



Article Info/Makale Bilgisi

Received/Geliş:05.04.2022 Accepted/Kabul:29.04.2022

DOI:10.30794/pausbed.1099169

Research Article/Araştırma Makalesi

Kuş, O. (2022). "İletişim Çalışmalarında İçerik Analizi ve Metin Madenciliğinin Kullanımı", *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2022 Sayı 51: Özel sayı 1, Denizli, ss. Ö177-Ö191.

İLETİŞİM ÇALIŞMALARINDA İÇERİK ANALİZİ VE METİN MADENCİLİĞİNİN KULLANIMI

Oğuz KUŞ*

Öz

İletişim çalışmaları çerçevesinde büyük veri kavramını göz önünde bulunduran yöntemsel yaklaşımlar gelişmektedir. Bu bağlamda içerik analizi ve metin madenciliğinin bir arada kullanıldığı gözlemlenmektedir. Bu yaklaşım gerçek zamanlı bulgular elde etmek, geniş veri setlerini değerlendirebilmek ve etkin sonuçlar elde etmek noktasında önemli bir potansiyel sunmaktadır. Bu çalışma tümdengelimci içerik analizi sürecine, metin madenciliğinin sistematik bir şekilde nasıl entegre edilebileceğini aşamalar halinde sunmayı amaçlamaktadır. Çalışmada metin madenciliği temelli tümdengelimci bir içerik analizi süreci önerilmektedir. Bu süreç farklı bağlamlarda, farklı programlarla, teknik olarak yüksek donanım gereksinimi gözetmeden dijital metin verisine uygulanabilecek yöntemsel bir yol haritası sunmaktadır. Bu bağlamda içerik analizinin aşamaları gözden geçirilmiş, tümdengelimci içerik analizi yaklaşımı tartışılmış, geleneksel içerik analizi yaklaşımlarının metin madenciliği süreçleriyle nasıl gerçekleştirilebileceğine yönelik araştırmacının geçmiş deneyimleri ve öncül araştırmalar ışığında bir süreç geliştirilmiştir.

Anahtar Kelimeler: İçerik analizi, Tümdengelimci içerik analizi, Büyük veri, Metin madenciliği.

THE USE OF CONTENT ANALYSIS AND TEXT MINING IN COMMUNICATION STUDIES

Abstract

Methodological approaches that consider the concept of big data are emerging in communication studies. In this regard, combined usage of content analysis and text mining were observed. This approach presents potential to obtain real-time findings, evaluate large data sets, and obtain effective results. This study aims to present how text mining can be systematically integrated into the deductive content analysis process. In the study, a deductive content analysis process based on text mining is proposed. This process can provide researchers with a methodological roadmap that can be applied to digital text data in different contexts, with different software, without high technical skill requirements. In this context, the stages of content analysis have been reviewed, the deductive content analysis approach was discussed, a process was developed in the light of the researcher's past experiences and preliminary research on how traditional content analysis approaches can be realized with text mining processes.

Keywords: Content analysis, Deductive content analysis, Big data, Text mining.

*Dr., İstanbul Üniversitesi, İletişim Fakültesi, İSTANBUL,
e-posta: oguz.kus@istanbul.edu.tr, (<https://orcid.org/0000-0002-2593-4980>)

GİRİŞ

Büyük veri ve büyük veriye bağlı olarak gelişen kavram ve yöntemler her geçen gün iletişim alanında gerçekleştirilen çalışmalarla daha bütünleşik bir hal almıştır. Özellikle yeni medya platformlarından veri toplamayı mümkün kılan API'ların (application processing interface), açık şekilde lisanslanmış veri toplama arayüzlerinin ve yazılım bilme zorunluluğunu belirli miktarda ortadan kaldırarak iş akışı şeklinde algoritmalar tasarlamayı mümkün kılan yazılımların yaygınlaşmasıyla iletişim alanında büyük veri perspektifine bağlı şekilde gelişen yöntemlerin kullanımı görünür hale gelmiştir. Bu noktada Helles ve Ørmen (2020:291) *“büyük verinin medya ve iletişim araştırmalarında kendi içinde makine öğrenimi teknikleriyle yakından ilişkili yöntemsel bir paradigma olarak görülmemesi gerektiğini, bunun yerine, medya ve iletişim alanında araştırma yapmanın mevcut yollarının çoğuna verimli bir şekilde entegre edilmiş yeni bir analitik teknikler seti olarak görülmesi gerektiğini”* vurgulamaktadır. Bu perspektif göz önünde bulundurulduğunda büyük veri tartışmalarının beraberinde getirdiği yöntemlerin iletişim çalışmalarında sıklıkla başvurulan içerik analizi yaklaşımına entegre edilebileceğini tartışmak mümkün hale gelmektedir. Hızla değişen toplumsal gündemlere iletişim perspektifinden yorumların getirilebilmesi noktasında büyük veri konsepti ile birlikte iletişim çalışmalarının kullandığı yöntem ve tekniklere entegre olan betimsel ve öngörüsül veri madenciliği yaklaşımlarının; otomatize bir şekilde büyük miktarda veriye ulaşma olanağının iletişim araştırmacılarına gündemle simültane araştırmalar gerçekleştirmek üzere ihtiyaç duydukları çevikliği kazandırabileceğinin altını çizmek mümkündür.

Büyük veri konseptine bağlı bir şekilde gelişen kavram ve yöntemlerin medya ve iletişim alanında yaygın kullanılan içerik analizine entegrasyonuna halihazırda ulusal ve uluslararası ölçekte yayımlanan araştırmalarda rastlanmaktadır. İletişim alanında otomatize içerik analizi süreçlerinin kullanıldığı erken dönem olarak betimlenebilecek çalışmalardan birisi Aytekin ve Sütcü (2013) tarafından gerçekleştirilmiştir. Aytekin ve Sütcü (2013) gerçekleştirmiş oldukları çalışmada sosyal medyada demokrasi olgusunun eleştirel bir biçimde tartışılıp tartışılmadığına yönelik bir analiz gerçekleştirmişler ve bu süreçte bir sınıflandırma algoritması olan Naive-Bayes algoritmasından faydalanmışlardır. Kuş (2016) tarafından gerçekleştirilen araştırmada ise sosyal ağlardan otomatize gönderi toplama süreçleri ve toplanan bu gönderilerin içerik analizi sürecine entegre edilmiş metin madenciliği ile analiz edilmesine yönelik bir yöntemin benimsendiğini görmek mümkündür. Araştırmacı, k-NN (k-Nearest Neighbour) sınıflandırma algoritmasını kullandığı bu yaklaşımla Suriyeli mültecilere yönelik haberlere yazılmış 7000'den fazla kullanıcı kaynaklı gönderinin duygu analizini gerçekleştirmiş ve dijital nefret söylemine yönelik bir kategorizasyon ortaya koymuştur. Atan (2018), hastanelerle ilgili 3 binin üzerinde haberi içerik analizi sürecine entegre ettiği metin madenciliği tekniği ile incelemiş; haberlerin duygu durumu ve hastane imajına etkisi arasındaki bağlantıyı aydınlatmayı amaçlamış ve insan değerlendirmesi ve makine değerlendirmesinin birbirlerine olan üstün ve zayıf yönlerini tartışmıştır. Güçdemir vd. (2020) tarafından gerçekleştirilen çalışma ise metin madenciliğinin iletişim çalışmalarında kullanıldığı yakın dönem araştırmalardandır. Araştırmacılar, veri toplama sürecinde API kullanarak Twitter'dan veri toplamış ve RStudio'da oluşturulan sözlük tabanlı bir algoritma aracılığı ile İstanbul Büyükşehir Belediyesi Seçimleri'ne aday olan liderlerin geliştirmiş oldukları söylemi siyasal iletişim bağlamında analiz etmişlerdir.

Metin madenciliğinin içerik analizi sürecine entegre edildiği uluslararası çalışmalar da mevcuttur. Osakwe ve diğerleri (2020), COVID-19 pandemisi sürecinde Twitter'ın API'nı kullanarak spesifik kelimeleri içeren tweetleri toplamışlar ve SAS Text Miner V.15.1 kullanarak tweetlerin betimsel metin madenciliğini gerçekleştirmişlerdir. Araştırmacılar bu yaklaşımla altı farklı kaygı temasına ulaşmışlardır. Yang ve Han (2021) konaklama endüstrisinin COVID-19 pandemisine tepkisini kullanıcı kaynaklı içerik üzerinden değerlendirmiş ve 94 binden fazla tweeti güdümsüz bir makinelerin öğrenmesi yaklaşımı olan konu modellemesi ile analiz etmişler ve bu endüstrinin COVID-19 sürecinde deneyimlediği zorlukları, endişeleri ve geliştirmiş oldukları tepkileri betimlemeyi amaçlamışlardır. Redek ve Godnov (2017) tarafından gerçekleştirilen araştırmada ise Avrupa Birliği üyesi 15 ülkedeki politikacıların göndermiş olduğu 43 bin tweet toplanmış ve konu modelleme/içerik analizi ve duygu analizini içeren bir metin madenciliği yaklaşımı benimsenerek ekonomik kriz sürecinde Twitter'ın politik kullanım tipolojisi anlaşılmaya çalışılmıştır.

