

2.1.1. Sınıflandırma Yöntemleri (Classification Methods)

Genel olarak iki tür sınıflandırma problemi vardır: ikili problem ve çok sınıflı problem. İkili problem, bir tahmin sonucunun evet veya hayır kararıyla belirlenmesi gereken bir durum iken, çoklu sınıflandırma problemi, tahmin edilen bir sonucun birden çok sonuç olarak belirlendiği bir durumdur. (Kraipeerapun, 2009) Sınıflandırma problemlerinin çözümü için İkili Sınıflandırma, Çok Sınıflı Sınıflandırma ve Çok Etiketli Sınıflandırma yöntemleri kullanılır. Ayrıntılar aşağıdaki alt başlıklarda listelenmektedir.

2.1.1.1. İkili Sınıflandırma (Binary Classification)

İkili sınıflandırma, adından da anlaşılacağı üzere verileri olası iki kategoriden birine sınıflandırma problemlerinin çözümünde kullanılır. Bu sınıflandırma yönteminde, sorulacak soruların sadece iki cevabı olmalıdır. Bu tür sınıflandırmalar için ünlü filozof Aristoteles'in mantığı geçerlidir. Bu mantığın temelinde, sorulara karşılık gelen iki ana cevaba dahil olup olmama durumu vardır.

Bu sınıflandırmaya örnek olarak: E-posta spam tespiti, belirli tıbbi durum tespiti, duygu analizi (yalnızca pozitif ve negatif kategori kabul edilmiştir) düşünülebilir.

2.1.1.2. Çok Sınıflı Sınıflandırma (Multi-class Classification)

Çok sınıflı sınıflandırmada, ikili sınıflandırmadan farklı olarak sınıf sayısı ikiden fazladır. Bu sınıflandırma problemlerinin çözümünde, verilecek cevap ikiden fazla değer içinden seçilerek cevaplanacaksa buna çok sınıflı sınıflandırma (Multi-class classification) problemi denir. Buradaki önemli noktalardan biri, verilen cevap maksimum bir sınıfa dahil olabilir.

Örneğin; elimizde bir e-ticaret sitesinde satılan ürüne yapılmış olan yorumlar olsun. Bu yorumların da 3 kategorisi olsun: Fiyat, kullanım ve kalite. Bu yorumlar belli bir sınıf altında toplanmak istendiği zaman çok sınıflı sınıflandırma yöntemi kullanılmalıdır.

2.1.1.3 Çok Etiketli Sınıflandırma (Multi-label Classification)

Çoklu etiketli sınıflandırma problemlerinin çözümünde de bir önceki yöntemde anlatılan çok sınıflı sınıflandırma gibi verilecek cevap, ikiden fazla değer içinden seçilerek verilecektir. Bu sınıflandırma yönteminin çok sınıflı sınıflandırma yönteminden farkı verilen cevabın birden fazla sınıfa dahil olabilesidir. Çoklu etiketli sınıflandırmanın amacı, tek bir örnek için bir dizi ilgili etiket ataması yapabilmektir.

Örneğin; bir önceki yöntemde bahsettiğimiz, satın alınmış olan yorumların kategorileri şu şekildedir: Fiyat, kullanım ve kalite.

Bu yorumların içinde birden fazla sınıfa dahil olan yorumlar olabilir. Sadece kullanımdan bahsedilebileceği gibi, fiyat ve kaliteden de aynı anda bahsedilen yorumlar içerebilir ve bu yorumları belirli bir sınıfa atamak istenmeyebilir. Bu tarz problemlerde doğru çıktı, çoklu etiketli sınıflandırma yöntemi kullanılarak elde edilmelidir. Bu çalışma boyunca yukarıda ayrıntıları paylaşılan sınıflandırma yöntemlerinden Çok Sınıflı Sınıflandırma kullanılmıştır. Bu sınıflandırma yönteminin kullanılmasının sebebi, ileride ayrıntıları verilecek olan veri setinin sınıflandırılmasının ikiden fazla değer içinden seçilerek maksimum bir sınıfa dahil olmasından kaynaklanmaktadır. Kısaca, kullanılacak olan veri seti beş farklı sınıftan oluşacak ve maksimum bir sınıfa dahil olacaktır.

Veri seti iki farklı kategoriden daha fazla kategoriye sahip olduğu için İkili Sınıflandırma yöntemi kullanılmamıştır. Veri seti içindeki herhangi bir verinin birden fazla kategoriye aynı anda dahil olma ihtimali olmadığı için de Çoklu Etiketli Sınıflandırma yöntemi kullanılmamıştır.

2.2. BERT (Bidirectional Encoder Representations from Transformers)

"Bidirectional Encoder Representations from Transformers" ifadelerinin baş harflerinden oluşan BERT algoritması, Ekim 2018'de Google tarafından geliştirilen doğal dil işleme (NLP) ön eğitimi için Transformer (derin öğrenme modeli) tabanlı bir makine öğrenimi tekniğidir. (Devlin vd., 2018)

Google, 2015 yılında Rankbrain (Schachinger, 2017) algoritmasını makine öğrenmesi ile destekleyerek arama sonuçlarında insan mantığına en yakın ve doğru cevapları filtrelemeyi sağlamıştır. 2019 yılı ile beraber BERT güncellemesini yayınlayarak, sorgu kelimelerini ayrı ayrı işlemek yerine tüm cümleyi incelemeye başlamış ve tüm