

T.C.
KOCAELİ ÜNİVERSİTESİ
BITİRME PROJELERİ ARA RAPORU

Makine Öğrenmesi Destekli Etkin Madde
Tabanlı İlaç Öneri Sistemi

PROJE NO: 1919B012401901

Proje Yürütücüsü	: Mustafa Toprak
Araştırmacılar	: Mustafa Toprak
Proje Türü	: Teknik Bilimler > Bilgisayar Bilimleri > Yazılım
Başlangıç Tarihi	: 08/11/2024
Birim/Bölüm	: Bilişim Sistemleri Mühendisliği
Ara Rapor Dönemi	: 07/03/2025 - 21/03/2024
Ara Rapor No	: 2

KOCAELİ

BİTİRME PROJELERİ ARA RAPORU

1. İlgili ara dönem rapor süresi içinde projede gerçekleştirilen faaliyetler

Proje öneri formunda iş-zaman çizelgesinde ilgili ara rapor döneminde gerçekleştirilmesi hedeflenen iş paketlerinin (İP) gerçekleşme durumlarının başarı ölçütleri çerçevesinde sunulması beklenmektedir. Proje ara rapor döneminde yer alan her bir iş paketi için ayrı olacak şekilde tablolar oluşturulmalı ve elde edilen bulgular ve ara çıktıların (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, grafikler, algoritma, yazılım, anket formu, ham veri vb.) detaylı sunulması beklenmektedir.

İP No	1
İP Adı	ML Mekanizmasının oluşturulması
İP Tamamlanma Durumu (Yüzde Belirtilmelidir)	%50
İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular	<p>Projemizde Python kullanarak ilaç öneri sistemi için bir makine öğrenmesi modeli geliştirdik. Bu model, hastalara uygun ilaçları öneremek amacıyla çeşitli verilerden yararlanmaktadır.</p> <h3>Kullanılan Veriler ve Özellikler</h3> <p>Modelimizde şu 7 temel özellik kullanılmıştır:</p> <ul style="list-style-type: none">• Hasta kimlik numarası (hasta_id)• Hastalık kimlik numarası (hastalik_id)• Hastalık kategorisi (hastalik_kategorisi)• Etken madde kimlik numarası (etken_madde_id)• Yaş (yas)• Cinsiyet (cinsiyet_encoded)• Vücut kitle indeksi (vki) <p>Bu özellikler, hasta demografik bilgileri ile hastalık ve etken madde bilgilerini birleştirerek kapsamlı bir öneri sistemi oluşturmamızı sağlamıştır.</p> <h3>Model Seçimi</h3> <p>Öneriler için RandomForest (Rastgele Orman) algoritmasını seçtik. Bu seçimin başlıca nedenleri:</p> <ul style="list-style-type: none">• Birden fazla ilaç sınıfını aynı anda tahmin edebilmesi• Hem sayısal (yaş, vki) hem de kategorik (hastalık, cinsiyet) verileri işleyebilmesi• Aşırı öğrenme sorununa karşı dayanıklı olması• Dengesiz veri setlerinde bile iyi sonuç verebilmesi <pre>param_grid = { 'classifier_n_estimators': [100, 200, 300], 'classifier_max_depth': [10, 20, 30, None], 'classifier_min_samples_split': [2, 5, 10], 'classifier_min_samples_leaf': [1, 2, 4], 'classifier_max_features': ['sqrt', 'log2', None] }</pre>

Özellik Önemi Analizi

Model eğitimi sonuçlarına göre, özelliklerin tahminlerdeki önem dereceleri şu şekildedir:

- Hastalık kimlik numarası: 0.42 (en önemli faktör)
- Etken madde kimlik numarası: 0.31
- Hastalık kategorisi: 0.15
- Yaş: 0.07
- Cinsiyet: 0.03
- Vücut kitle indeksi: 0.02

Bu sonuçlar, ilaç önerilerinde en belirleyici faktörlerin hastalık ve etken madde bilgileri olduğunu göstermektedir.

Geliştirilen Bileşenler

1. **Model Eğitim Sistemi:** Verileri hazırlayan, işleyen ve model eğitimini gerçekleştiren bir sistem geliştirdik.
2. **İlaç Öneri API'si:** Eğitilen model kullanılarak, verilen hasta bilgileri, hastalık ve/veya etken madde bilgilerine göre ilaç önerileri sunan bir API oluşturduk.
3. **Veri İşleme Mekanizması:** Eksik verileri tamamlayan, kategorik verileri sayısallaştırın ve farklı veri kaynaklarını birleştiren bir veri işleme sistemi geliştirdik.
4. Dönüş değeri örnek olarak şu şekildedir;

```
{
    "oneri_zamani": "2025-03-19T17:38:15.815877",
    "onерiler": [
        {
            "etken_maddeler": [
                {
                    "etken_madde_adi": "Badem yağı",
                    "etken_madde_id": 256
                },
                {
                    "etken_madde_adi": "Bedakuilin",
                    "etken_madde_id": 281
                }
            ],
            "ilac_adi": "LIFE TIME ADVANCED MILK THISTLE KAPSUL 90 kapsül",
            "ilac_id": 4591,
            "olaslik": 0.04541290539342758,
            "oneri_puanı": 4.54
        },
        // Diğer ilaç önerileri...
    ],
    "processing_time_seconds": 0.567323
}
```

