



**SAKARYA**  
ÜNİVERSİTESİ

**T.C**

**SAKARYA ÜNİVERSİTESİ**

**BİLGİSAYAR VE BİLİŞİM BİLİMLERİ FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ TASARIMI**

**Yapay Zekâ Destekli Yazı Dağıtım Sistemi**

**Danışman: Dr. Öğr.Üyesi Hüseyin Demirci**

**G191210060 - Alper Salih Aktaş**

**SAKARYA**

**Aralık ,2024**

# Proje Genel Tanımı

Bu proje, **yapay zekâ destekli bir yazı dağıtım sistemi** geliřtirmek amacıyla oluşturulmuřtur. Sistem, yüklenen PDF belgelerini analiz ederek içeriklerine göre uygun birimlere yönlendirir. Ayrıca, bu analiz sürecini açıklanabilir yapay zekâ ile destekleyerek sonuçların řeffaf bir řekilde raporlanmasını sağlar.

## Proje Hedefleri

- Yazıřmaların otomatik, hızlı ve doğru bir řekilde ilgili birimlere atanması.
- Analiz sonuçlarının önemli kelimeler ve güven skoru ile açıklanması.
- Sistemin kullanıcı dostu bir web arayüzü üzerinden çalışması.
- Sürekli veri güncellemeleriyle modelin performansının iyileřtirilmesi.
- Log kayıtlarının düzenli olarak tutulması ve gerektiğinde analiz edilmesi.

## Projenin Günümüz Problemlerine Katkısı

### Mevcut Problemler:

- ❖ **Belgelerin Manuel Yönlendirilmesi:** İşletmelerde belge yönlendirme süreçleri çoğunlukla manuel olarak yürütölmekte, bu da zaman kaybına ve hatalara neden olmaktadır.
- ❖ **Yanlış Yönlendirme Riski:** Yanlış birime yapılan yönlendirmeler iş süreçlerinde gecikmelere ve ek maliyetlere yol açmaktadır.
- ❖ **Veri Kayıtlarının Takip Edilememesi:** Belgelerin geçmiři ve dağıtım süreçleri genellikle yeterince kaydedilmemekte, bu durum denetim ve raporlama süreçlerini zorlařtırmaktadır.
- ❖ **Analiz Sürecinin Şeffaf Olmaması:** Karar mekanizmalarının řeffaf olmayıřı, güvenilirliři azaltmaktadır.

## Projenin Katkıları:

Otomasyon ve Hız: Belgelerin otomatik olarak analiz edilmesi ve yönlendirilmesi ile iş süreçleri hızlanacaktır.

Hata Oranının Azaltılması: Yanlış yönlendirmelerin önüne geçilerek süreçlerde doğruluk artırılacaktır.

Raporlama ve İzlenebilirlik: Analiz süreçleri kayıt altına alınarak şeffaflık ve izlenebilirlik sağlanacaktır.

İnsan Kaynaklarının Verimli Kullanımı: Belgelerin manuel kontrol ihtiyacı ortadan kaldırılarak çalışanların stratejik işlere odaklanması sağlanacaktır.

Mobil ve Web Uyumlu Çözüm: Her yerden erişilebilir bir platform sunularak esneklik sağlanacaktır.

Türkçe Dil Desteği: Yerel işletmelere uygun dil desteği ile geniş bir kullanım alanı oluşturulacaktır.

## Projede Kullanılacak Teknolojiler

Projeyi web tabanlı bir uygulama olarak Python programlama dili ve Flask framework'ü ile geliştireceğim.

- Makine öğrenimi işlemleri için **Scikit-learn** kütüphanesini kullanacağım.
- Veri işleme ve analiz için **Pandas** ve **NumPy** kütüphanelerinden faydalanacağım.
- Metin analizi ve özellik çıkarımı işlemleri için **TF-IDF** algoritması uygulanacak.
- PDF dosyalarının işlenmesi ve metin çıkarımı için **PyPDF2** kullanılacak.
- Web arayüzü tasarımı **HTML**, **CSS** ve **Flask'ın Jinja2** template sistemi ile oluşturulacak.
- Analiz sonuçlarının kaydedilmesi ve yönetimi için **CSV** dosya yapısı kullanılacak.

