T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

TÜTÜN BAĞIMLILIĞI ÖRNEĞİNDE HALK SAĞLIĞINDA YAPAY ZEKA TEKNOLOJİLERİNİN KULLANILMASI

HÜSEYİN KÜÇÜKALİ

DANIŞMAN
PROF. DR. MEHMET SARPER ERDOĞAN

HALK SAĞLIĞI ANA BİLİM DALI
HALK SAĞLIĞI DOKTORA PROGRAMI

İSTANBUL-2021

TEZ ONAYI

Bu çalışma 01.07.2021 tarihinde aşağıdaki jüri tarafından Halk Sağlığı Anabilim Dalı, Halk Sağlığı Doktora Programı, Doktora Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Prof. Dr. Mehmet Sarper ERDOĞAN
İstanbul Üniversitesi-Cerrahpaşa
Cerrahpaşa Tıp Fakültesi

Prof. Dr. Osman Erol HAYRAN
İstanbul Medipol Üniversitesi
Tıp Fakültesi

Doç. Dr. Günay CAN
İstanbul Üniversitesi-Cerrahpaşa
Cerrahpaşa Tıp Fakültesi

Prof. Dr. Ethem ERGİNÖZ
İstanbul Üniversitesi-Cerrahpaşa
Cerrahpaşa Tıp Fakültesi

Prof. Dr. Halim İŞSEVER
İstanbul Üniversitesi
İstanbul Tıp Fakültesi

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün safhalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığı beyan ederim.

Hüseyin KÜÇÜKALİ

TEŞEKKÜR

Bu tezin yazımı sürecinde bana rehberlik eden ve her zaman destekleyici olan Mehmet Sarper Erdoğan hocam başta olmak üzere, iyi bir eğitim almamı önemseyen Ethem Erginöz hocama, tez sürecinin uygun şekilde ilerlediğini kontrol eden Günay Can hocama, bana nasıl iyi bir biliminsanı olunacağını öğreten Osman Hayran hocama, beni halk sağlığı bilimiyle tanıştırıp doktora eğitimine teşvik eden Mustafa Taşdemir hocama ve öncesinden sonrasına her zaman esas desteğim olan eşim İrem Küçükali'ye teşekkür ederim.

Ayrıca bağımlılıklarla mücadeledeki öncülüğü, bilime verdiği değer ve bu araştırmaya destekleri için Türkiye Yeşilay Cemiyeti'ne, özel olarak bu araştırmadaki analize katkıları için Genç Yeşilay gönüllülerine teşekkür ediyorum.

Bu çalışma, Türkiye Yeşilay Cemiyeti tarafından Lisansüstü Tez Araştırma Bursu Destek Programı (2019/7) kapsamında desteklenmiştir.

İÇİNDEKİLER

TEZ ONAYI II
BEYAN III
TEŞEKKÜR IV
İÇINDEKILER V
TABLOLAR LISTESI VIII
ŞEKILLER LISTESI IX
KISALTMALAR LISTESI X
ÖZET XI
ABSTRACT XII
1. GİRİŞ VE AMAÇ 1
Amaç 3
2. GENEL BİLGİLER 4
2.1. İnfodemiyoloji 4
2.2. Twitter ve Başlıca Özellikleri 4
2.3. Yapay Zeka 5
2.4. Makine Öğrenmesi 6
2.5. Doğal Dil İşleme 7
2.6. Metin Sınıflandırmada Kullanılan Yöntemler 8
2.6.1. Ön İşleme 8
2.6.2. Özellik Çıkartma 8
2.6.3. Sınıflandırma Algoritmaları 9
2.6.4. BERT 10
3. GEREÇ VE YÖNTEM 12
3.1. Araştırma Türü 12
3.2. Evren ve Örnekleme 13
3.3. Araştırma için Geliştirilen Yazılımlar 13
3.4. Veri Toplama 14
3.4.1. Arama Stratejisi 15
3.5. Verilerin Saklanması 16
3.6. Değişkenler 16
3.7. İçerik Analizi 17

3.7.1. Tümevarım Analizi 17
3.7.2. Değerlendiriciler Arası Tutarlılık....17
3.7.3. Kitle Kaynağı ve Değerlendiricilerin Dahil Edilmesi....18
3.7.4. Değerlendirici Eğitimi....19
3.7.5. İçerik Kodlama Yazılımı ve Sohbet Robotu....20
3.7.6. Oyunlaştırma....22
3.7.7. Tümdengelim Analizi 24
3.8. Makine Öğrenmesi....24
3.8.1. Metin Ön İşleme....25
3.8.2. Dil Modeli....25
3.8.3. Eğitim ve Test....25
3.9. İstatistiksel analizler....27
3.10. Araştırma İzinleri....28
4. BULGULAR....29
4.1. Tanımlayıcı İstatistikler....29
4.2. Makine Öğrenmesi Modelinin Başarısı....34
4.3. İki Değişkenli Analizler....38
5. TARTIŞMA....45
5.1. Araştırmanın Ana Çıktıları....45
5.2. Sosyal Medyada Tütün Teşvikinin Mevcut Durumu....46
5.2.1. Tweet Özelliklerine Göre Dağılım....46
5.2.2. Zamana Göre Dağılım....46
5.2.3. Cihaz Türüne Göre Dağılım....46
5.2.4. Metinlerde Sık Geçen Kelimeler....47
5.3. Bilimsel Katkı....47
5.3.1. Bilgi Tarama Çalışmalarına Katkı....47
5.3.2. Tütünle İlgili Sosyal Medya İçerikleri İnceleyen Çalışmalara Katkı....48
5.3.3. Sağlık Davranışı Çalışmalarına Katkı....49
5.3.4. Sosyal Medya İçeriklerinde Makine Öğrenmesi Kullanan Çalışmalara Katkı....50
5.3.5. Türkçe Halk Sağlığı Literatürüne Katkı....50
5.4. Ulusal Tütün Kontrolüne Katkı....50
5.5. Halk Sağlığı Araştırma ve Uygulamalarında Kullanım Önerileri....52
5.5.1. Sağlığın Geliştirilmesinde Kullanım Önerileri....53

5.6. Araştırmanın Güçlü Yanları 54
5.7. Araştırmanın Kısıtlılıkları ve Gelecek Araştırmalar için Tavsiyeler 55
6. KAYNAKLAR 58
ETİK KURUL KARARI 66
TWITTER GELİŞTİRİCİ YETKİSİ 67
İTÜ DOĞAL DİL İŞLEME ARAÇLARI YETKİSİ 68
TARIM BAKANLIĞI BİLGİ EDİNME BAŞVURUSU CEVABI 69
İNTİHAL RAPORU İLK SAYFASI 70
ÖZGEÇMİŞ 71

TABLOLAR LİSTESİ

Tablo 3-1. Bağımlı değişken ve alt boyutları için değerlendiriciler arası tutarlılık..... 18
Tablo 3-2. İçerik analizinde kullanılan kodlama çerçevesi 24
Tablo 4-1. Tweet'lerin özgünlük durumuna göre dağılımı..... 32
Tablo 4-2. Tweet'lerin çoklu medya unsuru buldurma durumlarına göre dağılımı 32
Tablo 4-3. Tweet'lerin yayınlandıkları cihaz ve yazılım türlerine göre dağılımı..... 33
Tablo 4-4. Örnek tweet'lerin tütün kullanımını teşvik etmenin alt boyutlarına dağılımı34
Tablo 4-5. Tütün kullanımını teşvik etme durumunu öngören makine öğrenmesi modellerinin başarı düzeyleri..... 35
Tablo 4-6. Tütün kullanımını teşvik etmenin alt boyutlarını öngören makine öğrenmesi modellerinin başarı düzeyleri..... 37
Tablo 4-7. Tweet'lerin haftanın hangi kısmında yayınlandığına göre tütün kullanımı teşvik etme durumu..... 39
Tablo 4-8. Tweet'lerin gece veya gündüz yayınlanma durumuna göre tütün kullanımı teşvik etme durumu..... 41
Tablo 4-9. Tweet'lerin özgünlük durumuna göre tütün kullanımı teşvik etme durumu. 41
Tablo 4-10. Tweet'lerde çoklu medya unsuru bulunma durumuna göre tütün kullanımı teşvik etme durumu..... 42
Tablo 4-11. Tweet'lerin yayınlandıkları cihaz ve yazılım türlerine göre tütün kullanımı teşvik etme durumu..... 43
Tablo 4-12. Tütün kullanımı teşvik etme durumuna göre tweet'lerin yayılma özellikleri 44

ŞEKİLLER LİSTESİ

Şekil 3-1. Araştırma süreci 12
Şekil 3-2. Araştırma kapsamında geliştirilen tüm yazılımların iş akışı, bulut hizmetler ve veri tabanı ile bağlantıları 14
Şekil 3-3. Sohbet robotu iş akışı 21
Şekil 3-4. Sohbet robotu içerik atama ve kodlama ekran görüntüsü 22
Şekil 3-5. Sohbet robotu oyunlaştırma ekran görüntüsü..... 23
Şekil 4-1. Tweet'lerin günlere göre dağılımı..... 29
Şekil 4-2. Tweet'lerin saatlere göre dağılımı..... 30
Şekil 4-3. Tweet'lerin gün ve saatlere göre dağılımını gösteren ısı haritası..... 31
Şekil 4-4. Tweet'lerde en sık geçen kelimeler (n=50)..... 33
Şekil 4-5. "2. Model" tarafından öngörülen kategorilerin insan değerlendiriciler tarafından belirlenen kategorilerle karşılaştırılması 36
Şekil 4-6. Tütün kullanımını teşvik eden tweet yüzdesinin günlere göre dağılımı 38
Şekil 4-7. Tütün kullanımı teşvik eden tweet yüzdesinin haftanın günlerine göre dağılımı 39
Şekil 4-8. Tütün kullanımı teşvik eden tweet yüzdesinin saat aralıklarına göre dağılımı 40
Şekil 4-9. Tütün kullanımı teşvik eden ve etmeyen tweet'lerde en sık geçen kelimeler 44

KISALTMALAR LİSTESİ

API: Application Programming Interface (Uygulama Programlama Arayüzü)

BERT: Bidirectional Encoder Representations from Transformers

CDC: Centers for Disease Control and Prevention

DSÖ: Dünya Sağlık Örgütü

PPD: Pozitif Prediktif Değer

MKK: Matthews Korelasyon Katsayısı

ÖZET

Küçükali, H. (2021). Tütün Bağımlılığı Örneğinde Halk Sağlığında Yapay Zeka Teknolojilerinin Kullanılması. İstanbul Üniversitesi-Cerrahpaşa Lisansüstü Eğitim Enstitüsü, Halk Sağlığı Ana Bilim Dalı, Doktora Tezi. İstanbul.

Tütün endüstrisinin bir kaçış alanı olarak sosyal medyada yürüttüğü pazarlama faaliyetleri son yıllarda dikkat çekmektedir. İster ticari amaçlı ister bireysel olsun sosyal medyada tütünü teşvik eden içeriklere maruziyetin, takip eden süreçte tütün kullanımına etki ettiği gösterilmiştir. Tütün bağımlılığının önlenmesine yönelik mevcut politika ve sistemlerin yetersiz kaldığı bu sorunun çözümüne yapay zeka teknolojileri katkı sağlayabilir. Bu araştırmanın amacı sosyal medyada tütün kullanımını teşvik eden içerikleri otomatik olarak tespit edebilecek bir yapay zeka geliştirmektir.

Bu karma yöntemli araştırma nitel bir araştırmayı takip eden nicel bir araştırma şeklinde tasarlanmıştır. Ekim 2020'de Twitter'da yayınlanan tütün ile ilgili tweet'ler evreninden (n=177,684) olasılıklı yöntemle seçilen bir örnek (n=5000), gönüllü değerlendiricilerden oluşturulan bir kitle kaynağı tarafından içerik analizine tabi tutulmuş ve tütün kullanımını teşvik etme durumuna göre sınıflandırılmıştır. Oluşturulan bu veri seti kullanılarak, önceden Türkçe dilinde eğitilmiş bir dil modeli (BERT), gözetimli makine öğrenmesi yöntemiyle farklı senaryolar altında yeniden eğitilmiştir. Geliştirilen model kullanılarak evrendeki tüm tweet'ler için tütün kullanımını teşvik etme durumu öngörülmüş ve bunun bağımsız değişkenlere göre dağılımı incelenmiştir.

Geliştirilen makine öğrenmesi modeli, sosyal medya içeriklerinin tütün kullanımını teşvik etme durumunu %87,78 duyarlılık ve %81,11 pozitif prediktif değer ile öngörebilmektedir. Günde ortalama 2360,10 ± 599,43 adet tütün kullanımı teşvik eden tweet yayınlanmakta ve bunlar tütünle ilgili tweet'lerin %39,84'ünü oluşturmaktadır. Gece yayınlanan, özgün, başkalarından bahseden, mobil cihaz kaynaklı tweet'ler arasında tütün teşviki diğerlerine kıyasla daha sıktır.

Geliştirilen yapay zeka, tütünü teşvik eden sosyal medya içeriklerini izlemeyi mümkün kılan bir infosürveyans aracıdır. Bu araç tütün kontrolü politikalarını güçlendirebilir, özellikle sağlığın geliştirilmesi uygulamaları için yeni imkanlar oluşturabilir.

Anahtar Kelimeler: yapay zeka, makine öğrenmesi, sosyal medya, tütün kontrolü, birincil önleme

Bu çalışma, Türkiye Yeşilay Cemiyeti tarafından Lisansüstü Tez Araştırma Bursu Destek Programı (2019/7) kapsamında desteklenmiştir.

ABSTRACT

Küçükali, H. (2021). Artificial Intelligence in Public Health: In Case of Tobacco Addiction. İstanbul University-Cerrahpasa, Institute of Graduate Studies, Department of Public Health. PhD Thesis. İstanbul, Turkey.

Marketing activities of the tobacco industry on social media as an escape area has become salient in recent years. Whether they are commercial or personal, exposure to tobacco promoting contents in social media is shown to influence subsequent tobacco use. Artificial intelligence technologies may contribute to the solution of this problem where existing policies and systems are insufficient for the primary prevention of tobacco addiction. This study aims to develop artificial intelligence that can automatically identify tobacco promoting contents in social media.

This study is designed as a sequential mixed-methods study where qualitative analysis precedes quantitative analysis. A probabilistic sample (n=5000) from tobacco-related tweets that are published on Twitter in October 2020 (n=177,684) are selected and classified in terms of tobacco promotion via crowdsourced content analysis. The created dataset used in supervised machine learning to finetune a pre-trained Turkish language model (BERT) in multiple scenarios. Tobacco promotion status predicted for all tobacco-related tweets using the developed model and its distribution by independent variables is analyzed.

Tobacco promotion in social media contents is predicted by the model with a recall of 87.78% and precision of 81.11%. The mean daily count of tobacco promoting tweets is 2360.10 ± 599.43 and they constitute 39.84% of all tobacco-related tweets. Tobacco promotion is more frequent among tweets that are published in the night, are original, mention another user, are sourced from a mobile device than others.

Developed artificial intelligence is an infoveillance tool that makes it possible to monitor tobacco promoting social media contents. This tool can strengthen tobacco control policies and create new opportunities for health promotion practices in particular.

Key Words: artificial intelligence, machine learning, social media, tobacco control, primary prevention

This study was supported by the Turkish Green Crescent Society under their Graduate Research Scholarships Program (2019/7).

1. GİRİŞ VE AMAÇ

Son çeyrek asırda tütün kontrolünde elde edilen önemli kazanımlara rağmen tütün kullanımı halen Türkiye'de ölümler için en önemli ikinci risk faktörüdür. 2019 yılında gerçekleşen tüm ölümlerin %21,1'i tütünle ilişkilidir ve bu yüzde dünya genelinden (%15,4) yüksektir (Global Burden of Disease Collaborative Network, 2020). Eldeki en güncel istatistiklere göre Türkiye'de yetişkinlerin %31,6'sı (2016) 13-15 yaş arası gençlerin %17,9'u (2017) tütün kullanmaktadır (Centers for Disease Control and Prevention [CDC], 2019).

Özellikle 2004 yılında Tütün Kontrol Çerçeve Sözleşmesi'nin kabul edilmesi ve takip eden yıllarda hazırlanan ulusal tütün kontrolü program, strateji ve eylem planları doğrultusundaki uygulamalar tütün kontrolünde önemli başarılar elde edilmesini sağlamıştır (Saraçoğlu ve Öztürk, 2020). Yetişkinlerde 2008 yılında %31,2 olan tütün kullanımı 2012'de 27,1'e kadar azalmıştır (CDC, 2019).

Fakat politika ve uygulamalardaki eksiklikler 2016 yılında yetişkinlerde tütün kullanımının %31,6'ya kadar geri yükselmesiyle kendini göstermiştir. Diğer taraftan gençlerde tütün kullanımı ise 2003'de %8,4'den 2017'de %17,9'a kadar sürekli yükselme göstermiştir (CDC, 2019). Tütün bağımlılığının önlemesinde mevcut politika ve uygulamaların yetersiz kaldığı ve yenilikçi yaklaşımlara ihtiyaç olduğu görülmektedir.

Dünya Sağlık Örgütü (DSÖ) (2008) tarafından MPOWER olarak formüle edilen politika paketine göre tütün reklam ve tanıtımının engellenmesi tütün kontrolünün önemli bir ayağını oluşturmaktadır. Bu kapsamda birçok ülkedeki yasalar tütün ürünlerinin geleneksel medya kanalları (televizyon, radyo, gazete vb.) üzerinden pazarlanmasını engellemektedir. DSÖ kısmi yasakların etkisiz olduğunu söyleyerek; yasakların tüm medyayı, özellikle de yeni medyayı, kapsaması gerektiğine; aksi takdirde tütün şirketlerinin yüksek pazarlama bütçeleriyle yeni yollar bulduklarına dikkat çekmektedir (DSÖ, 2008).

Özellikle sosyal medya büyüklük, denetimsizlik ve anonimlik özellikleri dolayısıyla "illegalitenin kaçış alanı" olabilmektedir (Arklan ve Rençber, 2018). Uluslararası tütün şirketleri sosyal medya sayesinde söz konusu yasaların etrafından dolaşmaktadır. Son yıllarda uluslararası tütün şirketlerinin dünyanın dört bir yanında

“influencer” olarak adlandırılan sosyal medya ünlüleri aracılığıyla yürüttüğü pazarlama amaçlı faaliyetlerini ortaya çıkaran çok sayıda haber yayınlanmıştır (Davies ve Chapman, 2021; Kaplan, 2018; Kirkham, 2019; Rowell, 2020).

Genç yetişkinler üzerinde gerçekleştirilen bir izlem araştırması sosyal medyada sigara kullanımını teşvik eden içeriklere pasif maruziyetin altı ay sonraki sigara içme davranışını belirlediğini ve bu etkinin doz-yanıt ilişkisi gösterdiğini ortaya koymuştur (Zhu, 2017). Geniş çaplı başka bir izlem araştırması da bu etkiyi elektronik sigara ile ilgili medya maruziyeti için göstermiştir (Liu ve ark., 2020). Sosyal medyada tütün kullanımı teşvik eden içerikler yalnızca tütün şirketlerin tanıtım faaliyetlerinden ibaret değildir. Aynı zamanda sıradan kişilerin kar amaçlı olmayan yayınlarında da tütün kullanımı teşvik edilmektedir (Zhu, 2017).

Türkiye’de televizyon, video veya sinema filmi izleyen öğrencilerin %66,5’i bu ortamlarda tütün kullanan birini görmektedir (CDC, 2019). Sosyal medyada tütün kullanımını teşvik eden içeriklere maruziyetin ne ölçüde olduğu ise bilinmemekte ve izlenmemektedir.

4207 sayılı Tütün Ürünlerinin Zararlarının Önlenmesi ve Kontrolü Hakkında Kanun’da 2018 yılında yapılan değişiklikle diğer mecralarda olduğu gibi sosyal medyada reklam amacıyla tütün ürünlerinin kullanılması ve görüntülerine yer verilmesi yasaklanmıştır (4207 Sayılı Kanun, 1996/2018). Fakat ihlallerin tespit edilmesinde sosyal medyanın özelliklerinden kaynaklanan zorluklar söz konusudur.

2018-2023 Tütün Kontrol Strateji Belgesi ve Eylem Planı’nında tütün şirketlerinin reklam yasağı nedeniyle sosyal medya üzerinden reklam ve tanıtım yaptığına değinilmiş olup “Reklâm, Promosyon ve Sponsorluğun Önlenmesi” amacına yönelik belirlenen stratejilerden biri “Tütün endüstrisinin, gençlerin tütün ürünü kullanmalarını teşvik eden ve tütün ürünlerine erişimlerini kolaylaştıran kampanya ve programlarının tespit edilmesine yönelik bir sistem geliştirilmesi”dir (Sağlık Bakanlığı, 2018). Maalesef bu strateji için belirlenen hedeflere ulaşılamayıp böyle bir sistem geliştirilmediği anlaşılmaktadır (Bkz. Ek: Tarım Bakanlığı Bilgi Edinme Başvurusu Cevabı).

Sosyal medya içeriklerinin üretim miktarı ve tüketim hızı göz önüne alındığında, ticari amaçla veya bireysel olarak tütün kullanımını teşvik eden içeriklerin tespiti çok zordur. Tespitteki bu zorluk hem sosyal medyadaki tütün tanıtımlarının izlenmesini hem de sağlığın korunması ve geliştirilmesi için doğru müdahalelerinin geliştirilmesini

zorlaştırmaktadır. Her ne kadar yasada yer verilmişse de hantal bürokratik kurumlar tarafından şikayete dayalı olarak yapılacak tekil incelemeler, faili tespit edilebilirse uygulanacak cezalar ve erişim engellemelerinin soruna çözüm olmayacağı açıktır. Sosyal medyada tütün kullanımının teşviki, tütün bağımlılığının birincil önlemesinde önemli bir açık olarak karşımızda durmaktadır.

İnsan zekası ile başarılabilecek görevleri daha hızlı, daha yüksek miktarda ve sürekli gerçekleştirmeyi mümkün kılan yapay zeka teknolojileri bu sorunun çözümüne katkı sağlayabilir. Son yıllardaki gelişmelerle yapay zeka teknolojileri doğal dil (insan dillerini) anlama, metin sınıflandırma, çeviri, özetleme, soru cevaplama gibi geniş bir yelpazeye yayılan görevlerde çok başarılı hale gelmiştir (Kahn, 2021; Wolf ve ark., 2020). Bu bağlamda, sosyal medyada tütün kullanımını teşvik eden içeriklerin tespiti de bir metin sınıflandırma problemi olarak ele alınabilir. Yapay zeka bu içerikleri başarılı olarak tespit edebilirse, tütün bağımlılığının sürveyansı, önlenmesine ve sağlığın geliştirilmesi faaliyetlerine hizmet edebilir.

Amaç

Bu araştırmanın amacı sosyal medyada tütün kullanımını teşvik eden içerikleri otomatik olarak tespit edebilecek bir yapay zeka geliştirmektir. Bu amaç doğrultusunda araştırmanın hedefleri:

1. Sosyal medyada tütün kullanımının teşvikinin farklı şekillerini anlamak ve tanımlamak,
2. Bu tanımı kullanarak yüksek sayıda sosyal medya içeriğinin tütün kullanımını teşvik etme durumunu içeren bir veri seti oluşturmak,
3. Oluşturulan veri setini kullanarak içerikleri başarılı olarak sınıflandırabilen bir makine öğrenmesi modeli eğitmek,
4. Geliştirilen makine öğrenmesi modelini kullanarak bir aylık süre zarfında tütün kullanımını teşvik eden tüm sosyal medya içeriklerin tespit etmek ve bunun bağımsız değişkenlere göre dağılımını incelemektir.

