



Doğal Dil İşlemede Veri Ön İşleme Tekniklerinin Sosyal Medya Yorumlarında Kullanılarak Şeker Hastalığı Analizi

Sinan ALTUN^{1*}

¹Kahramanmaraş İstiklal Üniversitesi Yapı İşleri ve Teknik Daire Başkanlığı, Kahramanmaraş

¹<https://orcid.org/0000-0002-2356-0460>

*Sorumlu yazar: s.altun@yaani.com.tr

Araştırma Makalesi

Makale Tarihiçesi:

Geliş tarihi: 06.06.2023

Kabul tarihi: 22.10.2023

Online Yayınlanma: 11.03.2024

Anahtar Kelimeler:

Doğal dil işleme

Makine öğrenmesi

Sosyal medya

Şeker hastalığı

ÖZ

Sosyal medya araçları şeker hastaları tarafından yaygın olarak kullanılmaktadır. En yaygın şeker hastalıklarından ikisi olan Tip1 ve Tip2, sosyal medya sitesi Reddit'te iyi bir şekilde temsil edilmektedir. Bu çalışmada R/diabetes alt başlığında yorumları incelemek, yorum yapanların ilgi düzeyini ve şeker hastalığı için yaygın tedavilerle ilgili yorumları izlemek için doğal dil işleme araçlarını kullanılmıştır. 2019-2022 yılları arasında r/diabetes (n=110,483) alt başlığından gelen tüm yorumlar alınarak doğal dil işleme araçları kullanılarak işlendi. 2019-2022 yılları arasında alternatif tedaviler, yaşam tarzı değişiklikleri ile ilgili yorum hacmi dalgalı bir süreç göstermiştir. Çalışma 3 ana başlık altındaki yorumları inceleyecek şekilde yapılmıştır. Şeker hastalığı ile ilgili genel ifadeler (5 terim), tedavide kullanılan ilaçların etkin maddeleri (6 terim), alternatif tedavi yöntemleri ve yaşam stilleri (6 terim) başlıklarıdır. Terimlerin tamamına yakını yıllara göre yorumlarda yer alma sıklığı dalgalı bir seyir göstermekle beraber “bicycling” ve “Repaglinide” terimleri hariç 2022 yılında artış göstermişlerdir. “Hydrotherapy” teriminin yorumlarda yer almasında, bütünde yıllarda düzenli bir artış varken “Bicycling” ifadesine düzenli bir azalış söz konusudur. “Glizlazide” etken madde terimde, 2020 yılından 2021 yılına geçişte yüksek bir artış gözlemlenmiştir. Bu analizler şeker hastalığı tedavileriyle ilgili hasta ilgi düzeyi ve duyarlılığı ile ilgili olarak ek araştırma alanı önermektedir.

Analysis of Diabetes Using Natural Language Processing Methods in Social Media Comments

Research Article

Article History:

Received: 06.06.2023

Accepted: 22.10.2023

Published online: 11.03.2024

Keywords:

Natural language processing

Machine learning

Social media

Diabetes mellitus

ABSTRACT

Diabetics widely use social media tools. Two of the most common diabetes diseases, Type1 and Type2, are well represented on the social media site Reddit. In this study, we used natural language processing tools to analyze comments on the R/diabetes subthread and track commenters' levels of interest and comments about common treatments for diabetes. All comments from the sub-topic r/diabetes (n=110483) between 2019 and 2022 were retrieved and processed using natural language processing tools. Between 2019 and 2022, the volume of comments on alternative treatments and lifestyle changes showed a fluctuating process. The study was conducted to examine comments under 3 main headings. These are general statements about diabetes (5 terms), active ingredients of drugs used in treatment (6 terms), alternative treatment methods, and lifestyle changes (6 terms). Almost all of the terms showed an increase in 2022, except for the terms

"bicycling" and "Repaglinide", although the frequency of occurrence in the comments fluctuated according to the years. The term "Hydrotherapy" in the comments showed a steady increase in all years, while "Bicycling" showed a steady decrease. The active ingredient term "Glizlazide" increased greatly from 2020 to 2021. These analyses suggest additional areas of research about patient interest and responsiveness to diabetes treatments.

To Cite: Altun S. Doğal Dil İşlemede Veri Ön İşleme Tekniklerinin Sosyal Medya Yorumlarında Kullanılarak Şeker Hastalığı Analizi. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 2024; 7(2): 533-542.