Elbette içerik analizi ve metin madenciliğinin birlikte anılabileceği araştırmalar yukarıda bahsedilen çalışmalarla sınırlı değildir. Fakat bahsedilen örnekler içerik analizi ve metin madenciliği yaklaşımlarının iletişim çalışmalarını

da ilgilendiren mikro, mezo ve makro ölçekte değerlendirilebilecek farklı tematik kırılımların anlaşılmasında, değerli içgörülerin elde edilmesinde ve yüksek miktarda verinin mevcut olduğu durumlarda kullanılabilmesine işaret etmektedir. Diğer yandan, içerik analizi bağlamında metin madenciliği kullanılırken güdümlü, güdümsüz ve güdümlü/güdümsüz bir arada makinelerin öğrenmesi yaklaşımlarının kullanıldığını gözlemlemek mümkündür. Bu bağlamda konu modelmeden, duygu analizi ve sınıflandırma algoritmalarına uzanan bir çeşitlilikten söz etmek mümkündür. Bahsi geçen çalışmalarda içerik analizi sürecine entegre bir biçimde çalışan metin madenciliği algoritmalarını oluşturmak için araştırmacılar tarafından özgün programların yazıldığı, RStudio gibi yaygınlığı yüksek ve kütüphane temelli açık kaynak veri bilimi uygulamalarının kullanıldığı ve iş akışı mantığında çalışan Rapidminer gibi uygulamalardan faydalandığı da gözlemlenmektedir.

İletişim çalışmaları çerçevesinde büyük veri kavramıyla gelişen yönetsel yaklaşımlar, içerik analizi ve metin madenciliğinin bir arada kullanılmasının gerçek zamanlı bulgular elde etmek, geniş veri setlerini değerlendirebilmek ve etkin sonuçlar elde etmek noktasında fayda sağlama potansiyeli bulunduğu aşikardır. Fakat iletişim çalışmalarının sahip olduğu perspektifi göz önünde bulundurarak içerik analizi ve metin madenciliği süreçlerinin entegre bir şekilde kullanımına ve bu sürecin aşamalarına ışık tutan çalışmalar sınırlıdır. İçerik analizi ve metin madenciliği, yukarıda değinilenler de dahil olmak üzere farklı çalışmalarda kullanılıyor olsa dahi, iletişim çalışmaları bağlamında bu konuya bir yöntem tartışması çerçevesinde odaklanan çalışmalar sayıca azdır. Bu çalışma, iletişim çalışmalarında yaygın bir şekilde kullanılan tümdengelimci içerik analizi sürecine metin madenciliği süreçlerinin sistematik bir şekilde nasıl entegre edilebileceğini aşamalar halinde ortaya koymayı amaçlamaktadır. Bu bağlamda içerik analizinin aşamaları gözden geçirilmiş, tümdengelimci içerik analizi yaklaşımı tartışılmış, geleneksel içerik analizi yaklaşımlarının metin madenciliği süreçleriyle nasıl gerçekleştirilebileceğine yönelik araştırmacının geçmiş deneyimleri ve öncül araştırmalar ışığında bir süreç geliştirilmiştir.

1. GELENEKSEL İÇERİK ANALİZİ VE AŞAMALARI

Metin madenciliği tekniğinin içerik analizine entegrasyonunu tartışmaya başlamadan önce, içerik analizinin düşünce sistemi üzerine bir tartışma gerçekleştirmek elzemdir. Bu bağlamda içerik analizinin temel tanım ve işlevlerinden bahsederek, farklı içerik analizi perspektifleri hakkında fikri bir zemin oluşturmak iletişim çalışmalarında metin madenciliğini etkin bir şekilde kullanmayı mümkün kılma potansiyeli taşımaktadır.

Sallan Gül ve Kahya Nizam (2021:182) başlangıçta içerik analizinin iletişim alanında ortaya çıktığına fakat günümüzde sosyal bilimlerin farklı disiplinlerinde de kullanıldığının altını çizmektedir. İletişim çalışmaları bağlamında içerik analizi hakkında en eski tanımlardan bir tanesini ortaya koyan Berelson içerik analizini “*açık iletişim içeriğini objektif, sistematik ve niceliksel şekilde açıklamayı amaçlayan bir araştırma tekniği*” şeklinde betimlemektedir (Berelson’dan akt. Drisko & Maschi, 2016:3). Berelson’un perspektifinden değerlendirildiğinde içerik analizi nicel bir temelde yükselmektedir. Bu noktadan hareketle araştırmacının içeriği sayılabilir bir hale dönüştürme amacı taşıdığını belirtmek mümkün hale gelmektedir. Bu durum içerik analizi sürecinin sonunda içeriği bir değişken haline getirebilmektedir. Ayrıca objektifliğin ve tekrar edilebilirliğin içeriği sayısallaştırma girişiminden kaynaklandığını belirtmek mümkün hale gelmektedir. Berelson’un yaklaşımının da dâhil olduğu temel içerik analizi yaklaşımında genellikle araştırma amacının dışında başkaları tarafından üretilmiş metinlerin analiz edildiğine işaret edilmektedir (Drisko & Maschi, 2016:3). Bu yaklaşım metnin devamında veri kaynağı olarak betimlenecek olan kullanıcı kaynaklı dijital içeriklerin analizi hususunda yol göstericidir.

Krippendorf içerik analizini, kullanımları bağlamında metinlerden (veya anlamlı diğer materyallerden) tekrar edilebilir ve geçerli çıkarımlar elde etmeye olanak tanıyan bir araştırma tekniği olarak betimlemektedir. İçerik analizinin yeni içgörüler sağladığını, araştırmacının belirli bir olgu konusundaki anlayışını geliştirdiğini veya pratik aksiyonlara bilgi sağladığını vurgulamaktadır (Krippendorf, 2004:18). İletişim ve medya çalışmaları bağlamında içerik analizine sıklıkla başvurulmaktadır. İçerik analizi diğer disiplinlerle yoğun bir etkileşim halinde olan iletişim çalışmaları için değerli ve derinlemesine bulgular elde etme imkânı tanımaktadır. Bu bağlamda Riffe ve diğerlerinin (2019) görüşlerinin altı çizilmelidir. Riffe ve diğerleri (2019:20) içerik analizinin amacının ve prosedürünün sosyal bilimler yaklaşımını bilgi zeminine çektiğini vurgulamaktadır ve bunu insan davranışlarını ve zihinsel süreçleri tanımlayan ve açıklayan ilişkisel ifadeler oluşturmak için bir standartlar ve kılavuzlar sistemi olarak betimlemektedir. Riffe ve diğerlerinin sunmuş olduğu düşünsel perspektif yukarıda bahsedilen

araştırmacıların önemli bir bileşen olarak sunduğu sistematik olma haliyle örtüşmektedir. Diğer yandan, nicel yaklaşımlarla birer değişken haline indirgenen içerik arasında ilişki kurmak da mümkün hale gelebilmektedir. Öncül literatür bağlamında içerik analizinin nitel, nicel ve yorumsayıcı yönleri olabileceğini tartışılabilir (Drisko & Maschi, 2016). Ayrıca perspektif olarak içerik analizi tümdengelimci ve tümevarımcı şeklinde sınıflandırılabilir. Çalışmanın devamında güdümlü bir makinelerin öğrenmesi süreci kullanılacağından ve bu bağlamda bir eğitim seti geliştirilmesi gerektiğinden, tümdengelimci yaklaşımların eğitim seti geliştirmeyi daha ergonomik hale getirebileceğini tartışmak mümkündür.

Bu bağlamda metin madenciliğinin içerik analizine entegrasyonunda tümdengelimci içerik analizi perspektifini benimseyen bir tartışma yürütmek, hem bu tekniği daha kapsayıcı bir şekilde anlamayı mümkün kılmakta hem de metin madenciliğinin içerik analizi tekniğine entegre edildiği yönetsel bir yol haritası oluşturmak üzere etkin bir zemin sunmaktadır. Tümdengelimci bir perspektifi benimsemiş olan içerik analizine metin madenciliğini entegre etmek üzere içerik analizinin aşamalarının ortaya konulması elzemdir. Farklı araştırmacıların takip eden aşamalar üzerinde uzlaştığını belirten Piñeiro Naval (2020:3) içerik analizinin on farklı aşamadan oluştuğunu belirtmektedir:

1. Amaçların belirlenmesi
2. Araştırma sorularının ve/veya hipotezlerin oluşturulması
3. Araştırma evreninin belirlenmesi
4. Çalışma örnekleminin seçilmesi
5. Kod kitabının ve kodlama formunun tasarımı ve geçerliği
6. Kodlayıcıların eğitimi
7. Örneklemin tasarlanan araçlar kullanılarak kodlanması
8. Analizin nesnelliliğinin değerlendirilmesi: kodlayıcılar arası güvenilirlik
9. Verilerin istatistiksel analizi
10. Nihai araştırma raporunun taslağının hazırlanması

Hansen'in aşamalarına değinen Taylan (2011:74) ise içerik analizinin aşamalarını takip eden şekilde betimlemektedir.

1. Araştırma Problemini Tanımlama-Konu ve Sorun
2. Araştırma Evrenini Belirleme ve Örneklem Seçimi
3. Araştırma Kategorilerini Oluşturma ve Tanımlama
4. Kodlama Cetvelini Oluşturma
5. Kodlama Cetvelini Sınama ve Güvenirliğini Ölçme
6. Veri Girişi, Analizi ve Yorumlama

Her iki araştırmacının sunduğu adımlar gözden geçirildiğinde içerik analizi gerçekleştirmek üzere takip edilmesi gereken aşamalar farklılık gösteriyor olsa dahi, aşamaların özü göz önünde bulundurulduğunda içerik analizinin spesifik ve örtüşen bir izleği takip ettiği yorumunu gerçekleştirmek mümkündür. Ayrıca belirtilen aşamaların Elo & Kyngäs (2008) tarafından belirtilen ve *hazırlık, organize etme, raporlama* adımlarını içeren üç aşamalı yaklaşımla da örtüştüğü gözlemlenmektedir. Bu aşamalar üzerinden tümdengelimci içerik analizi perspektifini tartışmak önem arz etmektedir.

