

Bu ödev, "Müzik Projesi: Şarkı Sözü Benzerlik Analizi" kapsamında şarkı sözleri arasındaki tematik benzerliği ölçmeyi amaçlamaktadır. Şarkı sözlerinin yoğun anlam taşıyan metinler olması sebebiyle, bu metinler arasındaki tematik benzerliklerin belirlenmesi, müzik tavsiye sistemlerinden telif hakkı ihlali tespitine kadar geniş bir yelpazede uygulama alanı bulmaktadır. Projemizdeki temel problem, şarkı sözlerinin içerdiği derin anlamı ve tematik yapıları otomatik olarak kavrayarak, aralarındaki benzerlik derecesini nicel olarak ifade etmektir.

Bu amaca ulaşmak için, öncelikle bir veri kaynağı olarak Genius API gibi platformlardan elde edilen şarkı sözleri üzerinde ön işleme adımları uygulanmıştır. Özellikle, şarkı sözlerinden tekrar eden nakaratların temizlenmesi, metinlerin daha saf ve anlamlı bir temsili elde etmek için kritik bir adımdır. Ardından, temizlenmiş bu metinler üzerinde **Word2Vec** ve **TF-IDF** gibi farklı metin vektörleştirme modelleri eğitilmiştir.

Ödev boyunca, eğitilen bu modeller kullanılarak, seçilen bir şarkı sözü ile veri setindeki diğer tüm şarkı sözleri arasındaki tematik benzerlik, **Kosinüs Benzerliği (Cosine Similarity)** skoru ile hesaplanmıştır. Elde edilen benzerlik skorları sayesinde, her bir modelin belirlenen giriş şarkı sözüne en çok benzeyen ilk 5 şarkı sözü tespit edilmiştir. Bu sonuçlar, hem anlamsal açıdan subjektif olarak değerlendirilmiş hem de farklı modellerin sıralama tutarlılıkları Jaccard benzerliği kullanılarak karşılaştırılmıştır. Bu kapsamlı değerlendirme, farklı model yapılandırmalarının (örneğin Word2Vec'teki vektör boyutu, pencere genişliği, CBOW/Skip-gram mimarisi gibi parametreler) ve metin temsil yöntemlerinin (TF-IDF vs. Word2Vec) şarkı sözü benzerlik analizindeki performanslarını ve güçlü/zayıf yönlerini ortaya koymayı hedeflemektedir.

Sonuç olarak, bu çalışma, şarkı sözleri gibi yapılandırılmamış metin verileri üzerinde tematik benzerlik analizinin nasıl yapılabileceğini göstermekte, farklı doğal dil işleme modellerinin etkinliğini deneysel olarak karşılaştırmakta ve elde edilen bulguların pratik uygulamalarına yönelik çıkarımlar sunmaktadır.