1. Giriş

Kronik bir hastalık olan diyabette, vücut insülin üretmez veya ürettiğini uygun şekilde kullanamaz. Şekerin kan dolaşımından hücrelere getirilmesini insülin gerçekleştirir. Vücut insülin yapamazsa veya buna cevap vermezse, şeker kan dolaşımında kalır ve hücreler ihtiyaç duydukları şekere ulaşamazlar (Zhao ve ark., 2023). Damarlardaki bu yüksek kan şekeri düzeyinin bir sonucu olarak, sonunda kan damarlarında ve organlarında hasar meydana gelir. Diyabetin belirtileri arasında bulanık görme, aşırı susama, yorgunluk, sık idrara çıkma, açlık ve kilo kaybı sayılabilir. Diyabetli kişilerin hemoglobin A1c seviyelerini her üç ila altı ayda bir kontrol ettirmeleri gerekir. A1c, önceki iki ila üç ay boyunca ortalama kan şekerinin bir ölçüsüdür (Eldib ve ark., 2023). Üç ana diyabet türü vardır. Tip 1 diyabet genellikle çocukluk çağında teşhis edilir. Bu diyabet tipinde vücut çok az insülin yapar veya hiç yapmaz, bu nedenle günlük insülin enjeksiyonları gerekir. Tip 2 diyabet genellikle yetişkinlerde görülür. Bu diyabet tipinde pankreas, genellikle vücut insüline iyi yanıt vermediği için kan şekerini normal seviyede tutacak kadar insülin üretmez (Zheng ve ark., 2023).

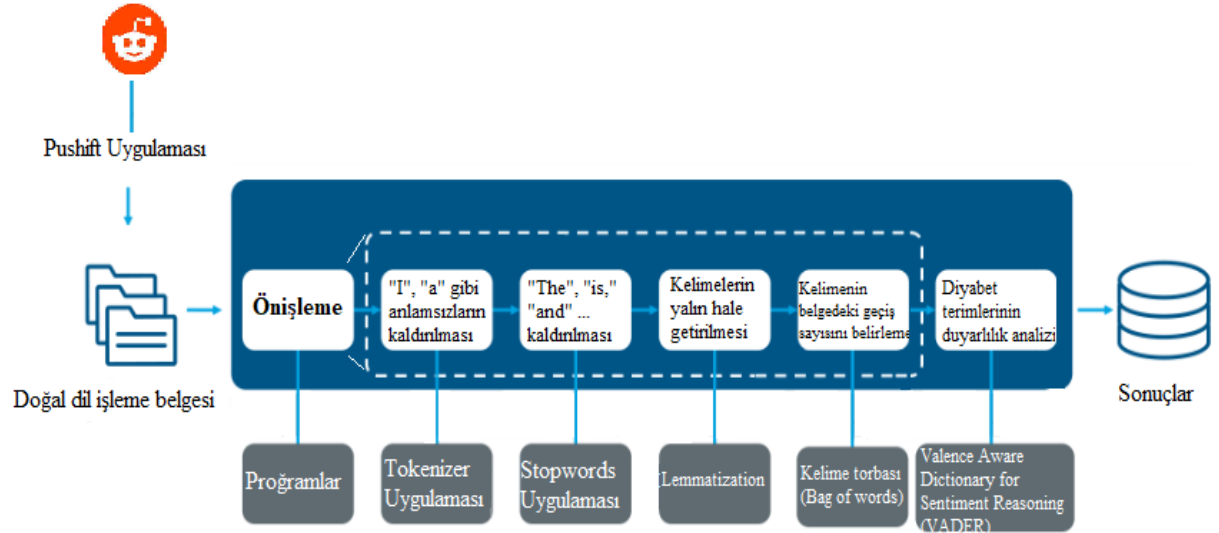
Dünyada, 537 milyon yetişkin (20-79 yaş) diyabetle yaşıyor. Bu sayının 2030'da 643 milyona ve 2045'te 783 milyona çıkacağı tahmin ediliyor (IDF Diabetes Atlas; 2021). Diyabetli her 4 yetişkinden 3'ü düşük ve orta gelirli ülkelerde yaşamakta olup, 2021'de 6,7 milyon ölüm diyabet kaynaklı olup bu 5 saniyede 1 ölüme denk gelmektedir. Şeker hastalığı, son 15 yılda 966 milyar dolarlık sağlık harcamasına neden olmuştur. Tip 2 diyabet hastalığı yönünden yüksek risk altına sokan Bozulmuş Glikoz Toleransına (IGT) sahip yetişkin sayısı 541 milyondur (IDF Diabetes Atlas, 2021).

Yetişkin hastaların atopik dermatit algılarının değerlendirilmesini, büyük sayıda veri içeren hasta anketi Doğal dil işleme (DDİ) yöntemi kullanılarak incelenmiştir (Falissard ve ark. 2020). Sosyal medyada birçok güncel konu tartışılabilen ve insanlar tecrübelerini paylaşabilmektedirler. Bu paylaşımlar arasında sağlık problemleri de büyük oranda konu başlığı oluşturabilmektedir. DDİ, şeker hastalarının hastalıklarıyla ilgili bilgilerini arttırmak için büyük miktarlarda sosyal medya yorumlarını etkili bir şekilde işleyebilir (Moulaei ve ark., 2022).