2. TÜMDENGELİMCİ İÇERİK ANALİZİNE YÖNELİK KAVRAMSAL BİR ÇERÇEVE

Tümdengelimci içerik analizi öncül bilgi birikimi üzerine kuruludur ve teori testi amacı gütmektedir. İçeriğin kategorizasyonu sürecinde araştırma amacı doğrultusunda öncül bilgiye dayanan yapılandırılmış/yapılandırılmamış kategorizasyon matrisi kullanılmalıdır (Elo & Kyngäs, 2008:109-111). Bu noktadan hareketle tümdengelimci içerik analizi perspektifinin araştırmacılara mevcut teorik zemini kullanarak, aynı tematik bağlama sahip farklı olay ve olgular hakkında sorgulamalar gerçekleştirmek üzere olanak tanıdığı belirtilebilir. Tümdengelimci içerik analizinin bu özelliğini örneklemek daha berrak bir perspektifin oluşturulmasını mümkün kılacaktır. Amerika Birleşik Devletleri'nde seçim sürecinde yanlış bilginin dijital dolaşımını inceleyen bir araştırmacının burada ortaya çıkan yanlış bilgi kategorilerini COVID-19 pandemisi sırasında ortaya çıkan yanlış bilginin kategorizasyonunda kullanması etkin sonuçlar vermeme riski taşırken; bu kategorileri Fransa'da seçim sürecinde yanlış bilginin dijital dolaşımını inceleyen bir çalışma bağlamında kullanması daha etkin sonuçlar ortaya çıkaracaktır. Farklı araştırmacıların görüşlerinden faydalanan Graneheim ve diğerleri (2017:30) tümdengelimci yaklaşımın teoriden veriye veya daha soyut ve genel bir seviyeden spesifik bir seviyeye doğru hareket ettiklerinin altını çizmekte ve bu bağlamda iki riskin üzerinde durmaktadır. Bunlar tümdengelimci bir perspektifle içerik analizi gerçekleştiren araştırmacıların kategorileri yalnızca mevcut olan bir teori veya model üzerine kurmaları ve kodlama süreci sonunda kategorizasyon dışında kalan içeriklere yönelik nasıl bir tavır takınılacağıdır. Böyle bir durumda yeni kategorilerin tümevarımcı bir perspektif benimsenerek geliştirilmesi gerekebilir (Elo & Kyngäs, 2008; Armat vd., 2018)

Tümdengelimci içerik analizi perspektifi ve nicel içerik analizi yaklaşımı örtüşen ve birbirini destekleyen yönleri sahiptir. McKibben ve diğerleri (2020:2), farklı araştırmacıların görüşlerini temel alarak, nicel içerik analizini tümdengelimci içerik analizi ile eşdeğer tutmakta ve bu yaklaşımın *a priori* doğasına işaret ederek araştırmacının mesajlardaki bilgiyi kodlamak üzere parametreleri belirlediği ve sonrasında içeriği belirli bir kodlama prosedürü çerçevesinde kategorize ettiğinin altını çizmektedir. Araştırmacılar parametrelerin mevcut teorik ve kavramsal çerçeve ve geçmiş araştırmalar üzerine inşa edildiğini belirtmektedir. Bu noktadan hareketle tümdengelimci perspektifte hazırlanan bir içerik analizi sürecinin ilgili alanda geniş bir literatür taraması ve ön okumayı içermesi elzemdir. Kategorilerin bir noktada önceden belirlendiği ve içeriğin çözümleme sürecinde değişkenlere dönüştüğü bir yapıda kategorileri belirlerken alandaki uyumlu örüntüleri göz önünde bulundurmamak kategorilerin veri setini kapsayan ve aynı zamanda birbirlerini dışarıda bırakan, kategoriler arası kesişimin minimize edildiği bir matrisin ortaya çıkmasını olanaklı kılmaktadır. Bu durum metnin devamında daha detaylı bir şekilde değinilecek olan metin madenciliği algoritmalarında eğitim seti oluşturulması hususuyla da yakından ilgilidir.

Tümdengelimci içerik analizi perspektifi iletişim çalışmaları çerçevesinde düşünüldüğünde, içerik analizinin sistematik ve nesnel olma hali hatırlanmalıdır. Tümdengelimci bir perspektif veri setini sistematik bir biçimde kategorize etmek üzere teorik bir zemin sunmaktadır. Teorik ve kavramsal zeminin mevcudiyeti bir noktada araştırmacının sistematik bir şekilde içerik çözümlemesini gerçekleştirmesine olanak tanıyan düşünsel bir araç haline gelmektedir. Bu düşünsel zemin ve sistematiklik ise, içerik analizi sürecinin ikinci önemli bileşeni olan objektifliğin daha etkin bir biçimde elde edilmesini sağlamaktadır. İçerik analizinin niceliksel olma hali ise metnin teoriden yola çıkılarak analiz edilmesiyle ortaya çıkmaktadır. Bu bağlamda metin çözümlenerek frekanslara ulaşılabileceği gibi raporlama aşamasında kategorilerin dağılımı veya kategoriler arasındaki ilişkinin sayısal bir şekilde tespitiyle de içerik analizinin niceliksel olma şartı yerine getirilebilir. Bu durum bir anlamda içeriği sayısal bir değişken haline dönüştürme ile de ilgilidir.

Piñeiro Naval (2020) tarafından betimlenen 10 aşamalı içerik analizi süreci göz önünde bulundurularak tümdengelimci içerik analizi sürecini açıklamak, metnin devam eden başlıklarında metin madenciliğini iletişim çalışmalarının benimsediği içerik analizi anlayışına entegre etmek açısından önem taşımaktadır. İlk iki aşama göz önünde bulundurulduğunda araştırma amaçlarının belirlenmesi gerekliliği ve araştırma sorusu/hipotezlerin oluşturulması gerekliliğinin altı çizilmektedir. Bekiroğlu ve Balcı'nın fikirlerine değinen Çilingir (2017:150), araştırma soruları oluşturmanın amaçsız uygulamaların önüne geçmekte önem arz ettiğine ve bu bağlamda literatür taraması gerçekleştirilmesi gerektiğine işaret etmektedir. Amaç belirleme ve araştırma sorusu/hipotez oluşturulması tümdengelimci bir yaklaşımı benimsemiş olan içerik analizini benimseyen çalışmalarda sürecin bütününe etkileyen aşamalarıdır. Araştırma amacı belirleme sürecinde, tümdengelimci içerik analizi yaklaşımını benimsemiş çalışmalarda iki farklı şekilde gerçekleştirilebilir:

1) Literatür taraması gerçekleştirilerek, alanda tespit edilen boşlukları gidermeye yönelik hedeflerin geliştirilmesi

2) Yeni gelişmiş veya gelişmekte olan güncel olay ve olguları anlamak üzere hedeflerin geliştirilmesi.

Bu noktadan hareketle ortaya konulan araştırma amacı, araştırma alanına spesifik bir bağlamda katkı sağlayan ve araştırma soruları veya hipotezler geliştirmek için yeterli zemini sunan bir özellik taşımalıdır. Araştırma soruları veya hipotezleri geliştirilirken, çatı araştırma amacını gerçekleştirmeye hizmet eden, alt aşamalar şeklinde planlanmalıdır. Tümdengelimci içerik analizi sürecinde geliştirilen araştırma soruları betimsel özellikler taşıyabilir. Bu bağlamda soruları cevaplarken veri görselleştirme ve bulguları sayısal olarak ifade etme önem kazanmaktadır. Hipotez geliştirme süreçlerinde ise literatürün titiz bir şekilde taranmasıyla en az iki değişkenin belirlenmesi ve içerik analizi süreci sonunda çözümlenen içeriğin bu değişkenlere dönüştürülmesi önem kazanmaktadır. Örneğin Kuş ve Doğu Öztürk (2021), dijital kullanıcı kaynaklı içerikte nefret söylemi ve yanlış bilgi arasındaki ilişkinin mevcudiyetini tümdengelimci bir yaklaşımla test etmiş; bu bağlamda içerik çözümlenerek Nefret Söylemi Var/Yok, Yanlış Bilgi Var/Yok şeklinde değişkenlere dönüştürülmüş, son aşamada iki değişken arasındaki ilişki test edilmiştir. Bu noktadan hareketle tümdengelimci içerik analizi nicel bir çıktıyı amaçlıyorsa, içerik analizi süreci sonunda betimsel araştırma sorularının sayısal bir şekilde raporlanması ve tartışılması; geliştirilen hipotezlerin ise test edilerek bu sonuçların raporlanması ve tartışılması önem arz etmektedir.

Araştırma amacı ve araştırma soruları veya hipotezleri belirlendikten sonra araştırmanın evrenini ve örneklemini belirlemek gerekmektedir. Bu çalışma çerçevesinde olasılıklı örnekleme ve olasılıklı olmayan örnekleme yöntemlerini detaylı bir şekilde tartışmaktan ziyade içeriğin kullanıcılar tarafından üretildiği durumlarda benimsenmesi gereken anlayış metin madenciliğinin içerik analizine entegrasyonu sürecinde daha fazla önem taşımaktadır. Çünkü çağdaş iletişim çalışmaları, kullanıcı kaynaklı içeriğin analizine yoğun bir şekilde odaklanmaktadır. Diğer bir husus ise çağdaş iletişim çalışmalarında, gelişmiş veri toplama teknikleriyle evrenin bütününe yönelik veri toplama teşebbüslerinde bulunulmasıdır. Bu husus örneklemin temsil gücü bakımından tartışılıyor olsa dahi toplanan veri hacminin büyüklüğü de güçlü bir potansiyel sunmaktadır. Neuendorf (2002:83), Fink ve Gantz'ın fikirlerine değinerek, sosyal bilimler geleneğinde bir örnekleme dayanarak evrene yönelik genelleme yapmanın önemli olduğunu; bu bağlamda olasılıklı örneklemin (rastgele) ideal olduğunu fakat zaman ve maliyet gibi kısıtlar söz konusu olduğunda olasılıklı olmayan tekniklerin de kullanılabilmesinin altını çizmektedir. Riffe ve diğerleri (2019:71) örneklemin hedefinin evreni temsil etmek olduğunu; birimlerin rastgele seçildiği olasılıklı örneklemin kullanıldığı yaklaşımlarda evren hakkında geçerli çıkarımlar yapılabileceğini, bu tipteki örneklemin, örnekleme hataları içerebileceğini fakat araştırmacıların bunları istatistiksel prosedürlerle hesaplayabileceğine işaret etmektedir. Bu noktada örnekleme tablosu kavramına değinilmelidir. Neuendorf (2002) ile Riffe ve diğerleri (2019) evrenin her biriminin listelendiği bir örnekleme tablosu geliştirilmesi gerektiğine işaret etmektedirler. Neuendorf (2002:84), çözümlenecek içeriğin birey veya gruplarca üretildiği durumlarda önce birey veya grupların örnekleme, ardından bu birey ve gruplar tarafından üretilen içeriğin örnekleme gerektiğine işaret etmektedir. Lacy ve diğerleri (2015: 4) anahtar kelime aramalarının dahil olduğu içerik listesi oluşturma süreçlerinin büyük oranda belirlenen arama terimleriyle ilgili olduğuna işaret etmektedir. Bu ifadenin, çalışmanın devamındaki tartışma dijital veri toplama tekniklerine de değiniyor olacağından önem taşıdığına altı çizilmelidir.