2. GENEL BİLGİLER

2.1. İnfodemiyoloji

Epidemiyoloji geleneksel olarak sağlık çalışanları tarafından toplanan verilere dayanarak yapılmaktayken, günümüzde bilgi teknolojilerinin günlük hayatın her alanını kuşatmasıyla insanların bıraktığı dijital izlerin de epidemiyolojik açıdan bir değer taşıdığı fark edilmiştir (Salathé ve ark., 2012).

Sosyal medya ve internet arama motorları başta olmak üzere dijital olarak erişilebilen veriler sağlıkla ilgili durumların dağılımı hakkında gerçek zamanlı bilgi üretmeye olanak imkan sağlamaktadır. Bu verilerin özellikle alışılmış sürveyans yöntemlerinin yetersiz kaldığı alanlarda tamamlayıcı rol oynayabileceği düşünülmektedir (Salathé ve ark., 2012).

İnternetteki bilgilerin kullanıldığı bu yeni epidemiyoloji yaklaşımının öncülerinden Eysenbach "infodemiyoloji" kavramını ortaya atmış ve "nihayetinde halk sağlığı ve kamu politikalarını yönlendirmek amacıyla özellikle internet olmak üzere elektronik bir ortam veya evrendeki bilginin dağılımının ve belirleyicilerinin incelenmesi bilimi" olarak tanımlamıştır (Eysenbach, 2009). İnfodemiyolojinin sürveyans amacıyla kullanılmasını da "infosürveyans" olarak adlandırmıştır. Daha sonraları Salathé daha geniş bir yaklaşımla, başka amaçlarla üretilen verilerin epidemiyolojide kullanılmasını "dijital epidemiyoloji" olarak adlandırmıştır (Salathé, 2018).

Güncel bir derlemeye göre her geçen yıl daha fazla sayıda araştırma başta bulaşıcı hastalıklar ve davranışsal risk faktörleri olmak üzere muhtelif sağlık olay ve durumlarını dijital veri kaynaklarını kullanarak incelemektedir (Shakeri Hossein Abad ve ark., 2021).

2.2. Twitter ve Başlıca Özellikleri

Twitter, araştırmacılar tarafından veri kaynağı olarak sıklıkla kullanılan bir sosyal medya sitesidir. Araştırmacılar tarafından çok tercih edilmesinin nedenleri, açık, temiz ve iyi belgelenmiş bir uygulama programlama arayüzüne ve zengin çeşitlilikteki geliştirici araçlara sahip olmasıdır. Bunların yanında Twitter verisini özellikle ilgi çekici kılan ise kullanıcıların düşüncelerinin gerçek zamanlı olarak kamuya açık hale gelmesi ve uluslararası zeminde toplumun geniş kesimlerini çok yönlü olarak temsil etmesidir (Russell, 2013, p. 5).

Tweet, Twitter platformunda yayınlanan her bir içeriğe verilen isimdir. Her tweet kendine özgü bir tanımlayıcı numara, metin ve oluşturulma zamanı gibi temel özelliklere sahiptir. Tweet'ler bir kullanıcı (user) tarafından oluşturulur. Resim ya da video gibi görseller (media), bağlantılar (url), kişi (mention) veya konu etiketleri (hashtag) içerebilir. Bazı tweetler daha önce gönderilmiş başka tweet'lere cevap niteliği taşır (reply). Bir tweet'in, asıl yayınlayan kullanıcıdan başka bir kullanıcı tarafından tekrar yayınlanmasına "retweet" ismi verilir.

Tüm bu özelliklerle birlikte tweet'lere ait beğenilme (favorite), yeniden yayınlanma (retweet) ve alıntılanma (quote) sayısı gibi istatistiklerin tamamı tweet'le ilişkili olarak Twitter sisteminde saklanır (Twitter, 2021b). Bu istatistikler bir tweet'in ne kadar ilgi çekici ya da popüler olduğuna dair bir ipucu verir (Russell, 2013, p. 27).

Twitter'da tweet yayınlamanın dışında diğer kullanıcılarla özel sohbetler yapmak için doğrudan mesaj (direct message) özelliği kullanılabilir. Bir doğrudan mesajın içinde metne ek olarak resim veya video da yer alabilir (Twitter, 2021a).

2.3. Yapay Zeka

Yapay zeka, kapsayıcı bir tanımla "hesaplama modellerini kullanarak zihinsel yetileri çalışan bilim dalıdır" (Charniak ve McDermott, 1986, p. 6). Bu alandaki çalışmaların başlıca amacı basitçe düşünen bir bilgisayar üretmektir. Başka bir ifadeyle beyinin işlevlerini bir hesaplama yöntemi olarak bilgisayara öğretmeyi amaçlar. Bu amacın detaylarına odaklanan doğal dil işleme gibi alt çalışma alanları yanında, bu amacı başka özelliklerle genişleten robotik gibi alt çalışma alanları da bulunmaktadır. Yapay zeka teknolojilerinin her sektörde birçok uygulama alanı bulunmaktadır; tıbbi tanı koyma, birbiriyle uyumlu bilgisayar bileşenlerinin konfigürasyonu, matematiksel işlemler, maden yataklarının varlığını tahmin etme, görüntü işleme ile varlık tespiti, elektronik donanım tasarımlarını doğrulama, tavla ya da satranç oynama bunlardan bazılarıdır. Yapay zeka teknolojilerinde bazı uygulamalar aynı işi yapan insanların performansından daha yüksek performans gösterecek şekilde ilerlemişken bazı alanlarda bu seviyeye henüz ulaşılamasa da sürecin bütününe dikkate değer katkı sağlayabilmektedir (Charniak ve McDermott, 1986, pp. 1–3).

Yapay zeka ifadesi, "zeka" kelimesini içerse de yapay zeka çalışmalarının amacı bilgisayarı yüksek bir zekaya ulaştırmaktan çok sıradan görevleri yerine getirebilecek zihinsel kapasiteyi sağlamaktır. Bu alanda üzerinde en çok durulan ve ilgi çekici konular

doğal dil ve görüntü gibi insanların çok büyük kısmının sahip olduğu sıradan zihinsel yetilerdir (Charniak ve McDermott, 1986, p. 6).

2.4. Makine Öğrenmesi

Makine öğrenmesi**,** bilinmeyen bir kurala, bir genellemeye ulaşmak veya bir veri setinde bulunan örüntüleri anlamak için bir model geliştirme işlemidir. Makine öğrenmesi çalışmaları, bilgisayarların açıkça programlanmalarına gerek olmaksızın öğrenmelerini sağlar (Hackeling, 2014, p. 8). Başlıca iki tür makine öğrenmesi yaklaşımı vardır: gözetimsiz ve gözetimli öğrenme.

Sonuçta doğru çıktının ne olduğu verilmeksizin veri noktalarının dağılımından hareketle algoritmanın verideki desenleri tespit etmesine dayanan makine öğrenmesi yaklaşımına "gözetimsiz öğrenme" adı verilir. Bunun için genellikle kümeleme adı verilen ve verileri benzerliklerine göre gruplara ayırmaya yarayan algoritmalar kullanılır. Örneğin benzer müşterileri gruplamak gibi problemler için gözetimsiz öğrenme tercih edilir (Taulli, 2019, p. 52).

Her bir gözlemin beraberinde istenen çıktının doğru değerinin (etiket) de verildiği bir veri seti kullanılarak öğrenme modeli oluşturulmasına "gözetimli öğrenme" adı verilir. Gözetimli yöntemle eğitilen bir model, yeni gözlemler için doğru çıktı değerlerini tahmin edebilir. Gözetimli öğrenmede genellikle sınıflandırma algoritması olarak adlandırılan algoritmalar kullanılır. Örneğin, bir e-postanın dolandırma amaçlı olup olmadığının belirlenmesinde gözetimli öğrenme kullanılır (Bird ve ark., 2009, p. 222). Gözetimli öğrenmede çok miktarda veri olması modelin daha doğru sonuçlar üretmesine katkıda bulunur. Bu yöntemle model oluşturmanın en büyük zorluğu çok miktarda etiketlenmiş veri bulunması ya da oluşturulmasıdır (Taulli, 2019, pp. 50–51). Bu araştırmada da gözetimli öğrenme yöntemi kullanılacaktır.

Bir makine öğrenmesi modelinin ne kadar iyi çalıştığı başarı değerlendirmesiyle anlaşılır. Gözetimli öğrenmede, makine öğrenmesi modelini oluşturmak için kullanılan veri "eğitim verisi" olarak adlandırılır (Taulli, 2019, p. 184). Tüm veri seti içinden verinin bir kısmı eğitim veri seti olarak belirlenirken bir kısmı da oluşturulan modelin tahmin başarısını değerlendirebilmek için test veri seti olarak ayrılır (Bird ve ark., 2009, p. 222).

Gözetimli öğrenme modellerinin başarısı, test veri setindeki her bir gözlem için çıktı değişkeninin gerçek değerleri ile model tarafından yapılan öngörü değerlerin

karşılaştırılmasıyla değerlendirilir (Bird ve ark., 2009, pp. 237–238). Değerlendirmede sıklıkla kullanılan ölçütler doğruluk, pozitif prediktif değer, duyarlılık ve F_1 skorudur. F_1 skoru, pozitif prediktif değer ve duyarlılığın harmonik ortalaması ile bulunan değerlendirme ölçütüdür (Bird ve ark., 2009, p. 240). Özellikle pozitif prediktif değer ve duyarlılığın dengesiz olduğu durumlarda araştırmacıya daha iyi bir değerlendirme imkanı sağlar (Hackeling, 2014, p. 81).

2.5. Doğal Dil İşleme

Doğal dil işleme, yapay zekanın bir alt dalıdır. Artık gündelik hayata girmeye başlayan sanal asistanlar (Örneğin; Siri, Alexa, Cortana…) doğal dil işleme teknolojilerinin en yaygın ve başarılı örnekleri arasındadır. Doğal dil işleme çalışmaları kelimeleri ve cümleleri yorumlayabilecek kuralları ortaya koyma amacıyla başlatılmıştır. İlerleyen zamanlarda dil verisinin yapılandırılmamış, yorumlanması zor ve tahmin edilemez olması bu alandaki çalışmaları, kurallarını tek tek bulmak yerine dili bir bütün olarak ele almaya yöneltmiştir (Taulli, 2019, p. 106).

Doğal dil işleme çalışmaları bir metindeki mekanlar, kişiler, kurumlar gibi unsurların tespit edilmesi (named entity recognition) (Galitsky, 2019, p. 375; Taulli, 2019, p. 110) veya duygunun tespit edilmesi (sentiment analysis) (Taulli, 2019, p. 52) veya metinlerin önceden belirlenmiş kategorilere ayrılması (text classification) (Bird ve ark., 2009, p. 221) gibi çok çeşitli görevleri konu edinir.

İstenmeyen e-postaları ayıklamak veya bir haber metninin konusunu belli bir konu listesi içinden seçmek gözetimli metin sınıflandırmasının kullanımlarına örnektir (Bird ve ark., 2009, p. 222).

Bu araştırmada da verilen bir sosyal medya içeriğinin tütünü teşvik edip etmediğinin tespit edilmesi bir metin sınıflandırması problemi olarak ele alınmıştır.

Metin sınıflandırması yapılırken dil verisinin hangi özelliklerinin sınıflandırma amacıyla kullanmaya en uygun olduğu incelenir. Bunun üzerinden dil işlemeyi otomatik olarak yapabilecek dil modelleri geliştirilir (Bird ve ark., 2009, p. 221). Basit bir yaklaşımla bir metinde geçen tüm kelimeler bir gözlemin özellikleri gibi incelenebileceği gibi, daha ileri yöntemlerde bir kelimenin kendisinden önce ve sonra gelen kelimelere bakılarak o kelimenin isim veya sıfat olma olasılıkları ile ilgili çıkarım yapılabilir,

bağlamın anlaşılması için kelimelerin metnin geneliyle ilişkisi de değerlendirmeye alınabilir (Bird ve ark., 2009, p. 230).

2.6. Metin Sınıflandırmada Kullanılan Yöntemler

Metin sınıflandırmasında kullanılan yöntemler üç ana aşamada incelenebilir: ön işleme, tahminde kullanmak için metinde özellik çıkarma ve çıkarılan özellikleri kullanarak metnin ait olduğu sınıfı tahmin etme.

2.6.1. Ön İşleme

Metin sınıflandırma işlemlerinin ilk aşamasıdır. Bu aşamada metinler; kelimelerine ayırma, bağlaçlar gibi gereksiz kelimeleri ayıklama, tüm harfleri küçük harflerle değiştirmek, kısaltmaları düzeltme, yazım yanlışlarını düzeltme, kelimeleri kök hallerine dönüştürme gibi bir dizi işlemden geçirilir, kelimeler tektipleştirilir ve metin temel yapıtaşlarına ayrılır (Kowsari ve ark., 2019).

2.6.2. Özellik Çıkartma

İkinci aşama kelimeler kullanarak üzerinde matematiksel işlem yapılabilecek özellikler (değişkenler) çıkartmaktır. Başlıca özellik çıkartma yöntemleri şunlardır:

Bag of Words (kelime torbası) yaklaşımında bir metin içerdiği kelimelerin toplamı olarak temsil edilir. Bu yaklaşımda kelimelerin anlamları veya sırası dikkate alınmaksızın kelimeler bir torbaya atılmış gibi değerlendirildiğinden sınırlı bilgi kullanan bir yaklaşımdır. Bir külliyattaki (incelenen tüm metinler, corpus) tüm kelimelerin olduğu bir dağarcık çıkarılır ve her bir kelime bir özellik olarak kabul edilir. Sonra her bir metinde her bir özellik için kelimenin bulunup bulunmamasına göre dikotomik veya sıklığına göre tam sayılı bir değer verilir (Hackeling, 2014, pp. 52–54). Sütunları kelime dağarcığındaki kelimeler ve satırları külliyattaki metinlerden oluşan bir kelime-metin matrisi oluşturulur (Jurafsky ve Martin, 2020). Böylece her bir metin kelime sayısı kadar boyutu olan bir uzayda bir vektöre dönüşür. Sonuçta Öklid uzaklığı olarak birbirine daha yakın vektörlerin (metinlerin) birbirine daha benzer olduğu kabul edilerek sınıflandırma yapılır (Hackeling, 2014, pp. 52–54). Bu yaklaşımda kelimelerin sırası dikkate alınmaması en önemli kısıtlılıktır. Özellikler tanımlanırken, her bir özelliğin tek bir kelimeden oluşması yerine art arda gelen kelimelerin 2'li, 3'lü birliktelikleri şeklinde oluşturulursa (ki bunlara "n-gram" adı verilir) bu kısıtlılık kısmen azaltılır ve metin daha iyi temsil edilmiş olur.

TF-IDF (Term Frequency - Inverse Document Frequency: Terim Sıklığı - Ters Doküman Sıklığı), yaklaşımı da benzer şekilde bir kelimenin bir metinde tekrarlama sıklığının; o kelimenin, o metin ile ne ölçüde ilgili olduğunu göstereceği varsayımına dayanır. Fakat kelime torbası yaklaşımından farklı olarak TF-IDF'de kelimenin metindeki sıklığı, kelimenin içinde geçtiği metin sayısına bölünür (Jurafsky ve Martin, 2020). Böylece her bir kelimenin metin için değeri daha iyi temsil edilmiş olur.

Fakat bu yöntemlerde kelimelerin anlamları ve bağlamları dikkate alınmamaktadır. Anlamı da dikkate almaya çalışan doğal dil işleme uygulamalarında kelime gömme yöntemi kullanılır.

Word Embedding (kelime gömme), kelimelerin anlambilim tarafından tanımlanan eş anlamlılık, benzerlik, ilişkililik, duygusal anlam (çağrışım) gibi bağlantılarını dikkate alan bir yöntemdir. Aynı anlama işaret eden kelimeler için eş anlamlılık, eş sesli kelimeler için anlam ayrımı, 'kedi' ve 'köpek' gibi farklı anlamda ama benzer kelimeler için benzerlik, 'kahve' ve 'fincan' gibi isimlendirildikleri şeyler ilişkili olan ya da aynı anlamsal alanda yer alan kelimeler için ilişkililik, 'masum' (olumlu çağrışım) ve 'saf' (olumsuz çağrışım) gibi anlamsal benzerlikleri olan ama farklı çağrışımlara sebep olan kelimeler için çağrışımlar bu yaklaşımda temsil edilir. (Jurafsky ve Martin, 2020). Bu ilişkiler önceden çok büyük külliyatlar üzerinde ileri algoritmalar ve denklemler kullanılarak hesaplanır ve her bir kelime için bir vektör tanımlanır. Kelimeler üzerinde işlem yapılabilir hale gelmiş olur. Örneğin, vektör(kral) - vektör(erkek) + vektör(kadın) ≈ vektör(kraliçe). Kelime gömme yaklaşımında benzer kelimeler birbirlerine yakın vektörler oluşturur (Bolukbasi ve ark., 2016). Böylece sınıflandırma işleminde kullanılabilirler.

Bu araştırmada hazır olarak kullanılan dil modeli de özelliklerin kelime gömme yaklaşımıyla oluşturulduğu bir modeldir.

2.6.3. Sınıflandırma Algoritmaları

Çok çeşitli sınıflandırma algoritmaları vardır, burada örnek olarak biri basit biri daha karmaşık olan iki algoritma ele alınacaktır.

Naif Bayes Metin Sınıflandırması'nda, bir metnin bir sınıfa ait olma olasılığı Bayes teoreminde belirtilen koşullu olasılık şeklinde hesaplanır. Eğitim veri setindeki etiketlerin sıklıkları Bayes teoremindeki ön olasılıklar olarak kullanılır ve özelliklerin

modele katkıları bu ön olasılığa bağlı olarak hesaplanır (Bird ve ark., 2009, p. 246). Naif olarak adlandırılmasının sebebi değişken olarak kelimelerin bağımsız olduğu varsayımına dayanmasıdır. Eğitilmesi görece basittir (Taulli, 2019, pp. 55–57).

Yapay Sinir Ağları (artificial neural networks), insan beynindeki nöronların çalışma şekline benzer şekilde çalıştıkları için bu şekilde adlandırılmışlardır. Tek bir nöron, bir çıktı üretmek için bir veya daha fazla girdiyi işleyen bir hesaplama birimi olarak düşünülebilir. Bu yöntemde bilgisayarda algılayıcı (perceptron) adı verilen benzer birimler kurulur. Fakat yalnızca bir algılayıcı birçok temel işlev için yetersizdir (Hackeling, 2014, pp. 155–156). Algılayıcılar birbirlerine seri ve paralel bağlanarak bir ağ oluşturulur.

Eğitim veri setindeki her bir metin için önceki aşamada çıkarılmış özellikler (örneğin, kelimelerin sıklıkları) girdi olarak; çıktı değişkeni ise çıktı olarak bu ağa verilir. Ağın amacı, eğitim verilerini kullanarak algılayıcılar arasındaki bağlantıların ağırlıklarını tespit etmektir. Bunun için algoritma önce bağlantılara rastgele ağırlıklar verir, sonra tanımlanmış bir hata formülüne göre ağdaki bağlantıların ağırlıklarını kademeli olarak güncelleyerek optimize eder. Nihayetinde ağ verilecek yeni bir girdiden istenen sonucu tahmin edebilecek hale gelir (Aggarwal ve Zhai, 2012, pp. 197–199).

Özetle, yapay sinir ağları, bir algılayıcılar ağı ve özellik-sınıf çiftlerine en uygun ağırlıkları kullanan bir sınıflandırma modelidir (Aggarwal ve Zhai, 2012, p. 315). Bu araştırmada da “transformer” olarak adlandırılan bir yapay sinir ağı türü kullanılmıştır.

2.6.4. BERT

BERT (Bidirectional Encoder Representations from Transformers) özellik çıkartma için kelime gömme ve eğitim için yapay sinir ağlarının ileri bir türü olan “transformer” yaklaşımı kullanılarak oluşturulmuş ve dili temsil eden bir modeldir. Çift yönlü olması nedeniyle kelime gömmeleri oluştururken kelimelerin hem solundaki hem sağındaki diğer kelimeleri dikkate alarak bağlamı daha iyi temsil eder. Maskeleme adı verilen yöntem kullanması eş sesli kelimeleri birbirinden ayrık, eş ve yakın anlamlı kelimeleri birbirine benzer olarak değerlendirmesini sağlar.

BERT, ön eğitim (pre-training) ve ince ayar (fine-tuning) olmak üzere iki adımda kullanılır. Ön eğitimde bir dildeki çok büyük külliyatlar kullanılarak modelin dili genel olarak temsil edebilir hale gelmesi sağlanır. Bu haliyle model sınıflandırma yapamaz. İnce ayar ise görece daha küçük eğitim verileriyle gerçekleştirilir ve artık dili genel olarak

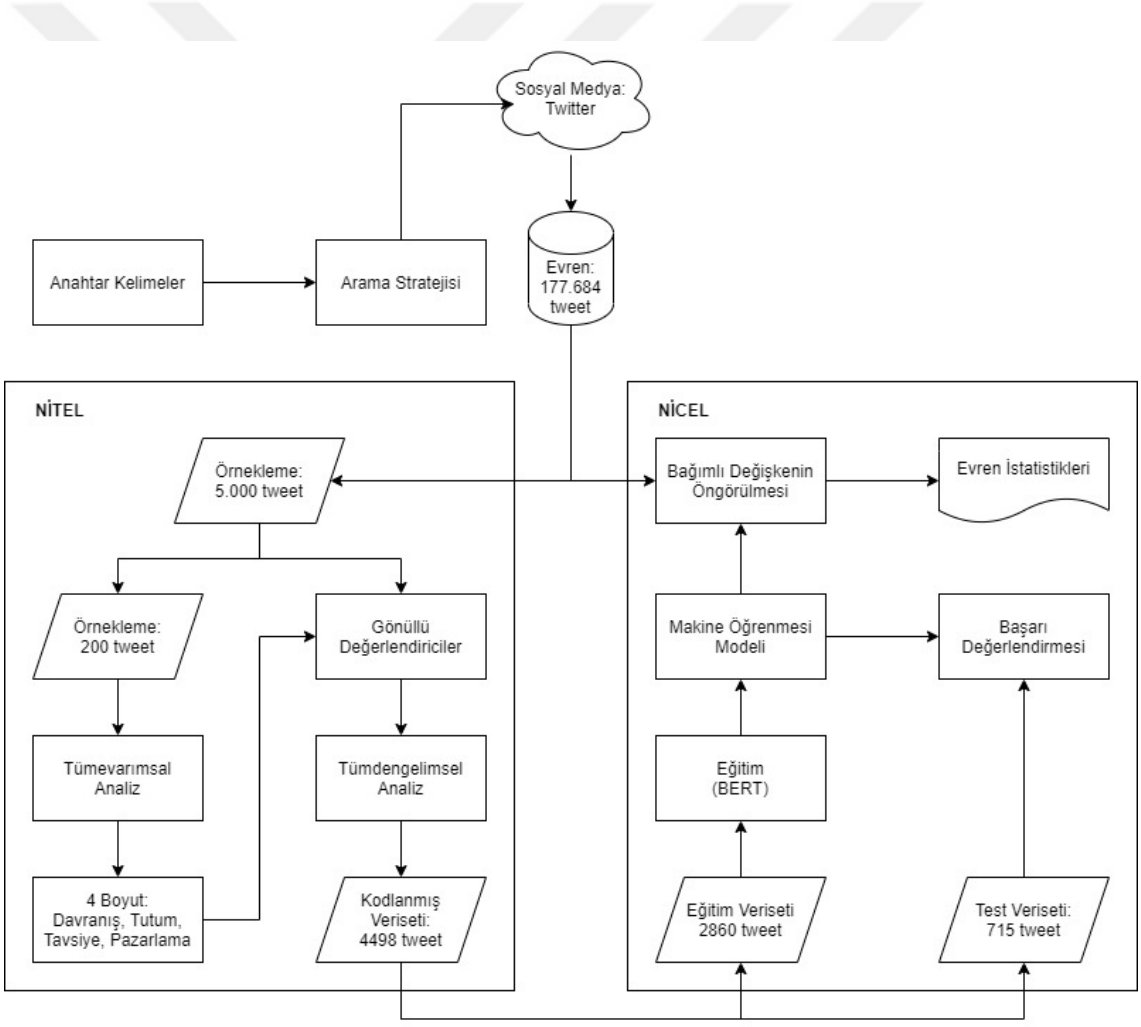
“anlamış” modelin spesifik bir sınıflandırma görevini öğrenmesini sağlar. BERT’in bu yapısı, uzun süren ve fazla hesaplama gücü gerektiren ön eğitimden bir kere geçmiş bir modelin, ince ayarlar ile farklı görevlerde kullanılmasına olanak sağlar (Devlin ve ark., 2018a).

