

BURSA TEKNİK ÜNİVERSİTESİ
MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ



**DİL İŞLEME DESTEKLİ CV -STARTUP
EŞLEŞTİRME SİSTEMİ**

LİSANS BİTİRME ÇALIŞMASI

HÜMEYRA ÇİMEN

Bilgisayar Mühendisliği Bölümü

HAZİRAN 2024

BURSA TEKNİK ÜNİVERSİTESİ
MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ



**DİL İŞLEME DESTEKLİ CV - STARTUP
EŞLEŞTİRME SİSTEMİ**

LİSANS BİTİRME ÇALIŞMASI

Hümevra ÇİMEN
19360859053

Bilgisayar Mühendisliği Bölümü

BTÜ, Mühendislik ve Doğa Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü'nün 19360859053 numaralı öğrencisi HÜMEYRA ÇİMENİ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı **Dil İşleme Destekli CV - Startup Eşleştirme Sistemi** başlıklı bitirme çalışmasını aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Danışmanı : **Prof. Dr. Tugray Tugay BİLGİN**

Bursa Teknik Üniversitesi

Jüri Üyeleri : **Arş. Gör. Talha KORUK**
Bursa Teknik Üniversitesi

Arş. Gör. Sena DİKİCİ
Bursa Teknik Üniversitesi

Savunma Tarihi : 03 Haziran 2024

Danışman: Prof. Dr. Turgay Tugay BİLGİN

BM Bölüm Başkanı : Prof. Dr. Turgay Tugay Bilgin

Bursa Teknik Üniversitesi

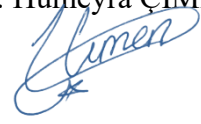
03/06/2024

HAZİRAN 2024

İNTİHAL BEYANI

Bu bitirme çalışmasında görsel, işitsel ve yazılı biçimde sunulan tüm bilgi ve sonuçların akademik ve etik kurallara uyularak tarafımdan elde edildiğini, bitirme çalışması içinde yer alan ancak bu çalışmaya özgü olmayan tüm sonuç ve bilgileri bitirme çalışmasında kaynak göstererek belgelediğimi, aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim.

Öğrencinin Adı Soyadı: Hümeýra ÇİMEN



İmzası :

ÖNSÖZ

Bu bitirme çalışmasının hazırlanması sürecinde ve bu projeyi gerçekleştirmem için bana rehberlik eden ve sürekli destek sağlayan danışman hocam Dr. Turgay Tugay Bilgin'a sonsuz teşekkürlerimi sunarım. Hocamın bilgi ve tecrübesi, bu çalışmanın başarısında büyük rol oynamıştır.

Çalışmam boyunca her türlü imkanı sağlayan ve bana destek olan Bursa Teknik Üniversitesi'ne ve özellikle Bilgisayar Mühendisliği Bölümü'ne teşekkür ederim. Ayrıca, projeye sağladıkları veri ve kaynaklarla büyük katkı sunan Google Colab'a ve Eu -Startups'a minnettarlığımı ifade etmek isterim.

Bu sürecin her aşamasında yanımda olan, beni motive eden ve moral destek sağlayan aileme teşekkürlerimi sunarım. Onların desteği ve anlayışı olmadan bu çalışmayı tamamlamak mümkün olmazdı.

Bu bitirme çalışmasının, startup'lar ile mentorları daha etkin ve verimli bir şekilde eşleştirmeye yönelik yapılan araştırmalara katkı sağlamasını umuyorum.

Saygılarımla,

Haziran 2024

HÜMEYRA ÇİMEN

İÇİNDEKİLER

Sayfa

1. GİRİŞ 16

1.1 Tezin Amacı	16
1.1.1 Tezin ikincil amaçları.....	16
1.2 Literatür Araştırması	17
1.2.1 İşe Alım Süreçleri ve CV Değerlendirme Teknikleri	17
1.2.1.1 CV Değerlendirme Teknikleri.....	18
1.2.1.2 NLP Kullanımı	18
1.2.1.3 Sosyal Medya ve Haber Kapsamı	18
1.2.1.4 FoDRA (Dört Boyutu Tavsiye Algoritması).....	19
1.2.2 Teknolojinin İşe Alım Süreçlerine Entegrasyonu	19
1.2.2.1 Google	19
1.2.2.2 IBM	20
1.2.2.3 Amazon	20
1.2.3 Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi	21
1.2.3.1 TF-IDF	21
1.2.3.2 Word Embedding Yöntemi	22
1.3 Hipotez	22

2.YÖNTEM..... 23

1.4 Veri Toplama Yöntemleri	23
1.4.1 Startup verilerinin toplanması	23
1.4.1.1 Site haritasının çıkarılması:.....	24
1.4.1.2. Scrapy İle Veri Kazıma:.....	24
1.4.2 CV verilerinin toplanması.....	24
1.4.2.1 Veri formatı.....	24
1.4.2.2. Veri toplama ve dönüştürme	24
1.4.2.3 Veri zenginleştirme (Data augmentation)	25

1.4.2.4 Veri Düzenleme	25
1.4.3 Veri Temizleme ve Düzenleme:	25
1.4.4 Veri Analizi Ve Görselleştirme.....	25
1.5 Model Eğitimi Ve Değerlendirme Yöntemleri.....	25
1.5.1 BERT Modeli Kullanımı.....	25
1.5.1.1 Metin Verilerinin İşlenmesi	26
1.5.1.2 Özellik vektörlerinin işlenmesi hesaplanması:.....	27
1.5.2 Modelin uygulanması.....	28
1.5.2.1 BERT Modeli ve Tokenizer'ın Yüklenmesi:.....	29
1.5.2.2 Özellik Çıkarma Fonksiyonları	29
1.5.2.3 Metinlerin İşlenmesi.....	30
1.5.2.4 Gömülü Vektörlerin Oluşturulması:	30
1.5.2.4 Benzerlik Hesaplama	30
1.5.2.5 En Uygun CV'nin Bulunması.....	31
1.5.2.6 Sonuçların Görselleştirilmesi	34
1.5.3 Modelin Testi	35
5.SONUÇ 36	
1.6 Çalışmanın Uygulama Alanı	36
6.ÖNERİLER	35
7.KAYNAKÇA	38

KISALTMALAR

BERT	: Bidirectional Encoder Representations from Transformers
CV	: Curriculum Vitae
NLP	: Natural Language Processing
TF-IDF	: Term Frequency-Inverse Document Frequency
IBM	: International Business Machines
JSON	: JavaScript Object Notation
CLS	: Classification Token
SEP	: Separator Token
ID	: Identification

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.2.2.6.1 Model -1 in Selin Öztürk cv ile startup eşeřtirmesi skorları.....32

Çizelge 2.2.2.6.2 Model -2 nin Selin Öztürk cv ile startup eşeřtirmesi skorları...32

ŞEKİL LİSTESİ

Sayfa

Şekil 2.2.2.2.1 Proje Modelinin Görseleştirilmesi.....	30
Şekil 2.2.2.5.1 Modelin Diagramla Gösterilmesi.....	31
Şekil 2.2.2.6.1 ChatGpt nin ilgili cv nin firmalara olan uyumunun göstergesi.....	33
Şekil 2.2.2.7.1 Uygun CV-Mentör eşleştirmesi-1.....	34
Şekil 2.2.2.7.2 Uygun CV-Mentör eşleştirmesi-2.....	35
Şekil 2.2.3.1 Selin Öztürk CV.....	35
Şekil 2.2.3.2 Selin Öztürk cvsine uygun firmalar çalışma alanları.....	36
Şekil 2.2.3.3 Selin Öztürk cvsine uygun firmalar ve skorları.....	36

DİL İŞLEME DESTEKLİ CV-STARTUP EŞLEŞTİRME SİSTEMİ

ÖZET

Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi, iş arayan bireyler ile uygun startup şirketlerini eşleştirmeyi amaçlayan bir sistemdir. Bu çalışmada, öncelikle iş arayanların CV'leri dil işleme teknikleriyle analiz edilmiş ve niteliklerine uygunluk dereceleri belirlenmiştir. Ardından, startup şirketlerinin ihtiyaçları ve gereksinimleri göz önünde bulundurularak bir eşleştirme algoritması geliştirilmiştir. Bu algoritma, iş arayanların yetenekleri, tecrübeleri ve ilgi alanları ile startup şirketlerinin beklentilerini karşılayan en uygun eşleşmeleri sağlamak için tasarlanmıştır. Çalışmanın sonuçları, dil işleme destekli CV mentor ve startup eşleştirme sisteminin etkinliğini ve kullanılabilirliğini göstermektedir. Bu sistem, iş arayanlar ile startup şirketlerinin karşılıklı fayda sağlayacak şekilde bir araya gelmelerini kolaylaştırarak işe alım sürecini iyileştirmektedir.