Piñeiro Naval (2020) tarafından ortaya konulan akış değerlendirildiğinde evren ve örneklem belirleme süreçlerini, veriyle etkileşime geçilen ve bu sürece hazırlık yapılan aşamaların takip ettiğini belirtmek mümkündür. Bu bağlamda kod kitabının ve kodlama formunun tasarımı ve geçerliği, kodlayıcıların eğitimi ve örneklemin tasarlanan araçlar kullanılarak kodlanması aşamaları gerçekleştirilmektedir. Bu çalışma çerçevesinde tümdengelimci içerik analizi göz önünde bulundurulduğundan dolayı kod kitabı ve kodlama formunun öncül araştırmalar ve literatür ışığında oluşturulması önem arz etmektedir. Bu aşamada araştırmacı, literatürü yeterli ölçüde tarayıp taramadığını gözden geçirmeli, kod kitabının içeriği çözümlenmeye ve araştırma sorularını cevaplamaya hizmet edecek kategorileri oluşturmak üzere literatürde örtüşen örüntüleri etkin bir şekilde tespit edip etmediğini sorgulamalıdır. Bilgisayar destekli içerik analizinin web kaynaklı verilerde kullanımı için bir model öneren Sjøvaag ve Stavelin (2012:224), araştırma soruları hazırlandığında pilot bir test gerçekleştirilmesi ve şemanın araştırma sorularını cevaplamak uygunluğunun anlaşılmasını önermektedir. Nitel bir perspektife sahip olan Fonteyn ve diğerleri (2008:167) ise araştırmacıların içeriği bağımsız olarak kodlayıp

çapraz doğrulama için toplantı yaptığını belirtmektedir. Bu perspektif temelinde oluşturulan kodlama cetvelinin geçerliğini sağlamak üzere örnekleme oluşturan veri setinden rastgele alınan verinin en az iki farklı araştırmacı tarafından değerlendirilmesini ve kodlama cetvelinin veriye yönelik kapsayıcılığı ve işlevinin test edilmesini önermek mümkündür. Kodlayıcıları eğitmek nitel içerik analizinin kullandığı yaklaşımlardandır. Bu yaklaşımda yorumsayıcılık yüksek olduğundan birden fazla kodlayıcı bulunuyorsa bu kodlayıcıların perspektif bakımından yakın bir düzleme getirilmesi içerik çözümleme sürecinin etkililiği bakımından önem taşımaktadır. Nicel içerik analizi sürecinde kodlayıcıları eğitirken, nitel içerik analizinden kimi yaklaşımları aktarmak mümkündür. Bu bağlamda Goodell ve diğerleri (2016:83) tarafından ortaya konulan beş aşamalı kodlayıcı eğitim süreci hatırlanmalıdır: 1) etik eğitim, 2) temel nitel araştırma yöntemlerinin gözden geçirilmesi, 3) kodlama kitabının gözden geçirilmesi, 4) takım kodlaması ve 5) grup tartışmalarını içeren bağımsız kodlama süreci. Benzer bir süreci tümdengelimci içerik analizi için de uygulama zeminine sahiptir. Goodell ve diğerlerinin (2016) belirttiği üzere kodlayıcılar araştırma gerçekleştirdikleri spesifik alanın gerektirdiği etik bilince sahip olmalıdır. Tümdengelimci bir araştırma gerçekleştirilirken, araştırma yönteminin aşamaları hakkında bilgi her kodlayıcıya açık bir şekilde aktarılmalı; ayrıca kodlama kitabının oluşturulmasında kullanılan teorik zemin hakkında bilgi verilmelidir. Bu süreç grup toplantıları ve okuma listeleriyle gerçekleştirilebilir. Son aşamada ise kodlamanın gerçekleştirilmesi önem arz etmektedir. Bu süreç benimsenen kod kitabı ve eğitim sürecindeki toplantılarda benimsenen yaklaşıma bağlı kalarak içeriği çözümleme sürecini kapsar. Bu noktada teoriden gelen kategorizasyonu hatırlamak, yapılandırılmış/yarı yapılandırılmış kodlama matrisi kullanmak, kodlayıcı eğitimi sonunda ortaya çıkan kurallar bütününe sadık kalmak, kararsızlık anında teoriyi tekrar gözden geçirmek ve düzenli toplantılar gerçekleştirmek önem arz etmektedir. Bireysel kodlama süreçlerinin sonunda gerçekleştirilen periyodik grup toplantıları, kategori dışında kalan bileşenlerin akıbetini değerlendirmek için de fırsatlar sunmaktadır.

İçerik analizi süreci mesajların özelliklerini nesnel bir şekilde belirlemeyi hedeflediğinden, güvenilirlik büyük önem arz etmektedir (Neuendorf, 2002:141). Bu bağlamda iletişim çalışmalarında kodlayıcılar arası hemfikir olmayı ölçmeye yönelik testlerin gerçekleştirildiği gözlemlenmektedir. Krippendorff's alpha, Cohen's kappa ve salt mutabakat kodlayıcılar arası hemfikir olmayı ölçmeye olanak tanıyan kimi yaklaşımlardır. Bunu gerçekleştirmek üzere kodlanan veriden bir örnek seçilmeli, minimum iki kodlayıcı sürece dahil olmalı ve birbirlerinin kodlamış olduğu içeriği kör bir şekilde tekrar kodlamalıdır. Potter ve Levine-Donnerstein, güvenilirlik ölçmek üzere kodlanan tüm veriden çekilecek örneğin %10 ve %100 arasında olabileceğine işaret etmektedir (akt. Krippendorff, 2004). Bu noktadan hareketle, kodlanan verinin hacmine göre, veriyi temsil eden rastgele bir örneğin ya da kodlanan verinin tümünün kodlayıcılar arası hemfikir olmayı anlamak üzere değerlendirmeye tabi tutulması gerekmektedir. Her ne kadar nitel yaklaşıma yönelik bir tespit olsa da tümdengelimci içerik analizi sürecine uygulanabilecek bir önerme de Vaismoradi ve diğerleri tarafından gerçekleştirilmektedir; araştırmacılar tümdengelimci veri analizlerinde sürece birden fazla kodlayıcı dahil oluyorsa kodlayıcılar arası güvenilirliğin yüksek olması gerektiğini belirtmektedir (akt. Elo vd., 2014:5). Bu durumun, tümdengelimci içerik analizi sürecinin geçmiş teori ve çalışmalar üzerine inşa edilerek yeni bir bağlam hakkında içgörü elde etme çabasından kaynaklandığını belirtmek mümkündür. Yüksek kodlayıcılar arası hemfikir olma ölçüm şekline göre farklılıklar göstermektedir: Kodlayıcılar arası hemfikir olma Cohen's kappa değeri ile ölçülüyorsa .81 değerinin; salt mutabakat kullanılıyor ise %90'lık bir anlaşmanın yüksek hemfikir olma düzeyine işaret ettiği belirtilmektedir (Graham vd., 2012). Krippendorff (2004:145), iyi uygulanmamış bir kodlayıcı şemasının, kodlayıcıların eğitiminin yetersiz olmasının, kodlayıcı yorgunluğunun, grupta yıkıcı davranışlarda bulunan bir kodlayıcının bulunmasının güvenilirliğe yönelik tehdit oluşturduğundan bahsetmektedir. Kodlayıcılar arası hemfikir olma tatmin edici düzeydeyse, kodlanmış olan veriler içerik analizi sürecinin devamında gerçekleştirilecek testlerin değişkenlerinin oluşturulması için kullanılabilir. Bu durumda gerekli testlerin gerçekleştirilmesi ve ardından raporlanması gerekmektedir. Kodlama sürecinin sonucunda ortaya farklı veri tipleri çıkabileceğinden, veri tipleri göz önünde bulundurularak değişkenler arasındaki ilişkinin test edilmesi elzemdir.

Tümdengelimci içerik analizinin temellerini oluşturan bu süreç, iletişim çalışmalarında içerik analizine metin madenciliğini entegre etmek üzere düşünsel bir çerçeve sunmaktadır. Bu bağlamda, devam eden başlıkta, bilgisayar destekli içerik analizi hakkındaki çalışmalar ve araştırmacının deneyimleri göz önünde bulundurularak iletişim çalışmalarında içerik analizi gerçekleştirirken metin madenciliğinin nasıl kullanılacağına yönelik bir yol haritası çizilmiştir.

