

Yapay Zeka Dersi Ödev-2: Metin Benzerliği ve Model Değerlendirme

Adın Soyadın

30 Mayıs 2025

1 Giriş

Bu raporda, TF-IDF ve Word2Vec modelleriyle CV ve iş ilanı metinleri arasında benzerlik analizi yapılmıştır. Giriş metni olarak CV_Index 0 seçilmiştir (Big data analytics ve database warehouse deneyimi).

2 Benzerlik Sonuçları

2.1 TF-IDF (Lemmatized)

Job_Posting_Index	Similarity Score
47635	0.45
123	0.34

2.2 TF-IDF (Stemmed)

Job_Posting_Index	Similarity Score
47635	0.43
123	0.32

2.3 Word2Vec (Lemmatized CBOW, window=2, dim=100)

Job_Posting_Index	Similarity Score
47635	0.99955
234	0.987

3 Anlamsal Değerlendirme

Model	Benzer Metinler	Skorlar	Ortalama
tfidf_lemmatized	47635, 123, ...	2,3,2,1,2	2.0
word2vec_lemmatized_cbow_win2_dim100	47635, 234, ...	5,4,4,3,4	4.0

4 Sıralama Tutarlılığı

	tfidf_lemmatized	word2vec_lemmatized_cbow_win2_dim100
tfidf_lemmatized	1.00	0.40
word2vec_lemmatized_cbow_win2_dim100	0.40	1.00

5 Yorumlar

Word2Vec modelleri, yüksek benzerlik skorları ve anlamlı eşleşmelerle TF-IDF'den daha iyi performans gösterdi. Lemmatized CBOW (window=2, dim=100) en yüksek anlamsal skoru (4.0) aldı. TF-IDF skorları daha düşük (0.45 civarı) ve eşleşmeler daha az anlamlı. Word2Vec'te CBOW modelleri, Skip-gram'e göre daha tutarlıydı. Vektör boyutu ve window büyüklüğü sonuçları az etkiledi.