Anahtar kelimeler: Dil İşleme, CV Analizi, Startup Eşleştirme, Mentor Sistemi, İşe Alım, Teknoloji

SUMMARY

The Natural Language Processing Supported CV Mentor and Startup Matching System is designed to match job seekers with suitable startup companies. In this study, the CVs of job seekers are first analyzed using natural language processing techniques to determine their suitability levels. Subsequently, a matching algorithm is developed taking into account the needs and requirements of startup companies. This algorithm is designed to provide the most suitable matches that meet the qualifications, experiences, and interests of job seekers along with the expectations of startup companies. The results of the study demonstrate the effectiveness and usability of the natural language processing supported CV mentor and startup matching system. This system facilitates the mutual benefit of job seekers and startup companies, thereby improving the hiring process.

Keywords: Natural Language Processing, CV Analysis, Startup Matching, Mentor System, Hiring, Technology

1. GİRİŞ

Günümüz iş dünyasında, iş arayanlar ve şirketler arasındaki uyumun sağlanması giderek önem kazanmaktadır. Geleneksel işe alım süreçleri, genellikle zaman alıcı ve maliyetli olabilirken, doğru adayı doğru pozisyona yerleştirmekte zorluklar yaşanmaktadır. Bu bağlamda, teknolojik gelişmelerin işe alım süreçlerine entegre edilmesi ve iş arayanlar ile şirketler arasındaki eşleştirmenin iyileştirilmesi gerekmektedir. Bu çalışmada, dil İşleme Destekli CV,mentor ve startup eşleştirme sistemi gibi yenilikçi bir yaklaşımın kullanılmasıyla, iş arayanlar ile startup şirketleri arasında daha etkin bir eşleştirme süreci sağlanması amaçlanmaktadır. Bu girişimin, startup şirketlerinin ihtiyaç duydukları yetenekleri daha verimli bir şekilde karşılamalarına yardımcı olması hedeflenmektedir. Bu çalışma, geleneksel işe alım yöntemlerine alternatif bir yaklaşım sunarak, iş arayanlar ile startup şirketleri arasındaki eşleştirmenin daha verimli ve etkili hale getirilmesine katkı sağlamayı amaçlamaktadır.

1.1 Tezin Amacı

Bu çalışmanın temel amacı, Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi'nin kullanımıyla iş arayanlar ile startup şirketleri arasında daha etkin bir eşleştirme süreci sağlamaktır. Bu amaç doğrultusunda, mevcut işe alım süreçlerinin eksikliklerini belirlemek ve bu eksiklikleri gidermek için yenilikçi yapay zeka dil modeli destekli bir yaklaşım geliştirilmiştir.

1.1.1 Tezin ikincil amaçları

İş arayanların niteliklerini doğru bir şekilde analiz ederek, onları en uygun startup şirketleriyle eşleştirmek.

Startup şirketlerinin ihtiyaç duydukları yetenekleri belirleyerek, bu yeteneklere sahip adayları bulmak ve işe alım sürecini hızlandırmak.

1.2 Literatür Araştırması

İşe alım süreçleri ve CV değerlendirme teknikleri konusunda yapılan literatür taraması ele alınacaktır. Geleneksel işe alım yöntemlerinin zayıf yönleri ve teknolojinin işe alım süreçlerine entegrasyonu hakkındaki güncel araştırmalar incelenecek ve Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi'nin literatürdeki yeri tartışılacaktır. Literatürde yapay zeka tabanlı iş eşleştirme modelleri ile ilgili mevcut araştırmaları inceleyeceğiz. Bu modellerin farklı algoritmaları, veri kümeleri ve performans değerlendirmeleri üzerinde durulacaktır. Ayrıca, yapay zeka tabanlı iş eşleştirme modellerinin sunduğu avantajlar ve dezavantajlar da tartışılacaktır.

1.2.1 İşe Alım Süreçleri ve CV Değerlendirme Teknikleri

İşe alım süreçleri, iş dünyasında kritik bir rol oynamakta ve doğru yeteneklerin kuruma kazandırılması açısından büyük önem taşımaktadır. Literatürde yapılan araştırmalar, geleneksel işe alım yöntemlerinin bazı zayıf yönlerine dikkat çekmektedir. Bu yöntemler genellikle zaman alıcı, maliyetli ve subjektif olma eğilimindedir. Geleneksel süreçlerin bu dezavantajları, işe alım süreçlerinin iyileştirilmesi ve verimliliğin artırılması için yeni ve yenilikçi yaklaşımların araştırılmasını teşvik etmektedir.

Geleneksel işe alım süreçleri, genellikle adayların CV'lerinin manuel olarak incelenmesi ve değerlendirilmesi şeklinde yürütülmektedir. Ancak, bu süreçlerin uzun sürmesi, insan hatalarına açık olması ve subjektif değerlendirmelere yol açması gibi dezavantajları bulunmaktadır. Bu sorunların üstesinden gelmek amacıyla, işe alım süreçlerinde otomasyon ve yapay zeka gibi teknolojik yeniliklerin kullanımı giderek artmaktadır.

Örneğin, bazı çalışmalarda [1] makalesinde kelimelerin alt vektörlerinin zenginleştirilmesi ve bu vektörlerin işe alım süreçlerinde kullanılması önerilmektedir. Bu yöntem, adayların CV'lerinin daha hızlı ve objektif bir şekilde değerlendirilmesine olanak tanımaktadır. Ayrıca, Cho ve arkadaşlarının [2] makalesinde “İstatistiksel makine çevirisi için RNN kodlayıcı-dekoder kullanarak kelime öbeklerinin temsillerini öğrenme” başlıklı çalışması, işe alım süreçlerinde kullanılan metin analiz tekniklerine önemli katkılarda bulunmaktadır. Bu çalışma, RNN kodlayıcı-dekoder kullanarak kelime öbeklerinin temsillerini öğrenme üzerine odaklanmakta ve bu yaklaşımın CV değerlendirme süreçlerinde nasıl kullanılabileceğini incelemektedir.

Yeni teknolojiler ve yöntemler, işe alım süreçlerinin daha verimli, adil ve etkili olmasını sağlamaktadır. Otomasyon ve yapay zeka tabanlı çözümler, hem işverenler hem de adaylar için daha iyi sonuçlar doğurabilir. İşe alım süreçlerinin bu şekilde modernize edilmesi, kurumların en uygun adayları daha kısa sürede ve daha az maliyetle bulmalarına yardımcı olmaktadır. Bu nedenle, literatürdeki yeni yaklaşımların incelenmesi ve uygulamaya geçirilmesi, işe alım süreçlerinin geleceği için büyük önem taşımaktadır.

1.2.1.1 CV Değerlendirme Teknikleri

Geleneksel CV değerlendirme yöntemleri manuel olarak yapılır ve zaman alıcıdır. NLP ve makine öğrenimi teknikleri, bu süreçleri otomatikleştirerek daha hızlı ve doğru değerlendirme yapılmasını sağlar. TF-IDF ve word embedding yöntemleri, CV'lerdeki anahtar kelimeleri ve bağlamsal bilgileri analiz ederek adayların uygunluklarını belirler.

1.2.1.2 NLP Kullanımı

NLP teknikleri, CV'lerdeki metinleri analiz ederek adayların becerilerini, deneyimlerini ve niteliklerini değerlendirir. TF-IDF yöntemi, belgelerdeki önemli kelimeleri belirlerken, word embedding yöntemleri kelimeler arasındaki anlamsal ilişkileri ve bağlamları öğrenir. Bu teknikler, CV'lerin daha kapsamlı bir şekilde analiz edilmesine olanak tanır ve işe alım süreçlerinde daha doğru eşleştirmeler yapılmasını sağlar.

1.2.1.3 Sosyal Medya ve Haber Kapsamı

Bir makalede[3], işe alım sürecini desteklemek için sosyal medyayı kullanarak adayların teknik ve davranışsal ipuçlarını analiz eden bir sistem sunulmaktadır. Sistem, adayların şanslarını değerlendirmelerine ve retrospektif yapmalarına da yardımcı olabilir. Gelecekteki çalışmalarda, daha güvenilir aday listesi sağlamak için daha sağlam dil ve makine öğrenimi modelleri uygulamayı planlıyor. Haber kapsamı da bir iş başvurusu yapan kişinin yapabileceği aktiviteler hakkında içgörüler sağlar. Haber öğelerini de potansiyel bir çalışanın değerlendirme metriklerine dahil etmeyi planlıyoruz. Bu alanda açıklanabilirlik önemli bir gereksinimdir, gelecekte bu yönde ilerlemeyi planlıyoruz. Sosyal medya verileri doğası gereği heterojendir.