Bu araştırmada da Türkçe dilinde geniş bir külliyat ile ön eğitimden geçirilmiş bir BERT modeli kullanılmıştır. (Bkz. 3.8.2 Dil Modeli)

3. GEREÇ VE YÖNTEM

3.1. Araştırma Türü

Bu araştırma karma yöntemli bir araştırma olup nitel analizi takip eden nicel analiz şeklinde tasarlanmıştır (QUAL → QUAN) (Schoonenboom ve Johnson, 2017). Öncelikle içerikler insan değerlendiriciler tarafından evrenden seçilen bir örnek üzerinde içerik analizi yöntemiyle sınıflandırılmıştır. Sonrasında yapılan sınıflandırma makine öğrenmesi yöntemleriyle bir modele dönüştürülerek evrendeki tüm içerikler için öngörüde bulunulmuş ve evren istatistikleri kesitsel yöntemle değerlendirilmiştir. Araştırma sürecinin detaylı akış şeması Şekil 3-1'de gösterilmiştir.



Şekil 3-1. Araştırma süreci

3.2. Evren ve Örnekleme

Araştırmanın evreni Twitter sosyal medya platformunda Türkçe dilinde 2 Ekim – 31 Ekim 2020 tarihleri arasında halka açık olarak yayınlanmış tütün ile ilgili en az bir anahtar kelime içeren tweet'lerdir.

Türkiye'de 13,6 milyon kişi tarafından kullanılan popüler bir sosyal medya platformu olması (Statista, 2021) ve bilimsel araştırma amaçlı veri toplamayı kolaylaştırıcı özellikler (Twitter, 2012) taşıması dolayısıyla Twitter veri kaynağı olarak tercih edilmiştir.

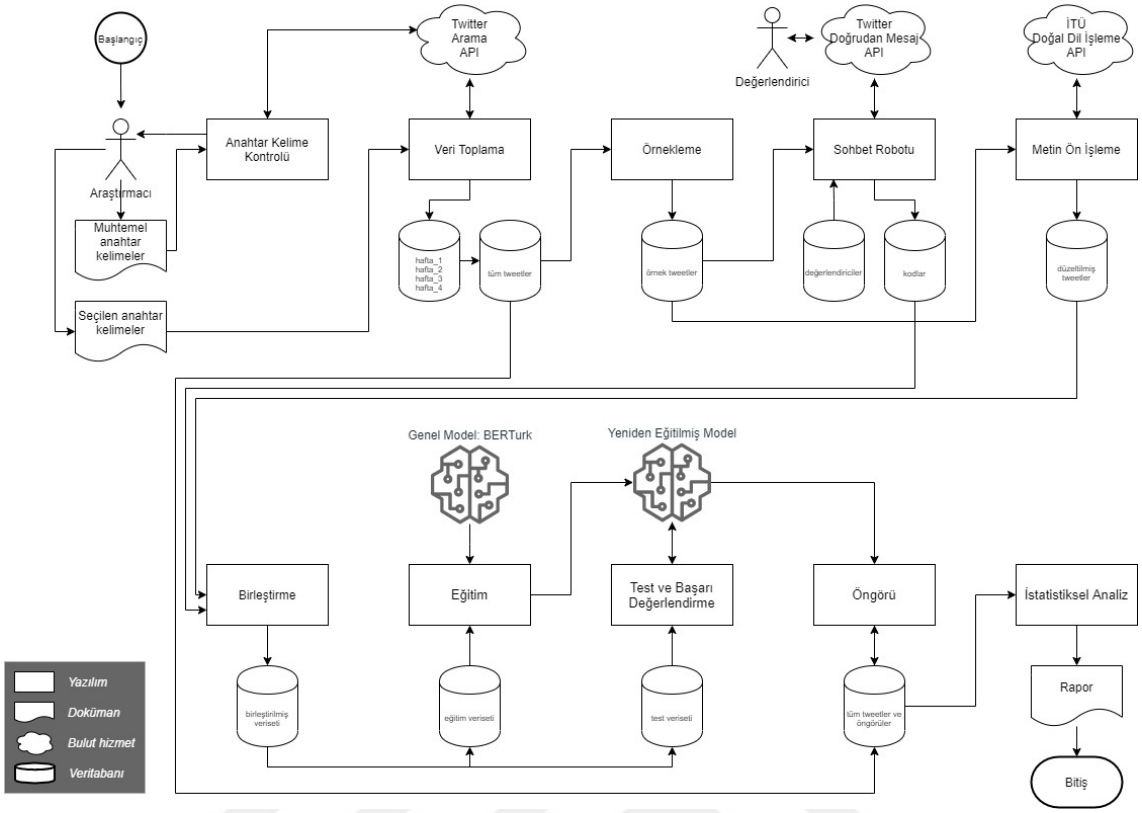
Literatürde makine öğrenmesi veya doğal dil işleme uygulamaları için mutabık olunan bir örnek büyüklüğü belirleme yöntemi bulunamamış olup önceki araştırmalarda (Küçükali ve ark., 2020) birkaç bin içeriğin değerlendirilmesi bir makine öğrenmesi modeli geliştirmek için yeterli görüldüğünden, olası kayıplar da göz önüne alındığında örnek büyüklüğü 5000 tweet olarak belirlenmiştir. Bu örnekte yer alacak tweet'ler her gün için yayınlanan tweet sayısıyla ağırlıklandırılarak tabakalı örnekleme yöntemiyle seçilmiştir.

3.3. Araştırma için Geliştirilen Yazılımlar

Araştırmacı tarafından, araştırmanın her bir aşaması için Python programlama dilinde yazılımlar geliştirilmiştir.

Python**,** yaygın kullanılan diğer programlama dillerine göre söz dizim ve gramer açısından yalın olmasının yanı sıra geliştirici ekosisteminin genişliği ve öğrenme kaynaklarının çokluğu sayesinde güçlü bir dil olma özelliği taşır (Taulli, 2019, p. 151). Özellikle makine öğrenmesi alanında Python dili için geliştirilmiş açık kaynaklı kütüphaneler, bu alanda en çok tercih edilen dillerden biri haline gelmesini sağlamıştır. Python'da geliştirilen yazılımların verimlilik ve kalitesi yüksek, bakımı kolaydır (Bird ve ark., 2009, p. XIII).

Bu yazılımların iş akışı, bulut hizmetleri ve veri tabanı ile bağlantıları Şekil 3-2'de gösterilmiştir. Ayrıca yazılımların geliştirilmesinde faydalanılan açık kaynak kodlu yazılım kütüphanelerine ekte yer verilmiştir.



Şekil 3-2. Araştırma kapsamında geliştirilen tüm yazılımların iş akışı, bulut hizmetler ve veri tabanı ile bağlantıları

Şekil 3-2'de gösterildiği gibi bazı yazılımlar çeşitli bulut hizmetlerine kısaca API olarak adlandırılan Uygulama Programlama Arayüzleri ile bağlanmaktadır. Uygulama Programlama Arayüzü (Application Programming Interface veya API), yazılımların diğer bir bilgisayar sistemi veya uygulamalarının özelliklerine veya verilerine erişimini sağlamak için oluşturulmuş fonksiyonlar ve süreçlere verilen isimdir (Oxford Dictionary, 2020).

3.4. Veri Toplama

Veriler Twitter'dan veri madenciliği (Russell, 2013) yöntemiyle toplanmıştır. Veri toplama aracı olarak, araştırmacı tarafından Python programlama dilinde geliştirilen bir yazılım kullanılmıştır (Şekil 3-2). Yazılım, yetkilendirilmiş kişilere Twitter'da herkese açık olarak yayınlanan tweet'lere erişme olanağı sağlayan API aracılığı ile veri toplamaktadır. Bunun için Twitter'ın sunduğu farklı seçenekler arasından sürüm 1.1'in ücretsiz (Standard) seçeneğinin Arama (Search) API'ı kullanılmıştır (Twitter, 2012).

3.4.1. Arama Stratejisi

Twitter'daki verilerin içinden tütünle ilgili tweet'leri süzebilmek için anahtar kelime belirlenmesi gerekmektedir. Anahtar kelimelerin belirlenmesi ve sadeleştirilmesinde Kim ve ark. (2016) tarafından önerilen yöntem izlenmiştir. Bunun için ilk olarak 3 grup altında anahtar kelimeler belirlenmiştir: (i) tütün ürünlerinin cins isimleri, (ii) geleneksel tütün ürünlerinin marka isimleri, (iii) elektronik sigara marka isimleri.

Tütün ürünlerinin cins isimlerini belirlemek için önceki araştırmalarda kullanılan kelimeler (Lienemann ve ark., 2017) incelenmiş, Türkiye'de yaygın olarak kullanılan türler için Türkçedeki karşılıkları ve farklı yazılış varyasyonları şu şekilde listelenmiştir: tütün, sigara, cigara, cugara, nargile, puro, pipo, e-sigara, esigara.

Geleneksel sigara marka isimlerini belirlemek için Türkiye'de tütün pazarında yaygın olarak bulunan beş üreticinin 34 alt markası yerel bir perakendeciden temin edilmiştir.

Elektronik sigara marka isimleri için uluslararası bir pazar raporunda (Grand View Research, 2020) yer alan 11 üreticinin internet sitelerinde yer alan 18 alt markasının isimleri kullanılmıştır.

Bazı marka isimlerinin çok yüksek oranda alakasız tweet süzülmesine sebep olduğu görülmüştür. Örneğin, bir tütün markası olan "Maltepe" anahtar kelimesi, kelimenin Maltepe ilçesi anlamında kullanıldığı tweetleri de dahil etmektedir. Araştırma kapsamında tütün ile alakalı tweet'lere odaklanabilmek için her bir marka ismi ile Twitter üzerinde tek tek arama yapılmış, ilk on sonuç arasında en az beşinin tütün ürünü anlamında kullanıldığı görülen kelimeler anahtar kelime olarak seçilmiştir. Twitter'da anahtar kelimelerden en az birinin bulunduğu tweet'leri süzebilmek için, kelimeler mantıksal "VEYA" (OR) işlemiyle birleştirilerek bir arama sorgusu yazılmıştır. Nihayetinde aşağıdaki arama sorgusu oluşmuştur:

> *tütün OR sigara OR cigara OR cugara OR nargile OR puro OR pipo OR "e-sigara" OR esigara OR Camel OR Marlboro OR Parliament OR Lark OR Rothmans OR "Pall Mall" OR "Tekel 2000" OR "Tekel 2001" OR "Samsun 216" OR Davidoff OR Juul OR IQOS OR HEETS*

3.5. Verilerin Saklanması

Yazılım, Twitter API'ına belirlenen anahtar kelimelerden oluşturulan arama sorgusunu ve arama ayarlarını göndermekte, yanıt olarak eşleşen tweet'lere ait verileri almaktadır. Standart sürümün kısıtlılıkları dolayısıyla (Lienemann ve ark., 2017, p. 18) araştırma için gereken veriler yazılımın haftada bir çalıştırılması suretiyle 5 defada toplanmıştır. Toplanan veriler analiz edilmek üzere MongoDB türünde bir veri tabanına kaydedilmiştir.

MongoDB verilerin yapılandırılmasında çok yüksek ölçüde esnekliğe izin veren doküman tabanlı bir veri tabanıdır. MongoDB doküman tabanlı olması sayesinde hem yapılandırılmış hem yapılandırılmamış verilerin saklanmasını ve hızlı erişilmesini mümkün kılmaktadır (Taulli, 2019, p. 27). Twitter API'ı gibi modern uygulama programlama arayüzlerinin çoğunun veri transfer notasyonu olan JSON (Javascript Object Notation) türündeki verinin saklanması ve sorgulanması için en iyi veri tabanlarından biridir (Russell, 2013, p. 226).

3.6. Değişkenler

Araştırmada tweet'lere ait aşağıdaki özellikler bağımsız değişken olarak analiz edilmiştir. Değişkenlerin orijinal isimleri de parantez içinde verilmiştir.

- Metin (full_text)
- Yayınlanma zamanı (created_at)
- Görsel içerme durumu (entities.media)
- İnternet bağlantısı içerme durumu (entities.urls)
- Etiket içerme durumu (entities.hashtag)
- Başka bir kullanıcıdan bahsetme durumu (entities.user_mentions)
- Orijinal veya alıntı olma durumu (retweeted_status, is_quote_status)
- Başka bir tweet'e karşılık olarak yayınlanma durumu (in_reply_to_status_id)
- Tweet'in yayınlandığı cihaz ve yazılımın türü (source)
- Tweet'i görebilecek potansiyel kitlenin büyüklüğü (user.followers_count)
- Yeniden yayınlanma sayısı (retweet_count)

- Beğenilme sayısı (favorite_count)

Bağımlı değişken ise "tütün kullanımı teşvik etme durumu" olup dikotomik (var/yok) niteliktedir. Bağımsız değişkenler için veriler Twitter API aracılığı ile toplanırken; bağımlı değişken değerleri örnek tweet'ler için, gönüllü insan değerlendiriciler tarafından metinlerden içerik analizi yoluyla üretilmiştir. Makine öğrenmesi modeli geliştirildikten sonra evrendeki tüm tweet'ler için bağımlı değişken değeri geliştirilen makine öğrenmesi modeli tarafından öngörülmüştür.

3.7. İçerik Analizi

3.7.1. Tümevarım Analizi

Araştırmanın nitel kısmını oluşturan içerik analizi için ilk olarak 5000 tweet'ten oluşan örnek içinden tekrar basit rastgele örnekleme yöntemiyle ile 200 tweet'ten oluşan daha küçük bir örnek seçilmiştir. Tweet'lerde tütün kullanımının hangi şekillerde teşvik edildiği tümevarım analizi yöntemiyle tespit edilmiştir. Bu analizde tütün kullanımını teşvik etme durumunun aşağıdaki 4 temanın farklı kombinasyonları şeklinde ortaya çıkabildiği tespit edilmiştir:

1. Tütün ürünü kullanma davranışı
2. Tütüne yönelik olumlu tutum
3. Tütün ürünü kullanmayı tavsiye etme
4. Belirli bir tütün ürününü pazarlama

Bu temalar 5000 tweet'in tümdengelim analizinde kullanılmak üzere kodlama çerçevesini (coding frame) oluşturmuş (bkz. Tablo 3-2) ve tütün kullanımını teşvik etmenin alt boyutları olarak (davranış, tutum, tavsiye, pazarlama) olarak kabul edilmiştir. "Tütün kullanımını teşvik etme durumu" değişkeni de bu dört alt boyuttan en az birinin mevcut olması olarak tanımlanmıştır.

3.7.2. Değerlendiriciler Arası Tutarlılık

5000 tweet'in tek değerlendirici tarafından analiz edilmesinin değerlendiriciden kaynaklanan bir bias (Lienemann ve ark., 2017, p. 17) doğurup doğurmayacağını tespit etmek için değerlendiriciler arası tutarlılık analizi yapılmıştır. Bunun için, bağımsız olarak iki değerlendirici, seçilen 200 tweet'i, dört alt boyutu içerip içermemesi yönünden analiz etmiştir. Değerlendiriciler arası tutarlılık *Krippendorff's α* istatistiği ile

incelenmiştir. Müstehcen ifadeler içeren 4 tweet hariç tutularak 196 tweet üzerinden hesaplanan α değerleri Tablo 3-1'de verilmiştir.

Tablo 3-1. Bağımlı değişken ve alt boyutları için değerlendiriciler arası tutarlılık

Değişken	α
Tütün ürünü kullanma davranışı	0,56
Tütüne yönelik olumlu tutum	0,32
Tütün kullanma tavsiyesi	0,65
Tütün ürünü pazarlama	0,56
Tütün kullanımını teşvik etme	0,58

Değerlendiriciler arası tutarlığının orta düzeyde kalması nedeniyle 5000 tweet'in her birinin iki ayrı değerlendirici tarafından değerlendirilmesi ve yalnızca tutarlı olan kodlamaların dikkate alınmasına karar verilmiştir. Nitel araştırmalarda geçerlilik "triangülasyon" olarak adlandırılan, yöntem veya değerlendiricilerin çeşitlendirilmesi yoluyla sağlanır. Bu araştırmada da değerlendiriciden kaynaklanabilecek biasları azaltıp ve triangülasyonu sağlamak için, tüm tweet'ler gönüllü değerlendiricilerden oluşan bir havuzdan basit rastgele yöntemle seçilen ikişer kişi tarafından değerlendirilmiştir.

3.7.3. Kitle Kaynağı ve Değerlendiricilerin Dahil Edilmesi

Tweet'lerin yüksek sayısı ve her tweet'in iki değerlendirici tarafından değerlendirileceği göz önünde bulundurularak içerik analizi kitle kaynağı (crowd sourcing) olarak adlandırılan yöntemle gerçekleştirilmiştir (Krauss ve ark., 2015). Merriam-Webster sözlüğüne göre kitle kaynağı: "gereken hizmet, fikir veya içeriklerin, çalışanlar veya tedarikçiler yerine büyük bir insan grubundan ve özellikle çevrimiçi topluluklardan katkı talep ederek elde edilmesi" olarak tanımlanmıştır (Merriam-Webster Dictionary). Bir çok yapay zeka geliştirme projesinde kitle kaynağına başvurulmaktadır. Teknoloji şirketi Amazon'un, insan zekası gereken işlere kitle kaynağı temin etmek için (1769'de Wolfgang von Kempelen'in ürettiği satranç oynayan otomatın isminden ilhamla) Mekanik Türk olarak adlandırılan ücretli bir hizmeti vardır (Amazon, 2021).

Benzer şekilde, bu araştırmada gereken çok sayıda değerlendirme gönüllü değerlendiricilerden oluşan bir kitle kaynağı sistemi kurularak temin edilmiştir.

Yapay zeka teknolojilerinin ham maddesi olan verinin, geliştirilecek yapay zeka modelinin yapacağı göreve uygun olması önemlidir. Bu nedenle gönüllü değerlendiricilerin sosyal medyanın hitap ettiği yaş grubundan, tütünle ilgili farklı boyutları tespit edebilecek okuryazarlığa ve tütün bağımlılığı konusunda duyarlılığa sahip kişiler olmalarının araştırma amacına hizmet edeceği düşünülmüştür. Buna uygun olarak gönüllüler değerlendiricileri dahil etmek için bir üniversitenin Genç Yeşilay Kulübü ile işbirliği yapılmıştır. Kulüp tarafından yapılan duyuru sonrasında projeye ait internet sitesine gelen başvurular arasından dahil etme kriterlerine uygun olan ve katılım şartlarını kabul edenler değerlendirici olarak seçilmiştir.

Değerlendirici olma kriterleri:

- 18 yaşından büyük olmak
- Herhangi bir tütün ürünü kullanmamak ve daha önce kullanmamış olmak
- Tütün endüstrisiyle ilişkisi olmamak
- Twitter hesabına sahip olmak

Araştırmaya 27 gönüllü değerlendirici katılmıştır. Değerlendiricilerin yaş ortalamaları 20,93 ± 2,60 olup, %85,19'u (n=24) kadın, %88,89'u üniversite öğrencisidir (n=24). Öğrencilerin hepsi Genç Yeşilay Kulübü üyesi olup, sağlık (n=14), eğitim (n=6) ve insan (n=4) bilimlerinde lisans düzeyinde eğitim görmektedirler.

3.7.4. Değerlendirici Eğitimi

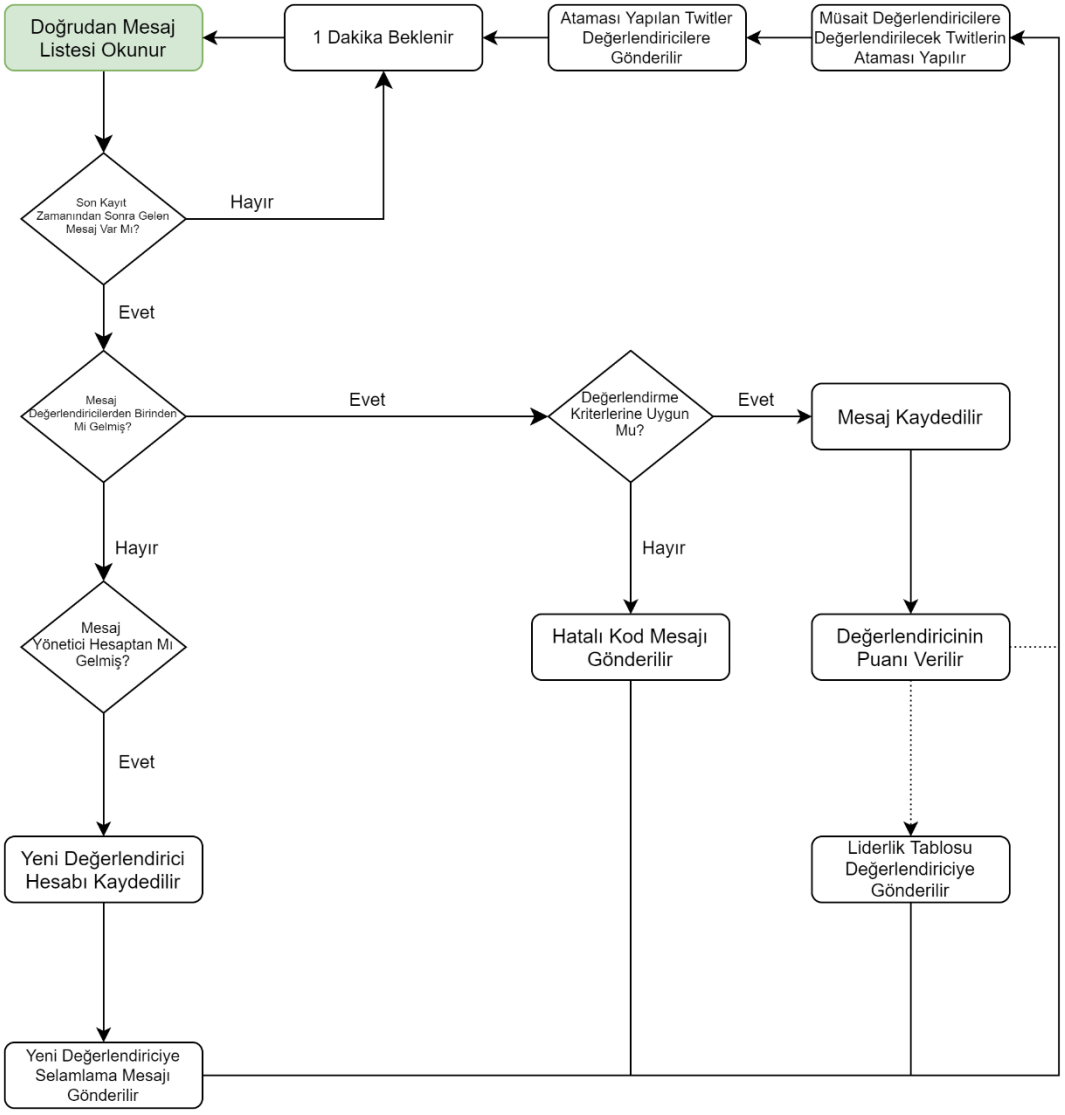
Kriterlere uyan gönüllü değerlendiricilere iki farklı günde yaklaşık ikişer saat uygulamalı eğitim yapılmıştır. Eğitimde projenin amacı ve yöntemi hakkında bilgi verildikten sonra; tweet'lerin kendilerine nasıl ulaştırılacağı, tweet'lerin hangi boyutlar altında değerlendirileceği, her bir boyut için hangi kategorilerde kodlama yapılabileceği ve ihtiyaç duyduklarında araştırmacıyla nasıl iletişime geçecekleri konularında bilgi verilmiştir. Eğitimde örnek tweet'ler üzerinde birlikte kodlama yapılmış, gönüllülerin soruları eğitim sırasında ve sonrasında cevaplandırılmıştır.

