Project Title: MBTI Personality Type Analysis and Prediction

Members:
Vibhu Krovvidi (vk2500)
Meghan Shah (ms5767)
Vrinda G Bhat (vgb2113)
Varalika Mahajan (vm2695)
Dhivyadarsan G M (dg3233)

Background & Context to the problem statement
Problem Statement:
Predicting a user's personality type, based on the Myers-Briggs personality test, using Machine Learning (ML) and Natural Language Processing (NLP) techniques.

Outline:
Personality tests are designed to elicit information from a person about their motivations, interests, and situational awareness. Personality is described using a combination of traits and features. Understanding the traits or features of an individual through the content they post can support a better understanding of their tastes, behavior, and habits.

Applications:
- Psychometric application, criminal investigations, hiring professionals, team division, job performance indicators, learning one's own personality type without potential self-survey bias, Linkedin profile scraping for candidate selection.

Identification & Description of Dataset
Dataset: (MBTI) Myers-Briggs Personality Type Dataset

Description:
The MBTI system classifies personalities along four dimensions:
- Introversion (I) – Extroversion (E)
- Intuition (N) – Sensing (S)
- Thinking (T) – Feeling (F)
- Judging (J) – Perceiving (P)

Each personality is assigned a four-letter code from these dimensions. The dataset provides 8675 unique text entries along with an associated MBTI type for the user. The objective of our project is to use this text data to construct a classification model that accurately predicts the personality type of an individual based on their comments.
To do this, we propose using:

Proposed ML Techniques: (Subject to the addition of other complex models)
Classification Models:
- K Nearest Neighbours, Naive Bayes, Logistic Regression, Support Vector classifier, Stochastic Gradient Classifier, Random Forests, XGBoost, RNNs