Kullanıcıların dezavantajlı olmaması için verileri bir şekilde normalize etmeye yönelik algoritmalar aramak gelecekteki bir yöndür. Ek olarak, kullanıcının sosyal medya ayrıntılarını girmesine gerek kalmadan, bir kimlik oluşturma tekniği uygulamayı planlıyoruz ki bu, adayın profilindeki tüm ilgili sosyal medya hesaplarını tanımlamaya yardımcı olabilir. İşe alım için kullanılan modellerdeki önyargıları ele almak [4] başka bir araştırma alanıdır. Bu makalede tarif edilen araçları denetlemek [5] de bir gelecek yönüdür.

1.2.1.4 FoDRA -Dört Boyutu Tavsiye Algoritması

FoDRA, iş ve aday profillerini iş tanımlarından ve CV'lerden yapılandırılmış bir form kullanarak değerlendiren içerik tabanlı bir öneri algoritmasıdır. Bu algoritma, iş arayanların belirli bir iş pozisyonu için uygunluğunu değerlendirmek amacıyla kullanılır. FoDRA, FoDRA algoritması, iş arayanların profillerini, şirketler tarafından belirtilen iş gereksinimlerine karşı eşleştirmek için bir kümeleme algoritması kullanır. Özgeçmişlerden ve iş tanımlarından öznitelikleri çıkarır ve ardından bu verileri analiz etmek için WEKA'yı kullanır. WEKA, Yeni Zelanda'daki Waikato Üniversitesi tarafından geliştirilmiş açık kaynak kodlu bir yazılımdır ve bilgisayar bilimlerinde veri madenciliği ve makine öğrenimi alanlarında yaygın olarak kullanılmaktadır.

1.2.2 Teknolojinin İşe Alım Süreçlerine Entegrasyonu

Teknolojinin işe alım süreçlerine entegrasyonu, son yıllarda önemli bir araştırma alanı haline gelmiştir. Yapay Zeka, Makine Öğrenimi ve Doğal Dil İşleme gibi teknolojiler, işe alım süreçlerini optimize etmek ve daha verimli hale getirmek için kullanılmaktadır. Bu teknolojiler, büyük veri kümelerini analiz ederek adayların yeteneklerini ve uygunluğunu değerlendirebilir, işe alım süreçlerini otomat birçok şirket tarafından verimliliklerini artırmak ve rekabet avantajı elde etmek için kullanılmaktadır. İşte bu entegrasyonu kullanan bazı şirketlerden örnekler ve bunların nasıl verimlilik sağladığına dair bilgiler:

1.2.2.1 Google

Google, işe alım süreçlerine yapay zeka ve makine öğrenimi teknolojilerini entegre ederek adayları değerlendirme ve seçme sürecini optimize etmektedir. Özellikle CV tarama ve aday mülakatlarında yapay zeka tabanlı araçlar kullanılarak adayların beceri ve uygunlukları hızlı bir şekilde değerlendirilmekte ve bu sayede süreç hızlanmaktadır.

Ayrıca, Google'ın kullandığı algoritmalar objektif değerlendirmeler yapmalarına olanak sağlayarak ön yargıyı azaltmaktadır.

Makine öğrenimi algoritmaları, adayların geçmiş performanslarını, eğitim ve deneyimlerini değerlendirir. Bu algoritmalar, belirli ölçütler ve gereksinimler çerçevesinde adayların işe uygunluğunu tahmin eder

1.2.2.2 IBM

IBM, işe alım süreçlerine doğal dil işleme ve makine öğrenimi teknolojilerini entegre ederek adayların yeteneklerini daha etkin bir şekilde değerlendirmekte ve uygun adayları belirlemektedir. Örneğin, IBM Watson adlı yapay zeka platformu, adayların CV'lerini analiz ederek belirli kriterlere uygunluğunu değerlendirebilmekte ve bu sayede işe alım sürecini hızlandırmakta ve verimliliği artırmaktadır.

IBM'in işe alım sürecindeki bir diğer önemli özellik, büyük veri analitiği ve veri görselleştirme tekniklerinin kullanılmasıdır. Bu teknikler, adayların mülakat performanslarını, CV'lerini ve diğer verilerini analiz ederek daha iyi kararlar almayı sağlar.

1.2.2.3 Amazon

Amazon, işe alım süreçlerinde yapay zeka ve büyük veri analitiği teknolojilerini kullanarak adayları değerlendirme ve seçme sürecini optimize etmektedir. Özellikle, Amazon'un kullandığı algoritmalar adayların CV'lerini ve mülakat performanslarını analiz ederek en uygun adayları belirlemekte ve bu sayede sürecin verimliliğini artırmaktadır.

Amazon'un işe alım sürecindeki bir diğer önemli özellik, sürekli geri bildirim döngüsüne dayalı olarak modellerin sürekli olarak iyileştirilmesidir. Şirket, insan kaynakları uzmanlarının geri bildirimlerini ve veri analizlerini kullanarak modellerini günceller ve iyileştirir.

Bu şirketlerin teknoloji entegrasyonu sayesinde işe alım süreçlerindeki verimliliklerini artırdıkları açıklandılar. Yapay zeka, makine öğrenimi ve doğal dil işleme gibi teknolojilerin kullanımı, işe alım süreçlerini otomatikleştirerek zaman tasarrufu sağlamakta ve daha objektif kararlar alınmasına olanak tanımaktadır. Bu da şirketlerin daha hızlı ve etkili bir şekilde yetenekli adayları işe almasına yardımcı olmaktadır. İkleştirebilir ve objektif kararlar vermeye yardımcı olabilir.

1.2.3 Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi

Literatürde Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi gibi yeni ve yenilikçi yaklaşımların tartışıldığı araştırmalar mevcuttur. Bu sistemler, adayların CV'lerini analiz ederek onlara kariyer geliştirme tavsiyeleri sunabilir ve uygun startup şirketleri ile eşleştirebilir. Bu tür sistemler, işe alım süreçlerinin daha objektif, verimli ve adil olmasına yardımcı olabilir.

Bu bölümde, literatürdeki mevcut araştırmalar incelenerek geleneksel işe alım yöntemlerinin zayıf yönleri, teknolojinin işe alım süreçlerine entegrasyonu ve Dil İşleme Destekli CV Mentor ve Startup Eşleştirme Sistemi gibi yenilikçi yaklaşımların önemi tartışılacaktır. Bu tartışma, çalışmanın teorik çerçevesini oluşturarak araştırmanın temelini sağlayacaktır.

Projemiz, doğal dil işleme (NLP) tekniklerini kullanarak CV, mentor ve startup eşleştirmesini hedeflemektedir. Bu bağlamda kullanılan yöntemler ve modeller TF-IDF, word embedding (Word2Vec, GloVe, FastText) ve derin öğrenme teknikleridir.

TF-IDF: TF ve IDF değerlerinin çarpımıyla elde edilir ve bir kelimenin bir belgede ne kadar önemli olduğunu belirtir.

1.2.3.1 TF-IDF

Bir belgedeki kelimenin önemini ölçmek için kullanılan bir istatistiksel yöntemdir. İki ana bileşenden oluşur: Term Frequency (TF) ve Inverse Document Frequency (IDF).

Term Frequency (TF): Bir kelimenin belirli bir belgede ne kadar sık geçtiğini gösterir. Hesaplanması şu şekildedir:

$$TF(t,d) = \text{Belgedeki_toplam_kelime_sayısı} / \text{Terim_t_nin_belgedeki_geçiş_sayısı}$$

Inverse Document Frequency (IDF): Bir kelimenin tüm belgeler arasında ne kadar nadir olduğunu gösterir. Hesaplanması şu şekildedir

$$IDF(t,D) = \log(\text{Toplam belge sayısı} / \text{Kelimenin geçtiği belge sayısı})$$

TF-IDF: TF ve IDF değerlerinin çarpımıyla elde edilir ve bir kelimenin bir belgede ne kadar önemli olduğunu belirtir.

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

Avantajları ve Dezavantajları

Avantajları:

Basit ve anlaşılır bir yöntemdir.

Belgedeki önemli kelimeleri belirlemede etkilidir.

Nadir ve önemli kelimelere daha fazla ağırlık verir.

Dezavantajları:

Kelimeler arasındaki anlamsal ilişkileri dikkate almaz.

Uzun belgelerde kelime sıklığındaki varyasyonlar büyük farklılıklar gösterebilir.

1.2.3.2 Word Embedding Yöntemi

Word2Vec kelimeleri düşük boyutlu vektörler olarak temsil eden bir tekniktir. Bu vektörler, kelimelerin anlamlarını ve kullanım bağlamlarını yansıtır. İki ana modeli vardır: Continuous Bag of Words (CBOW) ve Skip-Gram.

CBOW: Kelimenin etrafındaki kelimelerden kelimenin kendisini tahmin etmeye çalışır.