3. TÜMDENGELİMCİ İÇERİK ANALİZİ VE METİN MADENCİLİĞİNİ BİRLİKTE DÜŞÜNMEK

Büyük veri konseptinin getirdiği yeni teknolojiler veri toplamak için önemli fırsatlar sunmaktadır. Yeni medya platformlarının her geçen gün daha fazla kullanıcının ilgisini çekiyor olması, toplumsal konular hakkında bir tartışma alanı haline dönüşmesi ve gündelik meselelerin çözümü noktasında araçsallaşması, bu platformları iletişim araştırmacıları için bir veri kaynağı haline getirmektedir. Diğer yandan, büyük veri konseptinin getirdiği yeni yaklaşımlar buradan veri toplamayı mümkün kılan araçların geliştirilmesine, toplanan verinin yenilikçi ve disiplinlerarası süreçlerle çözümlenmesine olanak tanımaktadır. İletişim çalışmalarının başvurmuş olduğu yöntemlerde bir perspektif dönüşümüne sebep olan bu yaklaşımlar, daha önce elde edilmesi mümkün olmayan içgörülerin ve fark edilmeyen örüntülerin araştırmacılar tarafından saptanmasını mümkün kılmaktadır. İçerik analizine metin madenciliğinin entegrasyonunu gerçekleştirmek üzere bir süreç öneren bu çalışma, yukarıda tartışılan tümdengelimci içerik analizi süreçlerini göz önünde bulundurularak her aşamada nasıl bir yaklaşım benimsenmesi gerektiğine yönelik önerilerde bulunmaktadır.

Öncelikle, bu çalışma çerçevesinde benimsenen metin madenciliği yaklaşımından bahsetmek gerekmektedir. Bu bağlamda Waegel (2006:1) tarafından önerilen yaklaşım, metin madenciliğinin veri madenciliğinden farkını ortaya koymak, işlevlerini ve üstesinden gelebildiği zorlukları anlamak noktasında dikkate değerdir:

Metin madenciliği, '(büyük miktarda) metinsel veriden örtük, önceden bilinmeyen ve potansiyel olarak faydalı bilgilerin basit olmayan bir şekilde elde edilmesi' olarak tanımlanır. Önceden kategorize edilmiş verileri analiz etmek yerine, düzensiz, biçimlendirilmemiş ve genellikle parçalanmış metin koleksiyonlarından bilgi araması bakımından veri madenciliğinden farklılık göstermektedir. Metin madenciliği, yalnızca standart veri madenciliği, makine öğrenimi ve istatistikten değil, aynı zamanda hesaplamalı dilbilim ve doğal dil işlemeden de yararlanan disiplinler arası bir alandır.

Bu çerçevede düşünüldüğünde metin madenciliği, bulgular elde etmek üzere sıklıkla dijital metinlere başvuran ve bu süreçte içerik analizinden faydalanan iletişim çalışmaları için önemli bir potansiyel barındırmaktadır. Dijital metinler yapılandırılmamış bir formata sahip olduğundan kümeleme algoritmaları ile durum tespitine yönelik betimsel girişimler ve sınıflandırma algoritmalarıyla öngörücü bir yaklaşımı benimseyen çözümleme süreçleri, büyük miktarda metin verisinin mevcut olduğu durumlarda farklı yaklaşımları benimsemiş içerik analizi süreçlerini gerçekleştirmek için çözümler sunabilmektedir. Bu çalışmada sınıflandırma algoritmalarını kullanan, dolayısıyla güdümlü ve öngörücü olarak tanımlanabilecek metin madenciliği algoritmalarının içerik analizini gerçekleştirmek üzere kullanıldığı bir yaklaşım benimsenmiştir. Bu yaklaşım araştırmacılar tarafından manuel olarak kategorize edilmiş metinleri temel alarak, kategorize edilmemiş verilerin kategorilerini tahmin etmeye olanak tanımaktadır.

Metin madenciliği, tümdengelimci içerik analizi bağlamında düşünüldüğünde büyük veri kavramının sunmuş olduğu düşünsel zemin önemli derecede yol göstericidir. Mayer-Schönberger ve Cukier (2013) büyük veri kavramı hakkında düşünsel bir düzlem oluştururken "*kesinliğe takılıp kalmaktansa yönelimlere odaklanmak*" bileşeninden bahsetmektedir. Bu yönüyle büyük miktarda metin formatındaki veriyi analiz ederek, bu verideki yönelim ve gizli örüntüleri tespit etmenin bilgi düzleminde yaratabileceği katma değer yüksektir. İletişim çalışmalarının sıklıkla dijital platformlardan toplanan metin formatındaki verilere içerik analizi uyguladığı ve bu veriler toplanırken API'lardan faydalandığı gözlemlenmektedir. Lomborg ve Bechmann (2012:256) bir API'yı "Üçüncü parti geliştiricilerin mevcut bir servise eklemeler gerçekleştirmesini mümkün kılan bir back-end arayüzüdür. API aynı zamanda araştırmacıların bir sosyal medya servisinden ampirik analizler için veri topladığı bir arayüzdür." şeklinde tanımlamaktadır. Bu bağlamda API'nın metin madenciliğinin gerçekleştirileceği büyük miktardaki metin verisini toplamak üzere mühim bir rolü olduğu, dijital bir platform ve araştırmacının kullandığı veya geliştirdiği veri toplama programı ile etkileşime geçen bir geçit görevi gördüğü ve dijital platformdan verinin ne kadarının toplanabileceği konusunda bir karar mekanizması gibi çalıştığını belirtmek mümkündür. Sosyal bilimler alanında araştırmalar gerçekleştiren araştırmacılar kendi yazmış oldukları programları kullandığı gibi, açık kaynak/açık erişim şeklinde paylaşılan ve yazılım bilgisi gerektirmeyen programları da veri toplamak üzere kullanılmaktadır. Bu bağlamda sosyal medya platformlarından veri toplamak üzere Facepager (Jünger & Keyling, 2019: github.com/strohne/Facepager), Mozdeh Big Data Text Analysis (2020: mozdeh.wlv.ac.uk) veya R Studio'nun sahip olduğu kütüphanelerden (Klinkhammer, 2020: <https://arxiv.org/abs/2011.13327>) faydalanılabildiği gözlemlenmektedir.

Tümdengelimci içerik analizinin sunmuş olduğu düşünsel temel ve yukarıda tartışılan 10 aşamalı içerik analizi süreci göz önünde bulundurulduğunda araştırma amaçlarını ve hipotezlerini belirlerken metin madenciliğinin entegre edildiği bir içerik analizi süreci ve geleneksel bir içerik analizi süreci arasında fark bulunmadığını belirtmek mümkündür. Bu sebeple içerik analizine metin madenciliğinin entegrasyonu tartışmasına evren ve örnekleme belirlemeye yönelik aşamalara odaklanarak başlanmıştır. Bu bağlamda, süregelen örnekleme sürecine yönelik zorluklara değinmekte fayda bulunmaktadır. Dijital platformlardan API'larla etkileşime geçen araçlarla veri toplanırken, evrendeki verinin önemli bir miktarına ulaşıyor olsa dahi evrendeki verinin boyutu ve elde edilen verinin temsil gücü hakkında sınırlı miktarda bilgi sahibi olunması örneklemin temsil gücü hakkındaki tartışmaları beraberinde getirmektedir. Geleneksel içerik analizi yaklaşımları örnekleme tablosu kullanmakta ve burada evrende ulaşılabilen her bileşene yer vermektedir. Fakat dijital platformlardan toplarken evrendeki bileşen sayısına yönelik üç zorluk söz konusudur. Bunlar 1) veri toplama sürecinin API ve bununla bağlantıya geçen programlarla aracılanması sonucu opak bir sürecin mevcudiyeti, 2) evrendeki verinin potansiyel olarak hacmen büyük olması ve 3) evrendeki her bileşenin örnekleme tablosu oluşturabilmek adına kayıt altına alınamamasıdır. Kim ve diğerleri (2018:2) Twitter'dan toplanan veriden nasıl örneklem alınabileceğini tartışan çalışmasında kitle medyasında üretilen içeriğin alan ve organizasyonel kaynaklar bakımından sınırlı olduğunu; fakat sosyal medyanın esnek bir doğası bulunduğunu, 24 saat neredeyse sınırsız içerik sağlayıcı ile işlediğini belirtmektedir. Araştırmacılar, sosyal medyadaki verinin günlük/haftalık bir döngüsü olmadığını ve konuya göre değişkenlik gösterdiğini bu sebeple örneklemin temsil gücü yüksek günleri seçmek üzerinden yürütülmesinin de kırılğanlıklar yapabileceğine işaret etmektedir.

İletişim araştırmacılarının belirli konulara dair oluşan dijital gündemleri araştırdığı bilinmektedir ve bu bağlamda kullanıcı kaynaklı içerik yoğun bir şekilde kullanılmaktadır. Bu bağlamda geleneksel içerik analizinde Neuendorf (2002) ve Lacy ve diğerleri (2015) tarafından sunulan önermeler dikkate alınmalıdır. Neuendorf, çözümlenecek içeriğin birey veya gruplarca üretildiği durumlarda önce birey veya grupların örnekleme, ardından bu birey ve gruplar tarafından üretilen içeriğin örnekleme gerektirdiğini; Lacy ve diğerleri ise anahtar kelime aramalarının dahil olduğu içerik listesi oluşturma süreçlerinin büyük oranda belirlenen arama terimleriyle ilgili olduğunu belirtmektedir. Ayrıca araştırmacının geçmiş deneyimleri sosyal medyadan içerik toplarken iki aşamalı bir sürecin dikkate değer sonuçlar verdiğine işaret etmektedir. Bu noktadan hareketle birinci aşamada bir dijital platformda gündemlerin bir araya gelmesini mümkün kılan fonksiyonların örnekleme tablosu oluşturmak üzere bir araya getirilmesi gerekmektedir. Gündemleri kategorize etmek üzere Twitter'da hashtag'lerin kullanıldığı bilinmektedir, Reddit platformunda bu işlevi sub-reddit'ler görmektedir, YouTube'da ise spesifik bir kelimedeki arama gerçekleştirildikten sonra izlenme sayısı veya yüklenme sayısı gibi metriklerle göre listelenen videolar bu işlevi üstlenebilir. Gündemi kategorize etmek üzere kullanılan fonksiyonların tümünü içeren bir örnekleme cetveli oluşturularak evrene yönelik bir fikir edinilmesi mümkündür. Ardından buradan temsil gücü yüksek bileşenlerin seçilerek kullanıcı kaynaklı içeriğin toplanması gerekmektedir. Bunu Twitter'dan veri toplanan bir örnek üzerinden anlatmak süreci kristalleştirecektir. X gündemi hakkında bir araştırma yürütmek isteyen araştırmacı, o gündemle ilgili erişilebilen tüm hashtagleri örnekleme cetveline dahil etmelidir. Bu süreç #X şeklinde bir pilot veri toplama işleminden sonra elde edilen tweetlerdeki diğer hashtagleri tespit ederek gerçekleştirilebilir. Büyük veri perspektifinin benimsendiği bir bağlamda örnekleme cetveline dahil edilen hashtaglerin tümünden, kullanıcılar tarafından gönderilen tweetleri toplamak mümkündür. Örnekleme cetveline dahil edilen bileşen sayısı ve toplanan tweet sayısı arttıkça, örneklemin evreni temsil gücünün artma potansiyeli bulunduğunu tartışmak mümkündür.

