### TOBB ETÜ Yapay Zeka Mühendisliği Bölümü

**BİL 476/BİL 573 Dönem Projesi Ara Rapor**

**2023-24 Yaz Dönemi**

**Öğrenci Adı Soyadı: ÖZGE IŞIKLAR Date:18/07 /2024**

|  |
| --- |
| 1. Proje ismi |
| Avatar: The Last Airbender bölüm ve karakter bazlı clustering, data visualization |
| 2. Proje tanımı [[1]](#footnote-1) (kısa özet, problemin açıklanması, beklenen sonuçlar) |
| Bu projede sevilen animasyon dizisi Avatar: The Last Airbender için bir data analiz çalışması yapmak amaçlanmaktadır. Datasetten alınan bilgiler doğrultusunda karakterlerin en çok kullandığı kelimeler, bölüm bazlı kümeleme, kullanılan diyalogların analizi, IMDB ratingleri için analiz gibi konularda Avatar: The Last Airbender için data visualization ve clustering teknikleri kullanılarak bir çıktı verilecektir. |
| 3. Proje metodolojisi, daha önce benzer projelerde kullanılan metotlar hakkında bilgiler, literatür taraması |
| Bu projede, veri analizi ve görselleştirme teknikleri kullanılmaktadır. Seaborn ve Matplotlib kütüphaneleri ile grafikler oluşturulmuş, NetworkX ile karakter etkileşim grafiği yapılmıştır. WordCloud ve TextBlob ile kelime bulutu ve duygu analizi gerçekleştirilmiştir. Benzer projelerde duygu analizi için VADER lexicon kullanılırken bu projede TextBlob kullanıldı. Bu projede genellikle Matplotlib ve Seaborn kütüphaneleri kullanılırken başka projelerde Plotly gibi kütüphanelerin kullanıldığı gözlendi.  Bu proje özelinde literatür taraması:  Visualization of Text Document Corpus (Informatica 29, 2005)  Modeling narrative features in TV series: coding and clustering analysis  Using cluster analysis for data mining in educational technology research  Specificity of film quote use and functionıng in speech of various genres: Corpus-Assited Study |
| **4. Projede kullanılan veriler, dış bağlantılar hakkında bilgiler, verilerin tamamını veya birkaç örneği gösterebilirsiniz.** |
| Projede, Avatar: The Last Airbender dizisinin bölümlerine ve karakterlerine ait diyaloglar ve IMDB ratingleri kullanılmıştır. Veri seti:  - Bölüm numarası  - Kitap  - Bölüm  - Karakter  - Diyalog  - IMDB ratingi  Veri setinin tamamı CSV dosyası olarak sağlanmıştır. |
| 5. Projenin ayrıntılı planı, akış şeması, veri işleyişi vs., projede karşılaşılan sorunlar, projenin şu andaki durumu (varsa elde edilen sonuçlar veya ekran çıktıları), bundan sonra yapılacaklar |
| Proje Planı:  - Literatür Araştırması: 07.06.2024 (Tamamlandı)  - Veri Toplama ve İşleme: 14.06.2024 (Tamamlandı)  - EDA ve Data Visualization: 24.06.2024 (Tamamlandı)  - Diyalog Analizleri: 08.07.2024 (Tamamlandı)  - Ara Rapor Hazırlığı: 12.07.2024 (Tamamlandı)  - Clustering Çalışmaları: 24.07.2024 (Devam ediyor)  - Final Rapor Sunumu: 31.07.2024  Karşılaşılan Sorunlar:  -Veri setindeki bazı karakterlerin diyalog eksikliği  -Bazı verilerin kodlama sorunları nedeniyle okunamaması  Edilen Sonuçlar  - IMDB ratinglerinin bölüm ve kitap bazlı analizi yapılmıştır.  - Önemli karakterler için kelime bulutu ve duygu analizi gerçekleştirilmiştir.  - Karakterler arasındaki etkileşimler grafik ile görselleştirilmiştir.  Bundan Sonra Yapılacaklar:  - Clustering çalışmalarının tamamlanması.  - Final raporun hazırlanması ve sunumu  Proje çıktılarının ipynb dosyası için: [Link](https://drive.google.com/file/d/1PoGp0WLg3TOq5JgvgIw5UgzgeIm0v6bj/view?usp=sharing) |

1. Eğer gerekiyorsa ayrı bir sayfa kullanın [↑](#footnote-ref-1)