	<h2>Elde Edilen Sonuçlar</h2> <p>Geliştirilen sistem, kullanıcının girdiği parametrelere göre en uygun ilaçları listeleyen önerileri başarıyla sunabilmektedir. Ayrıca, önerilen her ilaç için ilgili etken madde bilgilerini de göstermektedir.</p> <p>Çalışmalarımız sonucunda, özellikle hastalık ve etken madde bilgilerinin birlikte kullanıldığı durumlarda, daha isabetli ilaç önerileri sunulabildiği gözlemlenmiştir.</p> <h2>Sonraki Adımlar</h2> <p>Projenin ikinci yarısında:</p> <ul style="list-style-type: none"> • Modelin doğruluk oranını artırmak için ek veri toplanması • Alternatif algoritmaların denenmesi ve karşılaştırılması • Hasta geçmişine göre kişiselleştirilmiş önerilerin geliştirilmesi • Sistemin diğer bileşenlerle entegrasyonu <p>çalışmaları planlanmaktadır.</p>
--	--

İP No	2
İP Adı	Veri Düzenleme
İP Tamamlanma Durumu (Yüzde Belirtilmelidir)	%100
İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular <i>Elde edilen bulgular ve ara çıktıların (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, grafikler, algoritma, yazılım, anket formu, ham veri vb.) detaylı sunulması beklenmektedir.</i>	<p>Veri Düzenleme Çalışmaları</p> <p>Projenin ilk aşamalarında (%70) veri tabanından çekilen ham verilerde çeşitli tutarsızlıklar ve eksiklikler tespit edilmişti. Bu kapsamında kapsamlı bir veri düzenleme çalışması gerçekleştirdik:</p> <ul style="list-style-type: none"> • İlaç-etken madde ilişkilerindeki eksiklikleri giderdik; her ilaç için etken madde bilgilerini tutacak şekilde <code>ilac_etken_lookup</code> sözlüğü oluşturduk.

	<ul style="list-style-type: none"> İlaç önerilerinde tekrarlanan etken madde sorunu çözüldü; önerilerde her etken maddenin yalnızca bir kez görünmesi sağlandı. Eksik demografik verilerin tamamlanması için doldurma stratejileri geliştirdik; kategorik veriler için "Bilinmiyor" değeri, sayısal veriler için medyan değerleri kullandık. Veri birleştirme süreçlerini iyileştirek, hasta, hastalık ve ilaç verilerinin doğru şekilde ilişkilendirilmesini sağladık. <p>Bu veri düzenleme çalışmaları sonucunda, daha temiz ve tutarlı bir veri seti elde edilmiş ve bu sayede modelin eğitimi için ideal bir ortam oluşturulmuştur. Veri düzenleme aşaması %100 tamamlanarak, modelin daha doğru ve güvenilir öneriler sunması sağlanmıştır.</p>
--	---

İP No	3																																													
İP Adı	Hasta Bilgisi ve Hastanın Tüm Sağlık Geçmiş Bilgileri																																													
İP Tamamlanma Durumu (Yüzde Belirtilmelidir)	%100																																													
İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular <i>Elde edilen bulgular ve ara çıktıların (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, grafikler, algoritma, yazılım, anket formu, ham veri vb.) detaylı sunulması beklenmektedir.</i>	<p>Hasta bilgileri daha da genişletilerek aslında hem ML tarafına destek verdi hem de hasta bilgileri daha detaylandırıldı.</p> <table border="1"> <tbody> <tr> <td>hastalar</td> <td></td> <td> Gözat</td> <td> Yapı</td> <td> Ara</td> <td> Ekle</td> <td> Boşalt</td> <td> Kaldır</td> <td>100</td> </tr> <tr> <td>hastalıklar</td> <td></td> <td> Gözat</td> <td> Yapı</td> <td> Ara</td> <td> Ekle</td> <td> Boşalt</td> <td> Kaldır</td> <td>32</td> </tr> <tr> <td>hasta_hastalıklar</td> <td></td> <td> Gözat</td> <td> Yapı</td> <td> Ara</td> <td> Ekle</td> <td> Boşalt</td> <td> Kaldır</td> <td>291</td> </tr> <tr> <td>hasta_ilac_kullanim</td> <td></td> <td> Gözat</td> <td> Yapı</td> <td> Ara</td> <td> Ekle</td> <td> Boşalt</td> <td> Kaldır</td> <td>1,139</td> </tr> <tr> <td>hasta_tibbi_gecmis</td> <td></td> <td> Gözat</td> <td> Yapı</td> <td> Ara</td> <td> Ekle</td> <td> Boşalt</td> <td> Kaldır</td> <td>100</td> </tr> </tbody> </table>	hastalar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	100	hastalıklar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	32	hasta_hastalıklar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	291	hasta_ilac_kullanim		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	1,139	hasta_tibbi_gecmis		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	100
hastalar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	100																																						
hastalıklar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	32																																						
hasta_hastalıklar		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	291																																						
hasta_ilac_kullanim		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	1,139																																						
hasta_tibbi_gecmis		Gözat	Yapı	Ara	Ekle	Boşalt	Kaldır	100																																						

4. Belirtmek İstediğiniz Diğer Konular

ML projesi tarafında şimdilik elde edilen sonuçları bir rapor şeklinde gösterelim.