3.7.5. İçerik Kodlama Yazılımı ve Sohbet Robotu

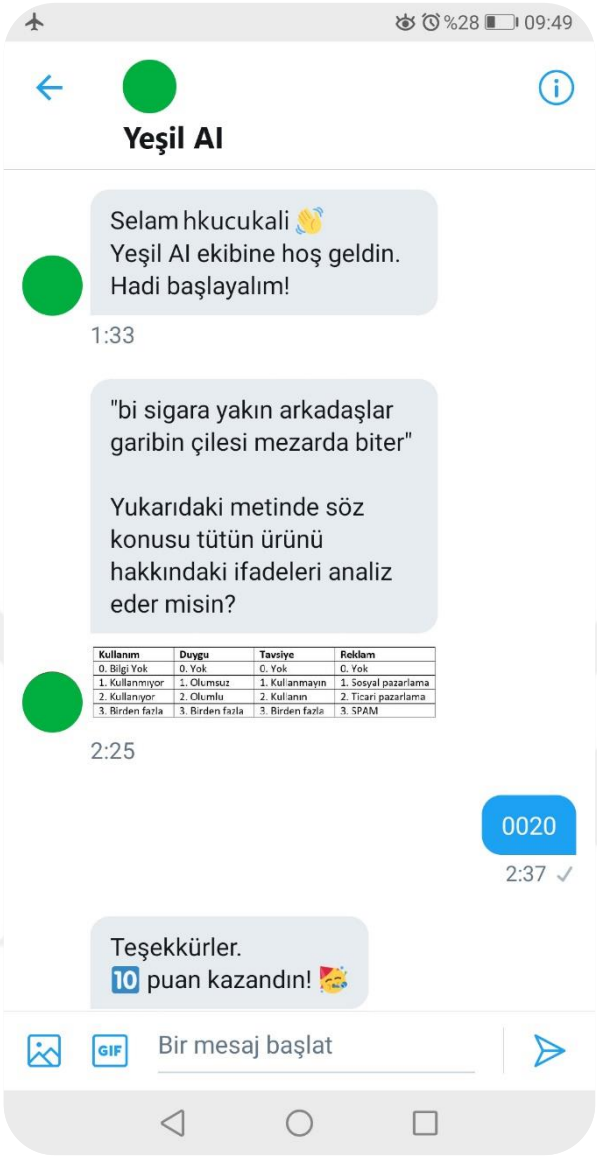
Evrenden seçilen örnekte yer alan tweet'leri gönüllülere rastgele şekilde tahsis etmek ve onlardan gelen cevapları kaydetmek için araştırmacı tarafından Python programlama dilinde geliştirilen bir yazılım kullanılmıştır (Şekil 3-2). Kullanımı kolaylaştırmak için, söz konusu yazılıma basit bir "sohbet robotu" özelliği eklenmiştir.

Sohbet robotu (chatbot, konuşma aracı, diyalog sistemi) sözlü veya yazılı doğal dili kullanarak kullanıcılar ile yazılım uygulaması arasında bir arayüz olarak çalışan bir bilgisayar sistemidir. Sohbet robotları, kullanıcıları sistemin karmaşık komut dizinlerini öğrenmek zorunluluğundan kurtarır. Karmaşık görevler için kullanıcı merkezli bir yaklaşım sağlar (Galitsky, 2019, pp. 13–14).

Bu sohbet robotu Twitter'ın Doğrudan Mesaj sayfası üzerinden gönüllü değerlendiricilerle otomatik olarak iletişim kurabilmektedir. Arkada çalışan yazılım basit rastgele örnekleme ile seçilen tweet'leri değerlendiricilere iletmekte, onların cevaplarını da araştırma veri tabanına kaydetmektedir. Sohbet robotunun iş akışı Şekil 3-3'te şematize edilmiştir. Sohbet robotunun çalışmasından örnek bir ekran görüntüsü ise Şekil 3-4'te verilmiştir.



Şekil 3-3. Sohbet robotu iş akışı



Şekil 3-4. Sohbet robotu içerik atama ve kodlama ekran görüntüsü

Yazılım, gönüllüleri rahatsız etmemesi için, müstehcen ifadeler içeren tweet'leri hariç tutup ve rastgele atamayı ondan sonra gerçekleştirecek şekilde programlanmıştır. Bunun için Türkçe küfürleri içeren açık kaynaklı bir liste (Oğuz, 2018) kullanılmıştır. Bu listedeki kelimeleri içeren tweet'ler (n=502) yazılım tarafından otomatik olarak hariç tutulmuştur.

3.7.6. Oyunlaştırma

Gönüllü değerlendiricileri motive etmek amacıyla süreç içinde oyunlaştırma yöntemleri kullanılmıştır. Gönüllülere yaptıkları her kodlama için puan verilmiştir. En

çok puan alan üç kişinin yer aldığı puan tablosu belirli periyotta hem gönüllüler arasında hem sosyal medyada halka açık olarak ilan edilmiştir. Her hafta en çok puan alan üç gönüllüye (kitap, müzik vb. hediye kartları şeklinde) ödüller verilmiştir. Sohbet robotunda bir oyunlaştırma unsuru olarak puan tablosu kullanımı Şekil 3-5'teki ekran görüntüsünde verilmiştir.



Şekil 3-5. Sohbet robotu oyunlaştırma ekran görüntüsü

3.7.7. Tümdengelim Analizi

Değerlendiricilerden kendilerine yazılım tarafından Twitter'ın Doğrudan Mesaj sayfası aracılığıyla mesaj şeklinde gönderilen tweet'leri okuyarak dört boyutta ve her bir boyut için dört kategoride kodlama yapmaları istenmiştir. Bu kategoriler Tablo 3-2'deki kodlama çerçevesinde gösterilmektedir. Dört boyutun en az birinde tütün lehine bir bulgu olması durumunda "tütün kullanımını teşvik etme durumu" yazılım tarafından "var" olarak, aksi takdirde "yok" olarak kodlanmıştır.

Tablo 3-2. İçerik analizinde kullanılan kodlama çerçevesi

Tütün kullanma davranışı	Tütüne yönelik tutum	Tavsiye etme	Pazarlama
0: Bilgi yok	0: Bilgi yok	0: Bilgi yok	0: Bilgi yok
1: Kullanmama	1: Olumsuz	1: Kullanmamayı tavsiye	1: Tütün ürünü pazarlama
2: Kullanma	2: Olumlu	2: Kullanmayı tavsiye	2: Tütün karşıtı kampanya
3: Her ikisi de	3: Her ikisi de	3: Her ikisi de	3: Her ikisi de

Hatırlatıcı olması için kodlama çerçevesi (Tablo 3-2) her bir mesajın altında tekrar gönderilmiştir (Lienemann ve ark., 2017, p. 17). Zihinsel açısından kodlamayı kolaylaştırmak için her bir boyut için bilgi olmadığını ifade eden kategori 0, tütün kullanımı aleyhine olan kategori 1, tütün kullanımı lehine olan kategori 2, hem lehine hem de aleyhine olduğunu belirten kategori 3 olarak numaralandırılmıştır. Örneğin: "Bugün ilk kez e-sigara içtim, pek bana göre değil, ama merak edenler bir denemeli." gibi bir cümle 2120 şeklinde kodlanacaktır.

3.8. Makine Öğrenmesi

Bağımsız ve bağımlı değişkenler için veri toplandıktan sonra, tüm veriler araştırmacı tarafından Python programlama dilinde geliştirilen yazılımlar ile düzenlenmiş ve analiz edilmiştir (Bkz. Şekil 3-2).

3.8.1. Metin Ön İşleme

Sosyal medya verisi yapılandırılmamış ve gürültülü bir veridir. Sosyal medya içeriklerinin metinlerinde sıklıkla eksik veya fazla harfler gibi yazım hataları bulunur. Bu nedenle makine öğrenmesi öncesinde metinler, söz konusu yazım yanlışlarının düzeltilmesi; bağlantı, konu etiketi, bahsetme gibi çoklu medya unsurlarının gizlenmesi gibi bir dizi işlemden geçirilerek formal yazı diline uygun hale getirilmiştir. Yazım yanlışlarının düzeltilmesi için İstanbul Teknik Üniversitesi Doğal Dil İşleme araştırma grubu tarafından geliştirilen API'ın "normalleştirme" (normalization) ismi verilen fonksiyonu kullanılmıştır (Eryiğit ve Torunoğlu Selamet, 2017).

Normalleştirme fonksiyonu üç bileşenden oluşmaktadır. Bunlardan ilki, Türkçe karakterler yerine benzer ASCII (American Standard Code for Information Interchange) karakterler kullanılan durumlarda kelimeyi doğru, yani Türkçe karakterli formuna dönüştürme bileşenidir. İkincisi özellikle sosyal medyada iletilerin kısaltılması için sesli harflerin eksik yazılması durumlarında, eksik sesli harfleri tamamlama bileşenidir. Üçüncüsü ise imla düzelticisidir (Eryiğit ve Torunoğlu Selamet, 2017).

3.8.2. Dil Modeli

Araştırmanın ana hedefi metin verisini kullanarak bağımlı değişkenlerin öngörülmesidir. Literatürde benzer amaçlarla kullanılmış diğer algoritmalar incelenmiş ve sınırlılıkları değerlendirilmiştir (Küçükali ve ark., 2020). Bunun üzerine Google tarafından 2018 yılında geliştirilen, literatürde kullanılan diğer yöntemlerden daha yeni ve doğal dil işleme görevlerinde daha başarılı olduğu gösterilmiş olan BERT yöntemi tercih edilmiştir (Devlin ve ark., 2018b).

Bu araştırmada, Bavyera Eyalet Kütüphanesi adına Stefan Schweter tarafından 2020'de Türkçe dilinde 35 GB'lık bir veri ile ön eğitimden geçirilmiş, 128 bin kelimelik dağarcığı olan, açık kaynaklı BERT modeli kullanılmıştır (Schweter, 2020). BERTurk isimli bu genel amaçlı model, araştırmanın tütün kullanım teşvikini saptama amacına uygun olarak tekrar eğitilmiştir.

3.8.3. Eğitim ve Test

Bu işlem için araştırmacı tarafından Python programlama dilinde bir yazılım geliştirilmiştir (Şekil 3-2). Veri seti, bağımlı değişkenin kategorilerine göre tabakalı

rastgele örnekleme yapılarak literatürde (Küçükali ve ark., 2020) önerildiği gibi 4:1 oranında eğitim (n=2860) ve test (n=715) veri setine ayrılmıştır.

Eğitim veri seti algoritmaya sokularak model tekrar eğitilmiştir. Algoritmanın binlerce metni işlemesi sıradan bir kişisel bilgisayarda uzun zaman alacağı için yazılım Google şirketinden kiralanan grafik işlemcili (Graphics Processing Unit veya GPU) sunucular üzerinde çalıştırılmıştır.

Daha sonra model, eğitim işleminde karşılaşmadığı verilere, yani test veri setine, uygulanarak bu veri setindeki her bir tweet için bir değer öngörülmüştür. Makine öğrenmesi modellerinin başarısı test veri seti için öngörülen değerlerin, altın standart olarak insan değerlendiriciler tarafından belirlenen değerlerle kıyaslanmasıyla belirlenmiştir.

"Tütün kullanımı teşvik etme durumu"nu öngörmek için açıklanan bu eğitim ve test işlemi önce değerlendiriciler tarafından yapılan tüm kodlamalar dahil edilerek, daha sonra yalnızca tutarlı kodlamalar dahil edilerek gerçekleştirilmiş ve iki ayrı model geliştirilmiştir.

Araştırmanın asıl amacı "tütün kullanımı teşvik etme durumu"nu belirlemek olsa da istenildiği takdirde dört alt boyutun bağımsız olarak değerlendirilebilmesi için de ayrı modeller geliştirilmiştir.

Daha başarılı modellere ulaşmak adına her bir boyut için veriler aşağıdaki 6 farklı senaryoya göre dahil edilerek 6'şar model geliştirilmiştir. Bu senaryolar:

1. Tüm kodlar, 4 kategori (bilgi yok, tütün lehine, aleyhine, her ikisi)
2. Tüm kodlar, 2 kategori (tütün lehine ve diğerleri)
3. Tüm kodlar, 2 kategori (tütün lehine ve tütün aleyhine)
4. Tutarlı kodlar, 4 kategori (bilgi yok, tütün lehine, aleyhine, her ikisi)
5. Tutarlı kodlar, 2 kategori (tütün lehine ve diğerleri)
6. Tutarlı kodlar, 2 kategori (tütün lehine ve tütün aleyhine)

Sonuçta tütün kullanımını teşvik etme durumunu tespit için 2 model ve tütün kullanımını teşvik etmenin 4 alt boyutunu ayrı ayrı tespit etmek için 6'şar senaryoda 24 model olmak üzere toplam 26 makine öğrenmesi modeli geliştirilmiştir.

3.9. İstatistiksel analizler

Öncelikle tweet'lere ait tanımlayıcı istatistikler hesaplanmıştır. Evrendeki tweet'lerin gün ve saatlere göre dağılımı ortalama ve standart sapma ile özetlenmiş, ayrıca histogram ve ısı haritası ile görselleştirilmiştir. Tweet'lerin, özgünlük durumuna; görsel, bahsetme, bağlantı, konu etiketi gibi çoklu medya unsuru bulundurma durumlarına ve yayınlandıkları cihaz türlerine göre dağılımı sıklık ve yüzde ile gösterilmiştir. Tweet'lerin ulaşabilecekleri potansiyel kitlenin büyüklüğü ortanca ve çeyreklikler arası aralık ile özetlenmiştir. Tweet'lerde kelimelerin kullanım sıklıkları tespit edilmiş ve en çok geçen 50 kelime belirlenerek kelimelerin boyutlarının sıklıklarıyla doğru orantılı olduğu bir kelime bulutu şeklinde görselleştirilmiştir.

Gönüllü değerlendiricilerin kodlamalarına dayanarak örnek tweet'lerin tütün kullanımını teşvik etme durumu ve alt boyutlarına dağılımı sıklık ve yüzde ile verilmiştir.

Makine öğrenmesi modellerinin başarısını ölçmek için doğruluk (accuracy), duyarlılık (recall), seçicilik, pozitif prediktif değer (precision), F_1 Skoru ve Pearson'ın phi (Matthews Korelasyon Katsayısı) ve Cohen'in kappa istatistikleri kullanılmıştır. İstatistiksel önemlilik için eşik değer 0,05 olarak kabul edilmiştir.

Son olarak geliştirilen makine öğrenmesi modeli kullanılarak araştırma evrenindeki tüm tweet'lerin tütün kullanımı teşvik etme durumu tespit edilmiş ve bağımsız değişkenler ile iki değişkenli olarak analiz edilmiştir. Tütün kullanımını teşvik eden tweet'lerin günlere, haftanın günlerine, hafta içi ve hafta sonuna, saatlere, gece ve gündüz saatlerine göre sıklık ve yüzdeleri hesaplanmış, histogram şeklinde görselleştirilmiştir. Yine özgünlük durumuna, çoklu medya unsurlarının her birini bulundurma durumuna, yazıldıkları cihazlara göre dağılımları sıklık ve yüzde ile verilmiştir. Bu değişkenler için gruplar arası farklılıklar ki-kare testi ile değerlendirilmiştir. Tütün kullanımı teşvik eden ve etmeyen tweet'lerin ulaşabilecekleri potansiyel kitlenin büyüklüğü ortanca ve çeyreklikler arası aralık ile özetlenmiştir. Gruplar arası farklılıklar Mann-Whitney U testi ile değerlendirilmiştir. Tütün kullanımı teşvik eden ve etmeyen tweet'lerde en sık geçen ilk 50 kelimeden oluşan kelime bulutları oluşturulmuştur.

Kelime bulutları oluşturulurken bağlaçlar ve sayılar gibi işlevsel olmayan kelimeler açık kaynaklı bir liste (Aksoy, 2018) kullanılarak hariç tutulmuştur.

Gelecek araştırmalarda tekrar kullanılmalarını kolaylaştırmak için bulgular sunulurken değişken isimlerinin yanında, Twitter tarafından tanımlanan ve günlük dilde de bu şekilde yaygınlaşan değişken isimleri de parantez içinde verilmiştir.

3.10. Araştırma İzinleri

Araştırmacının Twitter'ın Uygulama Programlama Arayüzü hizmetlerini kullanmak için Twitter şirketinden talep ettiği geliştirici yetkisi 25.04.2019 tarihinde onaylanmıştır.

Araştırma İstanbul Üniversitesi-Cerrahpaşa Girişimsel Olmayan Klinik Araştırmalar tarafından değerlendirilmiş ve etik açıdan uygun olduğuna oybirliği ile karar verilmiştir (06.05.2020 tarih ve 2020-06 sayılı karar).

İstanbul Teknik Üniversitesi'ndeki Doğal Dil İşleme Araştırma Grubu'ndan geliştirildikleri doğal dil işleme araçlarını kullanmak için 19.10.2020 tarihinde yetki alınmıştır.

4. BULGULAR

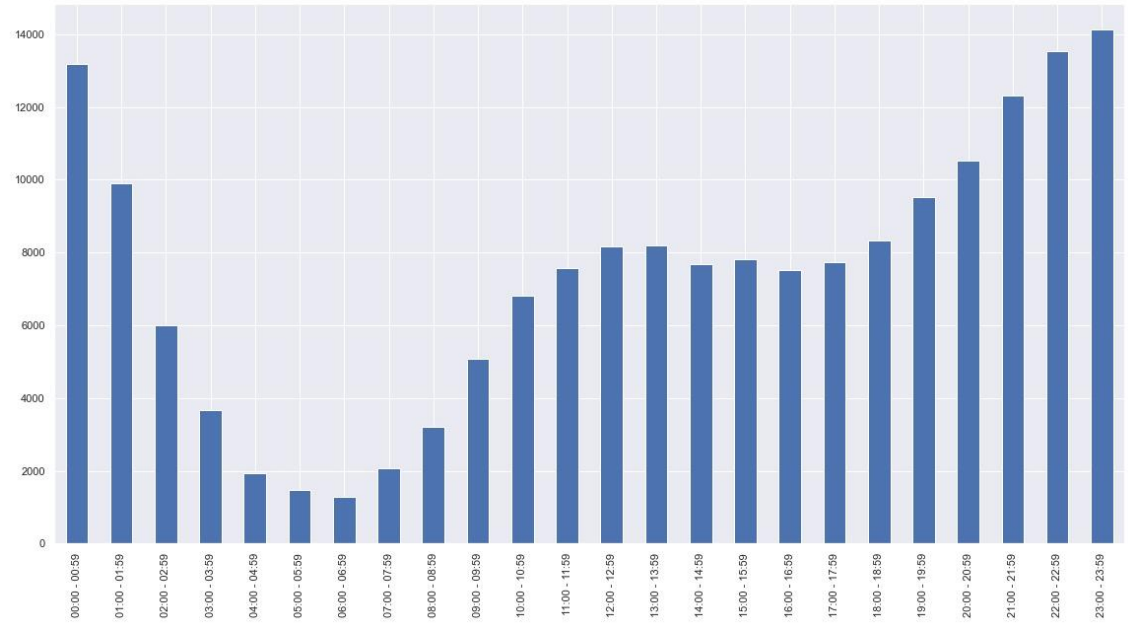
4.1. Tanımlayıcı İstatistikler

Araştırma döneminde tütünle ilgili anahtar kelimeleri içeren toplam 177.684 tweet tespit edilmiştir. Eksiksiz veri toplanan 26 gün üzerinden hesaplandığında günlük ortalama 6253,15 ± 1011,06 (en az: 4447, en çok: 8252) tweet yayınlanmıştır. Tweet'lerin günlere, saatlere ve gün-saat'e göre dağılımı sırasıyla Şekil 4-1, Şekil 4-2 ve Şekil 4-3'te gösterilmiştir.



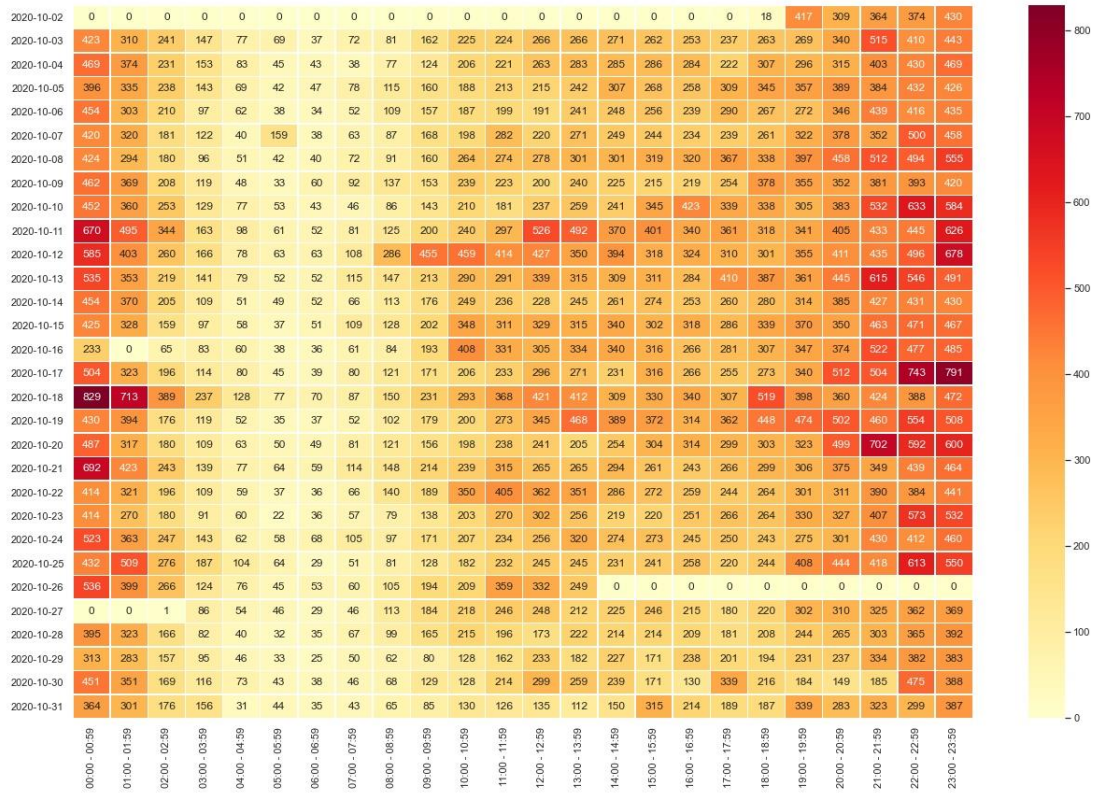
Şekil 4-1. Tweet'lerin günlere göre dağılımı

Bu tweet'ler 126.503 tekil hesap tarafından yayınlanmış olup hesapların %80'inde araştırma döneminde yalnızca bir tweet, %12,58'inde iki tweet yayınlanmıştır.