Reddit, Amerika Birleşik Devletleri'ndeki en popüler 6. web sitesi olup tartışılacak konuya özgün alt başlıklar içermektedir. Reddit sosyal medya platformunda "diabetes" isimli bir forum bulunmakta olup, şeker hastalıklarıyla ilgili hasta raporları ve yorumları için kaynaklardan bir tanesidir. DDİ, Reddit platformunda bulunan "diabetes" forumundaki yorumları işleyerek hastaların hızlı şekilde faydalanabileceği hale getirebilir. Bu çalışmada, hastaların şeker hastalığı tedavilerine olan ilgisini

anlamak için DDİ yöntemini kullanmaktadır. DDİ, yorumların yoğunluğunu ve ayrıca bu tür forumlarda tartışılan hasta tarafından bildirilen advers ilaç reaksiyonlarını irdelemektedir.

Şekil 1’de çalışmanın özeti yer almaktadır. Çalışmamız Pushift Uygulaması kullanılarak Reddit sosyal medya platformundan “r/diabetes” forumundan yorumların çekilmesi ile başlamaktadır. Bu veriler çalışmanın ham veri seti olup 4 aşamalı ön işleme tabi tutulacaktır. Bu 4 ön işleme aşamasının amacı verilerden en faydalı bilgilerin çıkartılmasıdır. Ön işlemenin ilk aşamasında “I”, “a” gibi anlam taşımayan edatların veri setinden kaldırılmasıdır. Bunlar herhangi bir anlam taşımayıp doğal dil işleme sonucunu olumsuz etkileyecektir. Türkçe ve diğer dillerde genel olarak tekrar eden ve işlenmesi sonucu etkilemeyecek “Stopwords” kelimeler vardır. Bunlar Türkçe’de “ve”, “veya” gibi kelimeler iken bizim veri setimizdeki cümlelerin dili olan İngilizce’deki “The”, “is” şeklindeki kelimelerdir. Gereksiz edat ve anlam ifade etmeyen terimlerden arındırılan cümlelerde eklerinde kaldırılması gerekmektedir. Bu işleminde ön işlemenin 3. adımını olan “Lemmatization” uygulamasında gerçekleştirilmektedir. Son aşamada ise işlenecek hale gelen yalın cümlelerdeki en sık tekrar eden kelimelerin “Bag of words (BOW)” ile sayılması işlemi yer almaktadır.



Şekil 1. Çalışma blok diyagramı

Hastalık hakkında sağlık okuryazarlığı zayıf diyabetli hastalar, zayıf uyum sergileyebilir ve fazla komplikasyon yaşayabilir. Bununla birlikte, sağlık hizmeti sağlayıcıları genel nüfus arasındaki diyabet kavramsal uçurumu tam olarak anlayamamıştır. Sosyal medyadaki diyabetle endişelerin şifresini çözmek, bu açığı kapatmaya yardımcı olabilecek bir çalışma yapılmıştır (Wang ve ark., 2022). Verilerin kalitesini kontrol ettikten sonra, metin madenciliği kullanarak insanların internetteki tartışmalarındaki eğilimleri analiz etmişlerdir.

Hasta eğitimi, diyabetin kendi kendine yönetiminin temel bir yönüdür. Oturumlar sırasında ve sonrasında kitle etkileşimi, Twitter analitiği, Tweepmap ve YouTube Studio kullanılarak gösterimler ve katılım oranı dahil olmak üzere sosyal medya ölçümleri kullanılarak analiz etmişlerdir. 22 “tweetorial” oturumu ve 5 quiz toplam 151 anket (tweetorial ve quiz oturumlarında) toplam 21.269 oy aldığı

makalede belirtilmektedir. Genel olarak, 1 saatlik tweet oturumları, %6,3 etkileşim oranıyla 1.821.088 gösterim kazandı (Thomas ve ark., 2021).

Dünyada 420 milyondan fazla diyabetli insan var. Dünya Sağlık Örgütü (WHO), diyabete bağlı ölümlerin sayısının 2030 yılına iki katına çıkacağını tahmin etmektedir. Diyabetle yaşamak, sağlık bilgilerini arama ve uygulama becerisini gerektirir. Çalışmada Kapsam belirleme incelemesinin yapılmış olup, diyabet hastalarının sosyal medya kullanımına ilişkin hangi kanıtların mevcut olduğunu ve sosyal medyanın sağlık bilgisi aramak veya duygusal destek elde etmek için kullanılıp kullanılmadığını ortaya çıkarmak amacı taşımaktadır (Zora, 2023).

(An Online Ethnographic Study, 2019) çalışma, Hindistan'da tip 2 diyabetli kişiler bir refah mekanizması olarak sosyal medyanın kullanımını ve faydalarını değerlendirmeyi amaçlamaktadır. Mevcut gözden geçirme makalesi, sosyal medya kullanımının tip 2 diyabet yönetiminde olumlu etkisini yansıtmakta olup, bu aynı zamanda sosyal desteğin teorik temelini ve uzun vadeli hastalıkları yönetmede biyopsiko-sosyal modeli kanıtlamaktadır.