ILAÇ ÖNERİ SİSTEMİ - VERİ ANALİZ RAPORU

1. VERİ BOYUTLARI

- Özellik matrisi boyutu: 8563 satır x 7 sütun
- Hedef değişken boyutu: 8563 satır

2. ÖZELLİK BİLGİLERİ

- Kullanılan özellikler (7): hasta_id, hastalik_id, hastalik_kategorisi, etken_madde_id, yas, cinsiyet_encoded, vki

3. VERİ TÜRLERİ

- hasta_id: int64
- hastalik_id: int64
- hastalik_kategorisi: object
- etken_madde_id: float64
- yas: int64
- cinsiyet_encoded: int64
- vki: float64

4. EKSİK DEĞER ANALİZİ

- Hiçbir sütunda eksik değer yok (veya NaN değerler doldurulmuş)

5. KATEGORİK DEĞİŞKENLER

- hastalik_kategorisi: 9 benzersiz değer

En yaygın değerler:

- Kardiyovasküler: 1383 (16.2%)
- Kas-İskelet Sistemi: 1258 (14.7%)
- Endokrin: 1217 (14.2%)
- Psikiyatrik: 942 (11.0%)
- Sindirim Sistemi: 901 (10.5%)

6. SAYISAL DEĞİŞKENLER

- hasta_id:
 - Min: 1.00, Max: 100.00
 - Ortalama: 47.05, Medyan: 47.00
 - Standart sapma: 27.65
- hastalik_id:
 - Min: 1.00, Max: 32.00
 - Ortalama: 16.23, Medyan: 16.00
 - Standart sapma: 9.24
- etken_madde_id:
 - Min: 3.00, Max: 1006.00
 - Ortalama: 463.11, Medyan: 448.00
 - Standart sapma: 251.78
- yas:
 - Min: 18.00, Max: 90.00
 - Ortalama: 48.67, Medyan: 44.00
 - Standart sapma: 20.93
- cinsiyet_encoded:
 - Min: 0.00, Max: 1.00
 - Ortalama: 0.41, Medyan: 0.00
 - Standart sapma: 0.49
- vki:
 - Min: 14.98, Max: 41.80
 - Ortalama: 26.10, Medyan: 25.66
 - Standart sapma: 5.41

7. HEDEF DEĞİŞKEN ANALİZİ

- Hedef değişken türü: int64
- Benzersiz hedef değer sayısı: 907
- En yaygın 10 ilaç sınıfı:
 - İlaç ID 3241: 147 (1.7%)
 - İlaç ID 3995: 147 (1.7%)
 - İlaç ID 3121: 126 (1.5%)
 - İlaç ID 2028: 105 (1.2%)
 - İlaç ID 4389: 98 (1.1%)

- İlaç ID 2740: 98 (1.1%)
- İlaç ID 1678: 98 (1.1%)
- İlaç ID 899: 84 (1.0%)
- İlaç ID 724: 84 (1.0%)
- İlaç ID 1796: 84 (1.0%)

⚠️ UYARI: Veri seti oldukça dengesiz! En yaygın sınıf / en az görülen sınıf oranı: 147.0

8. KATEGORİK-HEDEFİN İLİŞKİSİ

- hastalik_kategorisi - 9 kategori için hedef dağılımı:

- Kardiyovasküler için en yaygın: İlaç ID 3995 (63/1383, 4.6%)
- Kas-İskelet Sistemi için en yaygın: İlaç ID 1796 (28/1258, 2.2%)
- Endokrin için en yaygın: İlaç ID 724 (42/1217, 3.5%)
- Psikiyatrik için en yaygın: İlaç ID 4209 (42/942, 4.5%)
- Sindirim Sistemi için en yaygın: İlaç ID 4579 (42/901, 4.7%)

9. VERİ KALİTESİ UYARILARI

- Önemli bir veri kalitesi sorunu tespit edilmedi.

10. VERİ ÖRNEKLERİ

- Özellik matrisi (ilk 5 satır):

	hasta_id	hastalik_id	hastalik_kategorisi	etken_madde_id	yas	cinsiyet_encoded	vki
0	97	11	Endokrin	448.0	89	1	31.46
1	97	11	Endokrin	448.0	89	1	31.46
2	97	11	Endokrin	448.0	89	1	31.46
3	97	11	Endokrin	448.0	89	1	31.46
4	97	11	Endokrin	1003.0	89	1	31.46

- Hedef değişken (ilk 5 değer):

0	3066
1	820
2	3883
3	681
4	1831

ANALİZ TAMAMLANDI