Topic:

MBTI personality type prediction through social media posts.

Background:

Every person has MBTI personality type defined by Myer's Briggs test. MBTI type can be any of 16 combinations of:
Energy: Extrovert / Introvert
Information: Sensing / Intuition
Decision: Thinking / Feeling
Lifestyle: Judging / Perceiving

Dataset:

Kaggle has a dataset which has 45-50 most recent posts for 8600 users with their MBTI type. [This can be used as label] In this dataset we have 422,845 total data points.
https://www.kaggle.com/datasnaek/mbti-type

Features

- Bigrams
- Skip grams
- Part of speech
- Capital letters
- Bag of words

Goal:

Take in a set of input text of a person and output the predicted MBTI personality type. Compare different traditional ML classifiers along with Deep learning models which solves this problem. Analyze different metrics of every algorithm.

Methods:

1) NLP: We can use a lot of pre-processing for filtering data.
2) SVM Classifiers
3) Deep Learning (Word embeddings)
4) *Try: using Random forest for classification
5) Other binary classifiers.