Evreni temsil ettiği düşünülen bir örneklem belirlendikten sonra kod kitabının ve kodlama formunun tasarımı ve geçerliliği ve kodlayıcıların eğitimi süreci işletilmelidir. Bu aşamalar aynı zamanda içerik analizinde kullanılacak olan ve içerik analizini metin madenciliğine entegre etmeye olanak tanıyan sınıflandırma algoritmasının temellerini de oluşturmaktadır. Kod kitabının ve kodlama formunun tasarımı süreci dört aşamalıdır: 1) Toplanan metin verisinin bütününden onu temsil edecek bir miktarın rastgele bir biçimde çekilmesi, 2) Tümdengelimci bir perspektifle kategorilerin oluşturulması, 3) Rastgele çekilen verinin kodlanması, 4) Kodlanan veride kodlayıcıların hemfikir olma oranının tespit edilmesi, 5) Kodlayıcılar arası hemfikir olma düzeyi tatmin edici düzeyde ise bu verinin, algoritmayı eğitmek için kullanılması. Bu sürecin ilk üç aşamasını etkin bir şekilde yerine getirebilmek için insan kodlayıcıların geleneksel içerik analizindeki kurallar bütününe uygun bir şekilde eğitilmiş olması

önem taşımaktadır. Dördüncü ve beşinci aşama ise makine-kodlayıcı, yani içerik analizi gerçekleştirmek üzere oluşturulmuş olan algoritmanın etkinliği üzerinde belirleyici bir role sahiptir. Metin madenciliği sürecinin başarılı sonuçlar verebilmesi için eğitim seti olarak kullanılacak olan metinlerin çeşitliliği (eğitim setindeki metin çeşitliliği sonuçların daha etkin bir biçimde tahminine olanak tanıyabilir), belirlenen kategori sayısı (kategori sayısı azaldıkça tahmin gücü yükselebilir) gibi değişkenler önem taşımaktadır. Bu bağlamda incelenen gündemin, veri toplanılan dijital platformun ve içerik üreticisi ekosisteminin dahi rolü olduğunu tartışmak mümkündür.

Eğitim seti geliştirdikten sonra içerik analizi sürecine entegre edilecek sınıflandırma algoritmasının tasarlanması gerekmektedir. Bu çalışma kapsamında bu aşamada güdümlü bir derin öğrenme sürecinden bahsetmek mümkündür. Güdümlü derin öğrenme süreçlerinde eğitim seti el ile kodlanmalıdır¹ ve bir sınıflandırıcı oluşturarak çözümlenecek olan veri seti eğitim seti temel alınarak sınıflandırılmaktadır (Yuxin, 2021). Bu aşama geleneksel içerik analizindeki örneklemin tasarlanan araçlar kullanılarak kodlanması aşamasının metin madenciliği işlemindeki karşılığı olarak betimlenebilir. Bu aşama 1) eğitim setinin yüklenmesi ve ön işlemden geçirilmesi, 2) Analiz edilecek metinlerin yüklenmesi ve ön işlemden geçirilmesi, 3) Metinlere kategorileri atamak için kullanılacak sınıflandırma algoritmasının seçilmesi olmak üzere üç aşamadan oluşmaktadır. Bu süreçte eğitim setindeki metinlerin içerdiği kelimelerin ve seçilmiş olan örneklemden metinlerin içerdiği kelimelerle örtüşmesi etkinlik bakımından önemlidir. Bu bağlamda eğitim seti ve çözümlenecek olan veri seti bir dizi ön işlemden geçirilmektedir. Kapucugil İkiz ve Özdağoğlu (2015), kümeleme algoritmaları bağlamında bir ön işlem sürecine değinmiş, bu ön işlem sürecinde cümleleri kelimelere bölme, dolgu kelimelerini filtreleme, belirli harften daha az harf içeren kelimelerin filtrelenmesi, kelime eklerinin yuvarlanması, n-gram'ların oluşturulması, term-vector'lerin oluşturulması ve pruning aşamalarından bahsetmişlerdir. Benzer aşamaları Welbers ve diğerleri (2017) tarafından önerilen ön işlem sürecinde de gözlemlemek mümkündür. Bu aşamaların bir kısmının metin madenciliğini tümdengelimci içerik analizi çerçevesinde kullanılacak olan süreçte kullanmak da mümkündür. Bu çalışma bağlamında ön işlem süreci, İngilizce metin formatındaki veri setleri göz önünde bulundurulduğunda, metinlerin hepsinin büyük/küçük harfe dönüştürülmesi (transforming cases, her cümlenin kelimelere bölünmesi (tokenization), belirli harften daha az harf içeren kelimelerin filtrelenmesi (filter tokens), dolgu kelimelerin filtrelenmesi (filtering stopwords) ve kelime eklerinin yuvarlanması (stemming) aşamalarından oluşmaktadır. Bu süreçler hem sınıflandırma sürecinde işlevi olmayan kelimelerin çözümlenme dışında bırakılması hem de eğitim setindeki metinler ve çözümlenecek olan metinlerin içermiş olduğu kelimelerin karışıklık olarak uyumlanması noktasında önem arz etmektedir. Bu durum eğitim seti temel alınarak çözümlenecek metinlerin kategorilerinin tahmin edilmesi için önemlidir. Bu sürecin ardından kullanılacak sınıflandırma algoritmasına karar verilmelidir. Sınıflandırma algoritmaları tercih edilirken eğitim setinin hacmi belirleyici bir rol oynamaktadır. Support Vector Machine ve k-Nearest Neighbour algoritmalarının sınıflandırma süreçlerinde sıklıkla kullanıldığına rastlanmaktadır. Eğitim seti analiz edilecek veriden büyükse k-NN, SVM'e göre daha iyi performans sergilemektedir, fakat daha fazla analiz edilecek veri ve az eğitim verisi söz konusuysa SVM, k-NN'den daha iyi performansa sahiptir (Varghese, 2018: <https://bit.ly/3DFO3YA>). Bu bağlamda, araştırmacı uzun süredir belirli bir konu üzerinde otomatize yöntemlerle içerik analizi gerçekleştiriyor ve bu nedenle elinde diğer çalışmalarına da uygulayabileceği geniş bir eğitim seti bulunuyorsa k-NN, eğitim seti incelenecek verinin bir kısmı rastgele seçilerek oluşturuluyorsa SVM kullanması söz konusu olabilir.