Şekil 4-2. Tweet'lerin saatlere göre dağılımı

Tütünle ilgili en az bir anahtar kelime içeren tweet'lerin saatlere göre dağılımı gösteren Şekil 4-2 incelendiğinde tweet sıklığının 23:00-23:59 aralığında zirve yaptığı, gece saatlerinde gündüzlere göre daha çok tweet yayınlandığı görülmektedir.



Şekil 4-3. Tweet'lerin gün ve saatlere göre dağılımını gösteren ısı haritası

Not: Eksik veriler (sıfır değerleri) hakkında açıklama için bkz. Bölüm 5.7

Tütünle ilgili tweet'lerin özgünlük durumuna göre dağılımı Tablo 4-1'de gösterilmiştir. Tweet'lerin %37,63'ü özgün tweet olup, %34,37'si ise başka hesaplar tarafından yayınlanan tweet'lerin tekrar yayınlanmış halidir.

Tablo 4-1. Tweet'lerin özgünlük durumuna göre dağılımı

	Sayı	Yüzde
Özgün	66.863	37,63
Tekrar yayınlanan (retweet)	61.067	34,37
Başka tweet'e karşılık (reply)	43.829	24,67
Yorumlayarak alıntı (quote)	5.947	3,35
Toplam*	177.684	100

* 22 tweet hem "başka tweet'e karşılık" hem de "yorumlayarak alıntı" olduğundan toplam değeri satırların toplamına eşit değildir.

Tweet'lerin çoklu medya unsuru bulundurma durumlarına göre dağılımı Tablo 4-2'de verilmiştir. Görsel, bağlantı ve konu etiketinin aksine tweet'lerin çoğunluğunun (%57,81) "bahsetme" içermesi dikkat çekicidir.

Tablo 4-2. Tweet'lerin çoklu medya unsuru buldurma durumlarına göre dağılımı

	Sayı	Yüzde
Bahsetme (mention)	102.715	57,81
Görsel (media)	13.090	7,37
Bağlantı (url)	10.755	6,05
Konu etiketi (hashtag)	7.725	4,35
Toplam	177.684	100

Yayınlandıkları cihaz ve yazılım türlerine göre tweet'lerin dağılımı Tablo 4-3'te verilmiştir. Tweet'lerin %87,84'ünün taşınabilir cihazlar üzerinden yayınlandığı görülmektedir.

Tablo 4-3. Tweet'lerin yayınlandıkları cihaz ve yazılım türlerine göre dağılımı

	Sayı	Yüzde
Cep telefonu: Android	98.027	55,17
Cep telefonu: iPhone	57.465	32,34
Bilgisayar: Web (PC)	18.996	10,69
Bilgisayar: Mac	1.080	0,61
Tablet: iPad	581	0,33
Diğer	1.535	0,86
Toplam	177.684	100

Araştırma evrenindeki tweet'lerin ulaşabileceği potansiyel kitle değerlendirildiğinde; tweet'lerin en az 0, en çok 7.466.283 kullanıcıya ulaştığı görülmüş, ortanca kitle büyüklüğü 275, çeyreklikler arası aralık 867 olarak hesaplanmıştır.

Tweet'lerde en sık geçen ilk 50 kelime Şekil 4-4'teki kelime bulutunda gösterilmiştir. Tütün ürünlerinin isimleri dışında, diğer bağımlılık yapıcı maddeler (Örn. alkol, bira) ve sigara ile birlikte tüketilen gıdalardan (Örn. çay, kahve) ve para ile ilgili kavramlardan (Örn. lira, TL, vergi) bahsedilmesi dikkat çekmektedir.



Şekil 4-4. Tweet'lerde en sık geçen kelimeler (n=50)

4.2. Makine Öğrenmesi Modelinin Başarısı

Makine öğrenmesi modeli geliştirmek amacıyla gönüllü değerlendiriciler tarafından belirlenen kategoriler incelenmiş, örnek olarak seçilen 5000 tweet'in 502 tanesi müstehcen ifadeler nedeniyle hariç tutulmuştur. Çalışmada kalan 4498 tweet değerlendirici havuzundan basit rastgele örnekleme ile seçilen iki değerlendirici tarafından değerlendirilmiştir.

4498 tweet'in 3575'inin (%79,48) "tütün kullanımı teşvik etme durumu" iki değerlendirici tarafından tutarlı olarak kodlanmıştır. Bunların 2005 tanesi (%56,08) en az bir boyutuyla tütün kullanımını teşvik etmektedir.

Değerlendiricilerin yaptıkları kodlamalara dayanarak örnek tweet'lerin tütün teşvikinin 4 alt boyutundaki kategorilere dağılımı Tablo 4-4'te verilmiştir. Diğer alt boyutların aksine, tweet'lerin yüksek bir oranının (%35,35) tütün kullanma davranışını içermesi dikkat çekicidir.

Tablo 4-4. Örnek tweet'lerin tütün kullanımını teşvik etmenin alt boyutlarına dağılımı

	Davranış		**Tutum**		**Tavsiye**		**Pazarlama**	
Kategoriler	**Sayı**	**Yüzde**	**Sayı**	**Yüzde**	**Sayı**	**Yüzde**	**Sayı**	**Yüzde**
Bilgi yok	1437	31,95	2662	59,18	3889	86,46	4007	89,08
Tütün kullanımı aleyhine	156	3,47	218	4,85	61	1,36	17	0,38
Tütün kullanımı lehine	1590	35,35	275	6,11	124	2,76	75	1,67
Hem lehine hem aleyhine	13	0,29	10	0,22	0	0,00	4	0,09
Tutarsız	1302	28,95	1333	29,64	424	9,43	395	8,78
Toplam	4498	100	4498	100	4498	100	4498	100

Tweet'lerin tütün kullanımını teşvik etme durumunu öngörmek için geliştirilen 1. ve 2. Model'in başarı düzeyleri Tablo 4-5'te gösterilmiştir. Değerlendiricilerin tütün kullanımını teşvik etme durumunu tutarlı olarak kategorize ettiği veriler ile eğitilen 2. Model'in, bekleneceği üzere, tüm ölçütlere göre, tutarsız verilerin de dahil edildiği 1. Model'den daha başarılı olduğu görülmektedir.

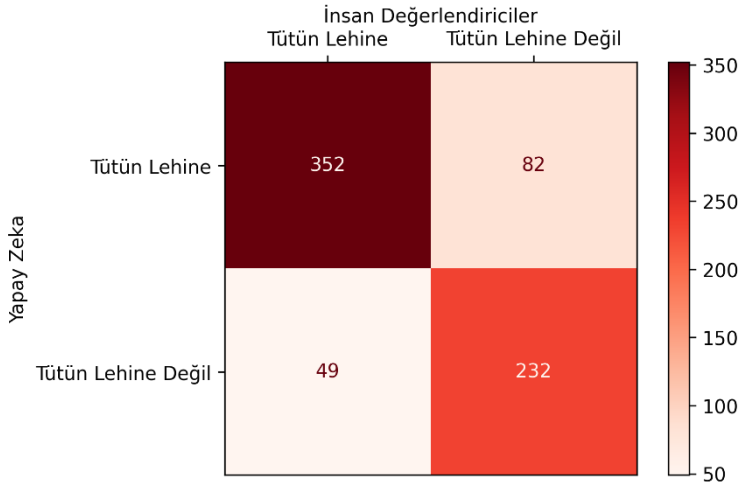
Tablo 4-5. Tütün kullanımını teşvik etme durumunu öngören makine öğrenmesi modellerinin başarı düzeyleri

	1. Model	**2. Model**
Doğruluk	%79,67	%81,68
PPD	%81,91	%81,11
Duyarlılık	%80,75	%87,78
Seçicilik	%78,35	%73,89
F1 Skoru	%81,33	%84,31
MKK	0,59	0,63
κ	0,59	0,62

PPD: Pozitif Prediktif Değer

MKK: Matthews Korelasyon Katsayısı

Şekil 4-5'te 2. Model tarafından öngörülen kategoriler insan değerlendiriciler tarafından belirlenen kategorilerle karşılaştırılmıştır. Bu model, sosyal medya içeriklerinin tütün kullanımı teşvik edip etmediğini %81,68 doğruluk ve %87,78 duyarlılık ile ayırt edebilmektedir ($\kappa = 0,623$).



Şekil 4-5. "2. Model" tarafından öngörülen kategorilerin insan değerlendiriciler tarafından belirlenen kategorilerle karşılaştırılması

$\kappa = 0{,}623$

Tütün kullanımını teşvik etme durumunun dört alt boyutunu değerlendirmek için her bir boyut için 6 senaryoda ayrı ayrı geliştirilen makine öğrenmesi modellerinin başarı düzeyleri Tablo 4-6'da gösterilmiştir. Bekleneceği üzere eğitiminde yalnızca tutarlı kodlamaların kullanıldığı modeller (4-6. senaryolar) tüm kodlamaların kullanıldığı modellere göre (1-3. senaryolar) genellikle daha başarılıdır.

15 ila 18. Modeller sosyal medya içeriklerinde tütün kullanma davranışını %77,81; tütüne yönelik tutumu %82,62; tütün ile ilgili tavsiyeyi %97,42; tütün pazarlamasını %98,66 doğrulukla öngörülebilmektedir.

Tablo 4-6. Tütün kullanımını teşvik etmenin alt boyutlarını öngören makine öğrenmesi modellerinin başarı düzeyleri

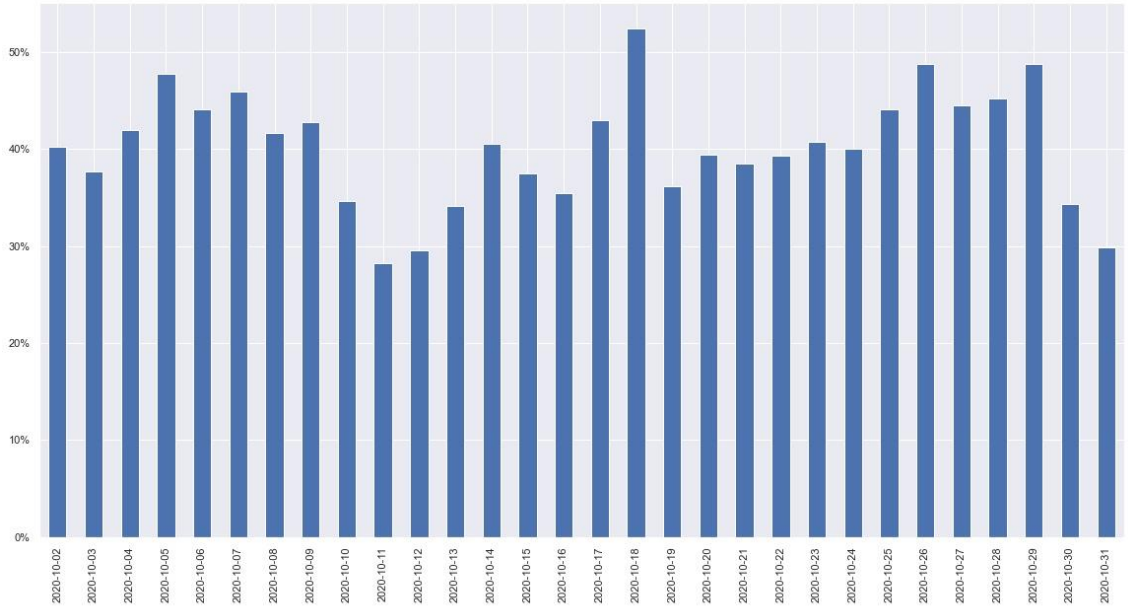
Model	Boyut	Doğruluk (%)	PPD (%)	Duyarlılık (%)	Seçicilik (%)	F_1 Skoru (%)	MKK	κ
Senaryo 1: Tüm kodlar, 4 kategori								
3	Davranış	73,61					0,54	0,54
4	Tutum	69,67					0,30	0,30
5	Tavsiye	92,11					0,46	0,46
6	Pazarlama	94,33					0,39	0,38
Senaryo 2: Tüm kodlar, 2 kategori (tütün lehine ve diğerleri)								
7	Davranış	82,33	80,07	81,63	82,92	80,84	0,64	0,64
8	Tutum	84,56	44,51	29,73	93,77	35,65	0,28	0,27
9	Tavsiye	97,28	74,29	62,65	98,95	67,97	0,67	0,67
10	Pazarlama	97,83	61,02	69,23	98,68	64,86	0,64	0,64
Senaryo 3: Tüm kodlar, 2 kategori (tütün lehine ve tütün aleyhine)								
11	Davranış	91,16	92,91	97,2	52,34	95,01	0,58	0,57
12	Tutum	82,34	86,99	82,63	81,92	84,75	0,64	0,64
13	Tavsiye	84,67	90,79	83,13	87,04	86,79	0,69	0,69
14	Pazarlama	91,36	94,12	92,31	89,66	93,2	0,81	0,81
Senaryo 4: Tutarlı kodlar, 4 kategori								
15	Davranış	77,81					0,60	0,60
16	Tutum	82,62					0,41	0,41
17	Tavsiye	97,42					0,68	0,68
18	Pazarlama	98,66					0,67	0,66
Senaryo 5: Tutarlı kodlar, 2 kategori (tütün lehine ve diğerleri)								
19	Davranış	81,88	81,76	81,76	81,99	81,76	0,64	0,64
20	Tutum	89,26	37,74	36,36	94,29	37,04	0,31	0,31
21	Tavsiye	98,28	82,35	56	99,62	66,67	0,67	0,66
22	Pazarlama	99,15	90	60	99,88	72	0,73	0,72
Senaryo 6: Tutarlı kodlar, 2 kategori (tütün lehine ve tütün aleyhine)								
23	Davranış	94	96,56	96,87	64,52	96,71	0,62	0,62
24	Tutum	78,79	76,56	89,09	65,91	82,35	0,57	0,56
25	Tavsiye	78,38	86,96	80	75	83,33	0,53	0,53
26	Pazarlama	94,74	93,75	100	75	96,77	0,84	0,83

PPD: Pozitif Prediktif Değer; MKK: Matthews Korelasyon Katsayısı.
Kabul edilen eşik değerlerin üstünde olan hücreler yeşil renk ile gösterilmiştir. (MKK ve κ için 0,6 ve diğerleri için %80) 1. ve 4. senaryolarda her bir boyut ikiden fazla kategoride incelendiği için PPD, Duyarlılık, Seçicilik ve F_1 Skoru değerleri hesaplanamamıştır.

Geliştirilen makine öğrenmesi modeli (2. Model) kullanılarak araştırma evrenindeki tüm tweet'lerin tütün kullanımı teşvik etme durumu öngörülmüştür. Sonuçta tweet'lerin %39,85'inin (n=70.803) tütün kullanımı teşvik ettiği ortaya çıkmıştır. Sonrasında tütün kullanımı teşvik eden ve etmeyen tweet'lerin bağımsız değişkenlere göre dağılımını incelemek için iki değişkenli analizler yapılmıştır.

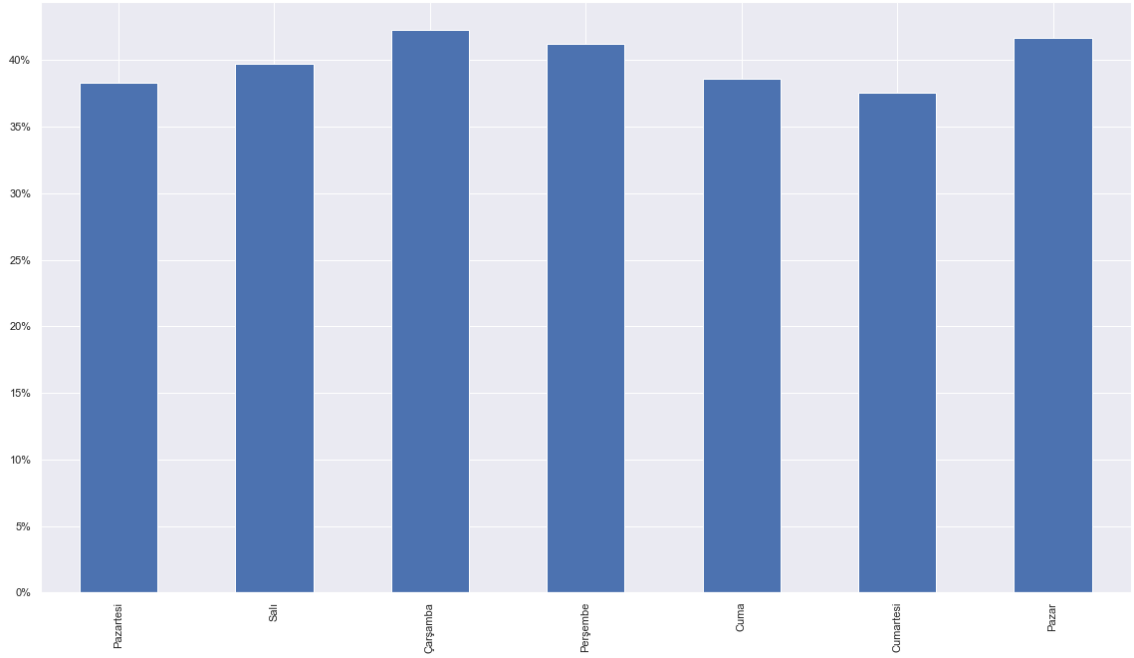
4.3. İki Değişkenli Analizler

Twitter'da günde ortalama 2360,10 ± 599,43 adet tütün kullanımı teşvik eden tweet yayınlanmaktadır. Tütün kullanımını teşvik eden tweet'lerin tütünle ilgili tweet'ler arasındaki yüzdesinin günlere göre dağılımı Şekil 4-6'da gösterilmiştir. Tütün kullanımı teşvik eden tweet yüzdesi 11.10.2020 tarihinde en düşük (%28,26), 18.10.2020 tarihinde en yüksek (%52,42) değere ulaşmıştır. Dağılımda tekrarlayan bir desen gözlenmemiştir.



Şekil 4-6. Tütün kullanımını teşvik eden tweet yüzdesinin günlere göre dağılımı

Tütün kullanımı teşvik eden tweet yüzdesinin haftanın günlerine göre dağılımı Şekil 4-7'de gösterilmiştir. Tütün kullanımı teşvik eden tweet yüzdesi Cumartesi günü en düşük (%37,53), Çarşamba günü en yüksek (%42,24) değerdedir.



Şekil 4-7. Tütün kullanımı teşvik eden tweet yüzdesinin haftanın günlerine göre dağılımı

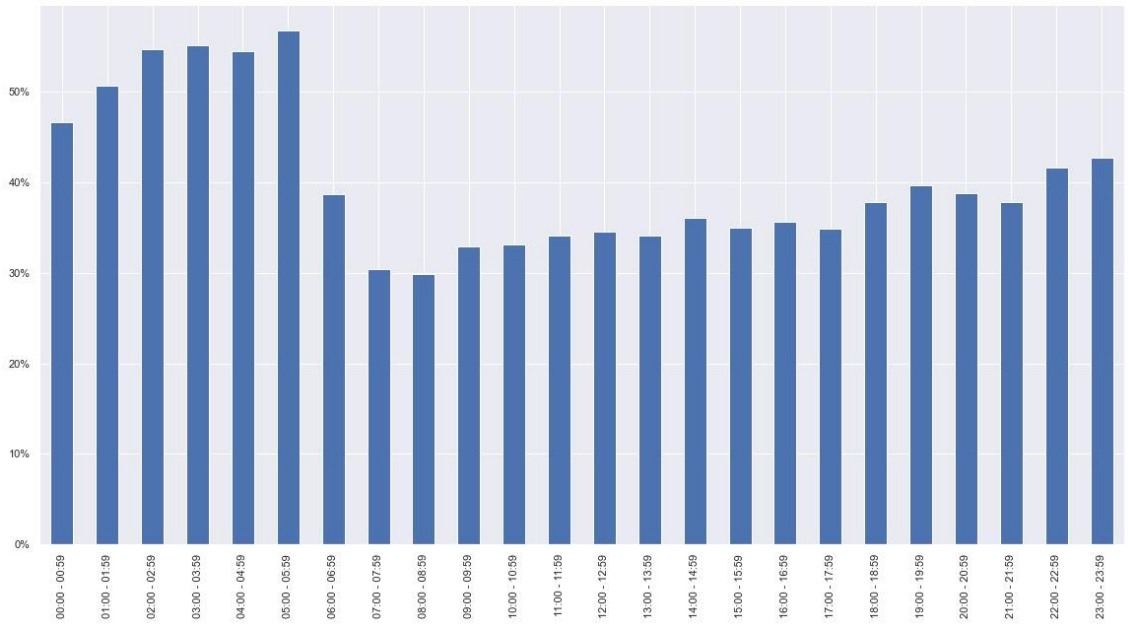
Hafta içi veya hafta sonu yayınlanan tweet'ler için tütün kullanımı teşvik eden tweet yüzdesi Tablo 4-7'de karşılaştırılmış olup önemli bir fark gözlenmemiştir.

Tablo 4-7. Tweet'lerin haftanın hangi kısmında yayınlandığına göre tütün kullanımı teşvik etme durumu

	Tütün Teşviki Yok		Tütün Teşviki Var		Toplam	
	Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde
Hafta içi	71.596	60,00	47.722	40,00	119.318	100
Hafta sonu	35.285	60,45	23.081	39,55	58.366	100
Toplam	106.881	60,15	70.803	39,85	177.684	100

χ^2=3,30 p=0,069

Gün içindeki saat aralıklarına göre tütün kullanımı teşvik eden tweet yüzdesi Şekil 4-8'de verilmiştir. Gece yarısından sabaha kadar olan saat aralıklarında (00:00-05:59) diğer saatlere göre tütün kullanımı teşvik eden tweet'lerin yüksek olduğu görülmektedir.



Şekil 4-8. Tütün kullanımı teşvik eden tweet yüzdesinin saat aralıklarına göre dağılımı

Tablo 4-8'de saat aralıkları gece ve gündüz olarak 12 saatlik iki kısımda incelenmiş, gece saatlerinde (%43,72) gündüz saatlerine (%34,30) kıyasla tütün kullanımı teşvik eden tweet'lerin istatistiksel olarak önemli düzeyde daha sık olduğu görülmüştür ($p<0,001$).