Bu teknolojiler, yapay zeka destekli yazı dağıtım sisteminin güvenilir, hızlı ve kullanıcı dostu bir yapıda çalışmasını sağlayacak.

# Veri İhtiyacı ve Sentetik Veri Çözümü

## 1.Gerçek Veri Bulma Zorlukları

**Veri Güvenliği ve Gizlilik:** Gerçek belgelerde hassas bilgiler yer aldığından bu verilerin paylaşımı ve kullanımı sınırlıdır.

**Veri Eksikliği:** Projenin gerektirdiği belge türleri ve senaryolar, mevcut veri kaynaklarında yeterli çeşitlilikte bulunamamıştır.

**Erişim Zorlukları:** Kurumsal veriye erişim karmaşık ve zaman alıcı bir süreçtir.

## 2.Neden Sentetik Veri?

**Esneklik:** İhtiyaç duyulan her kategoriye uygun içeriklerin oluşturulmasını sağlar.

**Çeşitlilik:** Rastgele değişkenlerle farklı senaryolar üretilerek model eğitimi için daha geniş bir veri havuzu sunulur.

**Zaman ve Kaynak Tasarrufu:** Gerçek veri toplama ve temizleme süreçleri atlanarak hızlı veri üretimi yapılır.

## Sentetik Veri Üretim Süreci

### ➤ Şablonların Oluşturulması:

Her departman için özgün belge şablonları hazırlandı. Bu şablonlar, belgelerin başlıklarını ve içeriklerini dinamik değişkenlerle tanımlar.

Örnek Şablon (İnsan Kaynakları):

```
Çalışan {name}, {reason} nedeniyle yıllık izin talebinde bulunmuştur.  
İzin süresi {date} - {date} tarihleri arasında kapsamaktadır. Talep edilen izin,  
{supervisor} tarafından onay beklemektedir.
```

### ➤ Rastgele Değişkenlerin Kullanımı:

Şablonlardaki dinamik alanlar için rastgele değerler üretilmiştir.

Kullanılan Değişkenler:

- Çalışan isimleri: **names**
- Şikayet türleri: **complaint\_types**
- Sözleşme türleri: **contract\_types**
- Tarihler: **dates**
- Yazılım adları: **software\_names**

Örnek Değişken Tanımları:

```
names = ["Ahmet Yılmaz", "Ayşe Demir", "Mehmet Kaya", "Fatma Özkan"]
reasons = ["yıllık izin", "sağlık sorunu", "ailevi nedenler"]
software_names = ["CRM Sistemi", "ERP Yazılımı", "Veritabanı Yönetim Yazılımı"]
```

### ➤ Belge Oluşturma:

#### ❖ Departman ve Belgelerin Tanımlanması:

9 departman için oluşturulan belge şablonları ve rastgele değişkenler kullanılarak 180 belge üretilmiştir.

#### ❖ CSV Dosyasına Kaydetme:

Oluşturulan veri, pandas kütüphanesi kullanılarak CSV formatında kaydedilmiştir:

```
df.to_csv("veri_seti_supervisor.csv", index=False, encoding="utf-8-sig")
```

## Veri Seti Özellikleri

| Özellik                | Detay                                  |
|------------------------|--|
| Departman Sayısı       | 9                                      |
| Belge Sayısı           | 360 (Her departman için 40 belge).     |
| Kullanılan Değişkenler | İsimler, tarihler, şikayet türleri vb. |
| Çıktı Dosyası          | veri_seti_supervisor.csv               |

## Örnek Bir Belge:

**Başlık:** Çalışan İzin Talebi

**İçerik:**

Çalışan Ahmet Yılmaz, sağlık sorunu nedeniyle yıllık izin talebinde bulunmuştur. İzin süresi 15 Ocak - 25 Ocak tarihleri arasını kapsamaktadır. Talep edilen izin, Elif Aydın tarafından onay beklemektedir.