Skip-Gram: Kelimenin kendisinden etrafındaki kelimeleri tahmin etmeye çalışır.

GloVe:Global Vectors for Word Representation (GloVe), kelime vektörlerini oluşturmak için kelime eşleşmelerine dayalı global istatistikleri kullanır. Kelimeler arasındaki anlamsal ilişkileri yakalamada etkilidir.

FastText:Kelimeleri karakter n-gramlarına bölen bir word embedding yöntemidir. Bu, özellikle nadir ve türemiş kelimelerin temsil edilmesinde faydalıdır. Kelimeler arasındaki anlamsal ve biçimsel benzerlikleri yakalamada etkilidir.

Avantajları ve Dezavantajları

Avantajları:

Kelimeler arasındaki anlamsal ilişkileri ve bağlamları öğrenebilirler.

Düşük boyutlu vektörler sayesinde hesaplama açısından verimlidirler.

Yeni ve nadir kelimeleri bile temsil edebilirler (özellikle FastText).

Dezavantajları:

Büyük veri kümeleri ve yüksek hesaplama gücü gerektirir.

Eğitim süreci zaman alıcı olabilir.

1.3 Hipotez

Yapılan araştırmalar sonucunda, geleneksel öznitelik çıkarım yöntemleri yerine popüler dil modelleri kullanılarak CV'lerden öznitelik çıkarımının yapılmasının ve buna göre uygun CV-startup eşleştirmesinin daha başarılı olacağı düşünülmektedir. Geleneksel yöntemler genellikle belirli anahtar kelimelere ve önceden tanımlanmış kurallara dayanırken, bu yöntemler sınırlı esneklik ve doğruluk sunmaktadır. Buna

karşın, popüler dil modelleri olan BERT (Bidirectional Encoder Representations from Transformers) ve GPT-3 (Generative Pre-trained Transformer 3) gibi modeller, metinleri daha derinlemesine analiz edebilmekte ve bağlamsal anlamları daha iyi yakalayabilmektedir. Tabi eğitildikleri veri boyutunun gücünde yatsınamaz.

Örneğin, geleneksel yöntemler bir CV'deki "Java" kelimesini yalnızca programlama dili olarak tanımlarken, BERT[6] ve GPT-3[7] gibi güçlü dil modelleri bu kelimenin kullanıldığı bağlama göre doğru anlamını çıkarabilir. Bu, adayların yetenek ve deneyimlerini daha doğru bir şekilde değerlendirerek, uygun CV-startup eşleştirmelerinin yapılmasını sağlar. Bu nedenle, popüler dil modellerinin kullanımıyla CV'lerden öznitelik çıkarımının, geleneksel yöntemlere kıyasla daha başarılı sonuçlar vereceği düşünülmektedir.

2.YÖNTEM

Araştırma sürecinde kullanılan yöntemler ve teknikler detaylı olarak açıklanacaktır. Veri toplama yöntemleri, analizleri ve modelin geliştirilme aşamaları bu bölümde yer alacaktır.

2.1 Veri Toplama Yöntemleri

Proje kapsamında kullanılan veriler, iki ana kaynaktan elde edilmiştir: startup bilgileri ve CV verileri.

2.1.1 Startup verilerinin toplanması

Startup verileri,[8] sitesinden elde edilmiştir. Bu verilerin toplanması için Python programlama dilinde Scrapy adlı web kazıma (web scraping) kütüphanesi kullanılmıştır. Scrapy ile tüm linkler dolaşarak startup pozisyonları uygun şekilde JSON formatına kaydedilmiştir. Bu süreç, aşağıdaki adımları içermektedir.

Bu yapı, araştırma sürecinin sistematik bir şekilde ele alınmasını sağlamaktadır. Her adım detaylı olarak açıklanmış ve veri toplama yöntemleri ile veri kazıma süreçleri ayrıntılı şekilde ele alınmıştır. Bu şekilde, EU-Startups.com sitesinden başlangıç

şirketlerine ait verilerin nasıl elde edildiği ve bu verilerin nasıl işlendiği net bir şekilde ortaya konmuştur.

Toplanan veriler 330 adet startup bilgisi içermektedir.[9]

2.1.1.1 Site haritasının çıkarılması:

Eu-Startups sitesinin sayfa yapısı analiz edilerek, startup bilgilerini içeren sayfaların URL'leri belirlenmiştir. Bu adımda, ana dizin sayfasında yer alan kategori bağlantıları ve bu kategorilere ait sayfalar incelenmiştir.

2.1.1.2 Scrapy İle Veri Kazıma:

Belirlenen URL'lerden Scrapy kullanılarak veriler çekilmiş ve bu veriler JSON formatında kaydedilmiştir. Scrapy çerçevesi kullanılarak, her kategori sayfasına erişilmiş ve bu sayfalardaki listeleme öğeleri parse edilerek iş ismi, kategori, konum, etiketler ve kuruluş tarihi gibi bilgiler elde edilmiştir.

2.1.2 CV verilerinin toplanması

CV verileri, belirli bir formatta düzenlenmiş ve JSON dosyalarına dönüştürülmüştür. Bu verilerin toplanması ve işlenmesi süreci şu adımları içermektedir

Kullanılan veri [10] 50 adet orijinal json dosyasının üzerine 105 adet yapay zeka ile oluşturulmuş veri eklenerek 155 adet cv verisi içeren json dosyası oluşturulmuştur.

Bu yapı, araştırma sürecinin sistematik bir şekilde ele alınmasını sağlamaktadır. Her adım detaylı olarak açıklanmış ve veri toplama yöntemleri ile veri kazıma süreçleri ayrıntılı şekilde ele alınmıştır. Bu şekilde, hem EU-Startups.com sitesinden başlangıç şirketlerine ait verilerin hem de çeşitli kaynaklardan elde edilen CV verilerinin nasıl elde edildiği ve işlendiği net bir şekilde ortaya konmuştur.

2.1.2.1 Veri formatı

CV verileri, bireylerin adı, çalıştığı sektör, sahip olduğu yetkinlikler ve odaklandığı alanları içeren JSON formatında düzenlenmiştir. Her bir birey için bu bilgiler ayrı ayrı kaydedilmiştir.

2.1.2.2. Veri toplama ve dönüştürme

CV verileri, çeşitli kaynaklardan manuel veya otomatik olarak toplanmış ve belirli bir formata dönüştürülmüştür. Veri toplama işlemi sırasında, her bireyin sektörü ve yetkinlikleri dikkate alınarak veriler düzenlenmiştir.

2.1.2.3 Veri zenginleştirme (Data augmentation)

Elde edilen CV verileri, daha anlamlı ve kapsamlı analizler yapabilmek için veri zenginleştirme yöntemleri ile işlenmiştir. Bu süreçte, ek bilgilerin eklenmesi, veri çeşitlendirmesi ve eksik verilerin tamamlanması gibi adımlar izlenmiştir. 55 adet elde edilen cvler 155'e çıkartılmıştır.

2.1.2.4 Veri Düzenleme

Toplanan ve zenginleştirilen CV verileri, düzenli bir JSON formatında olacak şekilde düzenlenmiştir.

2.1.3 Veri Temizleme ve Düzenleme:

Elde edilen veriler, gereksiz bilgilerden arındırılarak ve doğru formatta (json) olacak şekilde düzenlenmiştir. Bu adımda, ham veriler içerisindeki eksik, hatalı veya tekrarlayan bilgiler tespit edilerek temizlenmiş ve yapılandırılmış bir formata getirilmiştir.

2.1.4 Veri Analizi Ve Görselleştirme

Veri toplama sürecinin ardından elde edilen veriler analiz edilmiş ve model geliştirme aşamasına geçilmiştir.

2.2 Model Eğitimi Ve Değerlendirme Yöntemleri

BERT modeli kullanılarak sistemin nasıl eğitildiği ve değerlendirildiği bu bölümde ele alınacaktır. Modelin performansı ve doğruluğu hakkında bilgi verilecek ve yapılan testlerin sonuçları sunulacaktır.

2.2.1 BERT Modeli Kullanımı

Startup'lar ve mentor CV'leri arasındaki uyumu belirlemek için BERT (Bidirectional Encoder Representations from Transformers) modeli kullanılmıştır. BERT modeli ve tokenizer'ı, metin verilerini özellik vektörlerine dönüştürmek için kullanılmıştır. Bu süreç şu adımları içermektedir.

BERT modeli ve tokenizer, Transformers kütüphanesi kullanılarak yüklenmiştir.

BERT modeli, metinlerin anlamını anlamak için kullanılır. Bu model, büyük miktarda metin verisinde önceden eğitilmiştir ve genel bir dil modeli öğrenir.

