MBTI Personality Prediction

By, Alexander Santa Cruz, Cole Clark, Henry Le, and Kyle Nepo

Introduction

-Create an AI that is able to determine a twitter user's MBTI (A self reported personality based on an online test) based on aspects of their profile.

-Features include but are not limited to:
 -Follower count
 -Friend count
 -Retweet count

-Test multiple algorithms and see which is the most efficient and accurate
 -K-nearest neighbor, Random Forest, Naive Bayes, and Neural Networks

K-Nearest Neighbor

- K nearest neighbors using distance between points

- The classifier we used estimates based on majority neighbor classification

- Tried different values of K

- Tried with and without weights
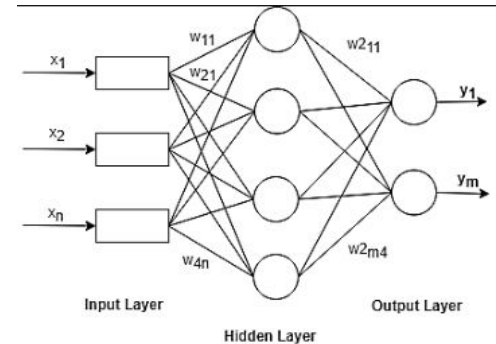
Random Forest

- Creates decision trees each using a subset of data

- Takes majority of trees prediction to make forest prediction

- Tried changing n_estimators

- Tried changing min_samples_split

Naive Bayes

- Often used in text classification
- Algorithm assumes predictors are conditionally independent
- All features contribute equally to the outcome
- Used GaussianNB from sklearn.naive_bayes

Neural Networks

- MLP classifier is a type of feedforward neural network design for classification task
- Consist of layers of neurons, where each layer performs specific transformation on the input data/trained data.
- Why use this algorithm?
  - Handles complex non-linear relationships in data, making it suitable for a wide range of classification.
  - Automatically learn relevant features from raw input data
  - Generalize well to unseen data
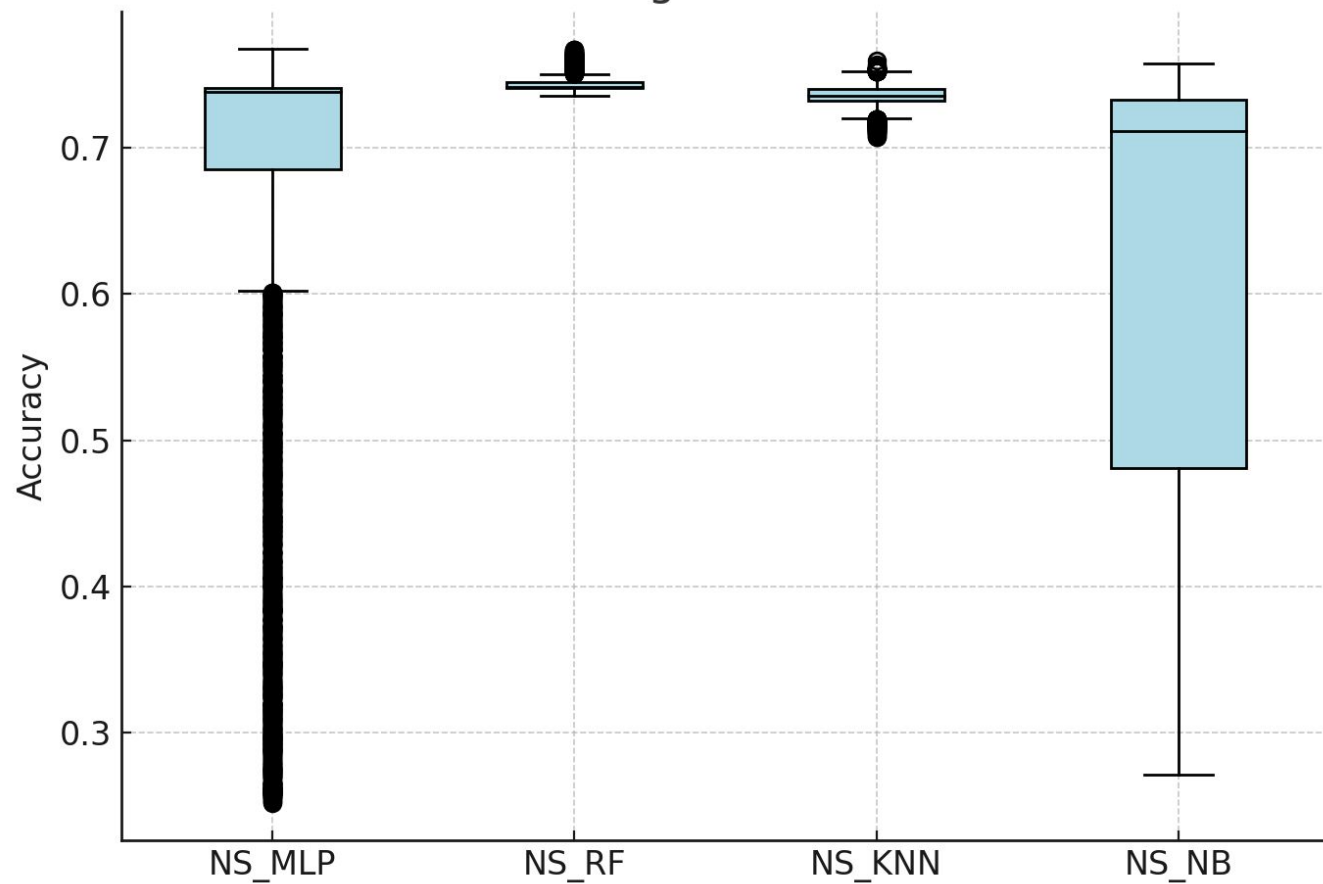- Slower process (Took 7 hours), but resulted in accurate readings