## Veri Analizi ve Model Eğitimi

### Metin Önileme ve Özellik Çıkarma:

Metin önileme işlemleri, metinlerin analiz edilmesini kolaylaştırmak ve makine öğrenimi algoritmalarına uygun hale getirmek amacıyla gerçekleştirilmiştir. Bu süreçte aşağıdaki adımlar izlenmiştir:

- Metin Temizleme: Gereksiz kelimeler ve bağlaçlar (stopwords) çıkarıldı.
- Büyük/Küçük Harf Dönüşümü: Tüm metinler küçük harfe dönüştürülerek standart hale getirildi.
- Özellik Çıkarma:
  - TF-IDF (Term Frequency - Inverse Document Frequency): Belgelerde geçen kelimelerin önem derecesi belirlendi.
  - Kelimeler, makine öğrenimi modellerinin anlayabileceği sayısal değerlere dönüştürüldü.

### Model Eğitimi:

Metin sınıflandırma problemleri için aşağıdaki algoritmalar kullanılmıştır:

- Multinomial Naive Bayes (MultinomialNB): Sözcük sıklığına dayalı verilerle çalışarak hızlı ve etkili sonuçlar sağlaması nedeniyle tercih edilmiştir.
- Destek Vektör Makineleri (SVM)

- Lojistik Regresyon
- Rastgele Orman (Random Forest)
- En Yakın Komşu (KNN)

Veriler, %80 eğitim ve %20 test olarak ayrılmıştır. Model, eğitim seti ile öğrenme sürecini tamamladıktan sonra test seti üzerinde doğrulanmış ve performansı değerlendirilmiştir. Performans ölçümleri doğruluk (accuracy), hassasiyet (precision), duyarlılık (recall) ve F1-Skoru metrikleri üzerinden analiz edilmiştir.

Tablo 1 Performans sonuçları

| Algoritma                      | Doğruluk (%) | Hassasiyet | Duyarlılık | F1-Skoru |
|--------------------------------|--------------|------------|------------|----------|
| Multinomial Naive Bayes        | 97           | 0.96       | 0.97       | 0.96     |
| Lojistik Regresyon             | 95           | 0.94       | 0.95       | 0.94     |
| Destek Vektör Makineleri (SVM) | 96           | 0.95       | 0.96       | 0.95     |
| Rastgele Orman (Random Forest) | 93           | 0.92       | 0.93       | 0.92     |
| En Yakın Komşu (KNN)           | 90           | 0.89       | 0.90       | 0.89     |

Test sonuçlarına göre Multinomial Naive Bayes algoritması %97 doğruluk oranı ile en iyi performansı göstermiştir. Algoritmanın hızı ve doğruluk oranı, özellikle metin sınıflandırma işlemlerinde avantaj sağlamıştır. Bu yüzden projemizin eğitimini Multinomial Naive Bayes algoritması ile yapacağız.

## Açıklanabilir Yapay Zekâ (Explainable AI):

Öncelikle, PDF dosyası PyPDF2 kütüphanesi kullanılarak metne dönüştürülür. Elde edilen metin, TF-IDF ile işlenerek modele iletilir. Modelin kararlarını açıklamak için TF-IDF ile öne çıkan kelimeler

ıkarıldı. Modelin atama yaparken belirttiđi gven skoru hesaplandı ve raporlandı. Kullanıcıya neden bu birime ynlendirme yapıldıđı grselleřtirildi ve raporlandı.

```
Belge ID: 1
Atanan Birim: Mali İřler

Önemli Kelimeler:
bte, bte detayları, detayları, yılı, 2024 yılı,
000, ekte yer, ekte, btesinde oranında, revize

Gven Skoru: 0.0418
```