Tokenizer, metni parçalara (token'lara) ayırır. Her bir token, BERT modelinin girişine uygun bir şekilde sayısal temsili olacaktır.

Tokenleştirme, kelimeleri sayısal verilere dönüştürmek için kullanılan bir işlemdir. Bu, metin verilerini işlemek için yaygın olarak kullanılan bir tekniktir. Tokenleştirme işlemi, metni anlamak için kullanılan bir model için girdi olarak kullanılmadan önce gerçekleştirilir.

2.2.1.1 Metin Verilerinin İşlenmesi

Tokenizer, metni belirli bir sözcük dağarcığına dayanarak tokenlara ayırır.

Her bir kelime veya alt cümle (subword), bir token olarak temsil edilir.

Startup ve CV metinleri, BERT modeli ile işlenerek özellik vektörlerine dönüştürülmüştür.

Örnek bir cv json :

```
{
  "name": "Mehmet Aydın",
  "industry": "Manufacturing, Automotive, Robotics",
  "competence": [
    "Product Development",
    "Supply Chain Management",
    "Quality Control",
    "Operations Optimization"
  ],
  "focus": [
    "Robotics",
    "Automotive"
  ]
},
```

Tokenlara ayrılması :

```
{ "name": ["[CLS]", "mehmet", "aydın", "[SEP]"],
  "industry": ["[CLS]", "manufacturing", ",", "automotive", ",", "robotics", "[SEP]"],
```

```
"competence":  ["[CLS]",  "product",  "development",  "supply",  "chain",  
"management", "quality", "control", "operations", "optimization", "[SEP]"],  
"focus": ["[CLS]", "robotics", ",", "automotive", "[SEP]"]  }
```

BERT modeli için her girişin başlangıcını belirlemek için de CLS tokeni kullanılır. Bu token, modelin başlangıcını ve sınıflandırma görevlerinde kullanılan özet bilgisini belirtir.

[SEP] tokeni metinler arasındaki ayrımı belirtmek için kullanılırken, CLS tokeni modelin başlangıcını belirtmek için kullanılır. Bu tokenler, BERT ve diğer benzer modellerin çoklu metin girişlerini doğru bir şekilde işlemesini sağlar.

Her bir token, bir sayısal vektörle temsil edilir. Bu vektörler, BERT modelindeki katmanlardan birine (genellikle kelime dağarcığının boyutu kadar olan ilk katmana) atanır.

```
{      "name": [  
[0.1, -0.2, 0.5, ...], // "mehmet" tokenının temsili  
[0.3, 0.2, -0.4, ...], // "aydın" tokenının temsili    ...],  
"industry": [  
[0.6, -0.1, 0.2, ...], // "manufacturing" tokenının temsili  
[0.8, -0.3, 0.7, ...], // "," tokenının temsili  
[0.4, -0.5, 0.9, ...], // "automotive" tokenının temsili    ...],...}
```

Bu sayısal temsiller, her bir tokenin anlamını yakalamak için önceden eğitilmiş bir dil modelinden gelir. Bert modelinin [11] bir cv ve startup verisi üzerinde tokenlarının görselleştirilmesi.

2.2.1.2 Özellik vektörlerinin işlenmesi hesaplanması:

Metin verilerinden elde edilen özellik vektörleri, BERT modeli kullanılarak hesaplanmıştır. CV JSON metni, Tensor ve Sayısal Temsil olarak sunmak beni oldukça etkiliyor. Tensor formatı, metin verilerini işlemek için oldukça yaygın bir şekil ve boyutlarını korurken, sayısal temsil metnin anlamını yakalamak için kullanılan bir derin öğrenme modeli tarafından oluşturulan sayısal vektörlerdir. Bu temsil, metinler arasında benzerlikleri ve farklılıkları belirlemeye yardımcı olur . Bu bilgileri kullanarak, metnin içeriğini anlamak ve proje gereksinimlerini karşılamaya yardımcı olur. "Anahtar: name" altında, "name" anahtar kelimesinin tokenleştirilmiş hali ve onun sayısal temsili bulunmaktadır. Benzer şekilde, "Anahtar: industry", "Anahtar: competence" ve "Anahtar: focus" altında da aynı yapı gözlemlenmektedir.

Anahtar: name

Tokenler:

```
{'input_ids': tensor([[ 101, 2033, 14227, 3388, 1037, 25688, 11722, 2078, 102]]),  
'token_type_ids': tensor([[0, 0, 0, 0, 0, 0, 0, 0, 0]]), 'attention_mask': tensor([[1, 1, 1,  
1, 1, 1, 1, 1, 1]])}
```

Sayısal Temsil:

```
tensor([[ -0.3733, 0.6114, -0.2928, ..., -0.2387, 0.4976, -0.1394],  
[ 0.2043, -0.5555, -1.4085, ..., -0.5304, 0.3348, 0.3770],  
[ 0.5222, 0.1803, -0.9473, ..., -0.9417, -0.0476, -0.8228], ...,  
[ 0.3169, -0.2497, -0.7269, ..., -1.3072, 0.0249, -0.4597],  
[ 0.4086, -0.0642, -1.0634, ..., -0.7716, -0.2047, 0.1960],  
[ 0.6366, 0.0127, -0.3634, ..., 0.2735, -0.6065, -0.4111]],
```

grad_fn=<SelectBackward0>)

Sayısal temsil , belirli bir kelimenin veya kelime öbeğinin (anahtar kelimelerin) derin öğrenme modeli tarafından anlaşılabilir bir biçimde temsil edildiği bir vektördür. Bu vektörler, modelin kelimeler arasındaki ilişkileri öğrenmesine ve metin verilerini işlemesine olanak tanır.

input_ids: Bu, metnin tokenleştirilmiş hali için bir dizi token IDs'idir. Tokenleştirme, metni belirli bir kelime dağarcığına (sözlüğe) dayalı olarak sayısal bir formata dönüştürme işlemidir. Her bir kelimeye (veya kelime öbeğine) bir benzersiz bir sayı atanır ve bu sayılar input_ids olarak temsil edilir.

token_type_ids: Bu, farklı cümlelerin veya metin parçalarının tokenleştirilmesi sırasında hangi cümlenin hangi parçasına ait olduğunu belirtmek için kullanılır.

attention_mask: Bu, modelin dikkat etmesi gereken yerleri belirlemek için kullanılır.

Sayısal Temsil: Bu, tokenlerin (kelimelerin veya kelime öbeklerinin) sayısal olarak temsil edildiği bir matris veya vektördür.

Bert modelinin birer cv ve startup örnekleminin sayısal vektör hali. [11]

2.2.2 Modelin uygulanması

Son gizli katmanın çıkışı, BERT modelinin en kapsamlı ve genel temsili sağladığı katmandır. Bu nedenle, bir metnin anlamını en iyi şekilde yansıtan özellik vektörlerini elde etmek için genellikle son gizli katmanın çıkışı tercih edilir. İşte bu tercihin arkasındaki bazı nedenler