Tablo 4-8. Tweet'lerin gece veya gündüz yayınlanma durumuna göre tütün kullanımı teşvik etme durumu

	Tütün Teşviki Yok		Tütün Teşviki Var		Toplam	
	Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde
Gündüz	48.041	65,70	25.086	34,30	73.127	100
Gece	58.840	56,28	45.717	43,72	104.557	100
Toplam	106.881	60,15	70.803	39,85	177.684	100

χ^2=1592,58 p<0,001
Gündüz: 06:00 - 17:59 , Gece: 18:00 - 05:59

Tweet'lerin özgünlük durumuna göre tütün kullanımı teşvik etme durumu Tablo 4-9'da verilmiştir. Özgün tweet'lerin %58,85'i tütün kullanımını teşvik ederken, tekrar yayınlanan tweet'lerin sadece %20,50'si teşvik etmektedir.

Tablo 4-9. Tweet'lerin özgünlük durumuna göre tütün kullanımı teşvik etme durumu

	Tütün Teşviki Yok		Tütün Teşviki Var		Toplam			
	Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde	χ^2 *	p *
Özgün	27.515	41,15	39.348	58,85	66.863	100	16.146,36	<0,001
Tekrar yayınlanan (retweet)	48.547	79,50	12.520	20,50	61.067	100	14.526,69	<0,001
Başka tweet'e karşılık (reply)	27.009	61,62	16.820	38,38	43.829	100	52,46	<0,001
Yorumlayarak alıntı (quote)	3.829	64,39	2.118	35,61	5.947	100	45,82	<0,001

* Her bir grup için diğer gruplar birleştirilerek karşılaştırılmıştır.
22 tweet hem 'başka tweet'e karşılık' hem de 'yorumlayarak alıntı' grubuna dahildir.

Çoklu medya unsuru bulunan ve bulunmayan tweet'lerin tütün kullanımı teşvik etme durumuna göre dağılımı Tablo 4-10'da verilmiştir. Tütün kullanımını teşvik eden tweet'lerin %60,34'ünün başka bir kullanıcıdan daha bahsetmesi dikkat çekicidir.

Tablo 4-10. Tweet'lerde çoklu medya unsuru bulunma durumuna göre tütün kullanımı teşvik etme durumu

		Tütün Teşviki Yok		Tütün Teşviki Var		Toplam			
		Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde	χ^2	p
Bahsetme (mention)	Var	32.243	30,17	42.726	60,34	74.969	42,19	15.900,97	<0,001
	Yok	74.638	69,83	28.077	39,66	102.715	57,81		
Görsel (media)	Var	9.496	8,88	3.594	5,08	13.090	7,37	904,71	<0,001
	Yok	97.385	91,12	67.209	94,92	164.594	92,63		
Bağlantı (url)	Var	7.997	7,48	2.758	3,90	10.755	6,05	962,94	<0,001
	Yok	98.884	92,52	68.045	96,10	166.929	93,95		
Konu etiketi (hashtag)	Var	5.860	5,48	1.865	2,63	7.725	4,35	830,39	<0,001
	Yok	101.021	94,52	68.938	97,37	169.959	95,65		
Toplam		106.881	100	70.803	100	177.684	100		

Tweet'lerin yayınlandıkları cihaz ve yazılım türlerine göre tütün kullanımını teşvik etme durumu Tablo 4-11'de incelenmiştir. Cep telefonu üzerinden, özellikle de iPhone uygulaması üzerinden yayınlanan tweet'ler arasında tütün kullanımını teşvik eden tweet'lerin yüzdesi (%45,19) diğer cihazlara göre daha yüksektir.

Tablo 4-11. Tweet'lerin yayınlandıkları cihaz ve yazılım türlerine göre tütün kullanımı teşvik etme durumu

	Tütün Teşviki Yok		Tütün Teşviki Var		Toplam	
	Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde
Cep telefonu: Android	59.563	60,76	38.464	39,24	98.027	100
Cep telefonu: iPhone	31.495	54,81	25.970	45,19	57.465	100
Bilgisayar: Web (PC)	13.028	68,58	5.968	31,42	18.996	100
Bilgisayar: Mac	905	83,80	175	16,20	1.080	100
Tablet: iPad	452	77,80	129	22,20	581	100
Diğer	1.438	93,68	97	6,32	1.535	100
Toplam	106.881	60,15	70.803	39,85	177.684	100

χ^2=2310,67 p<0,001

Tweet'lerin ulaştığı potansiyel kitle büyüklüğü ile yeniden yayınlanma ve beğenilme sayılarından oluşan yayılma özelliklerinin tütün kullanımını teşvik etme durumuna göre dağılımları Tablo 4-12'de verilmiştir. Potansiyel kitle yönünden iki grubun ortanca kitle büyüklükleri arasında fark istatistiksel olarak önemli olsa da sosyal medya bağlamında anlamlı değildir (U=3697980674,5; p<0,001).

Tablo 4-12. Tütün kullanımı teşvik etme durumuna göre tweet'lerin yayılma özellikleri

	Tütün Teşviki Yok		Tütün Teşviki Var	
	Ortanca	ÇAA	Ortanca	ÇAA
Potansiyel Kitle	282	949	265	757
Yeniden Yayınlanma*	1	79	0	0
Beğenilme*	0	2	2	7

ÇAA: Çeyreklikler Arası Aralık

* 6 ay sonra hala yayında olan 122.768 tweet üzerinden hesaplanmıştır (bkz. Araştırmanın Kısıtlılıkları).

Tütün kullanımı teşvik eden ve etmeyen tweet'lerde en sık geçen ilk 50 kelime Şekil 4-9'da sıklıklarıyla orantılı büyüklükte gösterilmiştir. Tütün kullanımını teşvik eden tweet'lerde tütünle birlikte tüketilen gıdalar (Örn. çay, kahve) ve tütün kullanılan ortamlardan bahseden kelimelerin (Örn. gece, balkon) yer aldığı görülmektedir. Tütün kullanımı teşvik etmeyen tweet'lerde ise diğer bağımlılık yapıcı maddeler (Örn. alkol, bira) ve finansal kelimeler (Örn. TL, lira, vergi) sık geçmektedir.

(a) (b)



Şekil 4-9. Tütün kullanımı teşvik eden ve etmeyen tweet'lerde en sık geçen kelimeler

a. Tütün kullanımını teşvik eden tweet'ler

b. Tütün kullanımını teşvik etmeyen tweet'ler

5. TARTIŞMA

5.1. Araştırmanın Ana Çıktıları

Araştırmanın "sosyal medyada tütün kullanımını teşvik eden içerikleri otomatik olarak tespit edebilecek bir yapay zeka geliştirme" amacı doğrultusunda belirlenen dört hedefe ulaşılmıştır.

1. ***Hedef: Sosyal medyada tütün teşvikinin farklı şekillerini tanımlamak:***

Araştırmanın nitel kısmında sosyal medyada tütün teşvikinin alt boyutları tanımlanmıştır. Buna göre sosyal medya içerikleri; tütün kullanımı davranışını sergileyerek, tütün hakkında olumlu tutum göstererek, tütün ürünü kullanmayı sözel olarak tavsiye ederek, spesifik bir marka veya adresin reklam ve pazarlamasını yaparak tütün kullanımı teşvik etmektedir. Teorik açıdan Davranış Tahmini Entegre Modeli'ne (Integrative Model of Behavioral Prediction) (Fishbein ve Yzer, 2003) paralellik gösteren bu temaların gelecek araştırmalara da yol göstereceği düşünülmektedir.

2. ***Hedef: Yüksek sayıda sosyal medya içeriğinin tütün kullanımını teşvik etme durumunu içeren bir veri seti oluşturmak***

Araştırmanın diğer bir ara çıktısı olarak 4498 sosyal medya içeriğinin tütün kullanımını teşvik etme durumu ve alt boyutlarının yer aldığı bir veri seti oluşturulmuştur. Bu veri seti gerek bu içerikleri farklı yönlerden analiz etmeyi gerekse daha iyi yapay zeka modelleri oluşturmayı amaçlayan gelecek araştırmalarda kullanılabilir. Bu nedenle araştırmada üretilen veriler açık veri olarak yayınlanacaktır.

3. ***Hedef: Başarılı bir makine öğrenme modeli geliştirmek:***

Bu çalışma sonucunda sosyal medyada tütün kullanımını teşvik eden içerikleri yüksek duyarlılık (%87,78) ve pozitif prediktif değer (%81,11) ile tespit edebilen bir makine öğrenmesi modeli geliştirilmiştir.

4. ***Hedef: Evren istatistiklerinin dağılımını incelemek:***

Geliştirilen model sayesinde bir aylık araştırma periyodunda sosyal medyada yayınlanan içeriklerin (n=177,784) ne kadarının tütün kullanımı teşvik ettiği (%39,84) ortaya konulabilmiştir. Böylece sosyal medya içeriklerinin tütün kullanımını teşvik etme durumuna göre dağılımı da incelenebilmiştir.

5.2. Sosyal Medyada Tütün Teşvikinin Mevcut Durumu

Ekim 2020'de sosyal medyada yayınlanan tütünle ilgili 177.684 Türkçe tweet'in %39,85'i tütün kullanımını teşvik etmektedir. Günde ortalama 2360,10 ± 599,43 adet tütün kullanımı teşvik eden tweet yayınlanmaktadır.

5.2.1. Tweet Özelliklerine Göre Dağılım

Tütün teşviki, diğerlerine kıyasla özgün tweet'lerde daha sık, yeniden yayınlanan tweet'lerde daha seyrek yer almıştır. Bunu destekler şekilde, tütünü teşvik eden tweet'lerin 6 ay içinde yeniden yayınlanma sayıları teşvik etmeyenlerden daha düşüktür. Bu bulgular kişilerin tütün kullanımını teşvik eden tweet'leri yeniden yayınlamadığına, diğer bir ifadeyle zararlı içerikleri yaymadığına işaret etmesi açısından sevindiricidir. Bu durum kişilerin tütün karşıtı tweet'leri tekrar yayınlamak konusunda daha istekli olmasından da kaynaklanıyor olabilir, fakat bunu kesin olarak söyleyebilmek için ayrı bir inceleme gerekmektedir. Diğer taraftan tütün kullanımını teşvik eden tweet'lerin beğenilme sayılarının etmeyenlere göre daha yüksek olması her ne kadar yaymıyorlarsa da kişilerin bu tweet'lerden daha çok etkilendiğine işaret etmektedir.

Kişiler arası ilişkilerin, bireylerin tütün hakkındaki sosyal norm algısına aracılık ettiği (Liu ve ark., 2020) göz önüne alındığında, başka kişilerden bahsederek onları da konuşmaya dahil eden tweet'lerin sıklığının, tütün kullanımını teşvik eden tweet'lerde etmeyenlere kıyasla daha yüksek olması endişe vericidir. Bunda tütün endüstrisi pazarlama stratejilerinin bir etkisi olup olmadığı gelecek çalışmalarda incelenebilir.

5.2.2. Zamana Göre Dağılım

Gece saatlerinde yayınlanan tweet'ler, gündüz yayınlananlara göre tütün kullanımı daha çok teşvik etmektedir. Özellikle gece yarısından sabaha kadar olan saat aralığı (00:00-05:59) dikkat çekmektedir. Bu aralıkta saatler ilerlerken, tütünle ilgili tweet sayısı azalmakta fakat bu tweet'ler içerisinde tütünü teşvik edenlerin yüzdesi artmaktadır. Halk sağlığı ve özellikle tütün kontrolü alanında çalışan devlet ve sivil toplum kuruluşları açısından bakılırsa mesai saatleri dışındaki bu zamana yönelik çözümler geliştirmeleri faydalı olabilir.

5.2.3. Cihaz Türüne Göre Dağılım

iPhone üzerinden yayınlanan tweet'ler arasında tütün teşvikinin diğer cihazların her birinden daha sık olduğu görülmektedir. Görece daha pahalı olduğu bilinen bu

markayı kullanan kişilerin tütün kullanıma davranışlarını daha fazla sergilemelerinden kaynaklanıyor olabilir. Tütün şirketlerinin, pazarlama faaliyetlerinde genellikle lüks hayat tarzı sergileyen influencer'ları kullanmayı tercih ettiği bilinmektedir (Kirkham, 2019). Araştırmada tespit edilen bu duruma tütün şirketlerinin etkisinin olup olmadığı ayrıca incelenmelidir. Halk sağlığı alanında çalışan kuruluşlar bu konuyla ilgili olarak sosyal medya üzerinden yapacakları sağlığın geliştirilmesi çalışmalarında iPhone kullanıcılarını öncelikli bir hedef olarak ele alabilirler.

5.2.4. Metinlerde Sık Geçen Kelimeler

Tütün kullanımını teşvik eden tweet'lerde sıklıkla tütünle birlikte tüketilen gıdalar (Ör. çay, kahve) ve tütün kullanılan ortamlardan bahseden kelimelerin (Ör. gece, balkon) geçmesi, bireylerin tütün kullanımlarını tetikleyen faktörlere işaret etmektedir. Bu faktörler, davranış değişikliği faaliyetlerinin tasarlanmasında dikkate alınabilir.

Tütün teşviki olmayan tweet'lerde ise diğer bağımlılık yapıcı maddelerin isimleri (Ör. alkol, bira) ve finansal kelimelerin (Ör. TL, lira, vergi) sık geçmesi, bu tweet'lerde tütünden bir bağımlılık olması yönüyle ve zararlarıyla bahsedilmesine yorulabilir. Çok sayıda araştırma tütünden bahseden sosyal medya içeriklerinde diğer riskli davranışlar ve bağımlılık yapıcı maddelerden bahsedildiğini tespit etmiş (Lienemann ve ark., 2017, p. 13), fakat içeriğin tütün kullanımı teşvik edip etmemesine göre bir ayrım yapamamıştır.

5.3. Bilimsel Katkı

5.3.1. Bilgi Tarama Çalışmalarına Katkı

Bilgi arama (information seeking) davranışından farklı olarak kişinin bilgiye, doğrudan amaçlamaksızın rastlantısal ve tekrarlayan şekilde maruz kalması "bilgi tarama" (information scanning) olarak adlandırılmaktadır. Bilgi taramanın insan davranışları üzerine etkisini gösteren çok sayıda araştırma vardır (Lewis, 2017). Yeni araştırmalar bilgi taramanın tütün kullanma davranışını da etkilediğini göstermektedir. Bigsby ve Hovick'in (2018) araştırmasında bilgi taramanın sigara içmek gibi çeşitli kronik hastalık risk faktörleri üzerinde aracı rol oynadığı gösterilmiştir.

Daha spesifik olarak Zhu'nun (2017) araştırması; sosyal medyada sigara kullanımını teşvik eden içeriklere maruz kalan genç yetişkinlerin 6 ay sonra sigara kullanımını artırdığını bir izlem araştırmasıyla ortaya koyması ve bilgi tarama ile davranış arasında doz-yanıt ilişkisini göstermesi açısından değerlidir. Benzer şekilde Liu ve

arkadaşları (2020) da bilgi taramanın gençlerin e-sigara kullanımına etkisini incelemiştir. Fakat her iki araştırmada da bilgi taramaya dair veriler beyana dayalı (self-reported) olarak toplanmıştır. Bu durum bilgi tarama araştırmalarında yaygın bir kısıtlılıktır (Lewis, 2017). Ayrıca hangi içeriklerin 'sigara kullanımı teşvik ediyor' (pro-smoking) olarak kabul edildiği belirtilmemiş, katılımcıların öznel değerlendirmesine bırakılmıştır. Bu araştırma, gelecek araştırmalarda tütün ile ilgili bilgi taramanın daha objektif olarak ölçülmesini sağlayacak bir araç niteliğindedir.

5.3.2. Tütünle İlgili Sosyal Medya İçerikleri İnceleyen Çalışmalara Katkı

Sosyal medya verileri üzerinde çalışılan araştırmalarda tütün ile ilgili anahtar kelimeleri içeren içerikler; reklam, sağlığa zarar, kullanım, politika ve yasalar, sigara bırakma, aromalar, kişisel görüşler, diğer risk davranışları, kullanım yeri gibi temalar yönünden incelenmiştir (Lienemann ve ark., 2017, p. 11). Bazı araştırmalarda (Myslín ve ark., 2013; van der Tempel ve ark., 2016) manuel olarak yapılan nitel veri analizi ile sosyal medya içeriklerinde; tütün ürünü kullanma davranışı, tutum ve pazarlama gibi temaların tespit edilmiş olması bu araştırmanın bulguları ile uyumludur. Fakat söz konusu araştırmalar, bu temalardan hareketle içeriklerin tütün kullanımı teşvik etme durumu değerlendirilmemiştir. İncelenen içerikler bir şekilde tütünle ilgili olsa da tütün kullanımını teşvik edip etmedikleri ayırt edilmeksizin analiz edilmiştir. Bu araştırma ise böyle bir ayrımı mümkün kılmaktadır.

Söz konusu araştırmalarda tematik analiz çoğunlukla araştırmacılar tarafından manuel olarak yapılmıştır. Bu da incelenebilecek içerik sayısını birkaç bin ile sınırlamaktadır. Facebook gibi bazı sosyal medya platformlarındaki içeriklerin, en fazla 280 karakterle sınırlı olan Twitter'daki içeriklere kıyasla daha uzun olacağı düşünülürse incelenebilecek içerik sayısı daha da azalacaktır. Diğer taraftan makine öğrenmesi yöntemleri kullanan araştırmalarda ise milyondan fazla içerik incelenebilmiştir. (Lienemann ve ark., 2017, p. 8). Bu araştırmada da sadece bir aylık bir dönemde yayınlanan yüz binden fazla tweet analiz edilmiştir.

Bazı araştırmalarda makine öğrenmesi yöntemleri kullanılmışsa da tütün kullanımını teşvik etme durumunu saptama değil; tütünle ilgili olma (Myslín ve ark., 2013), reklam içerme (Clark ve ark., 2016), sigara bırakma amaçlı e-sigara kullanımından bahsetme (Aphinyanaphongs ve ark., 2016) gibi temaları içerip içermediğini tespit etme ve metindeki duyguyu (Allem ve ark., 2017) değerlendirme amacıyla kullanılmıştır.

Ayrıca bazı araştırmalarda makine öğrenmesi yöntemleri, kullanıcıları e-sigara taraftarı olup olmama (Kavuluru ve Sabbir, 2016) durumuna göre sınıflandırmak veya sağlık eğitimi faaliyetleri için uygun hedef kitleleri tespit etmek (Chu ve ark., 2019) gibi amaçlarla da kullanmıştır.

Liu ve arkadaşları, 2019'da geleneksel medyanın yanı sıra Twitter içerikleri üzerinde yaptıkları araştırmada bireylerin tütün kullanma davranışlarını tespit etmek için manuel olarak yapılan tematik analizlerin yeterli olmadığına dikkat çekmiş ve makine öğrenmesi yöntemi kullanarak süreci otomatize etmişlerdir. Bu araştırmadan daha geniş kaynaklar ile çalışılmasına rağmen geliştirilen makine öğrenmesi modeli tweet'lerde tütün kullanma davranışını %71 duyarlılık ve %52 pozitif prediktif değer ile tahmin etmektedir (Liu ve ark., 2019, p. 573). Bu araştırmada geliştirilen 19. Model ise tütün kullanma davranışını %81,76 duyarlılık ve %81,76 pozitif prediktif değer ile tahmin ederek literatüre katkı yapmaktadır.

Ayrıca bu araştırmanın literatüre bir diğer katkısı, sosyal medya içeriklerinde yalnızca tütün kullanma davranışını tespit etmenin ötesine geçerek tütün teşvikini farklı alt boyutlarıyla (davranış, tutum, tavsiye ve pazarlama) ele almasıdır.

5.3.3. Sağlık Davranışı Çalışmalarına Katkı

Tümevarım analizi ile tespit edilen bu boyutlar davranış teorileriyle de paralellik göstermektedir (Liu ve ark., 2020; Yoo ve ark., 2016; Zhu, 2017). Zhu'nun (2017) araştırması sosyal medyadaki tütün kullanımını teşvik eden içeriklerin tütün kullanma niyet ve davranışına etkisini teorik olarak Davranış Tahmini Entegre Modeli'ne (Fishbein ve Yzer, 2003) dayandırarak açıklamış ve bunu ampirik bulgularla desteklemiştir. Bu modele göre sosyal medya içerikleri kişilerin tütün kullanımı niyet ve davranışını şu aracılar üzerinden etkilemektedir: tanımlayıcı normlar (tütün kullanımının ne kadar popüler olduğu), buyruksal normlar (tütün kullanımı hakkında yapılması söylenenler), tutum ve öz yeterliliktir (tütün kullanabilirlik) (Michie ve ark., 2014, pp. 211–214; Zhu, 2017). Araştırmada 4 altbaşlıkta gruplandırdığımız üzere içeriklerde yer alan "tütün kullanma davranışı", "tütüne yönelik tutum", "tütün kullanma tavsiyesi" ve "tütün ürünü pazarlaması" boyutları sırasıyla tanımlayıcı norm inancı, tutum, buyruksal norm inancı ve öz yeterlilik üzerinde etkili olabilir. Bu konu gelecek araştırmalarla aydınlatılmalıdır.

5.3.4. Sosyal Medya İçeriklerinde Makine Öğrenmesi Kullanan Çalışmalara Katkı

Benzer şekilde Cole-Lewis ve arkadaşlarının yalnızca e-sigaralara odaklanan araştırmasında on binden fazla tweet ile eğittikleri makine öğrenmesi modelleri tweet'lerde davranışı %84, tutumu %76, pazarlamayı %91 doğrulukla tahmin etmektedir (Cole-Lewis ve ark., 2015). Bu araştırmada geliştirilen modeller ise tweet'lerde davranışını %94; tutumu %82,62; pazarlamayı %99,15 doğrulukla tahmin edebilmektedir. Araştırmada duyarlılık ve pozitif prediktif değer raporlanmadığı ve dikotomik olarak incelenen değişkenlerde kategorilerin nasıl birleştirildiği belirtilmediği için daha detaylı bir karşılaştırma yapılamamaktadır.

Yapay zeka başarısında elde edilen üstünlükler, söz konusu araştırmada görece daha eski ve sınırlı bir metin sınıflandırma yöntemi olan "Support Vector Machine" algoritması kullanılırken bu araştırmada daha yeni bir yöntemin (BERT) kullanılması ve tüm tweet'lerin iki değerlendirici tarafından kodlanmasından kaynaklanıyor olabilir.

2020'de yayınlanan sistematik derlemede tütünle ilişkili sosyal medya içerikleri üzerinde kullanılan yöntemler incelenmiş olup, çoğunlukla eski ve bilinen yöntemlerin tercih edildiği ifade edilmiştir (Küçükali ve ark., 2020). Buna göre literatürde tütünle ilişkili sosyal medya içerikleri üzerinde kullanılan en yeni tarihli metin sınıflandırma algoritması 2013 tarihinde yayınlanan "Word2Vec"tir. Bu araştırmada ise 2018 tarihinde yayınlanan ve alternatiflerden daha başarılı bir metin sınıflandırma yöntemi olduğu gösterilmiş olan BERT modeli kullanılmıştır. Bu yönüyle araştırma, alandaki birikimi bir adım daha ileriye taşımıştır. Bununla birlikte yapay zeka teknolojileri çok hızlı gelişmektedir ve arada geçen üç yılda çok daha başarılı yöntemler geliştirilmiştir (Kahn, 2021). Teknolojinin hızını daha yakından takip edebilmek için halk sağlığı ve bilgisayar bilimi araştırmacıları arasında işbirliği artırılmalıdır.