## Log Dosyası:

Sistem, her analiz iřleminden sonra kayıt tutmak amacıyla log dosyasını kontrol etmektedir. Log dosyası yoksa oluřturulmakta ve dosyanın ierisine belge ID'si, atanan birim, nemli kelimeler ve gven skoru gibi bilgiler kaydedilmektedir. Bylece analiz gemiři takip edilebilir ve daha sonra yapılacak analizlerde referans alınabilir.

|   | A        | B                  | C                              | D           |
|---|----------|--------------------|--------------------------------|-------------|
| 1 | Belge ID | Atanan Birim       | nemli Kelimeler               | Gven Skoru |
| 2 |          | 1 İnsan Kaynakları | alıřan, terfi, katılım sađlam | 0.548642552 |
| 3 |          |                    |                                |             |

## Flask Rotaları:

Flask framework' ile iki temel rota tanımlanmıřtır. İlk rota olan "/" ana sayfayı oluřturur ve kullanıcıların PDF dosyalarını ykleyebilmesi iin bir form sunar. İkinci rota olan "/analyze" ise analiz srecini yrtr. Yklenen dosyanın iřlenmesi, analizi ve sonuların sunulması bu rota zerinden gerekleřtirilir.

## Web Arayz Tasarımı:

Modern ve kullanıcı dostu bir arayz tasarlandı. Mobil uyum (responsive) sađlandı.



index.html; PDF yükleme formu ve result.html; analiz sonuçlarını gösteren sayfa olmak üzere iki sayfa olarak tanımlandı. CSS kodları ise ayrıca yazılarak temiz bir kod görünümü sağandı (static/css/style.css).

**index.html:** PDF yükleme formu.

### Yapay Zekâ Destekli Yazı Dağıtım Sistemi

PDF Dosyanızı Yükleyin:

Dosya Seç

Dosya seçilmedi

Analiz Et

© 2024 Yapay Zekâ Destekli Yazı Dağıtım Sistemi. Tüm Hakları Saklıdır.

index.html

**result.html:** Analiz sonuçlarını gösteren sayfa.

### Analiz Sonucu

**Atanan Birim:** İnsan Kaynakları

**Önemli Kelimeler:** ['çalışan', 'terfi', 'katılım sağlamıştır', 'sağlamıştır', 'olarak', 'iletmiştir', 'katılım', 'yılmaz', 'ahmet yılmaz', 'detayları içerikleri']

**Güven Skoru:** 0.2735720578442212

© 2024 Yapay Zekâ Destekli Yazı Dağıtım Sistemi. Tüm Hakları Saklıdır.

result.html

# Sonuç ve Genel Değerlendirme

Bu proje, yapay zekâ destekli belge analiz ve yönlendirme sistemleri arasında modern ve etkili bir çözüm sunmaktadır. Modelin yüksek doğruluğu, açıklanabilirlik desteği ve kullanıcı dostu tasarımı sayesinde işletmelerin belge yönetim süreçleri daha verimli hale getirilmiştir.

## Öne Çıkan Başarılar:

- Yüksek Performanslı Model:** Seçilen Multinomial Naive Bayes %97 doğruluk oranı ile güvenilir sonuçlar vermiştir.
- Gelişmiş Raporlama:** Sonuçların detaylı raporlanması ve kaydedilmesi.
- Kullanıcı Deneyimi:** Modern ve mobil uyumlu web arayüzü.
- Ölçeklenebilirlik:** Sistem, ileride ek özellikler eklenerek genişletilebilecek bir mimariye sahiptir.

## İşlevsel Kazanımlar:

- ❖ Otomatik belge analizi ve yönlendirme süreçleri sayesinde iş verimliliği artırılmıştır.
- ❖ Karar destek mekanizmaları ile hataların önüne geçilmiştir.
- ❖ Kolay yönetilebilir ve şeffaf bir kayıt sistemi oluşturulmuştur.