**T.C.**

**KOCAELİ ÜNİVERSİTESİ**

**BİTİRME 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 : 03/03/2025 - 07/03/2024**

**Ara Rapor No : 1**

**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** | Verilerin Arttrılması |
| **İP Tamamlanma Durumu (Yüzde Belirtilmelidir)** | %100 |
| **İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular**  *Elde edilen bulgular ve ara çıktıların (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, grafikler, algoritma, yazılım, anket formu, ham veri vb.) detaylı sunulması beklenmektedir.* | Elde var olan ilaç verileri arttırmak adına tekrardan bir web scraping ile verilerimi arttırdım. İlk etapta var olan veri sayısı ilaç ve etken madde başına 300 adetti. Ancak bu verilerin uygulamanın ileri aşamalarında sorun yaratacağından veri arttırma aşamalarına gidildi. Şu an var olan veri sayısı ilaç tarafında binlerce iken etken madde bu ilaçlara denk bir değerdedir.  **Veri Artırma Süreci ve Detayları:**   * Veri kaynağı olarak RxList, Drugs.com ve TITCK (Türkiye İlaç ve Tıbbi Cihaz Kurumu) veritabanları kullanıldı. * Web scraping işlemi için Python'da BeautifulSoup ve Selenium kütüphaneleri kullanıldı. * Başlangıçtaki 300 ilaç verisinden 3,750+ ilaç verisine ulaşıldı. * Etken madde sayısı 240'tan 860'a çıkarıldı. * Her ilaç için eklenen veri alanları:   + Etken madde bileşimi ve miktarı   + Kullanım endikasyonları   + Yan etkiler ve oluşma olasılıkları   + Kontrendikasyonlar   + Etkileşime girdiği diğer ilaçlar   + İlaç formu (tablet, şurup, enjeksiyon vb.)   + Dozaj bilgileri     Görselde de göründüğü gibi şu an sadece .json dosyalarının sadece veritabanına aktarılmış kısmını görüyoruz. Ayrı olarak burada bulunan tablolar dışında etken madde tablosu da bulunmaktadır. |

|  |  |
| --- | --- |
| **İP No** | 2 |
| **İP Adı** | Veri Düzenleme |
| **İP Tamamlanma Durumu (Yüzde Belirtilmelidir)** | %70-%80 |
| **İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular**  *Elde edilen bulgular ve ara çıktıların (teknik rapor, liste, diyagram, analiz/ölçüm sonucu, grafikler, algoritma, yazılım, anket formu, ham veri vb.) detaylı sunulması beklenmektedir.* | **Veri Temizleme ve Yapılandırma:**   * Eksik verilerin tespit edilmesi ve tamamlanması için otomatik ve manuel kontroller gerçekleştirildi. * Etken madde ve ilaç isimlerindeki yazım hataları düzeltildi. * Veri standardizasyonu sağlandı (doz birimleri, yan etki sınıflandırmaları vb.)     Bu görselde bir ilacın türüne göre nasıl kullanacağı ne şekilde yan etkileri olacağı endikasyonları, farmakinetik, farmakodinamik, kontendikasyonlar, ilaç etkileşimleri, etki mekanizması ve daha görselde görünmeyen ancak devamı şu şekilde olan;    Bir tablo yapısına sahip. Bu tablo yapısının avantajı makine öğrenmesi sağlanırken hastaların bir ilacı kullanırken aldığı yan etkilere dayanarak kullanım yolları ve etki mekanizması başta olmak üzere aslında bir başka hastaya nasıl bir ilaç önereceğimizi belirleyecektir. Burada ilaç önerme kısmında aslında var olan hangisi daha çok kullanıyorsa onu önerelimin dışında hastanın sağlık açısından ilaç etkisini inceleyen bir yapı kurmayı hedefler. Bir sonraki İP kısmında ise aslında hastanın ne tür bilgilere sahip olduğunu gösterdikten sonra aslında bu veriler ile nasıl bağdaştırma kuracağımızı anlamış olacağız.  **Veri Zenginleştirme:**   * İlaçlar ATC (Anatomik Terapötik Kimyasal) kodlarına göre sınıflandırıldı. * İlaçların etkileşim seviyelerine göre risklilik derecelendirmesi yapıldı (düşük, orta, yüksek risk). |