Sosyal medya platformları, insanların iletişim kurmaları ve kendilerini ifade etmeleri için yeni yollar yaratmıştır. Bu nedenle, e-sağlıkla bilgilerin bu platformlarda nasıl üretildiğini ve yayıldığını araştırmak önemlidir. İsveç'teki insanlar diyabetle ilgili konular hakkında konuşmak için Twitter'ı kullandıklarında bilgi akışını ve içeriğini araştıran bir çalışma yapılmıştır. Hedeflerine ulaşmak için, 10 aylık bir süre boyunca toplanan Twitter verilerini keşfetmek, analiz etmek ve kümelemek için veri madenciliği ve görselleştirme tekniklerini kullanmışlardır (Oyebode ve ark., 2019).

Literatürde yer alan başka bir çalışmada eker hastalığı olan bireylerden web tabanlı de klinik içi anketler toplanmıştır. Katılımcıların diyabete özgü sosyal ağ sitesi davranışlarını değerlendirmek için tanımlayıcı ve korelasyon analizleri kullanılmıştır. 45 hasta web tabanlı anketi ve 167 klinik tabanlı anketi tamamlayarak ve bunlardan 40'ı diyabete özgü sosyal ağ sitelerini ziyaret etmiştir. Klinik tabanlı anket verileri, DM'ye özgü web sitelerini kullandıklarını bildiren hastaların, kullanmayanlara kıyasla evdeki glikoz değerlerini sık izlediklerini ve insülin uygulamasına iyi uyum sağladıklarını (her ikisi de $p < 0,05$) bulunmuştur (Nelakurthi ve ark., 2018).

En yaygın iltihaplı deri hastalıklarından ikisi egzama ve sedef hastalığı, sosyal medya sitesi Reddit'te aktif şekilde ve çokça tartışılmaktadır. R/psoriasis ve r/egzema alt dizilerindeki yorumları incelemek (birleşik kullanıcı tabanı>187,000), yorum yapanların ilgi düzeyini ve sedef hastalığı ve egzama için yaygın tedavilerle duygularını ve advers ilaç reaksiyonlarının tartışılmasını izlemek için doğal dil işleme araçlarını kullanılmıştır. Çalışmada, 2014- 2020 yılları r/egzama (n=196,571) ve r/psoriasis (n=123,144) alt dizinlerinden gelen yorumlar alındı ve doğal dil işleme araçları kullanılarak işlendiği anlatılmaktadır (Cummins ve ark., 2023b).

2. Materyal ve Metot

Spor, sağlık ve diğer konular olmak üzere çok çeşitli alt başlık konularına bölünen Reddit, aylık 430 milyondan fazla aktif kullanıcısı olan bir sosyal medya platformudur. Pushshift API sosyal medya

platformlarında bulunan yorum ve postların kaydedilmesi için kullanılan bir uygulamadır. Çalışmamızda “r/diabetes” alt başlığında bulunan, 2017'ten 2022'ye kadar olan yorumlar Pushshift API ile kaydedilmiştir. Boş ya da boşluktan başka karakter içermeyen yorumlar, otomatik mesajlar, Reddit moderatörleri veya gönderen kullanıcı tarafından kaldırılanlar ve Unicode olmayan karakterler içeren yorumlar, Unicode olmayan karakterler yorumlardan kaldırılarak veri seti ön işleme yapılmıştır.

Filtreleme sonucu elde edilen veri seti Doğal Dil İşleme (DDİ) ile işlenmiştir. Metinlerin tamamı küçük harfle yapılarak noktalama işaretleri kaldırılmıştır. Daha uzun bir kelimenin parçası olmayan ve İngilizce’ de anlam taşıyan "I", "a" gibi tek harfli kelimeler kaldırılmıştır. DDİ’ de işlenecek kelime/cümleyi tokenize etmek gerektiğinden her yorum bir kelime listesine dönüştürülerek tokenize edilmiştir.

Stopwords (Stopwords, herhangi bir dilde yaygın olarak kullanılan bir dizi sözcüktür. Örneğin, İngilizce’de "the", "is" ve "and" kolaylıkla durak kelimeler olarak nitelendirilebilir. NLP ve metin madenciliği uygulamalarında, durak kelimeler önemsiz kelimeleri elemek için kullanılır ve uygulamaların bunun yerine önemli kelimelere odaklanmasını sağlar.), DDİ verilerinde fazla anlam taşımayan ve gürültü oluşturan yaygın sözcükler olup, Genism ve Natural Language Toolkit’in (NLTK) durdurma sözcükleri listesinin yanı sıra “kullan” ve “beğen” sözcükleri kullanılarak durdurma sözcükleri çıkarıldı (Bird ve ark., 2009; Řehůřek ve Sojka 2010).

Kelimelerde bulunan eklerin kaldırılarak kelimenin yalın hale getirilmesine Lemmatization adı verilmektedir. NLTK'nin WordNet Lemmatizer'ı kullanılarak veri yorumlarda Lemmatization işlemi yapılmıştır. (Freihat ve ark., 2018). İşlenen yorumlar, bir genizm sözlüğü kullanılarak belgedeki bir kelimenin geçiş sayısını sayan bir kelime torbası (Bag of Words (BOW)) modeli oluşturmak için kullanıldı (Nabizadeh-ShahreBabak ve ark., 2021).