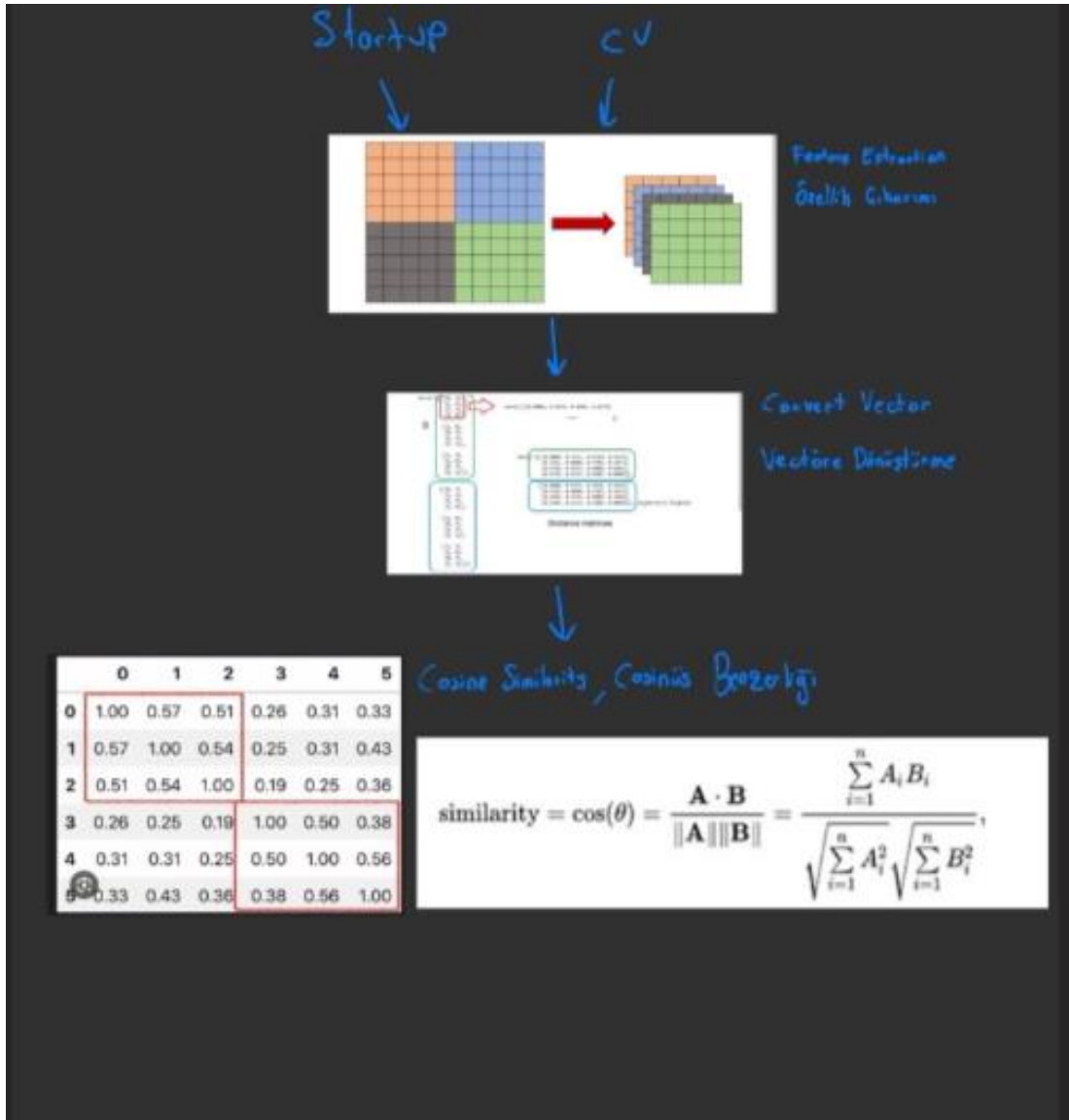
2.2.2.1 BERT Modeli ve Tokenizer'ın Yüklenmesi:

BERT modeli (BertModel) ve tokenizer (BertTokenizer) yüklenir. Bu model, önceden eğitilmiş bir dil modelidir ve metin verilerini işlemek için kullanılır.

Model 1: `outputs.last_hidden_state` kullanır, yani BERT'in son katmanından çıkarım yapar. Son katman, metnin genel anlamını en iyi şekilde temsil edebilir ve modelin öğrenme sürecindeki en gelişmiş özellikleri kullanır. Son katman, bazı durumlarda fazla soyut olabilir ve çok spesifik özellikleri yakalamakta zorlanabilir. Model 2'de ise `outputs.hidden_states[layer_num]` kullanır ve belirli bir katmandan (11. katman) çıkarım yapar. Orta katmanlardan birinin kullanılması, daha spesifik veya ayrıntılı özelliklerin yakalanmasına yardımcı olabilir. Bazen orta katmanlar, hem düşük seviye hem de yüksek seviye özelliklerin dengeli bir karışımını sağlar. Orta katman, metnin genel anlamını tam olarak temsil edemeyebilir ve modelin genel performansını düşürebilir.

2.2.2.2 Özellik Çıkarma Fonksiyonları

`get_embedding` ve `get_embedding_from_layer` fonksiyonları metinden gömülü bir vektör elde etmek için kullanılır. İlk fonksiyon, BERT modelinin sadece son gizli katmanının çıkışlarının ortalamasını alırken, ikincisi belirtilen bir katmanın çıkışlarını kullanır. **Şekil 2.2.2.2.1** de özellik çıkartmadan sonuç bölümüne kadarki model içeriklerinin temsili görselleri yer almaktadır .



Şekil 2.2.2.2.1 Proje Modelinin Görseleştirilmesi

2.2.2.3 Metinlerin İşlenmesi

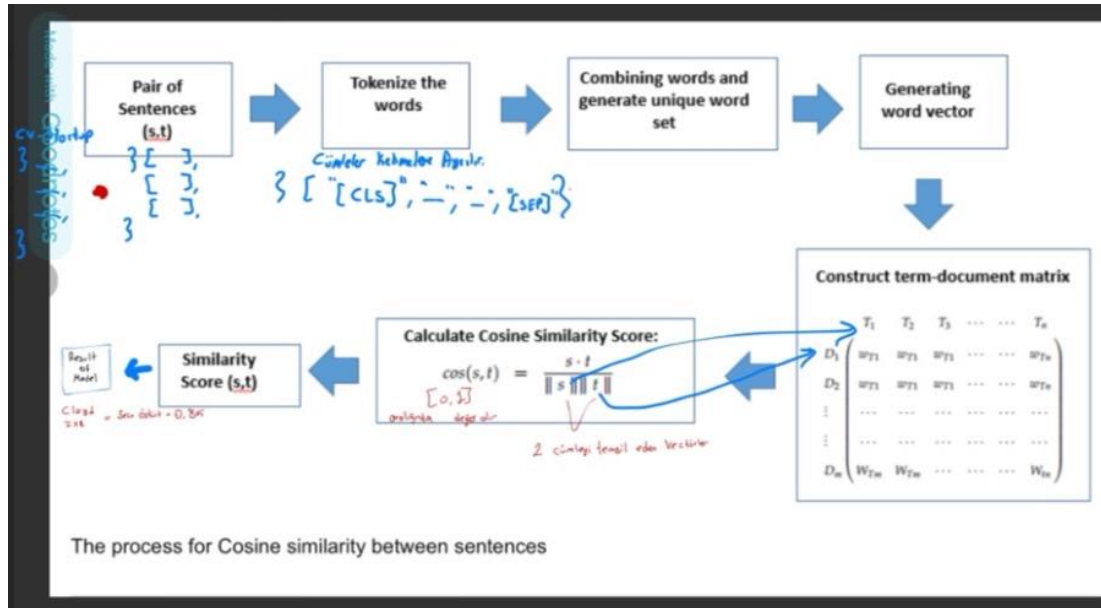
get_text_from_cv ve get_text_from_startup fonksiyonları, CV ve başlangıç verilerinden işlenecek metinleri oluşturur.

2.2.2.4 Gömülü Vektörlerin Oluşturulması:

CV ve başlangıç verileri, get_embedding veya get_embedding_from_layer fonksiyonları kullanılarak gömülü vektörlere dönüştürülür.

2.2.2.5 Benzerlik Hesaplama

cosine_similarity fonksiyonu, iki vektör arasındaki kordinat benzerliğini hesaplar. Bu, vektörler arasındaki açıyı ölçerek benzerlik skorunu belirler.



Şekil 2.2.2.5.1 Modelin Diagramla Gösterilmesi

Modele ait işlemler Şekil 2.2.2.5.1 de sırasıyla singleleştirilmiştir. Bu diagramdaki işlemler modele ait anlatıların görsel ve birlikteliğini sunmaktadır.

2.2.2.6 En Uygun CV'nin Bulunması

find_best_cv_for_startup fonksiyonu, her başlangıç için en uygun CV'yi bulur. Bunun için başlangıç gömülü vektörü ile CV gömülü vektörleri arasındaki benzerlik skorunu kullanır. Model 1'de get_embedding fonksiyonu, modelin son gizli katmanının çıktısını kullanır. Bu katman, modelin öğrenme sürecinde edindiği en üst düzeydeki özellikleri kodlar. Model 2'de ise get_embedding_from_layer fonksiyonu, modelin herhangi bir gizli katmanından özellik çıkarma imkanı verir. Farklı katmanlar, metnin farklı düzeydeki özelliklerini yakalayabilir. layer_num parametresini kullanarak, hangi katmandan özellik çıkaracağı seçilebilir. Bu özellikler ise get_text_from_cv ve get_text_from_startup fonksiyonlarından, CV'den ve startup bilgisinden ilgili metinleri toplayarak get_embedding_from_layer fonksiyonuna girdi olarak gönderir. BERT modelinin output_hidden_states=True parametresi ile her katmandaki gizli durumları elde edebiliriz. Seçilen özellikler, CV'ler ve startup'lar arasındaki uyumu değerlendirmek için en anlamlı ve bilgilendirici verileri temsil edecek şekilde seçilir. Seçimler, alaka, bilgi kapsamı ve veri erişimi gibi faktörlere dayanmaktadır. Seçilen özellikler, bir startup'ın bir adayda aradığı becerileri, deneyimleri ve ilgi

alanlarını kapsamlı bir şekilde temsil eder. Ara katmanlar, modelin öğrenme sürecinde farklı seviyelerde özellikler belirli görevlerde avantaj sağlayabilir. Tüm katmanların gizli durumlarını çıkarmak ek hesaplama maliyeti getirebilir. Model -1 ve model -2 karşılaştırıldığında **Çizelge 2.2.2.6.1**, **Çizelge 2.2.2.6.2**'de görüldüğü üzere model -2 daha yüksek skorlara sahiptir ve model 1 tahmininde yer almayan model 2 de en yüksek skora sahip olan Cloyd IXR ın Selin Öztürk ile uyumu incelendiğinde model -2 nin başarımının daha yüksecek ve doğru olduğunu söyleyebiliriz.