5.3.5. Türkçe Halk Sağlığı Literatürüne Katkı

Dolaylı bir katkı olarak bu araştırma veri madenciliği, makine öğrenmesi, doğal dil işleme, sohbet robotu, oyunlaştırma ve kitle kaynağı gibi teknoloji kavramlarını Türkçe halk sağlığı literatürüne tanıtmaktadır.

5.4. Ulusal Tütün Kontrolüne Katkı

Türkiye'deki istatistiklere göre yetişkinler arasında son otuz günde herhangi bir sigara reklam ve tanıtımı fark edenlerin yüzdesi 2008'den 2016'ya kadar sürekli

yükselmiştir. Televizyon, video ya da sinema filmi izleyen öğrencilerin %66,5'i bu ortamlarda tütün kullanan birini görmektedir (CDC, 2019). Fakat gençlerin hem daha sık kullandığı hem de daha çok etkilendiği sosyal medya üzerinde tütün kullanımını teşvik eden içeriklere maruziyetleri bilinmemektedir. Bu araştırma, Twitter'da günde ortalama 2360,10 ± 599,43 adet tütün kullanımı teşvik eden tweet yayınlandığını ortaya koymaktadır. Bu araştırmada geliştirilen makine öğrenmesi modelleri toplumun bu içeriklere ne kadar maruz kaldığını tespit etmeyi hedefleyen araştırmalarda bir araç olarak kullanılabilir.

Türkiye'de konunun paydaşı olan tüm kamu kurumlarının katılımıyla hazırlanan, 3 ana amaç altında 205 faaliyetin belirlendiği 2018-2023 Tütün Kontrol Strateji Belgesi ve Eylem Planında "Tütün ürünlerine talebin azaltılması" amacının "Reklâm, promosyon ve sponsorluğun önlenmesi" bileşeni kapsamında aşağıdaki faaliyetlerin belirlendiği görülmektedir (Sağlık Bakanlığı, 2018):

> "3.1.İnternet üzerinden yapılan reklam ve satışların izlenmesi ve önlenmesi"
>
> "3.4.Reklam ve örtülü reklam ihlalinin izlenmesini sağlayacak sistemin ilgili kurumların eşgüdümü ile güçlendirilmesi"
>
> "3.6.Tütün endüstrisinin, gençlerin tütün ürünü kullanmalarını teşvik eden ve tütün ürünlerine erişimlerini kolaylaştıran kampanya ve programlarının tespit edilmesine yönelik bir sistem geliştirilmesi"

Belgede 3.1 nolu faaliyet için sadece "sayfaya erişimin engellenmesi sürecinin hızlandırılması"nın planlandığı görülmektedir (Sağlık Bakanlığı, 2018, p. 126). Bu yaklaşım satışların yapıldığı internet siteleri için etkili olabilir fakat sosyal medya platformlarındaki içerikler yayınlandıktan sonra çok kısa süre içinde tüketilmektedir. Bir içeriğin yarılanma ömrünün Twitter için 6-18 dakika, Facebook için 78-90 dakika olduğu (SocialTalent, 2015) göz önünde bulundurulduğunda, erişim engelleme süreçlerinin hızlandırılmasının bu içeriklerin potansiyel zararlarını engellemek için tek başına yeterli olmayacağı anlaşılmaktadır.

3.4 ve 3.6 nolu faaliyetlerde bahsedilen sistemler için "mevzuat değişiklikleri yapılması" ve "birimler kurulması" planlanmıştır (Sağlık Bakanlığı, 2018, pp. 128–129). Eylem planında bu faaliyetlerden sorumlu olduğu belirtilen Tarım ve Orman Bakanlığı'ndan, bu araştırma için Bilgi Edinme Hakkı Kanunu kapsamında alınan bilgiye

göre belirlenen hedeflere henüz ulaşılamadığı anlaşılmaktadır (Bkz. Ek: Tarım Bakanlığı Bilgi Edinme Başvurusu Cevabı). Bu sistemlerden beklenen izleme ve tespit fonksiyonlarının, şikayete istinaden veya çalışanların manuel olarak değerlendirmesiyle gerçekleştirilmesi mümkün ve etkili değildir. Bu araştırmada geliştirilen makine öğrenmesi modelleri söz konusu sistemlere entegre edilerek ulusal tütün kontrol stratejisine hizmet edebilir.

Diğer taraftan bu araştırma, sosyal medyada tütün teşvikinin yalnızca tütün endüstrisi tarafından ve spesifik bir markanın pazarlanması şeklinde değil; bireysel kullanıcıların paylaştıkları davranış, tutum veya tavsiyeler yoluyla da gerçekleşebileceğini dikkate almaktadır. Geleneksel medyada içerikler sadece kurumlar tarafından üretilirken sosyal medyada sıradan vatandaşlar da içerik üreticisi haline gelmiştir. Dolayısıyla tütün kontrolü stratejileri yalnızca endüstrinin doğrudan yaptığı reklamları yasaklamaya odaklanmamalı, bireylerin ürettikleri içerikler için de (hak ve özgürlüklere saygılı olmak kaydıyla) uygun sağlığın geliştirilmesi yaklaşımları izlenmelidir.

5.5. Halk Sağlığı Araştırma ve Uygulamalarında Kullanım Önerileri

Yapay zeka teknolojileri tıbbi bilimlerde özellikle radyoloji, dermatoloji ve patoloji gibi tıbbi tanı alanlarında kendini ispat etmişse de halk sağlığında kullanımı henüz emekleme aşamasındadır (Morgenstern ve ark., 2021). Yine de bulaşıcı hastalıkların sürveyansından, kronik hastalık yönetimine geniş bir yelpazeye yayılmış erken uygulamalar mevcuttur (Küçükali, 2021).

Bu araştırmada geliştirilen yapay zeka önceki bölümde de belirtilen sürveyans ve denetim amaçlarına hizmet edebilir. Özetlemek gerekirse, önerilen yöntem:

İnfodemiyoloji ve infosürveyans

Tütün ürünlerinden bahseden ve tütün kullanımını teşvik eden sosyal medya içeriklerinin sıklığının ve bunlara maruziyetin belirlenmesi ve zaman içindeki değişiminin izlenmesinde kullanılabilir (Lienemann ve ark., 2017, p. 17).

Denetim ve yaptırım

Tütün endüstrisinin reklam kampanyaları, influencer pazarlama uygulamalarının tespit edilmesinde kullanılabilir. Tütün ürünlerini yasa dışı olarak satan sosyal medya hesaplarının ve internet adreslerinin tespit edilmesinde kullanılabilir.

5.5.1. Sağlığın Geliştirilmesinde Kullanım Önerileri

Akla ilk gelen bu uygulamaların ötesinde bu araştırmada geliştirilen yapay zeka, sağlığın geliştirilmesi yaklaşımıyla tütün kontrolüne çok daha zengin bir katkı sağlayabilir. Potansiyel kullanım alanları Ottawa Bildirgesi'nde (DSÖ, 1986) belirlenen eylem türlerine göre gruplanarak aşağıda listelenmiştir.

Bireysel Beceriler Geliştirme

Özellikle yeni nesil e-sigaralar olmak üzere tütün ürünleriyle ilgili yanlış bilgiler, tütünü teşvik eden içerikler üzerinden tespit edilip düzeltilebilir (Lienemann ve ark., 2017, p. 17).

Destekleyici Çevreler Oluşturma

Sigarayı bırakmış kişiler kendilerini tekrar başlamaktan alıkoymak için, sosyal medya kullanırken bu yapay zeka ile desteklenmiş bir filtreden yardım alabilir.

Yine, tütün kullanmayan ve sosyal medyada çeşitli tütün ürünlerinin kullanımına ilişkin içerikler görmekten rahatsız olan kişiler ve tüm çocuklar için böyle bir filtre kullanabilir.

Sosyal medya şirketleri bu yapay zekayı doğrudan kendi sistemlerine entegre edebilir ve tütün kullanımı teşvik eden içerikleri (i) göstermemeyi kullanıcılara bir seçenek olarak sunabilir, (ii) içeriklerin beraberinde otomatik uyarı mesajı gösterebilir, (iii) bu içerikleri kendi öne çıkarma algoritmalarından, (iv) "influencer"ların gelir elde etme mekanizmalarından çıkarabilir.

Halk sağlığı kurumları, sigara kullanan fakat sigaraya yönelik olumsuz tutum ve rahatsızlık bildiren kişileri sigara bırakma hizmetlerine yönlendiren müdahale çalışmaları yapılabilir.

Toplumsal Hareketi Güçlendirme

Sosyal medyadaki tütün karşıtı içerikler halk sağlığı kurumları ve gönüllüleri tarafından yeniden yayınlama ve beğeni gibi yollarla desteklenebilir veya ödüllendirilebilir.

Sosyal medyada yürütülecek sağlık eğitimi çalışmaları, tütün kullanımını teşvik eden içeriklerin dağılım özelliklerine göre daha stratejik olarak planlanabilir.

Tütün şirketlerinin gençlere pazarlama yapmamaları; sosyal medya şirketlerinin tütün ürünlerinin reklamını önlemeleri için yapılacak sağlık savunuculuğu faaliyetlerine istatistik ve bilgi sağlayabilir (Rowell, 2020).

Değerlendirme

Ayrıca bu yapay zeka, yürütülecek sağlığın geliştirilmesi faaliyetlerinin etkisini değerlendirmeye (Emery ve ark., 2014), faaliyet öncesi ve sonrası durumları karşılaştırmaya imkan sağlayabilir.

5.6. Araştırmanın Güçlü Yanları

Bu araştırma, halk sağlığı alanında Türkiye'deki ilk yapay zeka uygulamasıdır. PubMed, Web of Science, Yüksek Öğrenim Kurumu Ulusal Tez Merkezi, TÜBİTAK ULAKBİM TR Dizin ve Dergipark veri tabanlarında halk sağlığı ve yapay zeka ile ilgili anahtar kelimeler kullanılarak yapılan aramalarda Türkiye'den ilişkili bir sonuç elde edilememiştir.

Genel olarak doğal dil işleme, özelde metin sınıflandırma uygulamaları metinlerin yazıldıkları dile özgüdür. Farklı konularda İngilizce diline yönelik geliştirilmiş ve kullanıma hazır çok sayıda makine öğrenmesi modeli vardır fakat her dil için ayrı modeller geliştirilmesi gerekmektedir. Bu araştırma Türkçe dilinde bir çözüm sunmaktadır.

Bu araştırmanın en güçlü yanı karma yöntem izlenmesidir. Makine öğrenmesinden önce yapılan nitel analiz sosyal medyada tütün kullanımının farklı boyutlarını veriye dayalı anlamayı ve değerlendirmeyi mümkün kılmıştır. Araştırmada geliştirilecek makine öğrenmesi modeli için gereken eğitim veri seti de bu boyutlar ışığında ve yine nitel veri analizi yöntemine uygun olarak gerçekleştirilmiştir.

Literatürde sosyal medyadaki içeriklerde öne çıkan temaları keşfetmeyi amaçlayan bazı araştırmalar vardır (Lienemann ve ark., 2017, p. 11) fakat bu araştırma diğerlerinden farklı olarak doğrudan tütün kullanımı teşvik etme durumunu tespit etmektedir.

Sosyal medyadan veri toplamak amacıyla anahtar kelime olarak sadece cins isimler değil marka isimleri de dahil edilmiş, literatürdeki iyi uygulama önerileri (Kim ve ark., 2016) doğrultusunda, anahtar kelimeler yapılan ön aramalarla konuyla daha ilgili sonuçlar verecek şekilde sadeleştirilmiştir.

Daha önceki araştırmalarda farklı amaçlarla tütünle ilgili içerikler üzerinde makine öğrenmesi kullanılmışsa da genellikle eski ve iyi bilinen yöntemler tercih edilmiştir. Bu çalışmada 2018 yılında yayınlanan ve doğal dil işleme çalışmaları için dönüm noktası kabul edilen (Kahn, 2021) BERT algoritmasının kullanılmış olması çalışmayı daha güçlü kılmaktadır.

Eğitim veri setinin doğrudan söz konusu içeriklerin hitap ettiği genç yaş grubundan seçilen değerlendiriciler tarafından oluşturulmuş olması, tüm tweet'lerin iki değerlendirici tarafından kodlanması, model eğitimi öncesinde tweet'lerdeki yazım yanlışlarının düzeltilmesi ve eğitimin çok sayıda senaryo altında tekrarlanması araştırmanın diğer güçlü yanlarıdır.

Araştırmada yalnızca halka açık olarak yayınlanan tweet'lerin değerlendirilmiş olması, kişisel verilerin toplanmaması, metin içinde geçen kullanıcı adlarının gizlenmesi ve araştırma sonuçlarında kişilere değil içeriklere yönelik çıkarımlar yapılması potansiyel etik sorunların önüne geçmektedir.

Son olarak, araştırmanın yöntem ve bulgularının sosyal medya üzerine yapılan araştırmalar için literatürde belirtilen standartlara (Lienemann ve ark., 2017, pp. 17–18) uygun olarak raporlanmış olması karşılaştırılabilirlik ve tekrarlanabilirliği güçlendirmektedir.

5.7. Araştırmanın Kısıtlılıkları ve Gelecek Araştırmalar için Tavsiyeler

Araştırmada sosyal medya verisinin kendisinden kaynaklanan bazı kısıtlılıklar söz konusudur. Bunların başında sosyal medya metinlerinde dilin doğru kullanılmaması gelmektedir. Bu kısıtlılık İstanbul Teknik Üniversitesi Doğal Dil İşleme ekibi tarafından geliştirilen yazılım (Eryiğit ve Torunoğlu Selamet, 2017) kullanılarak modelin eğitilmesi öncesinde mümkün olduğu kadar düzeltilmiştir.

Veri toplamada Twitter tarafından sağlanan API kullanılmıştır. Bu API'ın hangi sınırlar içerisinde çalışacağı Twitter şirketi tarafından kurallara bağlanmıştır (Twitter, 2012). Bu kurallar araştırma açısından bazı kısıtlılıklar içermektedir.

Twitter API belli bir dönemde yayınlanan tüm tweet'lere erişmeye imkan vermeyip sadece belirtilecek anahtar kelimeleri içeren bir örneğe erişim imkanı vermektedir. Bu nedenle tüm tweet'ler içinde tütün kullanımını teşvik eden tweet'lerin sıklığı bulunamamıştır. Yine de, sosyal medyanın büyüklüğü ve içerdiği konuların yüksek

çeşitliliği göz önüne alındığında, tütün kullanımı teşvik eden tweet'lerin tütün ile ilgili tweet'ler içindeki yüzdesinin incelenmesi daha bilgilendirici olurdu.

Türkçenin sondan eklemeli yapısı nedeniyle farklı ek kombinasyonları ile bir kelimenin çok sayıda çekimi yapılabilir. Fakat Twitter API anahtar kelimelerin farklı çekimlerini içeren tweet'leri vermemekte ve kullanılabilecek anahtar kelime sayısını da sınırlamaktadır. Yine de bu sorunun araştırmada geliştirilen modellerin başarısını çok etkilemesi beklenmemektedir. Twitter'ın bir haftadan daha eski tweet'lere erişmek için yüksek bir ücret talep etmesi nedeniyle araştırmada veri toplama işlemi haftalık periyotta ve beş defada gerçekleştirilmiştir. İki veri toplama aralığında yayınlanan ve hemen silinen tweet'ler olduysa bunlar gözden kaçırılmış olabilir. Bu araştırmanın asıl amacı tütünle ilgili tüm tweet'leri en iyi şekilde temsil etmek değil başarılı bir makine öğrenmesi modeli geliştirmektir. Tütünle ilgili tweet'lerin tamamını dahil etmek isteyen gelecek araştırmalar ücretli Twitter API'lar ve geriye dönük periyodik aramalar yerine yayınlanan tweet'leri canlı olarak toplayan yazılımlar kullanabilir (Lienemann ve ark., 2017, p. 16).

Sunuculardaki yaşanan beklenmeyen bağlantı sorunları nedeniyle 2, 16, 26 ve 27 Ekim tarihlerinde belli saatlerde veri alınamamıştır (Bkz. Şekil 4-3). Bu durumun makine öğrenmesi modelinin başarısını etkilemesi beklenmemektedir. Sadece günlük ortalama tweet sayısı hesaplanırken, ortalama ve standart sapmayı etkilememesi için bu günler hariç tutularak 26 gün üzerinden hesaplama yapılmıştır.

Makine öğrenmesi için gereken eğitim veri seti gönüllü değerlendiriciler yardımıyla oluşturulurken müstehcen ifadeler içeren tweet'ler hariç tutulmuştur. Müstehcen ifadelerin, bir tweet'in tütün kullanımını teşvik etme durumunu etkileyip etkilemediği bilinmemektedir.

Twitter'da yeniden yayınlanma ve beğenilme sayıları tweet'in yayınlandığı andan itibaren sürekli değişmektedir. Genel olarak bu sayıların veri toplama tarihine yakın tarihte yayınlanan tweet'lerde daha düşük ve daha eski tweet'lerde ise daha yüksek olması beklenir. Bu durum geriye dönük aramalarda bu değişkenlerin tutarlılığını etkileyebilir. Bu sayıların araştırma döneminden altı ay sonra artık değişmeyecekleri varsayılarak bu sayılar altı ay sonra yapılan yeni bir sorgulamada tekrar toplanmıştır. Fakat arada geçen süre zarfında bazı tweet'lerin silindiği görülmüştür. Silinen tweet'lerin yüzdesi tütün kullanımı teşvik eden (%33,07) ve etmeyenler (%29,48) arasında istatistiksel olarak

önemli düzeyde farklılık göstermekteydi ($p<0,001$). Dolayısıyla bu sayılarla ilgili istatistikler (Tablo 4-12) yorumlanırken bu kısıtlılık göz önünde bulundurulmalıdır.

Bu araştırmada yalnızca Twitter'da yayınlanan içerikler kullanılmıştır. Geliştirilen modelinin diğer sosyal medya platformlarındaki başarıları değerlendirilmeli, farklı platformları da kapsayacak yeni araştırmalar yapılmalıdır.

Bu araştırma, tweet'lerin tütün kullanımını teşvik etme durumunu yalnızca metin kısmını dikkate alarak tespit etmektedir. Önceki araştırmalarda fotoğraf ve videolarda (Dhanwal ve ark., 2020) sigara (Kharade ve ark., 2021) ve nargile (Zhang ve ark., 2018) gibi tütün ürünlerini tespit eden makine öğrenmesi modelleri geliştirilmiş olduğu için bu araştırma metinlere odaklanmıştır. Gelecek araştırmalar iki yöntemi birleştiren daha bütüncül yaklaşımlar izleyerek daha iyi sonuçlar elde edebilir (Lienemann ve ark., 2017, p. 18).

Makine öğrenmesi modellerinin başarısı çeşitli mühendislik yaklaşımlarıyla iyileştirilebilmektedir. Bu araştırmada pratikte kullanılabilir bir başarı düzeyi elde edilmiştir. Yine de gelecekte mühendislerin dahil edildiği disiplinler arası ekipler daha başarılı modeller ve etkili uygulamalar geliştirebilir.

6. KAYNAKLAR

Aggarwal, C. C., ve Zhai, C. (2012). *Mining Text Data*. Springer US. https://doi.org/10.1007/978-1-4614-3223-4

Aksoy, A. (2018, June 29). *Türkçe Dolgu Sözcükleri*. https://github.com/ahmetax/trstop

Allem, J.-P., Ramanujam, J., Lerman, K., Chu, K.-H., Boley Cruz, T., ve Unger, J. B [Jennifer B.] (2017). Identifying Sentiment of Hookah-Related Posts on Twitter. *JMIR Public Health and Surveillance*, *3*(4), e74. https://doi.org/10.2196/publichealth.8133

Amazon. (2021, March 5). *Mechanical Turk*. https://www.mturk.com/

Aphinyanaphongs, Y., Lulejian, A., Brown, D. P., Bonneau, R., ve Krebs, P. (2016). Text Classification for Automatic Detection of E-Cigarette Use and Use for Smoking Cessation from Twitter: A Feasibility Pilot. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, *21*, 480–491.

Arklan, Ü., ve Rençber, H. (2018). İllegalitenin Kaçış Alanı Olarak Sosyal Medya. In E. Doğan & E. Geçğin (Eds.), *Current Debates in Public Relations Cultural & Media Studies* (pp. 31–52). IJOPEC Publication Limited.

Bigsby, E., ve Hovick, S. R. (2018). Understanding Associations between Information Seeking and Scanning and Health Risk Behaviors: An Early Test of the Structural Influence Model. *Health Communication*, *33*(3), 315–325. https://doi.org/10.1080/10410236.2016.1266575

Bird, S., Klein, E., ve Loper, E. (2009). *Natural language processing with Python*. O'Reilly.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., ve Kalai, A. (2016, July 22). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. http://arxiv.org/pdf/1607.06520v1

Centers for Disease Control and Prevention. (2019, December 2). *Global Tobacco Surveillance System Data*. https://www.cdc.gov/tobacco/global/gtss/gtssdata/index.html

Charniak, E., ve McDermott, D. V. (1986). *Introduction to Artifical Intelligence*. Addison-Wesley Publishing Company.