Elde edilen 1000 kelimelik liste gözden geçirilerek, kelime sayısı şeker hastalığı ile ilgisine göre 90 kelimedenden aza indirildi. Düzenli ifadeler (Regular expressions (RE)) alt dizinde seçilen ilgili kelimelerin her birini içeren yorumları belirlemek için kullanılabilir. Sözcüklerin bireysel RE analizinin işlenmesine ek olarak, RE analizi için aynı ilacın marka adları aynı anlama geldiğinden bazı sözcükler de RE grupları olarak alınmıştır. İlaç isimleri içeren incelemelerin ilgisini belirlemek için Valence Aware Dictionary for Sentiment Reasoning (VADER) kullanılarak duyarlılık analizi yapılmıştır (Hutto ve Gilbert 2014). VADER, -1 en olumsuz, +1 en olumluyu, 0 tamamen etkisiz durumu ifade etmektedir. Bu çalışmada, incelemelerin VADER puanları -0,988 ile 0,981 arasında hesaplanmıştır. Zaman içindeki duyarlılık puanları ve bu puanların grafik hesaplaması, Kruskal-Wallis sıralama toplamı testleri kullanılarak yapılmıştır. Güven aralığı ise Wilson skor aralıklarına dayanmaktadır. İki taraflı p değerleri <0,05 istatistiksel olarak anlamlı kabul edilmiştir.

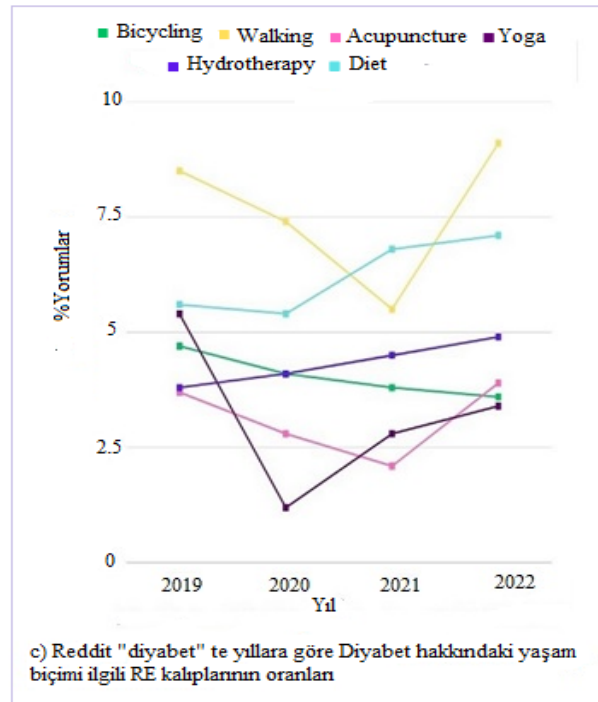
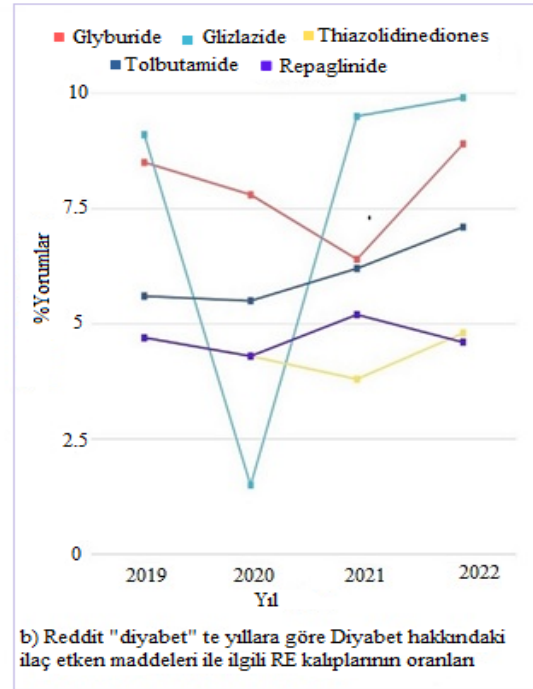
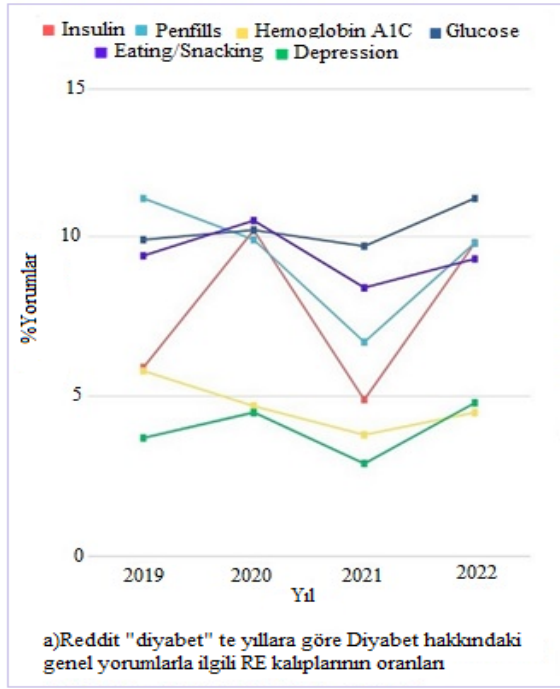
3. Bulgular ve Tartışma