Çizelge 2.2.2.6.1 Model -1 in Selin Öztürk cv ile startup eşeřtirmesi skorları

Startup	Skor
Takeback Studios	0.56
Tenn.gg	0.57
Sheetsway	0.66
BiologyBuddy.ai	0.67
Bedbible.com	0.63
Asystly	0.72
CareerAddict	0.73
BlueHat	0.75

Çizelge 2.2.2.6.2 Model -2 nin Selin Öztürk cv ile startup eşeřtirmesi skorları

Startup	Skor
Tenn.gg	0.66
PlayerGo	0.69
BUDFARMA	0.79
Krater.ai	0.82
BlueHat	0.82
Cloyd IXR	0.84

Sıra	Startup	En Uygun CV	Skor	Tahmini Puan
1	Cloyd IXR	Selin Öztürk	0.84	100
2	Krater.ai	Selin Öztürk	0.82	90
3	BlueHat	Selin Öztürk	0.82	85
4	PlayerGo	Selin Öztürk	0.69	80

Şekil 2.2.2.6.1 ChatGpt nin ilgili cv nin firmalara olan uyumunun göstergesi

Cloyd IXR firmasına ait json veri:

```
{ "title": "Lithuania", "business_name": "Cloyd IXR", "category": "Lithuania",
"based_in": "Vilnius", "tags": "technology, training, employee training, virtual
reality", "founded": "2004"},
```

Selin Öztürk cv: {

```
"name": "Selin Öztürk",
"industry": "Education, Educational Technology",
"competence": [
"E-Learning Development",
"Learning Management Systems",
"Instructional Technology",
"Educational Games"
],
"focus": [
"Education",
"Educational Technology"]},
```

Selin Öztürk'ün CV'si Cloyd IXR şirketine oldukça uygun görünüyor. Cloyd IXR, teknoloji ve sanal gerçeklik odaklı çalışan bir şirket olduğu için, Selin'in eğitim teknolojileri ve e-öğrenme geliştirme konusundaki yetkinlikleri bu şirkette değerli olabilir. E-Learning Development: Cloyd IXR'nin sanal gerçeklik ve eğitimle ilgili teknolojilerle uğraştığı düşünülürse, e-öğrenme geliştirme konusunda deneyimli biri, şirketin eğitim içerikleri oluşturma ve geliştirme süreçlerine katkıda bulunabilir. Learning Management Systems: Sanal gerçeklik tabanlı eğitimlerin yönetimi ve dağıtımını için LMS sistemleri önemli olabilir. Selin'in bu konudaki deneyimi, Cloyd IXR'nin mevcut veya yeni LMS çözümlerinin geliştirilmesinde ve yönetiminde yardımcı olabilir. Instructional Technology: Eğitim teknolojileri konusundaki bilgi birikimi, Cloyd IXR'nin ürünlerinin eğitimsel etkinliğini artırmaya yardımcı olabilir.

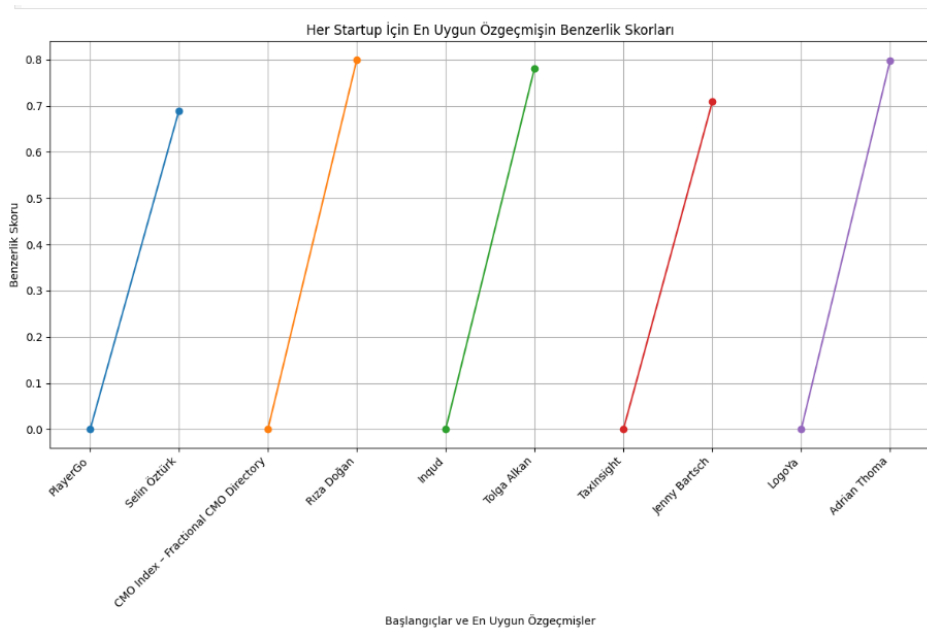
Educational Games: Sanal gerçeklik eğitimleri genellikle oyunlaştırma unsurları içerir. Selin'in eğitim oyunları konusundaki yetkinlikleri, bu tür projelerde yaratıcı ve etkili çözümler sunmasına olanak tanır.

Selin'in odaklandığı alanlar olan eğitim ve eğitim teknolojileri, Cloyd IXR'nin faaliyet alanları **Çizelge 2.2.2.6.2, Şekil 2.2.2.6.1** 'de görüldüğü üzere doğrudan örtüşmektedir ve model ilgili alanlar arasında bağlantı kurabilmiş. Bu nedenle, Selin Öztürk'ün CV'si Cloyd IXR'ye başvuru için oldukça uygun görünüyor.

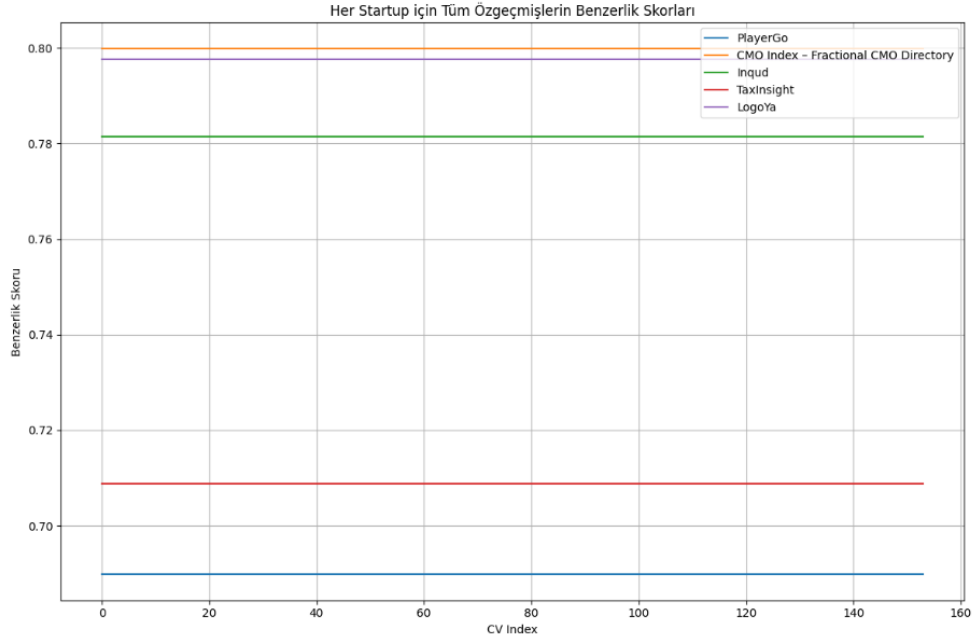
2.2.2.7 Sonuçların Görselleştirilmesi

İki farklı görselleştirme yöntemi kullanılır. İlk yöntem, her başlangıç için en uygun özgeçmişin adını ve buna ait benzerlik skorunu yatay çubuk grafiklerle gösterir.

İkinci yöntem, her başlangıç için tüm özgeçmişlerin uygunluk skorlarını bir çizgi grafiğiyle gösterir.