|  |  |
| --- | --- |
| **İP No** | 3 |
| **İP Adı** | Hasta Bilgisi ve Hastanın Tüm Sağlık Geçmiş Bilgileri |
| **İP Tamamlanma Durumu (Yüzde Belirtilmelidir)** | %90 |
| **İP Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular**  *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 Kapsamında Yapılan Çalışmalar ve Elde Edilen Bulgular**  Bu iş paketi kapsamında, ilaç öneri sisteminin kişiselleştirilmiş ve güvenli öneriler sunabilmesi için gerekli olan hasta verilerinin toplanması, yapılandırılması ve analizi gerçekleştirilmiştir. Hasta verilerinin mahremiyeti gözetilerek anonim hale getirilmiş gerçek hasta verileri ve sentetik veri setleri üzerinde çalışılmıştır.  **Veri Toplama ve Yapılandırma**  Hasta bilgisi ve sağlık geçmişi ile ilgili aşağıdaki veriler yapılandırılmış bir formatta toplanmıştır:   1. **Demografik Bilgiler**:    * Yaş, cinsiyet, kilo, boy, BMI değerleri    * Yaşam tarzı faktörleri (sigara kullanımı, alkol tüketimi, fiziksel aktivite düzeyi)    * Meslek ve çalışma koşulları (kimyasal maruziyeti, vardiyalı çalışma vb.) 2. **Tıbbi Geçmiş**:    * Kronik hastalıklar ve tanı tarihleri    * Geçirilmiş önemli akut hastalıklar    * Cerrahi operasyonlar ve prosedürler    * Alerji ve intolerans kayıtları (özellikle ilaç alerjileri)    * Aile sağlık geçmişi ve genetik yatkınlıklar 3. **İlaç Kullanım Geçmişi**:    * Şu anda kullanılan ilaçlar (reçeteli ve reçetesiz)    * Geçmiş ilaç kullanımları ve tedavi süreleri    * İlaç yan etki deneyimleri ve şiddetleri    * Tedavi uyumu ve ilaç kullanım davranışları 4. **Laboratuvar ve Tetkik Sonuçları**:    * Kan biyokimyası ve hematoloji değerleri    * Karaciğer ve böbrek fonksiyon testleri    * İlaç metabolizmasıyla ilişkili özel testler (CYP enzim varyantları)    * Görüntüleme sonuçları ve klinik ölçümler   **Veri Modeli ve İlişkisel Yapı**  Hasta verileri, ilaç verileri ile entegre edilebilecek şekilde aşağıdaki ilişkisel veri modeli çerçevesinde yapılandırılmıştır:  Hasta {  hasta\_id: String, // Benzersiz hasta tanımlayıcısı  demografik\_bilgi: {  yas: Number,  cinsiyet: String,  kilo: Number,  boy: Number,  bmi: Number,  sigara\_kullanimi: Boolean,  alkol\_tuketimi: String, // "Yok", "Hafif", "Orta", "Ağır"  fiziksel\_aktivite: String // "Düşük", "Orta", "Yüksek"  },  kronik\_hastaliklar: [ // Hastanın kronik hastalıkları  {  hastalik\_kodu: String, // ICD-10 kodu  hastalik\_adi: String,  tani\_tarihi: Date,  seviye: String // "Kontrol altında", "Aktif", "Şiddetli"  }  ],  alerjiler: [ // Hastanın alerjileri  {  alerjen\_turu: String, // "İlaç", "Gıda", "Çevresel"  alerjen\_adi: String,  reaksiyon: String,  siddet: String // "Hafif", "Orta", "Şiddetli", "Yaşamı tehdit eden"  }  ],  ilac\_kullanim\_gecmisi: [ // Hastanın kullandığı ilaçlar  {  ilac\_id: String, // İlaç veri tabanındaki ID  baslangic\_tarihi: Date,  bitis\_tarihi: Date,  doz: String,  kullanim\_amaci: String,  yan\_etki\_deneyimi: [  {  yan\_etki: String,  siddet: String, // "Hafif", "Orta", "Şiddetli"  baslangic: Date,  bitis: Date  }  ],  uyum\_duzeyi: String // "Düşük", "Orta", "Yüksek"  }  ],  lab\_sonuclari: [ // Hastanın laboratuvar sonuçları  {  test\_adi: String,  tarih: Date,  sonuc: Number,  birim: String,  referans\_aralik: String  }  ]  }  **Veri Modellerinin Tablo Görüntüsü**    Burada görünen işaretli tablolar aslında bir hastanın geçmiş sağlık verilerini ve hastalıklarını tutuyor. Bu verilere dayanarak aslında bir ilacın etkisini tespit etmede bu hasta verileri ve ilaç verileri birbiri arasında bağ kurulmaya başlandığında o zaman bu veriler daha anlamlı bir hale gelmiş olacaktır. |

**(İŞ PAKETLERİ (İP) TABLOLARI SAYISI ARTITILABİLİR)**

**2.** **Proje Çalışma Takvimine Uygun Yürümüyorsa Gerekçeleri**

**İş-Zaman Çizelgesinde planlanan iş paketlerinin tamamlanamama durumu söz konusu ise sebebi detaylı olarak açıklanmalı ve buna yönelik alınmış olan tedbirlerin (ek süre, yöntem değişikliği, B planını devreye sokulması vb.) detaylı şekilde sunulması beklenmektedir. Proje takvimine uygun şekilde gerçekleşiyorsa bu kısım boş bırakılmalıdır.**

|  |
| --- |
|  |

**3.** **Bir Sonraki Dönemde Yapılması Planlanan Çalışmalar**

**3.1. Proje öneri formunda iş-zaman çizelgesinde bir sonraki ara ve/veya sonuç raporu dönemine kadar gerçekleştirilmesi hedeflenen iş paketleri sunulmalıdır.**

|  |  |
| --- | --- |
| **İP No** |  |
| **İP Adı** |  |
| **İP zaman aralığı** |  |

|  |  |
| --- | --- |
| **İP No** |  |
| **İP Adı** |  |
| **İP zaman aralığı** |  |

|  |  |
| --- | --- |
| **İP No** |  |
| **İP Adı** |  |
| **İP zaman aralığı** |  |

**3.2. Mevcut dönemde bitirilemeyip sonraki dönemde devam edilmesi gereken çalışmalar var ise kısaca bilgi verilmelidir.**

|  |
| --- |
|  |

**4. Belirtmek İstediğiniz Diğer Konular**

|  |
| --- |
|  |