Filtreleme işlemleri yapıldıktan sonra, “diabetes” alt başlığında 110.483 yorum toplanmıştır. Alt başlıkta, yıllara göre farklı yorumcu sayıları şu şekildedir: 2019: 4.005, 2020: 5.060, 2021: 5.275, 2022: 5.098. Tablo 1’ de yıllara göre yorumcu başına düşen yorum sayısı görülmektedir.

Tablo 1. r/diabetes yıllık yorum sayısı, benzersiz yorumcu sayısı ve benzersiz yorumcu başına yorum sayısı

Yıl	Yorum sayısı	Yorumcu sayısı	Yorumcuya düşen yorum sayısı
2019	23.633	4005	5.9
2020	30.867	5060	6.1
2021	27.432	5275	5.2
2022	28.551	5098	5.6

Şekil 2’ de “diabetes” alt başlığında yer alan terimlerin yıllara göre değişim oranı görülmektedir. Şekil 2. (a)’ da şeker hastalığı ile ilgili, yorumlarda geçen en sık ifadeler görülmektedir. 6 terimin yorumlarda yer alma sıklığı yıllara göre hep dalgalı olarak gerçekleşmiştir. Dalgalanmanın en fazla olduğu ifade ‘Insulin’ dir. Terimler arasında en fazla orana %11,2 ile 2022 yılında ‘Glucose’ sahiptir. Terimler arasında en düşük orana 2021 yılında %2,9 ile ‘Depression’ sahiptir. ‘Glucose’ terimi bütün yıllarda yüksek oranda yorumlarda bulunduğu söylenebilir. Ayrıca bu terimdeki dalgalanma diğer terimlere göre daha azdır. ‘Penfills’ ve ‘Hemoglobin A1C’ terimlerinde, 2019’dan 2021’e kadar düzenli bir azalış söz konusudur. 2022 yılında ise bu iki terim yorumlarda yer alma oranı artmıştır. Aslında 6 terimin hepsinde 2021 yılından 2022 yılına geçişte artış söz konusudur. Terimlerin yorumlarda yer alma oranı 2022 yılında şu şekilde gerçekleşmiştir: Insulin=%9,8, Penfills=%9,8, Hemoglobin A1C=%4,5, Glucose=%11,2, Eating/Snacking=%9,3, Depression=%4,8. Şekil 2 (b)’ de şeker hastalığı tedavisinde kullanılan ilaç etken maddelerinin, 2019-2022 yılları arasındaki yorumlarda yer alma oranları görülmektedir. Yıllara göre dalgalanma en fazla ‘Glizlazide’ terimindedir. ‘Tolbutamide’ teriminde 2019’dan 2020’ye geçişte önemsenmeyecek bir düşüş olmakla beraber düzenli bir artış söz konusudur. 2022 yılında ‘Repaglidine’ terimi %5,2’den %4,6’ ya düşmüş diğer 4 terim ise yükselmiştir. ‘Glizlazide’ terimindeki 2019 yılından 2020 yılına geçişte oransal düşüş %83’ü bulmuştur. Şekil 2 (c)’ de şeker hastalığının tedavisinde alternatif yöntem terimlerinin yıllara göre oransal değişimi görülmektedir. Genel terimler ve tedavide kullanılan ilaçların etken madde adlarındaki dalgalanma burada yoktur. ‘Hydrotherapy’ teriminde düzenli bir artış varken, ‘Bicycling’ teriminde düzenli bir azalış söz konusudur. 2022 yılında ‘Bicycling’ hariç bütün terimler artış göstermiştir. 2019 yılından 2020 yılına geçerken 6 terimden ‘Hydrotherapy’ hariç hepsi düşüş göstermiştir. 2019 yılından 2020 yılına geçişte ‘diet’ teriminde küçük bir düşüş olsa da düzenli artış vardır. 2022 yılında terimlerin yorumlardaki oransal olarak şu şekilde gerçekleşmiştir: Bicycling=%3,6, Walking=%9,1, Acupuncture=%3,9, Yoga=%3,4, Hydrotherapy=%4,9, Diet=%7,1.