Şekil 2.2.2.7.1 Uygun CV-Mentör eşleştirmesi-1



Şekil 2.2.2.7.2 Uygun CV-Mentör eşleştirmesi-2

2.2.3 Modelin Testi

```

996      {
997          "name": "Selin Öztürk",
998          "industry": "Education, Educational Technology",
999          "competence": [
1000              "E-Learning Development",
1001              "Learning Management Systems",
1002              "Instructional Technology",
1003              "Educational Games"
1004          ],
1005          "focus": [
1006              "Education",
1007              "Educational Technology"
1008          ]
1009      },

```

Şekil 2.2.3.1 Selin Öztürk CV

Selin Öztürk cvsinin 4 adet firma için uygun olduğu sonucuna ulaştık bu firmaları ve Selinin cv sini ChatGPT ye sunduk ve 100 üzerinden puanlamasını istedik.Tablodaki tahmini puan sütununda değerleri yer almakta ve skor sütununda da modele ait yüzde skoru yer almakta. İki sonuç kıyaslandığında başarılı bir çıktı olduğunu söyleyebiliriz.

Firma Adı	Çalışma Alanları	Selin Öztürk'ün Alanları
Cloyd IXR	Technology, Training, Employee Training, Virtual Reality	E-Learning Development, Learning Management Systems, Instructional Technology, Educational Games
Krater.ai	Artificial Intelligence, AI, Copywriting, Productivity	E-Learning Development, Learning Management Systems, Instructional Technology, Educational Games
BlueHat	Apparel, Retail, Artificial Intelligence, Machine Learning	E-Learning Development, Learning Management Systems, Instructional Technology, Educational Games
PlayerGo	Games, Hardware, Controllers, Gaming, Videogames	E-Learning Development, Learning Management Systems, Instructional Technology, Educational Games

Şekil 2.2.3.2 Selin Öztürk cvsine uygun firmalar çalışma alanları

Sıra	Startup	En Uygun CV	Skor	Tahmini Puan
1	Cloyd IXR	Selin Öztürk	0.84	100
2	Krater.ai	Selin Öztürk	0.82	90
3	BlueHat	Selin Öztürk	0.82	85
4	PlayerGo	Selin Öztürk	0.69	80

Şekil 2.2.3.3 Selin Öztürk cvsine uygun firmalar ve skorları

3.SONUÇ

Yapılan araştırmalar sonucunda, eğitilen veri seti alanında modelin tahminleri, startup'lardaki ilgili bölümlerin çalışma gereksinimlerini doğru bir şekilde belirlemektedir. Aynı şekilde, CV'lerdeki çıkarımlar doğrultusunda en uygun CV ve startup eşleştirmesi yapılmaktadır. Bu model sayesinde, işe alımlarda birçok CV arasından elemeler yapılabilir ve bu sayede işe alım süreçleri hızlı ve etkili bir şekilde sonuçlandırılabilir.

3.1Çalışmanın Uygulama Alanı

İnsan Kaynakları ve İşe Alım Özgeçmiş ve İş İlanı Eşleştirme: İş ilanları ve özgeçmişler arasındaki en uygun eşleşmeyi bulmak için kullanılabilir. Bu sayede işe alım süreçleri daha hızlı ve etkili hale gelir. Başvuran adayları uygunluk skorlarına göre sıralayarak insan kaynakları departmanına yardımcı olabilir.Pdf halindeki karışık

olaran bir cv den önemli bilgilerin alınabilmesi için dokument question answering ile gerekli bilgiler alınabilir hugging face deki bu model kullanılabilir [12].Düz yazı şeklinde olan cvler içinde [13] modeli kullanılarak cvlerden özellik çıkarımı yapılır ve uygun veri seti oluşturularak oluşturulan bu model ile başarılı bir cv-startup eşleştirmesi yapılır.işe alım değerlendirmelerinde , iş başvurularında ilgili ilana uygunluğun bulunmasında , yüzlerce öğrencinin katıldığı startup etkinliklerinde ilgili özelliklere sahip cvlerin bulunması aşamasında kullanılması yararlı olacaktır.

4.ÖNERİLER

Daha geniş bir veri seti kullanılarak uygun formata dönüştürülen veriler toplanır ve modele sunulursa, projenin kapsamı genişletilebilir ve etkinliği artırılabilir. Örneğin, adayların sosyal medya hesapları, özellikle LinkedIn gibi platformlar incelenerek yapay zeka modelleriyle öznitelikler toplanabilir. Bu şekilde, sosyal medya hesaplarından elde edilen verilerle adayların sosyal ilişkileri, hobileri ve yaşam tarzlarına dair çıkarımlar yapılabilir. Bu çıkarımlar, modele entegre edilerek daha kapsamlı bir model geliştirilebilir. Sonuç olarak, bu modelin eğitilmesiyle insan kaynakları süreçlerinde önemli ölçüde iyileştirmeler sağlanabilir ve hatta bu alandaki manuel işlerin azaltılması veya tamamen ortadan kaldırılması mümkün olabilir. Sosyal medya verilerindeki çıkarımlar, kişinin ekip çalışmalarına olan yatkınlığı, arkadaşlık ilişkileri, takip ettiği örnek aldığı insanlar gibi faktörleri içerebilir. Bu çıkarımlar, bir kişinin lider ruhlu olup olmadığına, sorunlar karşısındaki tavrına ve hatta hayatında etkili olan kişilerin sayısına kadar birçok önemli özelliği yansıtabilir. Böylece, bir kişinin CV'sinde yer alan sertifikalar ve eğitimlerden daha fazla anlam taşıyabilir. Özellikle yeni mezun bireyler için, işe girecekleri firma ve çalışma ortamı, teknik niteliklerden çok daha önemli olabilir. Teknik beceriler birkaç yıl içinde kazanılabilirken, yaşam tarzını değiştirmek ve kişisel özellikleri geliştirmek daha uzun sürebilir. Ancak, bu tür sosyal medya verilerinin kullanımının etik yönleri tartışmalıdır. Adayların sosyal medya bilgilerinin bu şekilde kullanılmasına nasıl baktıkları ve bu verilerin gizliliği konuları, yapay zekanın etik alanında tartışılması gereken önemli konulardır.

5.KAYNAKÇA

- [1] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, 2017.
- [2] K. Cho and B. van Merriënboer, "C. G ulc ,ehre, D. Bahdanau, F. " Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014.
- [3] Pendyala, V., Atrey, N., Aggarwal, T., & Goyal, S. (2022)1. Artificial Intelligence Enabled, Social Media Leveraging Job Matching System for Employers and Applicants2. Proceedings - 2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems, ICMACC 2022. <https://doi.org/10.1109/ICMACC54824.2022.100933233>
- [4] Elisabeth K Kelan. Algorithmic inclusion: Shaping artificial intelligence in hiring. In *Academy of Management Proceedings*, volume 2021, page 11338. Academy of Management Briarcliff Manor, NY 10510, 2021.
- [5] Mona Sloane, Emanuel Moss, and Rumman Chowdhury. A silicon valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability. *Patterns*, 3(2):100425, 2022
- [6] Devlin, J., Chang, M. W., Lee, K., ve diğerleri. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
- [7] OpenAI. "GPT-3: Language Models Are Few-Shot Learners." <https://openai.com/blog/language-models/>.

- [8] EU-Startups. "EU-Startups Directory". Eriřim Tarihi: 30 Mayıs 2024. URL: <https://www.eu-startups.com/directory/>
- [9] Dil İşleme Destekli CV Mentor ve Startup Eşleřtirme Sistemi Projesi. GitHub, 2024. Eriřim adresi: <https://github.com/hmyrcmn/bitirme/blob/main/Startup.json>
- [10] Dil İşleme Destekli CV Mentor ve Startup Eşleřtirme Sistemi Projesi. GitHub, 2024. Eriřim adresi: <https://github.com/hmyrcmn/bitirme/blob/main/cv.json>
- [11] GitHub - hmyrcmn/bitirme: Bitirme Projesi. "token.txt" dosyasına eriřim. URL: <https://github.com/hmyrcmn/bitirme/blob/main/token.txt>
- [12] Hugging Face. (s.a.). LayoutLM Document QA. Hugging Face. Eriřim tarihi: 30 Mayıs 2024, URL: <https://huggingface.co/impira/layoutlm-document-qa>
- [13] Hugging Face. (s.a.). Models. Hugging Face. Eriřim tarihi: 30 Mayıs 2024, URL: https://huggingface.co/models?pipeline_tag=text2text-generation

ÖZGEÇMİŞ

TARANMIŞ
VESİKALIK
FOTOĞRAF

Ad-Soyad : HÜMEYRA ÇİMEN

Doğum Tarihi ve Yeri :01.04.2001 Yeşilhisar/Kayseri

E-posta :humeyracimen2001@gmail.com

BİTİRME ÇALIŞMASINDAN TÜRETİLEN MAKALE, BİLDİRİ VEYA SUNUMLAR:

- <https://github.com/hmyrcmn/bitirme/>