**Rapor**

**Amacımız:**

Kullanıcıların yüklediği özgeçmişlere doğal dil işleme ve öğrenme algoritmaları kullanarak iş veri kümesi oluşturmak ve bu verileri TF-IDF, KNN ve Kosinüs benzerlikleri kullanarak analiz etmek. Böylece, kullanıcılara uygun iş tavsiyeleri sunmak istiyoruz.

**Web Scraping:**

Verilerin ön işleme sürecinin ardından, web scraper aracılığıyla kariyer.net de yazılım sektöründe ilanlar toplandı. İlanların bağlantıları, başlıkları, şirket isimleri ve genel nitelikler gibi bilgiler toplandı. İş ilanlarında skill adı altında bir alan olmadığı için gereksinimler kısmında bulunan alanlar çekildi ve birçok skill başlıklarının bulunduğu liste üzerinden paragraf taranarak skill niteliğindeki alanlar belirlendi. Veri seti, Görsel-1'de gösterildiği gibi oluşturuldu..



Görsel-1

TF-IDF ve KNN:

İş ilanlarını ve kullanıcının CV'sini karşılaştırmak için Cosine Similarity ve KNN algoritmalarını kullanır. Her iki algoritma da kullanıcının CV'si ile iş ilanlarının beceri kümesi arasındaki benzerliği ölçer ve en yüksek benzerlik skoruna sahip iş ilanlarını önerir. KNN algoritması içinde tekrardan TF-IDF uygulanır.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Görsel-2

KNN Skor Aralığı:

KNN (k-Nearest Neighbors) algoritması, öklidyen mesafesine veya başka bir mesafe ölçüsüne dayalı olarak benzerlikleri değerlendirdiğinden, skorlar genellikle 0'dan büyük olacaktır. KNN skorlarının referans aralığı, skorların doğrudan mesafe ölçüsüne bağlı olarak değişir.

Kodda, KNN algoritması sklearn kütüphanesi kullanılarak uygulanır ve öklidyen mesafesi kullanır. Öklidyen mesafesi, iki nokta arasındaki düz çizgi mesafesidir ve 0'dan daha büyük veya eşit değerler alır. İki nokta aynıysa, mesafe 0 olacaktır. İki nokta ne kadar uzakta olursa, mesafe de o kadar büyük olur.

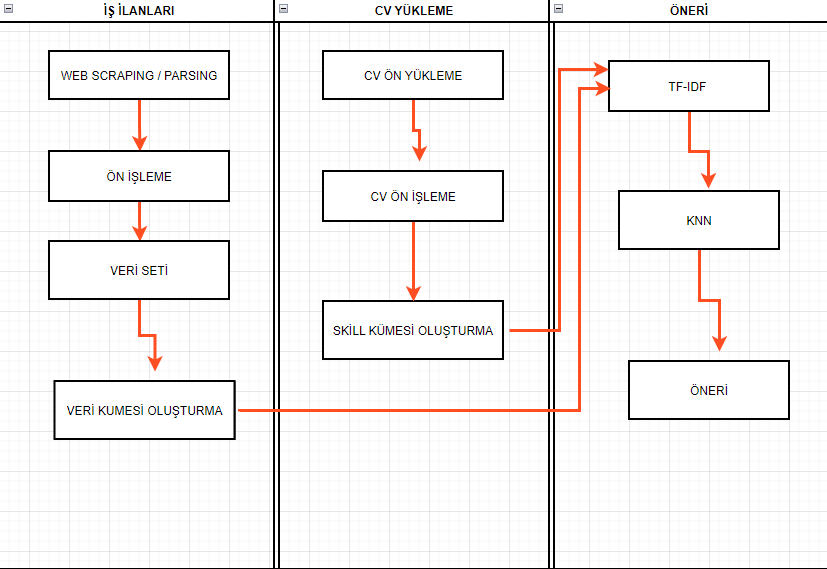
Öneri sistemi bağlamında, KNN skorları mesafe değerlerinin tersine dönüştürülür, böylece daha büyük skorlar daha benzer iş ilanlarını temsil eder. Bu durumda, KNN skorlarının referans aralığı 0'dan büyük olacaktır. Daha düşük bir KNN skoru, kullanıcının CV'sine daha yakın bir iş ilanı anlamına gelir. Ancak, bu skorlar öznel olup, doğrudan yorumlanabilir bir anlam taşımaz ve yalnızca önerilen iş ilanları arasında karşılaştırma yapmak için kullanılabilir.(Görsel-3)

Görsel-3

**Öneri Sistemi Uygulanan Adımlar**

Kullanıcıların sisteme yükleyeceği özgeçmişler için uyguladığımız adımlar:

1. **Veri toplama:** Web scraping ile kariyer.net sitesinden iş ilanları ve beceri gereksinimleri gibi ilgili veriler toplanır.
2. **Veri ön işleme:** Toplanan veriler temizlenir, gereksiz kısımları çıkarılır ve işlemeye hazır hale getirilir. Bu adımda, verilerin içeriği ile ilgili gereksiz bilgiler kaldırılır, veriler düzenlenir ve standartlaştırılır.
3. **Özellik çıkarma:** Veriler üzerinde yapılabilecek işlemleri daha kolay hale getirmek için, verilerden ilgili özellikler çıkarılır. İş ilanlarındaki gereksinimlerde skillerin olduğu listede olan ve yazılım alanındaki genel araç, program ve tekniklerin kullanımına göre özellikler çıkarılır.
4. **Özellik vektörü oluşturma:** KNN algoritması, özellikleri vektör şeklinde kullanır. Bu nedenle, verilerdeki her bir özellik bir vektör elemanı olarak kodlanır.
5. **TF-IDF işlemi:** Kullanılan özelliklerin vektörlerini oluşturduktan sonra, TF-IDF (Term Frequency-Inverse Document Frequency) işlemi uygulanır. Bu işlem, yazılımsal skillerin yüklenen CV içerisindeki sıklığına, aynı özelliğin tüm CV’ler içindeki sıklığına da bakarak önemini belirler. Önemi belirlerlen skiller için TF-IDF skoru bulunur.
6. **Veri kümesi oluşturma:** KNN algoritması için bir veri kümesi oluşturulur. Bu veri kümesi, iş ilanlarının vektörleri ve bu ilanların özelliklerini içerir.
7. **KNN modeli eğitimi:** Veri kümesi oluşturulduktan sonra, KNN modeli eğitilir. Bu adımda, veri kümesindeki her bir iş ilanı için, diğer iş ilanlarından ne kadar benzer olduğuna dair bir benzerlik skoru hesaplanır.
8. **Öneri yapma:** Son olarak, kullanıcının CV'si de vektörleştirilerek ve TF-IDF işlemi uygulanarak bir vektör oluşturulur. Kullanıcının CV'si ile ilgili benzerlik skorları hesaplanır ve en yüksek skorlara sahip iş ilanları kullanıcıya önerilir.



Görsel-4