Chu, K.-H., Allem, J.-P., Unger, J. B [Jennifer B.], Cruz, T. B., Akbarpour, M., ve Kirkpatrick, M. G. (2019). Strategies to find audience segments on Twitter for e-cigarette education campaigns. *Addictive Behaviors*, *91*, 222–226. https://doi.org/10.1016/j.addbeh.2018.11.015

Clark, E. M., Jones, C. A., Williams, J. R., Kurti, A. N., Norotsky, M. C., Danforth, C. M., ve Dodds, P. S. (2016). Vaporous Marketing: Uncovering Pervasive Electronic Cigarette Advertisements on Twitter. *PloS One*, *11*(7), e0157304. https://doi.org/10.1371/journal.pone.0157304

Cole-Lewis, H., Varghese, A., Sanders, A., Schwarz, M., Pugatch, J., ve Augustson, E. (2015). Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning. *Journal of Medical Internet Research*, *17*(8), e208. https://doi.org/10.2196/jmir.4392

Davies, R., ve Chapman, M. (2021, February 20). Tobacco giant bets £1bn on influencers to boost 'more lung-friendly' sales. *The Guardian*. https://www.theguardian.com/business/2021/feb/20/tobacco-giant-bets-1bn-on-social-media-influencers-to-boost-lung-friendlier-sales

Devlin, J., Chang, M.-W., Lee, K., ve Toutanova, K. (2018a, October 11). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/pdf/1810.04805v2

Devlin, J., Chang, M.-W., Lee, K., ve Toutanova, K. (2018b, October 11). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. http://arxiv.org/pdf/1810.04805v2

Dhanwal, S., Bhaskar, V., ve Agarwal, T. (2020). Automated Censoring of Cigarettes in Videos Using Deep Learning Techniques. In P. K. Kapur, G. Singh, Y. S. Klochkov, & U. Kumar (Eds.), *Asset Analytics, 2522-5162. Decision analytics applications in industry* (pp. 339–348). Springer. https://doi.org/10.1007/978-981-15-3643-4_26

Dünya Sağlık Örgütü. (1986). *Ottawa Charter for Health Promotion*. https://www.euro.who.int/__data/assets/pdf_file/0004/129532/Ottawa_Charter.pdf

Dünya Sağlık Örgütü. (2008). *MPOWER: a policy package to reverse the tobacco epidemic*. https://apps.who.int/iris/bitstream/handle/10665/43888/9789241596633_eng.pdf

Emery, S. L., Szczypka, G., Abril, E. P., Kim, Y., ve Vera, L. (2014). Are you Scared Yet? Evaluating Fear Appeal Messages in Tweets about the Tips Campaign. *Journal of Communication*, *64*(2), 278–295. https://doi.org/10.1111/jcom.12083

Eryiğit, G., ve Torunoğlu Selamet, D. (2017). Social media text normalization for Turkish. *Natural Language Engineering*, *23*(6), 835–875. https://doi.org/10.1017/S1351324917000134

Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, *11*(1), e11. https://doi.org/10.2196/jmir.1157

Fishbein, M., ve Yzer, M. C. (2003). Using Theory to Design Effective Health Behavior Interventions. *Communication Theory*, *13*(2), 164–183. https://doi.org/10.1111/j.1468-2885.2003.tb00287.x

Galitsky, B. (2019). *Developing Enterprise Chatbots*. Springer International Publishing. https://doi.org/10.1007/978-3-030-04299-8

Global Burden of Disease Collaborative Network. (2020). *Global Burden of Disease Study 2019 (GBD 2019) Results.* Institute for Health Metrics and Evaluation (IHME). http://ghdx.healthdata.org/gbd-results-tool

Grand View Research. (2020). *E-Cigarette And Vape Market Size & Share Report*. https://www.grandviewresearch.com/industry-analysis/e-cigarette-vaping-market

Hackeling, G. (2014). *Mastering machine learning with scikit-learn: Apply effective learning algorithms to real-world problems using scikit-learn*. *Community experience distilled*. Packt Publ. http://www.gbv.de/dms/tib-ub-hannover/798957859.pdf

Jurafsky, D., ve Martin, J. H. (2020). *Speech and language processing*. https://web.stanford.edu/~jurafsky/slp3/6.pdf

Kahn, J. (2021, January 20). *A.I. breakthroughs in natural-language processing are big for business*. https://fortune.com/2020/01/20/natural-language-processing-business

Kaplan, S. (2018, August 24). Big Tobacco's Global Reach on Social Media. *The New York Times*. https://www.nytimes.com/2018/08/24/health/tobacco-social-media-smoking.html

Kavuluru, R., ve Sabbir, A. K. M. (2016). Toward automated e-cigarette surveillance: Spotting e-cigarette proponents on Twitter. *Journal of Biomedical Informatics*, *61*, 19–26. https://doi.org/10.1016/j.jbi.2016.03.006

Kharade, A., Abhishek, K., Dwibedi, D., Mehta, S., Meruga, H., Gangula, P., Narayana, D., ve Borse, R. (2021). Image Analytics to Detect Cigarette in an Image Using Deep Learning. In S. N. Merchant, K. Warhade, & D. Adhikari (Eds.), *Lecture Notes in Electrical Engineering, 1876-1100: Volume 703, Advances in Signal and Data Processing: Select Proceedings of ICSDP 2019* (pp. 659–678). Springer.

Kim, Y., Huang, J., ve Emery, S. (2016). Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *Journal of Medical Internet Research*, *18*(2), e41. https://doi.org/10.2196/jmir.4738

Kirkham, C. (2019, November 5). Exclusive: Philip Morris suspends social media campaign after Reuters exposes young 'influencers'. *Reuters Media*. https://www.reuters.com/article/us-philipmorris-ecigs-instagram-exclusiv/exclusive-philip-morris-suspends-social-media-campaign-after-reuters-exposes-young-influencers-idUSKCN1SH02K

Kowsari, Meimandi, J., Heidarysafa, Mendu, Barnes, ve Brown (2019). Text Classification Algorithms: A Survey. *Information*, *10*(4), 150. https://doi.org/10.3390/info10040150

Krauss, M. J., Sowles, S. J., Moreno, M., Zewdie, K., Grucza, R. A., Bierut, L. J., ve Cavazos-Rehg, P. A. (2015). Hookah-Related Twitter Chatter: A Content Analysis. *Preventing Chronic Disease*, *12*, E121. https://doi.org/10.5888/pcd12.150140

Küçükali, H. (2021). Halk Sağlığında Yapay Zeka. *SD (Sağlık Düşüncesi Ve Tıp Kültürü) Dergisi*(58).

Küçükali, H., Küçükali, İ., Rakkad, N., ve Rakkad, A. (2020, December 18). *Using artificial intelligence to identify tobacco-related contents in social media: a systematic scoping review of the literature*. 4th International 22nd National Public Health Congress, Online, Turkey. https://uhsk.org/2020/

Lewis, N. (2017). Information Seeking and Scanning. In *The International Encyclopedia of Media Effects.* American Cancer Society. https://doi.org/10.1002/9781118783764.wbieme0156

Lienemann, B. A., Unger, J. B [Jennifer B.], Cruz, T. B., ve Chu, K.-H. (2017). Methods for Coding Tobacco-Related Twitter Data: A Systematic Review. *Journal of Medical Internet Research, 19*(3), e91. https://doi.org/10.2196/jmir.7022

Liu, J., Lochbuehler, K., Yang, Q., Gibson, L. A., ve Hornik, R. C. (2020). Breadth of Media Scanning Leads to Vaping among Youth and Young Adults: Evidence of Direct and Indirect Pathways from a National Longitudinal Survey. *Journal of Health Communication*, *25*(2), 91–104. https://doi.org/10.1080/10810730.2019.1709925

Liu, J., Siegel, L., Gibson, L. A., Kim, Y., Binns, S., Emery, S., ve Hornik, R. C. (2019). Toward an Aggregate, Implicit, and Dynamic Model of Norm Formation: Capturing Large-Scale Media Representations of Dynamic Descriptive Norms Through Automated and Crowdsourced Content Analysis. *Journal of Communication*, *69*(6), 563–588. https://doi.org/10.1093/joc/jqz033

Merriam-Webster Dictionary. *Definition of crowdsourcing*. https://www.merriam-webster.com/dictionary/crowdsourcing

Michie, S., West, R., Campbell, R., Brown, J., ve Gainforth, H. (2014). *ABC of behaviour change theories*. Silverback Publishing.

Morgenstern, J. D., Rosella, L. C., Daley, M. J., Goel, V., Schünemann, H. J., ve Piggott, T. (2021). "AI's gonna have an impact on everything in society, so it has to have an impact on public health": A fundamental qualitative descriptive study of the implications of artificial intelligence for public health. *BMC Public Health*, *21*(1), 40. https://doi.org/10.1186/s12889-020-10030-x

Myslín, M., Zhu, S.-H., Chapman, W., ve Conway, M. (2013). Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of Medical Internet Research*, *15*(8), e174. https://doi.org/10.2196/jmir.2534

Oğuz, Ö. (2018). *Türkçe Küfür Karalistesi*. https://github.com/ooguz/turkce-kufur-karaliste

Oxford Dictionary. (2020). *Definition of API*. Oxford University Press (OUP). https://www.lexico.com/definition/api

Rowell, A. (2020, January 23). Big Tobacco wants social media influencers to promote its products – can the platforms stop it? *The Conversation*.

https://theconversation.com/big-tobacco-wants-social-media-influencers-to-promote-its-products-can-the-platforms-stop-it-129957

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. O'Reilly Media. https://books.google.com.tr/books?id=_VkrAQAAQBAJ

Sağlık Bakanlığı. (2018, May 30). *Tütün Kontrolü Strateji Belgesi ve Eylem Planı 2018-2023*. https://hsgm.saglik.gov.tr/depo/birimler/tutun-mucadele-bagimlilik-db/haberler/tutun_eylem_plani/Tutun_Kontrolu_Strateji_Belgesi_ve_Eylem_Plani.pdf

Salathé, M. (2018). Digital epidemiology: What is it, and where is it going? *Life Sciences, Society and Policy*, *14*(1), 1. https://doi.org/10.1186/s40504-017-0065-7

Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., Cattuto, C., Khandelwal, S., Mabry, P. L., ve Vespignani, A. (2012). Digital epidemiology. *PLoS Computational Biology*, *8*(7), e1002616. https://doi.org/10.1371/journal.pcbi.1002616

Saraçoğlu, S., ve Öztürk, F. (2020). Türkiye'de Tütün Kontrol Politikaları ve Tütün Tüketimi Üzerine Bir Değerlendirme. *Politik Ekonomik Kuram*, *4*(1), 20–44. https://doi.org/10.30586/pek.730271

Schoonenboom, J., ve Johnson, R. B. (2017). How to Construct a Mixed Methods Research Design. *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie*, *69*(Suppl 2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Schweter, S. (2020). *BERTurk - BERT models for Turkish* [Computer software]. Zenodo.

Shakeri Hossein Abad, Z., Kline, A., Sultana, M., Noaeen, M., Nurmambetova, E., Lucini, F., Al-Jefri, M., ve Lee, J. (2021). Digital public health surveillance: A systematic scoping review. *Npj Digital Medicine*, *4*(1), 41. https://doi.org/10.1038/s41746-021-00407-6

SocialTalent. (2015). *The Short Half-Life of Your Social Recruiting Posts*. https://www.socialtalent.com/blog/recruitment/the-short-half-life-of-your-social-media-posts

Statista. (2021). *Twitter: most users by country*. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Taulli, T. (2019). *Artificial Intelligence Basics*. Apress. https://doi.org/10.1007/978-1-4842-5028-0

4207 Sayılı Kanun, Resmi Gazete (1996 & rev. 2018). https://www.resmigazete.gov.tr/eskiler/2018/12/20181205-8.htm

Twitter. (2012). *Standard API v1.1* [Tweet]. Twitter. https://developer.twitter.com/en/docs/twitter-api/v1

Twitter. (2021a). *About Direct Messages* [Tweet]. Twitter. https://help.twitter.com/en/using-twitter/direct-messages

Twitter. (2021b). *Tweet object* [Tweet]. Twitter. https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet

van der Tempel, J., Noormohamed, A., Schwartz, R., Norman, C., Malas, M., ve Zawertailo, L. (2016). Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *International Journal of Public Health*, *61*(2), 249–256. https://doi.org/10.1007/s00038-016-0791-2

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. von, Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., . . . Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Yoo, W., Yang, J., ve Cho, E. (2016). How social media influence college students' smoking attitudes and intentions. *Computers in Human Behavior*, *64*, 173–182. https://doi.org/10.1016/j.chb.2016.06.061

Zhang, Y., Allem, J.-P., Unger, J. B [Jennifer Beth], ve Boley Cruz, T. (2018). Automated Identification of Hookahs (Waterpipes) on Instagram: An Application in Feature Extraction Using Convolutional Neural Network and Support Vector Machine Classification. *Journal of Medical Internet Research*, *20*(11), e10513. https://doi.org/10.2196/10513

Zhu, Y. (2017). Pro-smoking information scanning using social media predicts young adults' smoking behavior. *Computers in Human Behavior*, *77*, 19–24. https://doi.org/10.1016/j.chb.2017.08.004

ETİK KURUL KARARI

İÜC Tarih ve Sayı: 22/05/2020-26926

T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
REKTÖRLÜĞÜ
Girişimsel Olmayan Klinik Araştırmalar Kurulu Başkanlığı

Sayı :74555795-050.01.04-
Konu :2020-06 sayılı Etik Kurul Kararı

Sayın Hüseyin KÜÇÜKALİ
İÜC Cerrahpaşa Tıp Fakültesi Halk Sağlığı Anabilim Dalı
Halk Sağlığı Doktora Programı Öğrencisi

İlgi :27.04.2020 tarihli proje başvurunuz.

Yürütücülüğünü üstlendiğiniz **"Tütün Bağımlılığı Örneğinde Halk Sağlığında Yapay Zeka Teknolojilerinin Kullanılması"** başlıklı **"Doktora Tezi"** başvurunuz etik kurulumuz tarafından 06.05.2020 tarihli toplantıda değerlendirilmiş ve etik açıdan uygun olduğuna oybirliği ile karar verilmiştir.

ÖNEMLİ AÇIKLAMA
COVİD-19 salgının nedeniyle, kurulumuz toplantısını on-line platformda gerçekleştirilmiş olup toplantısında yapılandeğerlendirmeler sonucunda; elektronik belge sistemi (EBYS) ile uygunluğu bildirilen yazının ilgili makamlarca "Etik Kurul Onayı" yerine geçen belge olarak değerlendirilmesi hususunda gereği için bilgilerinizi rica ederi

e-İmzalı
Doç. Dr. Ayşem KAYA
Başkan

ETİK KURUL ÜYELERİ
Doç. Dr. Ayşem KAYA (Başkan)
Prof. Dr. Nurten KAYA (Başkan Yrd)
Prof. Dr. Sibel ÖZYAZGAN
Prof. Dr. AYFER ÖZBAŞ
Doç. Dr. Günay CAN
Doç. Dr. İlker YÜCESİR
Doç. Dr. Okay ABACI
Avukat Seren ÇİÇEK (Sivil Üye)
Emk. Lab.Teknk Şenel ÜRESİN (Sivil Üye)

21/05/2020 V.Haz.Kont.İş. : C.SÖNMEZTÜRK

Doğrulamak İçin:http://dogrulama.istanbulc.edu.tr/enVision.sorgula/belgedogrulama.aspx?V=BECD7VS8J

Ayrıntılı bilgi için irtibat : Canan SÖNMEZTÜRK
İstanbul Üniversitesi-Cerrahpaşa Rektörlüğü, 34320 Avcılar-İstanbul
Tel : 0212 404 03 00 Faks : 0212 404 07 01
Elektronik Ağ : www.istanbulc.edu.tr

Bu belge 5070 sayılı Elektronik İmza Kanununun 5. Maddesi gereğince güvenli elektronik imza ile imzalanmıştır.

TWITTER GELİŞTİRİCİ YETKİSİ

Account Application Approved

Twitter Developer Accounts <developer-accounts@twitter.com> 25 Nisan 2019 16:26
Alıcı: Hüseyin Küçükali

Developer

Account Application Approved

Your Twitter developer account application has been approved!

Thanks for applying for access. We've completed our review of your application, and are excited to share that your request has been approved.

Sign in to your developer account to get started.

Thanks for building on Twitter!

developer.twitter.com | **@twitterdev**

Twitter, Inc. 1355 Market Street, San Francisco, CA 94103

İTÜ DOĞAL DİL İŞLEME ARAÇLARI YETKİSİ

RE: API Erişim Talebi

Dilara Torunoğlu <torunoglud@itu.edu.tr>
19.10.2020 Pzt 17:48
Kime: Hüseyin KÜÇÜKALİ <hkucukali@medipol.edu.tr>
Merhabalar,

Kullanıcı adı ve şifre işlemleriniz aşağıdaki gibidir:

Kullanıcı adı: hkucukali
Şifre: [redacted]

Sisteme giriş yaptıktan sonra elde edeceğiniz token bilgisi ile araçlarımızı kullanabilirsiniz.

İyi çalışmalar,

Dilara

From: Hüseyin KÜÇÜKALİ [mailto:hkucukali@medipol.edu.tr]
Sent: 14 Ekim 2020 Çarşamba 10:28
To: Dilara Torunoğlu <torunoglud@itu.edu.tr>
Subject: Ynt: API Erişim Talebi

Merhaba Dilara Hanım,
Lisans anlaşmasını imzaladım, ekte gönderiyorum.

Değerli çalışmalarınız için teşekkür eder, kolaylıklar dilerim.

-
Dr. Hüseyin Küçükali
Tıp Fakültesi Halk Sağlığı Ana Bilim Dalı
Araştırma Görevlisi

İSTANBUL MEDİPOL ÜNİVERSİTESİ
Kavacık Güney Kampüs, Göztepe Mah.
34815 Beykoz, İstanbul

+90 216 681 5100

Gönderen: Dilara Torunoğlu <torunoglud@itu.edu.tr>
Gönderildi: 13 Ekim 2020 Salı 09:05
Kime: Hüseyin KÜÇÜKALİ <hkucukali@medipol.edu.tr>
Konu: RE: API Erişim Talebi

Merhabalar,

Api'ye erişim izni için ekte bulunan belgeyi imzalayıp taratıp tarafımıza iletirseniz, kullanıcı adı ve şifre işlemleri başlatılacak sonrasında size verilecek olan token bilgisi ile api'ye erişiminiz sağlanacaktır.

Sunucularımızdaki yoğunluk nedeni ile kullanıcı bazlı kota sınırlandırılması bulunmaktadır. Sınır oldukça yüksektir ancak, verilerinizi tarattırdıktan sonra sonuçları lokalinize kayıt ederek devam etmenizi tavsiye ederiz.

İyi çalışmalar Dilerim,

İTÜ DDİ Grubu adına,

Dilara Torunoğlu

İTÜ

Lect. Dilara Torunoğlu Selamet
torunoglud@itu.edu.tr

ISTANBUL TECHNICAL UNIVERSITY
COMPUTER ENGINEERING

From: Gülşen Eryiğit [mailto:gulsenc@itu.edu.tr]
Sent: 12 Ekim 2020 Pazartesi 17:39
To: 'Hüseyin KÜÇÜKALİ' <hkucukali@medipol.edu.tr>; gulsen.cebiroglu@itu.edu.tr
Cc: 'Dilara Torunoğlu' <torunoglud@itu.edu.tr>
Subject: RE: API Erişim Talebi

Merhaba Hüseyin bey,
Araçlarımıza ilginiz için teşekkür ederiz.
Dilara hanım Hüseyin bey kayıt açma işlemlerini başlatabilir miyiz?

İTÜ

Assoc. Prof. Dr. Gülşen Eryiğit
Department Vice Chair
gulsen.cebiroglu@itu.edu.tr

ISTANBUL TECHNICAL UNIVERSITY
ARTIFICIAL INTELLIGENCE AND DATA ENGINEERING
İTÜ Ayazağa Kampüsü Bilgisayar Ve Bilişim Fakültesi Sarıyer / İstanbul
P. 0212 285 3503

TARIM BAKANLIĞI BİLGİ EDİNME BAŞVURUSU CEVABI

CİMER Başvuru Cevabı

cimer_bilgi@tccb.gov.tr <cimer_bilgi@tccb.gov.tr>
22.04.2021 Per 11:07
Kime: huseyinkucukali@hotmail.com <huseyinkucukali@hotmail.com>

CİMER | CUMHURBAŞKANLIĞI İLETİŞİM MERKEZİ

Sayın HÜSEYİN KÜÇÜKALİ,

T.C. Cumhurbaşkanlığı İletişim Merkezi (CİMER)' ne 16.04.2021 tarihinde yapmış olduğunuz 2101749869 sayılı başvurunuz 22.04.2021 tarihinde PİYASA DÜZENLEME VE KONTROL DAİRE BAŞKANLIĞI tarafından cevaplanmıştır:

CİMER aracılığıyla Başkanlığımıza iletilmiş olan başvurunuz ile halk sağlığı ve tütün kontrolü alanında yaptığınız çalışmalarda kullanmak üzere 2018-2023 Tütün Kontrolü Eylem Planı'nda yer alan 3.4 ve 3.6 nolu faaliyetlere ilişkin Bakanlığımızın 2018-2019-2020-2021 yıllarında ayrı ayrı yaptığı çalışmalara ilişkin tarafınıza bilgi verilmesi istenilmiştir.

2018-2023 Tütün Kontrolü Eylem Planı "Reklam, Promosyon ve Sponsorluk" başlığı altında yer alan "**3.4. Reklam ve örtülü reklam ihlalinin izlenmesini sağlayacak sistemin ilgili kurumların eşgüdümü ile güçlendirilmesi**" faaliyetine ilişkin olarak, Eylem Planı çerçevesinde Sağlık Bakanlığı koordinasyonunda çalışmalar devam etmektedir. Ayrıca, görev alanımız kapsamına giren reklam ihlallerine yönelik gerek Başkanlığımız personeli tarafından yapılan denetimler sonucunda tespit edilen gerekse diğer kamu kurum ve kuruluşları ile tüzel kişilikler ve vatandaşlar tarafından Başkanlığımıza intikal ettirilen tespitlere istinaden; *4207 sayılı Tütün Ürünlerinin Zararlarının Önlenmesi ve Kontrolü Hakkında Kanun, 4733 sayılı Tütün, Tütün Mamulleri ve Alkol Piyasasının Düzenlenmesine Dair Kanun* ve reklam ihlallerine ilişkin ayrıntılı düzenlemeler içeren *Tütün Mamulleri ve Alkollü İçkilerin Satışına ve Sunumuna İlişkin Usul ve Esaslar Hakkında Yönetmelik* hükümleri doğrultusunda Başkanlığımızca idari yaptırımlar uygulanmaktadır.

2018-2023 Tütün Kontrolü Eylem Planı "Reklam, Promosyon ve Sponsorluk" başlığı altında yer alan "**3.6. Tütün endüstrisinin, gençlerin tütün ürünü kullanmalarını teşvik eden ve tütün ürünlerine erişimlerini kolaylaştıran kampanya ve programlarının tespit edilmesine yönelik bir sistem geliştirilmesi**" faaliyeti ile ilgili olarak da söz konusu Eylem Planı çerçevesinde ilgili kurumlarla eşgüdümlü çalışmalar sürdürülmektedir. Ayrıca konu ile ilgili gerek Başkanlığımız personeli tarafından yapılan denetimler sonucunda tespit edilen gerekse diğer kamu kurum ve kuruluşları ile tüzel kişilikler ve vatandaşlar tarafından Başkanlığımıza intikal ettirilen tespitlere istinaden yukarıda yer verilen mevzuat hükümleri kapsamında idari yaptırımlar uygulanmaktadır.

Diğer taraftan, gençlere yönelik tütün mamulü kullanımını teşvik edici ve özendirici reklam-tanıtım faaliyetleri ile kampanya ve promosyon uygulamalarını önlemeye yönelik hem satış noktalarında hem de 81 İl genelinde düzenlenen özellikle gençlerin katılım sağladığı festival ve fuar alanlarında ilgili diğer kamu kurum ve kuruluşları ile eşgüdümlü olarak denetimlerin yapılması sağlanmaktadır.

Yukarıda belirtilen 3.4 ve 3.6 nolu faaliyetler kapsamında gerçekleştirilen çalışmalara ilişkin Daire Başkanlığımız uhdesinde yıllar itibariyle ayrıştırılmış istatistiki veri bulunmadığından *4982 sayılı Bilgi Edinme Hakkı Kanunu*'nun "*İstenecek Bilgi veya Belgenin Niteliği*" başlıklı 7 inci maddesi kapsamında tarafınıza daha ayrıntılı bilgi verilememiştir.

Bilgilerinize sunulur.

İNTİHAL RAPORU İLK SAYFASI

TÜTÜN BAĞIMLILIĞI ÖRNEĞİNDE HALK SAĞLIĞINDA YAPAY ZEKA TEKNOLOJİLERİNİN KULLANILMASI

ORİJİNALLİK RAPORU

%3	%3	%0	%1
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	nek.istanbul.edu.tr:4444 İnternet Kaynağı	%1
2	tubav.yt.com.tr İnternet Kaynağı	<%1
3	havanikoru.org.tr İnternet Kaynağı	<%1
4	openaccess.hku.edu.tr İnternet Kaynağı	<%1
5	polen.itu.edu.tr İnternet Kaynağı	<%1
6	Submitted to TechKnowledge Turkey Öğrenci Ödevi	<%1
7	Submitted to Beykent Universitesi Öğrenci Ödevi	<%1
8	www.fixsem.com İnternet Kaynağı	<%1
9	cdn.istanbul.edu.tr İnternet Kaynağı	<%1

ÖZGEÇMİŞ