Results

Introversion vs. Extroversion
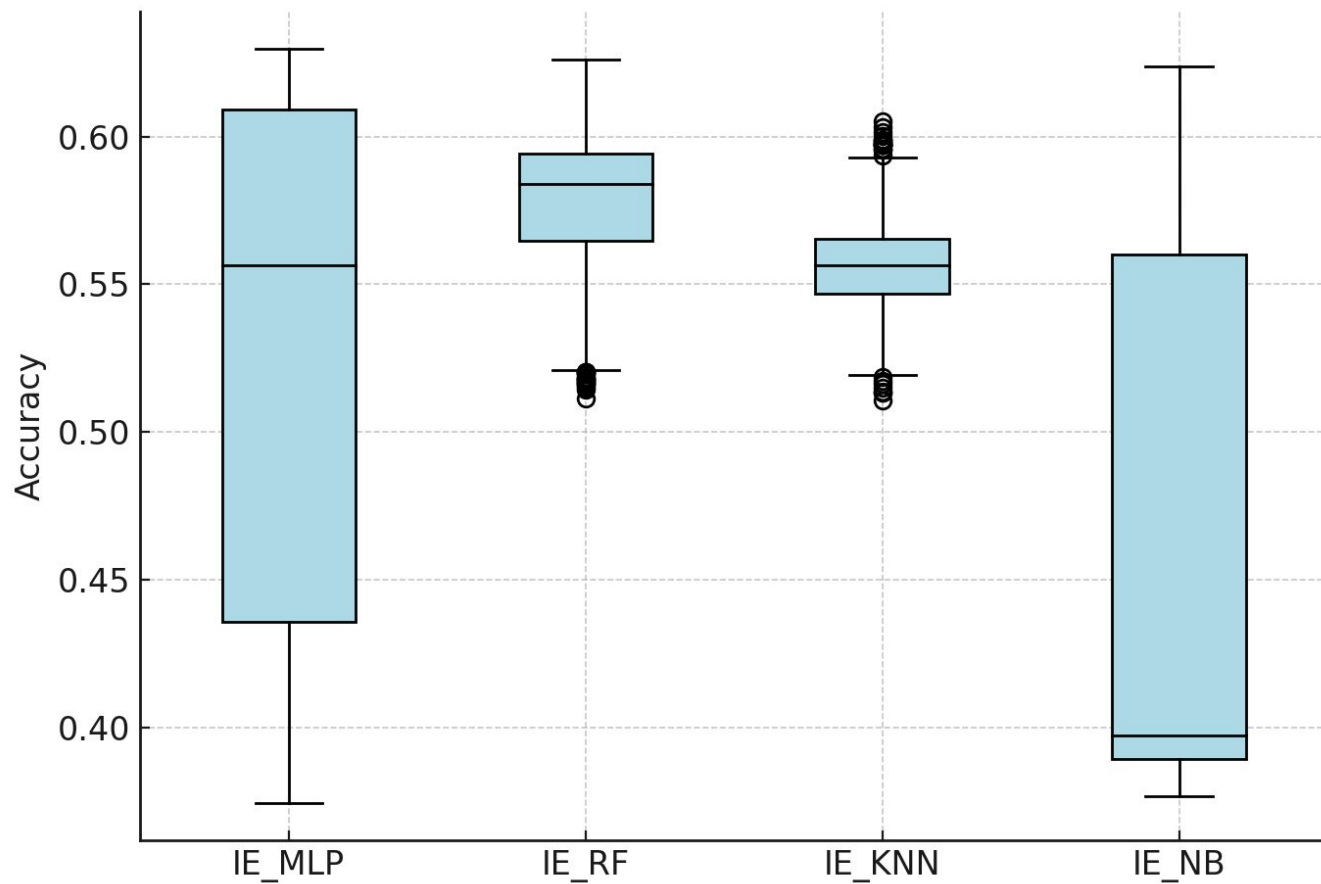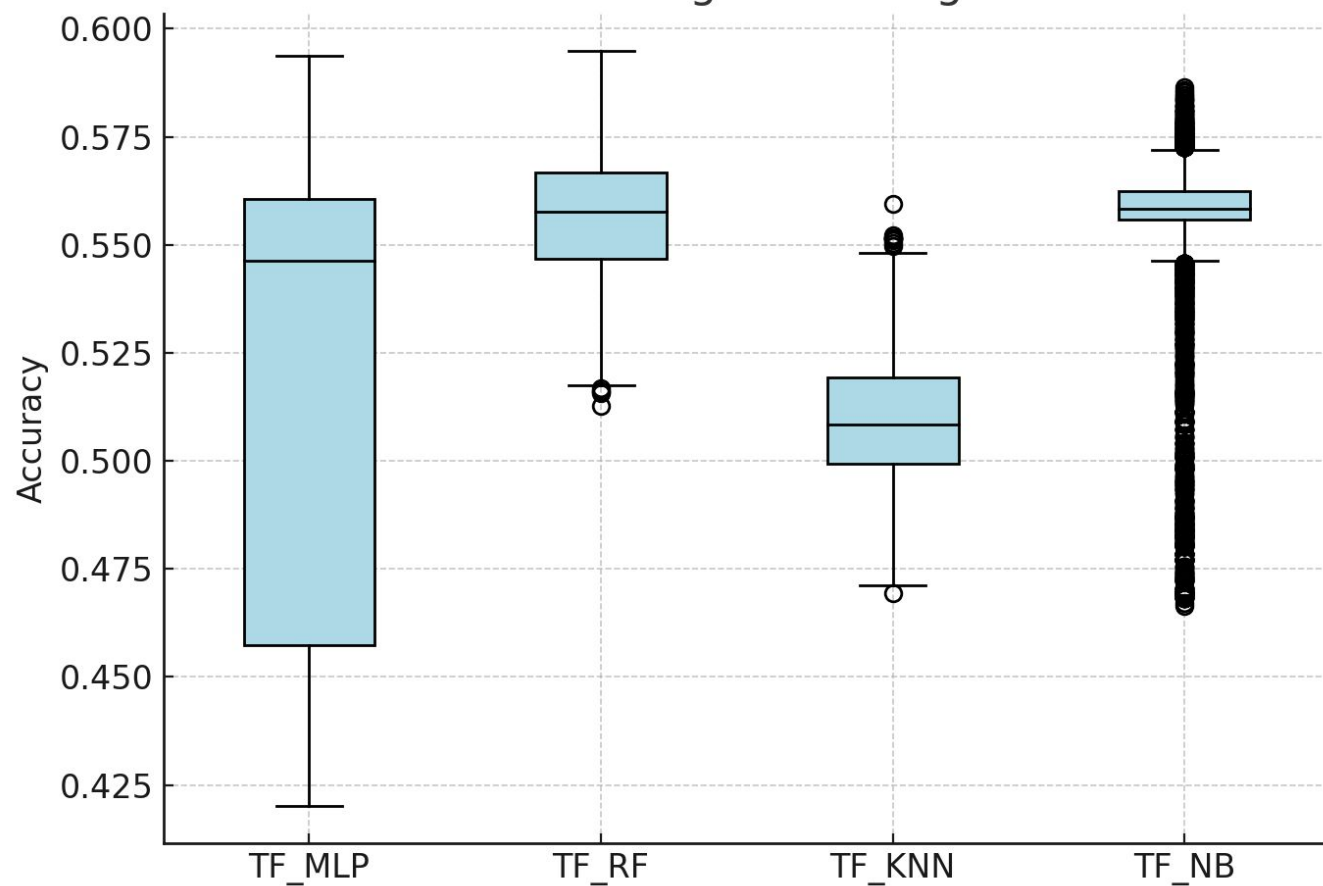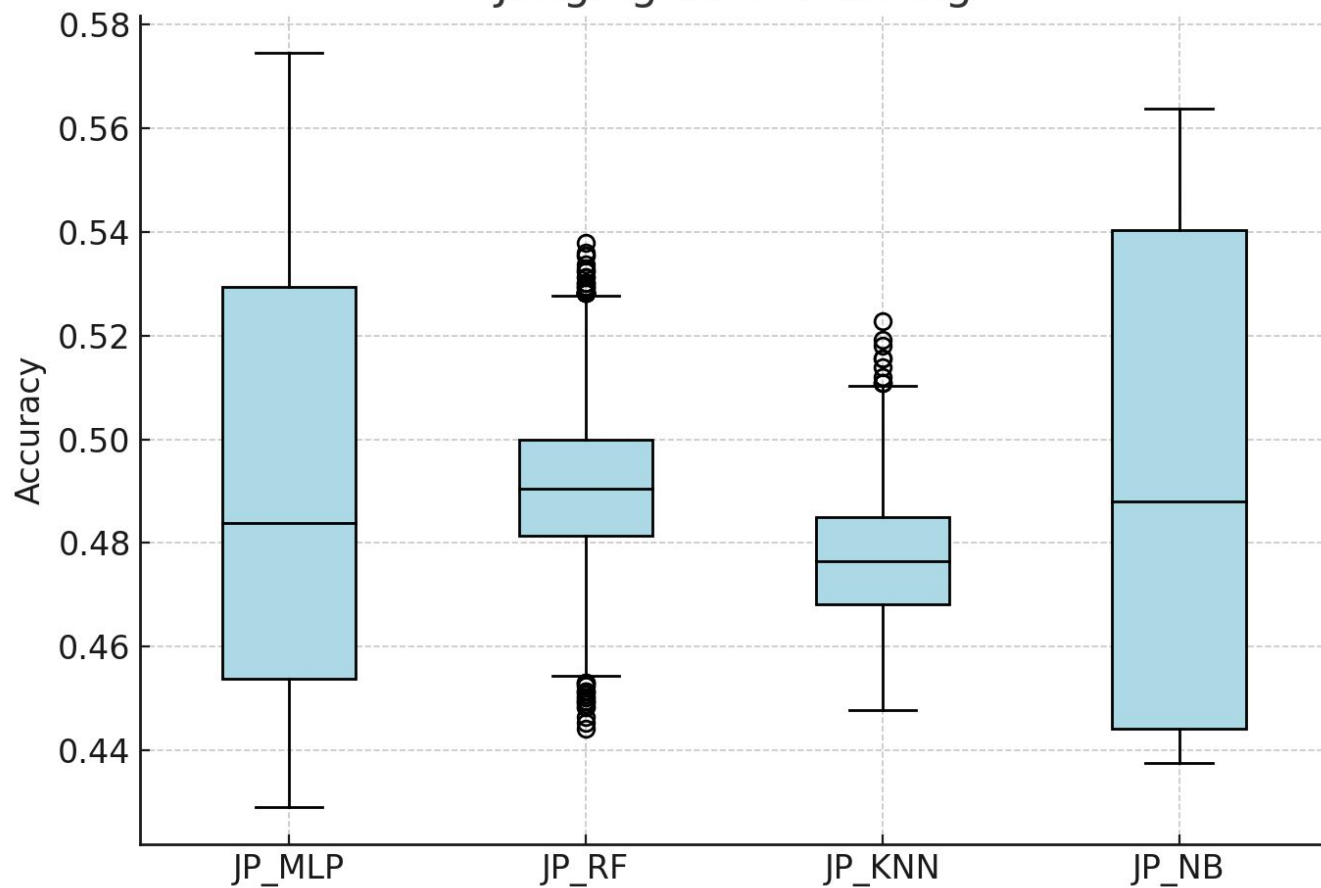
Sensing vs. Intuition

Thinking vs. Feeling

Judging vs. Perceiving

Conclusion

Random Forest Superiority

- Highest Mean Accuracy
- Lowest Standard Deviation
- Small IQR

Possible Reasonings

- Handles Diverse Data
- Robust Against Overfitting

Future Endeavors

- Mental Health Applications

- Constrained Neural Networks