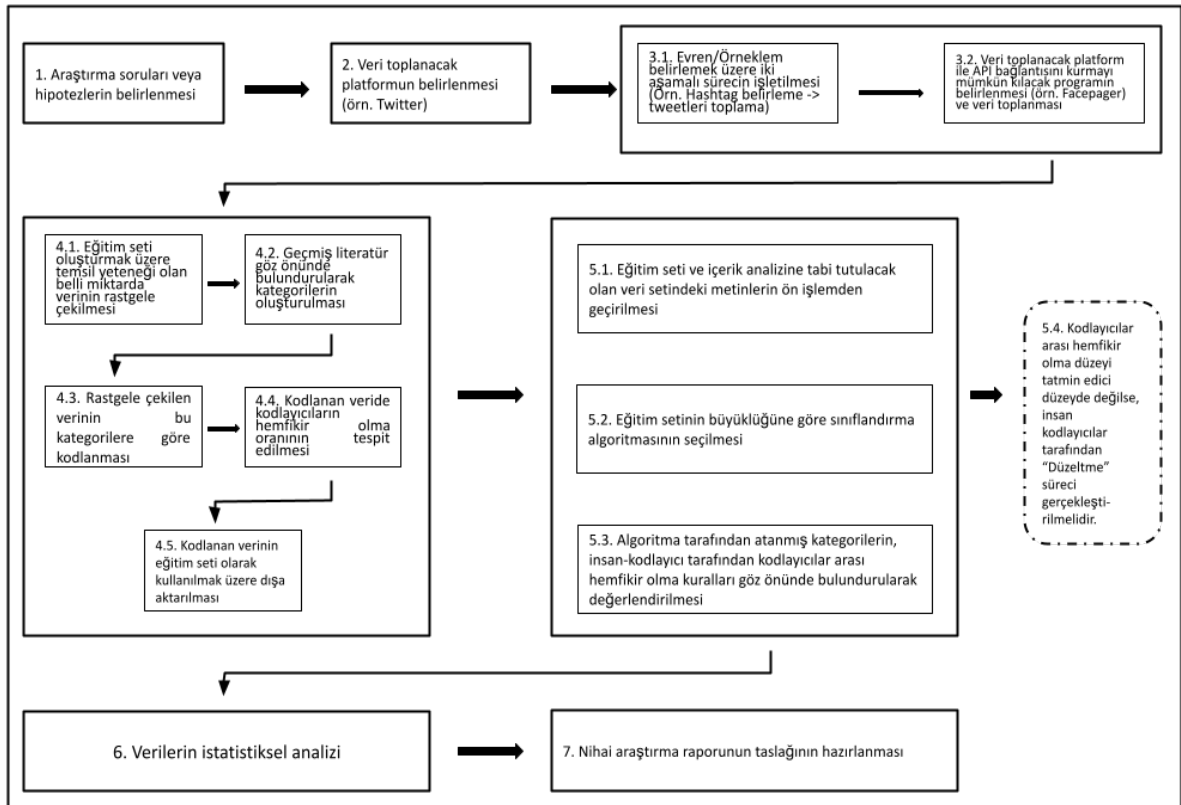
Metin madenciliği sürecinin sonucunda analizin nesnellüğünün değerlendirilmesi ve kodlayıcılar arası güvenilirlik konularına odaklanması elzemdir. Bu bağlamda algoritma bir kodlayıcı olarak kabul edilmelidir. Algoritmanın otomatik bir şekilde kategorize ettiği içeriklerden geleneksel içerik analizindeki kurallar göz önünde bulundurularak rastgele metin çekilmeli ve sorumlu araştırmacı tarafından algoritmanın atamış olduğu değerler kodlayıcılar arası hemfikir olma kuralları göz önünde bulundurularak tekrar değerlendirilmelidir. Eğer makine kategorizasyonu ve araştırmacı kategorizasyonu arasında tatmin edici düzeyde bir tutarlılık söz konusuysa değişkenler arasındaki ilişkiyi tespit etmeye yönelik testlere ve raporlama aşamasına devam edilebilir. Fakat kodlayıcılar arası hemfikir olma düzeyi tatmin edici değilse, bu durumda araştırma yarı-otomatik içerik analizi yaklaşımı göz önünde bulundurularak, insan kodlayıcıların aktif katılımı ile devam edebilir. Yarı-otomatik içerik analizi sürecinde Yan ve diğerleri (2014) tarafından ortaya konulmuş olan perspektif benimsenerek, insan kodlaması ve makine kodlaması aşamalarından sonra sürece düzeltme aşaması dahil edilebilir. Böylece sınıflandırma algoritmasının kategorizasyonda gerçekleştirebileceği hatalar, insan kodlayıcıların müdahalesiyle

1 Bir önceki aşamada bu süreç gerçekleştirilmiştir.

düzeltilir. Kodlayıcılar arası güvenilirlik süreci ardında geleneksel içerik analizine benzer şekilde testlerin gerçekleştirilmesi ve sonuçların raporlanması söz konusudur. Bu noktada raporlama aşamasında yenilikçi veri görselleştirme yaklaşımlarının veya sosyal ağ analizi gibi tekniklerin verileri raporlarken kullanılabilmesi belirtilebilir.

SONUÇ VE DEĞERLENDİRME

Çalışmada önerilen metin madenciliği temelli tümdengelimci içerik analizi süreci farklı bağlamlarda, farklı programlarla, teknik olarak yüksek donanım gereksinimi gözetmeden dijital metin verisine uygulanabilecek yöntemsel bir yol haritası sunmaktadır. Çalışma çerçevesinde gerçekleştirilen tartışma, büyük veri konseptinin sunduğu teknikler, araçlar ve yaklaşımların içerik analizi sürecine katkı sunabileceğine de işaret etmektedir. Metin madenciliğinin içerik analizine sınıflandırma algoritmaları aracılığı ile entegrasyonu, dijital platformlardan toplanan büyük miktardaki veriden etkin içgörüler elde etmek noktasında fırsatlar sunmaktadır. Bu çalışma kapsamında, geleneksel içerik analizi perspektifinin sunmuş olduğu düşünsel zemin üzerine inşa edilmiş, makine-kodlayıcıları içeren, otomatize bir içerik analizi sürecinden bahsedilmiştir. Bu yönüyle çalışma, iletişim çalışmalarında özellikle sosyal medyadan toplanan verilerin içeriğinin etkin bir şekilde analiz edilebilmesi için bir pencere açmaktadır. Metin madenciliği sürecinin, tümdengelimci içerik analizi konseptine entegrasyonunu, şematize etmek bu çalışmanın gerçekleştirdiği tartışmanın sunduğu katma değerini somutlaştırmasını sağlayacaktır. Bu bağlamda, iletişim çalışmalarında metin madenciliği temelli tümdengelimci bir içerik analizi gerçekleştirirken aşağıdaki aşamaların takip edilmesi mümkündür.



Şekil 1. Kuş'un metin madenciliği temelli tümdengelimci içerik analizi süreci

Bu çalışma çerçevesinde ortaya konulan metin madenciliği temelli tümdengelimci içerik analizi modeli, geleneksel içerik analizinin ana süreçlerinin yanısıra metin madenciliği süreçleri ve bu tekniğin getirdiği alt aşamaları içermektedir (Şekil 1). Ayrıca ihtiyaç halinde Yan ve diğerleri (2014) tarafından geliştirilmiş olan yarı-otomatize içerik analizinin Düzeltilme fazından faydalanmaktadır. Çalışma çerçevesinde ortaya konulan, metin madenciliği temelli tümdengelimci içerik analizi modelinin aşamalarını takip eden şekilde özetlemek mümkündür.

1. Araştırma soruları veya hipotezlerin geliştirilmesi
2. Veri toplanacak platformun tespit edilmesi
3. İki aşamalı veri toplama süreci ve API bağlantısı
 - 3.1. Evren/örneklem belirleme
 - 3.2. API bağlantısını kurmak (Facepager veya Mozdeh Big Data Text Analysis gibi programlarla)
 4. Metin madenciliği sürecinde kullanılacak sınıflandırma algoritması için eğitim seti geliştirmek
 - 4.1. Temsil gücü yüksek belli miktarda verinin, veri setinin bütününden rastgele çekilmesi
 - 4.2. Geçmiş literatür ışığında kategorilerin oluşturulması
 - 4.3. Rastgele çekilen verinin kategorizasyonu
 - 4.4. Kodlanan veride kodlayıcılar arası hemfikir olma düzeyinin belirlenmesi
 - 4.5. Kodlanan verinin eğitim seti olarak kullanılmak üzere dışa aktarılması
 5. Sınıflandırma algoritmasının oluşturulması (Bu aşamada Rapidminer, Knime ve R Studio kullanılabilir)
 - 5.1. Eğitim seti ve çözümlenecek metinlerin ön işlemden geçirilmesi
 - 5.2. Sınıflandırma algoritmasının seçimi
 - 5.3. Algoritma kategorizasyonu ve insan-kodlayıcı arasındaki hemfikir olma düzeyinin hesaplanması
 - 5.4. Algoritma kategorizasyonu ve insan-kodlayıcı arasındaki hemfikir olma düzeyi yeterli olmadığı takdirde Düzeltme aşamasının uygulanması
6. Araştırma amacı doğrultusunda verinin istatistiksel analizi
7. Nihai araştırma raporunun hazırlanması

Bu çalışma temelde Piñeiro Naval (2020) içerik analizi aşamalarını ve Yan ve diğerleri (2014) tarafından ortaya konulan yarı-otomatize içerik analizi modelini göz önünde bulundurmaktadır. Geleneksel içerik analizi üzerine Neundorf, Krippendorf ve Riffe ve diğerleri tarafından gerçekleştirilen tartışmaların sunmuş olduğu düşünsel zemin ve süreç önerileri, bu çalışmanın kavramsal zeminini oluşturmaktadır. Çalışma, bu perspektifleri göz önünde bulundurarak metin madenciliğini geleneksel tımdengelimci içerik analizine entegre etmektedir. Çalışmayı sonlandırırken bu entegrasyon sürecine yönelik araştırmacı veya araştırma grubunun konumunu ilgilendiren kimi önemli konuları aydınlatmak önemlidir.,

Metin madenciliği ve geleneksel içerik analizinin entegrasyonuna yönelik bir süreç önerisinde bulunan bu çalışmada makine-kodlayıcılar, insan-kodlayıcılar yerine geçen bir bileşen olarak düşünülmemektedir. Ampofo ve diğerleri (2015: <https://bit.ly/3LGjHrM>), algoritmik metin analizlerine oldukça yoğun bir şekilde başvurulmasının insan-kodlayıcıların tespit etmekte daha donanımlı olduğu (örn. kimi terimlerin kullanıldığı iletişimsel bağlama göre farklılaşması) değerli içgörülerin gözden kaçmasına sebep olabileceğini not etmektedir. Bu noktadan hareketle, metin madenciliği temelli tımdengelimci içerik analizinde insan-kodlayıcılar, otomatik kategorizasyon sürecinin mihenk taşı olan eğitim setini oluşturup, eğitim setinin etkinliğini ve sınıflandırma algoritmalarının çıktılarını değerlendirecek olan karar vericiler olarak konumlanmalıdırlar. Eğitim seti kodlanırken insan-kodlayıcı ekibinin hemfikir olma düzeyi içerik analizinin objektifliği ve güvenilirliği noktasında fikir edinilmesini mümkün kılmaktadır. Birden fazla araştırmacının dahil olduğu ve kodlayıcılar arası hemfikir olma süreciyle test edilerek oluşturulan eğitim seti, sınıflandırma algoritmasının önyargılı olmasını engellemek üzere fırsatlar da sunmaktadır. Son aşamada ise insan-kodlayıcılar makine kodlayıcıyı, kodlayıcılar arası hemfikir olma düzeyi aracılığı ile denetlemektedir. Bu durum makine kodlayıcıların, yani metin madenciliği sürecinin, içerik analizi destek bileşeni şeklinde konumlanabileceğine işaret etmektedir.