Şekil 2. Reddit belirtilen düzenli ifade modellerini içeren r/diabetes alt başlığındaki yorumların yıllara göre yüzdesi (p=0,038)

4. Sonuç

Bu çalışma, diyabet tedavi seçenekleri hakkında fikirlerin yanı sıra yan etki raporlarına ilgi ve duyarlılık düzeylerini yapay zeka ile araştırılması imkanı sunmaktadır. Sonuçlar, r/diabetes hakkında yorum yapan Reddit kullanıcılarının bir dizi tedaviye aşına olduğunu gösterdi. Çoğu durumda, açıklanan hacimdeki artışlar ve azalmalar, mevcut verilere ve fikir birliğine dayalı olarak bu

tedavilerin hekimler tarafından tercih edilme derecesi ile tutarlı görünmektedir. Alt dizinde biyolojik ve oral ilaçlarla ilgili anahtar kelime aramaları, ilacın adını içeren yorumların yüzdesinin genel olarak dalgalı bir seyir işlerken 2022 yılında oransal olarak artış göstermiştir. İncelenen tüm yıllar boyunca yaygın olarak kullanılan ilaçlara ve alternatif tedavi yöntemlerine olan ilginin devam ettiğini gösteriyor.

Son yıllarda, r/diabetes alt dizini, alternatif tedaviler, yaşam tarzı ve diyet yaklaşımlarıyla ilgili terimler içeren yorumların yüzdesinde genel olarak artış yaşanmıştır. Bu değişiklikler, bu konulara olan ilgideki gerçek bir artışı yansıtabilir veya Reddit kullanıcılarının bu seçeneklere olan ilginin artması nedeniyle göreceli bir artışı yansıtmış da olabilir.

Alt dizinlerdeki yorumların hacmi, Reddit kullanıcılarının mevcut önerilen yönetimi anladığını gösteren hastalık yönetimi ve dermatoloji topluluğunda zaman içinde beklenebilecek önerilerdeki değişiklikleri yansıttığı söylenebilir. Bir ilaç kullanılmadan önce, tartışmalar ilacın beklentisine ve önceki ilaçlara göre algılanan faydasına odaklanabilir; bu, ilaç önceki seçeneklere göre bir gelişme olsa bile, fiili kullanımdan sonra azalabilir. Benzer şekilde, Repaglinid gibi uzun yıllardır FDA tarafından onaylanan ilaçlar, muhtemelen zamanla daha yeni ve daha etkili ilaçlarla karşılaştırıldığında etkiyi azalttıkları için zamanla daha az duyarlı hale gelebilirler.

Bu çalışma, birçok hastanın madde kullanımıyla ilgili semptomları kendi kendine bildirdiğini göstermektedir, bu nedenle DDI aracı, potansiyel madde kullanımı ve yaşanabilecek semptomlar için binlerce forum yorumunu aramak için değerli bir araç olabilir. Uyuşturucu kullanımının ve olası semptomların bilindiği durumlarda, izleme incelemeleri ve ilaç açıklamaları, uyuşturucu kullanımının ve olası semptomların daha iyi tanımlanmasına yardımcı olur. Buna ek olarak, daha önce ilk klinik denemeler sırasında tespit edilemeyen olası nadir semptomları belirlemek için büyük verilerdeki kalıpları belirlemek için daha geniş çapta uyuşturucu kullanımına ve olası semptomlara odaklanan araçlar kullanılabilir. Hasta tarafından bildirilen toplu sonuç göstergelerini belirlemek için bu tür verileri araştırmak, ek bir güçlü kullanım durumu sağlar.

Bu çalışmanın bulguları, r/diabetes alt dizini içinde bildirilen yorumlarla sınırlıdır ve bu nedenle tüm şeker hastalarını genellemez. Bu alt başlık içinde yorum yapan bireyler, hastalık yönetimine daha fazla para harcıyor olabilir ve bu koşullara sahip ortalama şeker hastalarından daha fazla hastalık etkisine sahip bir topluluğu temsil ediyor olabilir. Reddit kullanıcılarının bu sosyal medya forumuna katılabilmeleri için teknik bilgiye de sahip olmaları gerekmektedir. DDİ, büyük miktarlarda yapılandırılmamış verinin işlenmesine izin vermesine rağmen, sosyal medyanın standartlaştırılmamış doğası, alternatif yazımlar veya kelime seçimleri veya belirli konuların alternatif forumlara geçişi nedeniyle yakalanmamış veriler gibi doğal sınırlamalara sahiptir. Torba kelimeler metodolojisi yaklaşımı (BOW), en sık kullanılan terimleri tanımlayarak bu sınırlamayı hafifletmeye yardımcı olmuştur. Diğer bir sınırlama, çalışma yılları boyunca belirli ilacın hangi hastalıklara ve hangi biçimde uygun olacağını veya bir operasyonun hangi koşullarda gerekli olduğunu belirlemek için her ilaç satış

veya reçete hacmine ilişkin bulunmaması ve yorum hacimlerini her bir ilaç için reçete sayısı gibi bir ölçüyle doğrudan karşılaştırma olanağını sınırlamasıdır.

Veri analizi için öncelikle DDİ'ye dayanan herhangi bir çalışma, manuel inceleme yoluyla tanımlanabilecek daha ince ayrıntıların ve anlamların bazılarını kaçırma eğilimindedir. Bu çalışmada kullanılan BOW yaklaşımı, tek tek kelimeleri yakalar ancak bir cümlenin veya yorumun tam anlamını sağlamaz, bu da yorumun amaçlanan anlamının yorumlanmasını sınırlamaktadır.

