**2. Gereç ve Yöntem**

Bu araştırma, hasta güvenliği ve inovasyon kavramlarının sosyal medya nasıl kullanıldığının belirlenmesi, kavramların yıllara ve lokasyonlara göre nasıl evrildiğinin ortaya konulması ile karar vericilere bir perspektif sunulması için tasarlanmıştır.

Araştırma, bir programlama dili olan Python (KAYNAK) kullanılarak gerçekleştirilmiştir. Bu bağlamda Twitter’dan #innovation, #innovations, innovation, patient safety ve patientsafety anahtar kelimeleri kullanılarak web scrapping (KAYNAK) yöntemi ile tweetler elde edilmiştir. Veri setinin analiz edilerek anlamlandırılmasında Topic Modeling (KAYNAK) yöntemi kullanılmıştır.

Elde edilen 114.460 adet inovasyon ve 455.033 adet hasta güvenliğine ilişkin tweet analiz edilmiştir. Toplam iki adet duplikasyon tweeti araştırma dışında bırakılmış ve toplam 569.491 adet tweet ile araştırma gerçekleştirilmiştir.

Elde edilen tweetler öncelikle bir temizleme işleminden geçirilmiş, metin semboller, fotoğraflar, piktoğraf, bayrak, harita resimleri, url linkleri vb. karakterlerden temizlenmiştir. Metin içerinden bulunan ve ‘#’ ve ‘@’ sembolleri ile başlayan, iki harf ve daha az harften oluşan kelimeler metinden çıkarılmıştır. Metin içinde herhangi boş bir değer olmadığı görülmüştür.

Veri setinde bulunan tarih formatından elde edilen yıl, ay, gün, gün adı ve saat değişkenleri ayrı birer özellik olarak veri setine eklenmiştir. Yapay zekâ alt bileşenlerinden olan doğal dil işleme (Natural language processing, NLP)’de sıklıkla kullanılan stopwords ve lemmatizer (KAYNAK) kullanılarak veri setindeki gereksiz kelimeler temizlenmiş ve çekimli kelimelerin indirgeme işlemi yapılmıştır.

Çalışma kapsamında İngilizce dilinde atılan 338.441 adet tweet analize tabi tutulmuştur. Bununla birlikte, elde edilen bütün tweetlere ilişkin sayısal bilgi de çalışmada verilmiştir.