Bir diğer husus ise bu çalışmanın özellikle iletişim alanındaki çalışmaları göz önünde bulundurarak metin madenciliği temelli bir içerik analizi süreci sunmasıdır. Bu yönüyle güncel iletişim çalışmaları göz önünde bulundurulmuş ve dijital platformlardan otomatize bir şekilde toplanan yüksek hacimli dijital metin verisine yönelik bir veri toplama süreci tasarlanmıştır. Fakat bu çalışmada sunulan modelin farklı platformlardan toplanan veya dijitalleştirilen analog metinler üzerinde de uygulanma potansiyeli söz konusudur. Diğer yandan, eğitim seti oluştururken veri setinin tümünden rastgele seçilen metinlerle ilgili altı çizilmesi gereken hususlardan bahsetmek önemlidir. Bu bağlamda rastgele çekilecek miktar üzerine spesifik bir bildirimde bulunulmamasının sebebi eğitim seti oluşturmak üzere rastgele çekilecek olan veri miktarının konu, veri miktarı, araştırma grubunun öncül çalışmaları ve kullanılacak olan algoritmaya göre değişme potansiyelidir. Bu noktadan hareketle bu aşama ana hatlarıyla tanımlanmıştır. Çalışmanın önermiş olduğu sürecin araştırma bağlamı, öncül literatür, yeni gelişmeler ve araştırma hedefleri doğrultusunda sahip olduğu aşamaların dönüştürülmesini ve yeni aşamaların eklenmesini gerektirecek ve bu yeni aşamaların entegrasyonuna imkan tanıyacak modüler bir yapısının olduğunu belirtmek mümkündür. Son olarak çalışmanın değindiği otomatize yaklaşımların İngilizce metinlerde daha etkin sonuçlar verdiğinin altı çizilmelidir. İngilizce dışındaki metinlerde farklı aşamaların sürece dahil olması veya insan-kodlayıcının üstlenmiş olduğu görevin daha fazla önem kazanması söz konusudur.

KAYNAKÇA

- Ampofo, L., Collister, S., O'Loughlin, B., Chadwick, A., Halfpenny, P. J., & Procter, P. J. (2015). Text mining and social media: When quantitative meets qualitative and software meets people. (?).10.2013). <https://bit.ly/3LGjHrM>
- Armat, M., Assaroudi, A., Rad, M., Sharifi, H., & Heydari, A. (2018). "Inductive and Deductive: Ambiguous Labels in Qualitative Content Analysis", *The Qualitative Report*, 23/1, 219-221.
- Atan, S. (2018). "Haberlerin Kurumsal İmaja Etkisi ve Türkiye'deki Hastaneler Hakkında Medyada Yer Alan Haberlerin Metin Madenciliği ile Analizi", *Journal of Communication Theory & Research/İletişim Kuram ve Araştırma Dergisi*, 46, 222-240.
- Aytekin, Ç., & Sütcü, C. S. (2013). "Sosyal Medya'da Demokrasi Algısı:"Demokrasinin Renkleri" Üzerine Fikir Madenciliği Araştırması", *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*, 19, 28-44.
- Çilingir, A. (2017). "İletişim Alanında İçerik Analizi Yöntemi Kullanılarak Yapılan Yüksek Lisans ve Doktora Tezleri Üzerine Bir İnceleme", *Erciyes İletişim Dergisi*, 5/1, 148-160.
- Drisko, J. & Maschi, T. (2016). *Content Analysis*, Oxford University Press, New York.
- Elo, S. & Kyngäs, H. (2008). "The qualitative content analysis process", *Journal of Advanced Nursing*, 62/1, 107-115.
- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). "Qualitative content analysis: A focus on trustworthiness", *SAGE open*, 4/1, 1-10.
- Fonteyn, M. E., Vettese, M., Lancaster, D. R., & Bauer-Wu, S. (2008). "Developing a codebook to guide content analysis of expressive writing transcripts", *Applied Nursing Research*, 21/3, 165-168.
- Goodell, L. S., Stage, V. C., & Cooke, N. K. (2016). "Practical qualitative research strategies: Training interviewers and coders", *Journal of Nutrition Education and Behavior*, 48/8, 578-585.
- Graneheim, U. H., Lindgren, B. M., & Lundman, B. (2017). "Methodological challenges in qualitative content analysis: A discussion paper", *Nurse education today*, 56, 29-34.
- Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*, Center for Education Compensation Reform (CECR), US.
- Güçdemir, Y., Çakar-Mengü, S., & Günay, K. (2020). "Büyük Veri Metin Madenciliği Bağlamında İstanbul Büyükşehir Belediyesi Seçimlerinde Aday Liderlerin Tweet Kullanımları Üzerine Bir İnceleme", *Connectist: Istanbul University Journal of Communication Sciences*, 59, 161-191.
- Helles, R., & Ørmen, J. (2020). "Big data and explanation: Reflections on the uses of big data in media and communication research", *European Journal of Communication*, 35/3, 290-300.
- Sallan Gül, S., & Kahya Nizam, Ö. (2021). "Sosyal bilimlerde içerik ve söylem analizi", *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 42, Özel Sayı 1, 181-198.

- Jünger, J. & Keyling, T. (2019). Facepager. (05.04.2020) <https://github.com/strohne/Facepager>
- Kapucugil İkiz, A., & Özdağoğlu, G. (2015). "Text mining as a supporting process for VoC clarification", *Alphanumeric Journal*, 3/1, 25-40.
- Kim, H., Jang, S. M., Kim, S. H., & Wan, A. (2018). "Evaluating sampling methods for content analysis of Twitter data", *Social Media+ Society*, 4/2, 1-10.
- Klinkhammer, D. (2020). Analysing Social Media Network Data with R: Semi-Automated Screening of Users, Comments and Communication Patterns. (26.11.2020) <https://arxiv.org/abs/2011.13327>.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*, Sage publications, California.
- Kuş, O. (2016). "Dijital Nefret Söylemini Anlamak: Suriyeli Mülteci Krizi Örnek Olayı Bağlamında BBC World Service Facebook Sayfasına Gelen Yorumların Metin Madenciliği Tekniği İle Analizi", *İstanbul Üniversitesi İletişim Fakültesi Dergisi*, 51, 97-121.
- Kuş, O., & Öztürk, İ. D. (2021). "Çin'de Üretilen Aşılar Yönelik YouTube Tartışmaları Çerçevesinde Yanlış Bilgi ve Nefret Söylemi İlişkisi Üzerine Bir İnceleme", *Türkiye İletişim Araştırmaları Dergisi*, 38, 96-117.
- Lacy, S., Watson, B. R., Riffe, D., & Lovejoy, J. (2015). "Issues and Best Practices in Content Analysis", *Journalism & Mass Communication Quarterly*, 92/4, 791-811.
- Lomborg, S., & Bechmann, A. (2014). "Using APIs for data collection on social media", *The Information Society*, 30/4, 256-265.
- Neuendorf, K. A. (2002). *The Content Analysis Guidebook*, Sage Publications, Thousand Oaks.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Büyük Veri: Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim*, Paloma, İstanbul.
- Mozdeh Big Data Text Analysis (2020). Mozdeh Big Data Text Analysis. (01.02.2021) mozdeh.wlv.ac.uk
- McKibben, W. B., Cade, R., Purgason, L. L., & Wahesh, E. (2020). "How to conduct a deductive content analysis in counseling research", *Counseling Outcome Research and Evaluation*, 1-13.
- Osakwe, Z. T., Ikhapho, I., Arora, B. K., & Bubu, O. M. (2020). "Identifying public concerns and reactions during the COVID-19 pandemic on Twitter: A text-mining analysis", *Public Health Nursing*, 38/2, 145-151.
- Piñeiro-Naval, V. (2020). "The content analysis methodology. Uses and applications in communication research on Spanish-speaking countries", *Communication & Society*, 33/3, 1-15.
- Redek, T., & Godnov, U. (2018). "Twitter as a political tool in EU countries during the economic crisis: A comparative text-mining analysis", *Društvena istraživanja: časopis za opća društvena pitanja*, 27/4, 691-711.
- Riffe, D., Lacy, S., Watson, B. R., & Fico, F. (2019). *Analyzing Media Messages: Using Quantitative Content Analysis in Research*, Routledge, New York.
- Sjøvaag, H., & Stavelin, E. (2012). "Web media and the quantitative content analysis: Methodological challenges in measuring online news content", *Convergence*, 18/2, 215-229.
- Taylan, H. H. (2011). "Sosyal Bilimlerde Kullanılan içerik analizi ve söylem analizinin karşılaştırılması", *Bingöl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (BUSBED)*, 1/2, 63-76.
- Varghese, D. (2018). *Comparative Study on Classic Machine learning Algorithms*. (06.12.2018) <https://bit.ly/3DFO3YA>
- Waegel, D. (2006). *The Development of Text-Mining Tools and Algorithms*, Ursinus College, Pennsylvania.
- Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). "Text analysis in R", *Communication Methods and Measures*, 11/4, 245-265.
- Yang, M. & Han, C. (2021), "Revealing industry challenge and business response to Covid-19: a text mining approach", *International Journal of Contemporary Hospitality Management*, 33/4, 1230-1248.
- Yan, J. L. S., McCracken, N., & Crowston, K. (2014). "Semi-automatic content analysis of qualitative data", *iConference 2014*, 1128-1132.
- Yuxin, M. (2021). "Big Data Text Mining Techniques in Journalism and Communication", *7th International Conference on Humanities and Social Science Research (ICHSSR 2021)*, Atlantis Press, 256-259.

Beyan ve Açıklamalar (Disclosure Statements)

1. Bu çalışmanın yazarları, araştırma ve yayın etiği ilkelerine uyduklarını kabul etmektedirler (The authors of this article confirm that their work complies with the principles of research and publication ethics).
2. Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir (No potential conflict of interest was reported by the authors).
3. Bu çalışma, intihal tarama programı kullanılarak intihal taramasından geçirilmiştir (This article was screened for potential plagiarism using a plagiarism screening program).