TOBB ETU
Department of Computer Engineering
2022-2023 Fall Semester

BİL 470/570 Term Project

Project Members: Dursun Karaca Erdemir, İrem Bigat **Date:**03/10/2022

1. Project name
Myers-Briggs Personality Type Prediction Using Machine Learning Techniques
2. What is the problem that you will be investigating? Why is it interesting?
The problem we want to investigate involves the identification of people's personalities. It is a widely researched subject in the field of psychology, and many systems were created to determine a person's personality and classify them under sets of certain character traits. Some of the most popular systems are 'Big five-factor model' [1], 'Myers-Briggs Type Indicator (MBTI)' [2, 3] and 'Caliper Profile' (Mainly used for job applicant evaluation). Personality type prediction with machine learning is a relatively new sub-branch but some papers and projects are available. We decided to use 'Myers-Briggs Type Indicator (MBTI)' as the personality system in our models mainly due to its combinational and binary structure. Personality types consist of 4 letters and all of which take binary values (I/E, S/N, T/F, P/J). There are 16 different types in total and we felt this was the ideal test for using machine learning approaches.
3. What reading will you examine to provide context and background?
We will examine papers in the literature, specifically, we choose three papers [4, 5, 6] for the beginning. The methods and algorithms used in each differ significantly. We will sift through tens of methods and algorithms and combine them with our own intuition and hope for the best.
4. What data will you use? If you are collecting new data, how will you do it?
We are going to use a Kaggle dataset [7]. It has more than 8000 rows and has one target and one other attribute. We will need to add more features using relevant methods.

5. What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations? You don't have to have an exact answer at this point, but you should have a general sense of how you will approach the problem you are working on.

We have quite a few methods and algorithms in mind. We don't plan to use any code, and implement our project under the guidance of published papers. We will start on an empty notebook and then fill it, most of the implementation of our selected methods are shown in class and are present in slides, and the papers we follow will help us choose and decide which algorithm comes when.

For now, we plan to use methods like Word Cloud for feature addition and one hot encoding to adjust features properly.

As we are predicting a categorical result, we plan to use models such as Logistic Regression, Decision Tree, Gradient Boost and maybe other models we learn at the lecture.

We may combine our models and procure an ensemble model at the final stages of our project.

6. How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)?
Quantitatively, what kind of analysis will you use to evaluate and/or compare your results?

We plan to use an unseen portion of our data to test our model(s), some more than once while using various methods then plotting it with their matching performance metric of our choice.

Currently, we plan to compare the f1 scores of our models as our data seem to have class imbalance. One class has way more samples than the rest.

7. References

1) De Raad, B., & Mlacic, B. (2015). Big five factor model, theory and structure. International encyclopedia of the social & behavioral sciences, 2, 559-566.
2) Varvel, T., & Adams, S. (2003, June). A Study Of The Effect Of The Myers Briggs Type Indicator. In 2003 Annual Conference (pp. 8-124).
3) Boyle, G. J. (1995). Myers-Briggs type indicator (MBTI): some psychometric limitations. Australian Psychologist, 30(1), 71-74.
4) Hernandez, R., & Knight, I. S. (2017). Predicting Myres-Briggs Type Indicator with text classification. Curran Associates Inc., New York.
5) Ontoum, S., & Chan, J. H. (2022). Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning. arXiv preprint arXiv:2201.08717.
6) Amirhosseini, M. H., & Kazemian, H. (2020). Machine learning approach to personality type prediction based on the myers–briggs type indicator. Multimodal Technologies and Interaction, 4(1), 9.
7) Mitchelle, J.; Myers-Briggs Personality Type Dataset. Includes a Large Number of People's MBTI Type and Content Written by Them. Available online: https://www.kaggle.com/datasnaek/mbti-type (accessed on 30 September 2022).