Bu makale, hastalığı ve hastaların tedavi deneyimlerini daha iyi anlamak için diğer tıp forumlarında uygulanabilecek bir model olarak şeker hastalığına başarıyla uygulanmış bir DDİ aracı sunmaktadır. Sosyal medya geri bildirim, klinisyenlerin ve araştırmacıların hasta deneyimini, yapılandırılmış anketlerin sınırlamaları veya klinisyenleri veya araştırmacıları memnun etmeye çalışan hastaların önyargısı olmadan, hastaların kendi sözleriyle daha iyi anlamalarına olanak tanıdığı için benzersiz bir şekilde değerlidir. Sosyal medya forumlarındaki binlerce yorumu analiz etmek, hastalık ve tedavi müdahaleleriyle ilgili duygu ve endişelerdeki eğilimleri etkili bir şekilde belirleyebilir.

Çıkar Çatışması Beyanı

Makale yazarı herhangi bir çıkar çatışması olmadığını beyan eder.

Araştırmacıların Katkı Oranı Beyan Özeti

Yazar makaleye %100 oranında katkı sağlamış olduğunu beyan eder.

Kaynakça

Bird S., Klein E., Loper E. Natural language processing with python 2009.

Cummins JA., Zhou G., Nambudiri VE. Natural language processing for large scale analysis of eczema and psoriasis social media comments. JID Innovations 2023b; 00210.

Eldib AH., Dhaver S., Al-Badri M., Salah T., Kibaa K., Elenani O., Tomah S., Gardner H., Hamdy O. Magnitude of A1C improvement in relation to baseline A1C and amount of weight loss in response to intensive lifestyle intervention in real-world diabetes practice: 13 years of observation. Journal of Diabetes 2023.

Exploring the Role of social media as a support mechanism among persons with diabetes: An online ethnographic study. Journal of Content, Community & Communication 2019; 10(9).

Falissard B., Simpson EL., Guttman-Yassky E., Papp KA., Thyssen JP., Gadkari A., Saba G., Gautier L., Abbe A., Eckert L. Correction to: qualitative assessment of adult patients' perception of atopic dermatitis using natural language processing analysis in a cross-sectional study. Dermatology and Therapy 2020; 10(2): 307–310.

Freihat AA., Abbas M., Bella G., Giunchiglia F. Towards an optimal solution to lemmatization in arabic. Procedia Computer Science 2018; 142: 132–140.

- Hutto C., Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media 2014*; 8(1): 216–225.
- International Diabetes Federation. *IDF diabetes Atlas, 10th edn.* Brussels, Belgium: 2021. Available at: <https://www.diabetesatlas.org>
- Moulaei K., Dinari Z., Dinari F., Jahani Y., Bahaadinbeigy K. The role of social networks in diabetes self-care: A cross-sectional study. *HealthScienceReports 2022*; 5(3).
- Nabizadeh-ShahreBabak Z., Karimi N., Khadivi P., Roshandel R., Emami A., Samavi S. Detection of COVID-19 in X-ray images by classification of bag of visual words using neural networks. *Biomedical Signal Processing and Control 2021*; 68: 102750.
- Nelakurthi AR., Pinto AM., Cook CB., Jones LC., Boyle M., Y, J., Lappas T., He J. Should patients with diabetes be encouraged to integrate social media into their care plan? *Future Science OA 2018*; 4(7).
- Oyebode O., Orji R. Detecting factors responsible for diabetes prevalence in nigeria using social. *Media and Machine Learning 2019*.
- Řehůřek R Sojka P. Software framework for topic modelling with large corpora. *proc. LREC 2010 Workshop New Chall. NLP Framew. Valletta, Malta: ELRA 2010*: 45–50
- Thomas RL., Alabraba V., Barnard S., Beba H., Brake J., Cox AJ., Bowker R., Edwards DM., Epps A., Fletcher-Salt T., Holmes P., Kar P., Kausar N., Kelly B., Leveridge M., Newland-Jones P., Ng SM., Puttana A., Stewart R. Use of social media as a platform for education and support for people with diabetes during a global pandemic. *Journal of Diabetes Science and Technology 2021*; 17(2): 353–363.
- Wang R., Hong Y., Li C., Li S., Liu J., Wu C., Chiu CJ. Using social media data in diabetes care: bridging the conceptual gap between health providers and the network population. *BMC Primary Care 2022*; 23(1).
- Zhao H., Zhuo L., Sun Y., Shen P., Lin H., Zhan S. Response to: Critical commentary on the association between thiazolidinedione use and dementia risk in patients with type 2 diabetes. *PubMed 2023*.
- Zheng J., Shen S., Xu H., Zhao Y., Hu Y., Xing Y., Song Y., Wu X. Development and validation of a multivariable risk prediction model for identifying ketosis-prone type 2 diabetes. *PubMed 2023*.
- Zora H. The use of social media-communities by diabetes patients- a scoping review. *Population Medicine 2023*; 5(Supplement).