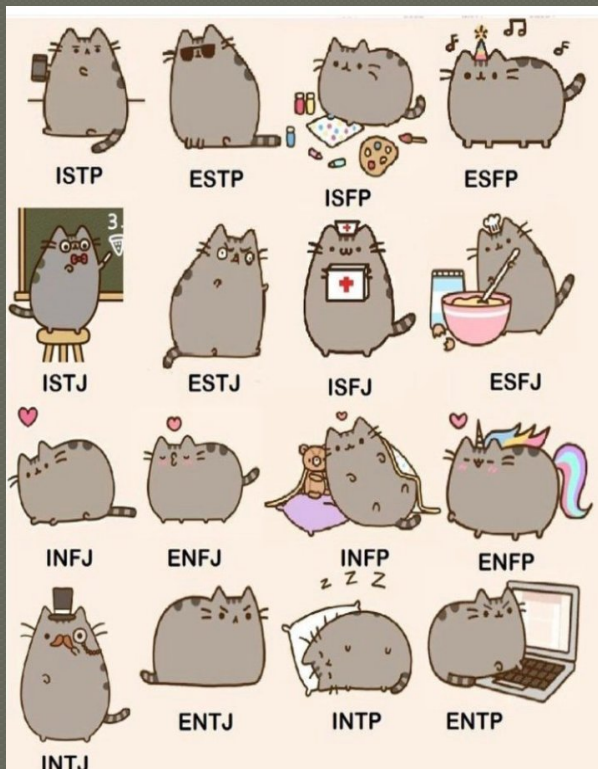


# Personality Prediction Using Machine Learning

Thejas S  
Anujith K

# Introduction

MBTI is a test that uses a questionnaire to classify the personality using 4 scales, into 16. Even though MBTI is criticized as pseudo science, it can be applied broadly to personalize Ads or to recommend videos in an OTT platform.



The shortcoming of a traditional MBTI test are-

- >The questionnaire is forced choice and will be time consuming.
- >Answers can be biased due to individual expectations.

So we are trying to use an efficient and accurate ML algorithm to scan media posts by individuals and predict these MBTI parameters.

# Dataset

---

- We are using a Dataset provided by Kaggle.com. It consists of 8675 rows and 2 columns with the MBTI personality codes and posts of users from PersonalityCafe forum.

<https://www.kaggle.com/datasnaek/mbti-type>

- Reason for using this dataset is because the MBTI for the users are already labeled. So the result can be analyzed for accuracy.

# Plan

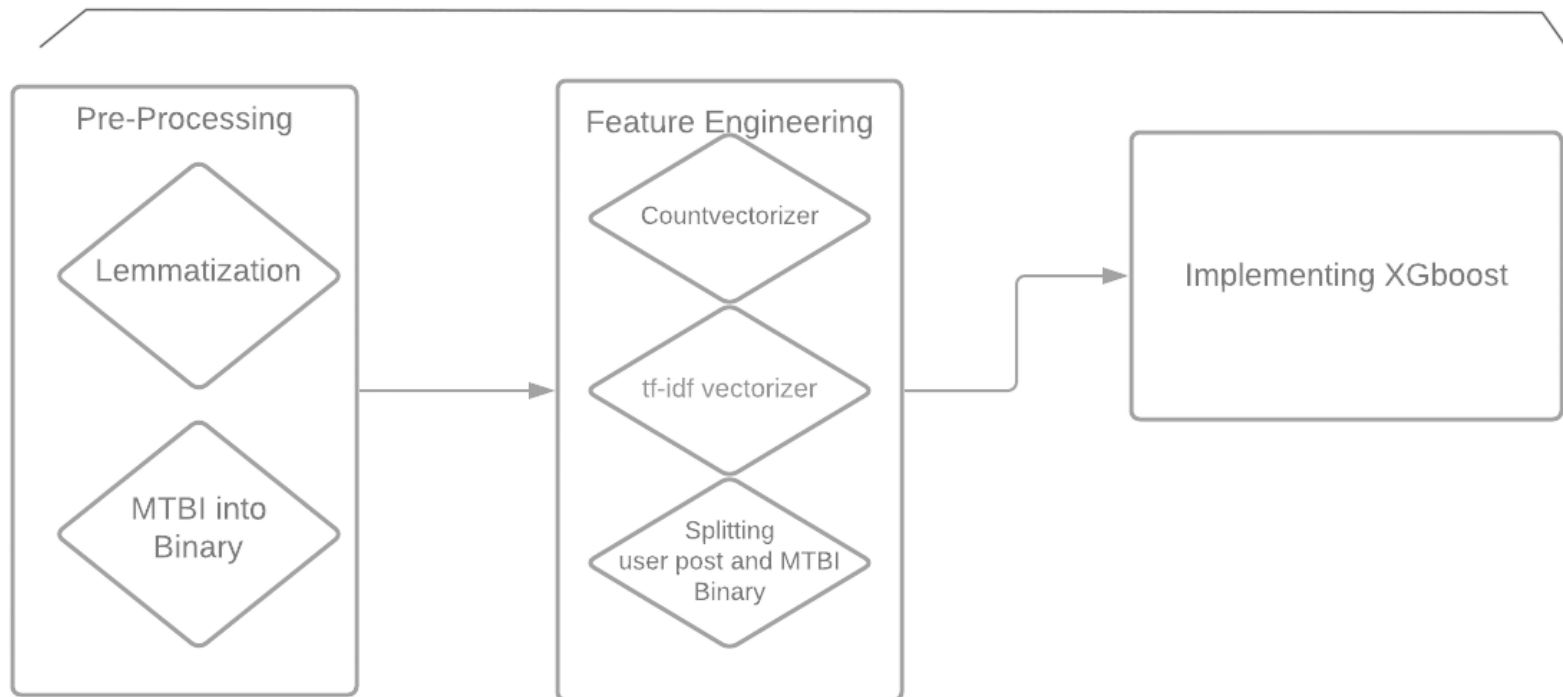
---

- Pre-processing data: The raw data gets converted to an easier to work with form. (Letters gets converted to lowercase, URLs are removed, MBTI references are removed and its lemmatized.)
- Splitting the dataset into two variables to work with, one set of binary MBTI and the other set will be posts.
- Feature vector generation: We plan to use CountVectorizer and TF-IDF to weigh the words and remove junk.
- Running a tweaked XGboost for an optimal result, in terms of accuracy.
- Comparing our accuracy with other implementations that have already being done like Recurrent neural networks and logistic regressions and analyzing what could have been changed for an improved prediction.

# Midway expectation

For Midway, we are planning to complete the pre-processing, feature engineering and first iteration of XGboost implementation.

## Midway goal



# Work division

---

Website management, Pre-processing and XGboost implementation would be done by Thejas

Feature engineering and analysis of our models to others would be done by Anujith

Final report will be done by us both

# References

---

- Abidin, N. H. Z., Remli, M. A., Ali, N. M., Phon, D. N. E., Yusoff, N., Adli, H. K., & Busalim, A. H. (2020). Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*.
- Hernandez, R. K., & Scott, I. (2017). Predicting Myers-Briggs type indicator with text. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.
- J, M., 2017. *(MBTI) Myers-Briggs Personality Type Dataset*. [online] Kaggle.com. Available at: <<https://www.kaggle.com/datasnaek/mbti-type>>

---

**THANK YOU**