

## [1] GİRİŞ

Bu çalışmada, metin tabanlı veri işleme teknikleri kullanılarak CV (özgeçmiş) metinleri ile iş ilanı açıklamalarının eşleştirilmesi hedeflenmiştir. Özellikle, doğal dil işleme (NLP) alanında yaygın olarak kullanılan TF-IDF (Term Frequency–Inverse Document Frequency) ve cosine similarity (kosinüs benzerliği) yöntemleri yardımıyla, iki farklı metin kümesi arasında içerik benzerliği analizi yapılmıştır. Proje kapsamında geliştirilen sistem, kullanıcıların iş ilanlarına daha uygun pozisyonları kolayca belirlemelerine yardımcı olabilecek bir altyapı sunmaktadır. Bu proje, gözetimsiz öğrenme temelli olup herhangi bir etiketli veri kullanılmadan yalnızca metin içeriklerinden anlamlı sonuçlar elde etmeye odaklanmaktadır.

## [2] VERİ SETLERİNİN TANITIMI

Projede iki temel veri seti kullanılmıştır:

- CV Verisi (resume\_data.csv): Bu veri seti bireylerin kariyer hedeflerini içeren career\_objective adlı metin sütununu içermektedir. Her satır bir özgeçmiş temsil eder.
- İş İlanı Verisi (postings.csv): İşverenlerin yayınladığı iş ilanlarını içeren description adlı sütun bu çalışmada esas alınmıştır. Bu sütun, işin gereklilikleri, sorumlulukları ve aranan nitelikleri içeren metinlerden oluşmaktadır.

Veri setleri anonimdir ve herhangi bir kişisel bilgi barındırmamaktadır. Yalnızca doğal dil içeriği üzerinden analiz yapılmıştır.

<https://www.kaggle.com/datasets/saugataroyarghya/resume-dataset>

<https://www.kaggle.com/datasets/arshkon/linkedin-job-postings>

## [3] VERİ ÖN İŞLEME

Ham metinlerin makine tarafından işlenebilir hale getirilmesi için çeşitli ön işlemler uygulanmıştır:

- Metinlerin küçük harfe dönüştürülmesi
- Noktalama işaretlerinin ve özel karakterlerin kaldırılması
- Türkçe stopword'lerin (gereksiz kelimeler) silinmesi
- Kelimelerin kök hallerine indirgenmesi (lemmatization)
- Bazı örneklerde stem işlemi ile alternatif köklemeler denenmiştir

Bu işlemler sayesinde, metinlerin içerik yapıları sadeleştirilmiş ve daha sağlıklı benzerlik hesaplamaları yapılabilmektedir.

## [4] VEKTÖRLEŞTİRME YÖNTEMİ: TF-IDF

TF-IDF yöntemi, bir kelimenin belirli bir metin içinde ne kadar önemli olduğunu ölçen bir tekniktir. Bu çalışmada her CV ve iş ilanı metni, TF-IDF vektörlerine dönüştürülmüştür. Bu vektörler, belgeler arasındaki içerik benzerliğini karşılaştırmak için temel veri yapısı olarak kullanılmıştır.

## [5] BENZERLİK HESAPLAMA: COSINE SIMILARITY

Metin vektörleri oluşturulduktan sonra, her CV için tüm iş ilanlarıyla olan cosine similarity değeri hesaplanmıştır. Bu sayede, her CV için en yakın (yani en yüksek benzerlik skoruna sahip) iş ilanı belirlenmiştir. Cosine similarity, iki metnin yön benzerliğini ölçerek içerik olarak ne kadar benzeştiğini belirlemeye yarayan bir yöntemdir.

## [6] SONUÇLAR VE ANALİZ

Elde edilen eşleşmeler cv\_job\_matches.csv adlı çıktı dosyasına kaydedilmiştir. Her satırda, bir CV'nin en yakın eşleştiği iş ilanı ve bu eşleşmenin cosine similarity skoru yer almaktadır. Çoğu eşleşme anlamlı skorlarla belirlenmiştir. Ancak, bazı CV'ler boş veya çok kısa içerik barındırdığı için benzerlik skoru çok düşük (hatta 0.00) olarak çıkmıştır. Bu durumlar, metin ön işleme sürecinde yeterli bilgi içermeyen örnekler olarak yorumlanmıştır.

## [7] SONUÇ VE ÖNERİLER

Bu projede, CV ve iş ilanları arasında otomatik eşleşme sağlanabilecek bir temel sistem başarıyla geliştirilmiştir. Kullanılan yöntemler basit ama etkili olup, büyük ölçekli veri kümelerinde de çalışabilecek niteliktedir. Gelecekte bu sistem aşağıdaki yönlerde geliştirilebilir:

- Daha derin bağlam anlayışı için Word2Vec, BERT gibi ileri düzey modellerin entegrasyonu
- Sadece "career\_objective" değil, CV'deki diğer bölümlerin de (eğitim, beceriler) dahil edilmesi
- Görselleştirme araçlarıyla eşleşmelerin sunulması
- Geri bildirim mekanizmasıyla sistemin zamanla öğrenmesi