The Myers Briggs personality assessment puts a personality type to an individual. There are four components to an individual’s personality. For each component, an individual takes on one out of two attributes. Since each of the four components has two attributes, then there are sixteen different personality types that an individual can be labeled as. For the first component there is introversion vs. extroversion, for the second there is intuition vs. sensing, for the third thinking vs feeling, and for the last component there is judgment vs perception. The objective of this project is to investigate the effectiveness of the perceptron to label each component of an individual’s personality type.

The dataset being used for this project is from Kaggle’s website and is titled MBIT Personality Classification Challenge. Each row represents an anonymous individual. For each individual there is a personality type feature and multiple features containing the posts of that individual. This dataset will be altered to better fit the goals of the project.

The first action that needs to be taken is to “split” each component of the personality type into separate features. For example, the introversion vs extroversion component will be one feature and the intuition vs sensing component will be another feature, etc. The next action will be to combine all the post features into just one feature called post. Once these actions have been completed, every word will be split into its own feature with its frequency of use being the measurement of the feature.

As can be noticed from the previous action, there will be a large number of features. To reduce the number of features, the highest frequency words for an attribute will only be used. It has not been decided what number of words will be allowed yet. For example, say that the first component is being looked at first, then in the training set fifty of the highest frequency used words for each individual would be chosen and then combining all of these words will be the features.

A perceptron will be trained for each component of the personality type. Then, for the test set, the featured words will be counted for each individual. The perceptron will then make a decision on each component. The accuracy of this algorithm will then be assessed.