Ocean, and sub-Antarctic islands. *Environmental Modelling & Software*, 140, 105015. <https://doi.org/10.1016/j.envsoft.2021.105015>
- Ooms, J. (2023). cld3: Google's Compact Language Detector 3 (Version 1.6.0).
- Pebesma, E. J. (2018). Simple features for R: standardized support for spatial vector data. *R J.*, 10(1), 439.
- Psyllidis, A. (2020). Sensing the city through new forms of urban data. In *Seeing the City: Interdisciplinary Perspectives on the Study of the Urban*, 56-69. Amsterdam University Press.
- Senaratne, H., Mobasher, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139-167. <https://doi.org/10.1080/13658816.2016.1189556>
- Stock, K. (2018). Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*, 71, 209-240. <https://doi.org/10.1016/j.compenvurbsys.2018.05.007>
- Taşkanat, T., Karaağaç, A., Beşdok, E., & Bostancı, B. (2018). Kentsel sorunların yönetimi için bir gönüllü coğrafi bilgi mobil uygulaması geliştirilmesi. *Geomatik*, 3(1), 84-91. <https://doi.org/10.29128/geomatik.371144>
- Tennekes, M. (2018). tmap: Thematic Maps in R. *Journal of Statistical Software*, 84, 1-39. <https://doi.org/10.18637/jss.v084.i06>
- Tsou, M. H., Zhang, H., & Jung, C. T. (2017). Identifying data noises, user biases, and system errors in geo-tagged twitter messages (Tweets). *Social and Information Networks*. <https://doi.org/10.48550/arXiv.1712.02433>
- URL-1: <https://medium.com/rakuten-rapidapi/top-10-social-media-apis-twitter-facebook-instagram-and-many-more-5c13262c61fe>
- URL-2: <https://twitter.com/XDevelopers/status/1621026986784337922>
- URL-3: <https://www.add.scar.org/>
- URL-4: <https://www.comnap.aq/antarctic-facilities-information>
- URL-5: <https://iaato.org/information-resources/data-statistics/>
- Wang, L., & Gan, J. Q. (2017). Prediction of the 2017 French election based on Twitter data analysis. *2017 9th Computer Science and Electronic Engineering (CEECE)*, 89-93. <https://doi.org/10.1109/CEECE.2017.8101605>

