Analyzing Message Popularity in Group Chats

Research Background

Being influential in a group chat can seem its not on your side and your messages may not be able to get read or liked.

Various ways on how you write your messages could contribute to these factors.

Our goal is to figure out how popularity links to the styles of a written message.

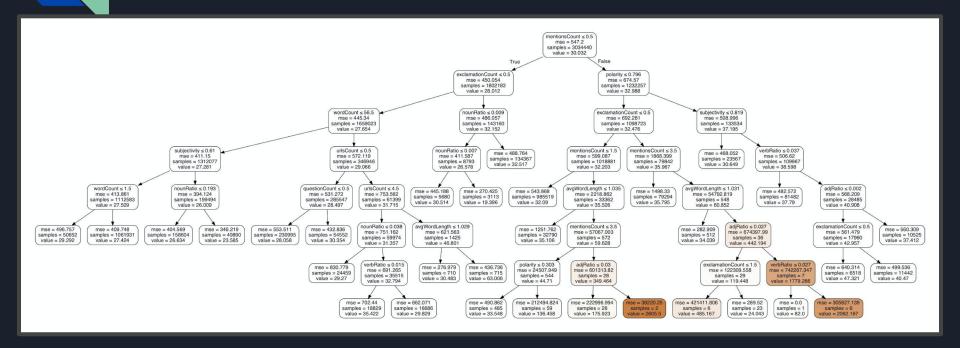Dataset: FreeCodeCamp on Kaggle w/ 5 million lines

Our Strategy

1. Converting each message into a set of message features.

   E.g. "hello world!" → {word_count: 2, ave_word_length: 5, verb_ratio: 0.5......}

1. Train a machine learning model that can take in the features and will be able to predict a message's readBy (popularity).
2. Plot out the decision tree inside of machine learning model
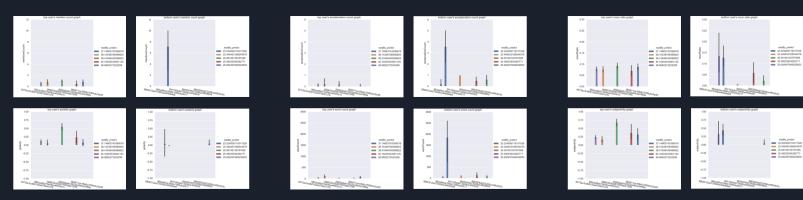3. Manually analyze the features on top of the decision tree and come to a conclusion.

Research Result

Top features: Mention Count, Exclamation Count, Polarity, Word Count, Noun Ratio, Subjectivity

Research Result

Feature Analysis:

Conclusion:

- Exclamation Count and Word Count has negative influence over message popularity
- Polarity and Subjectivity have positive influence over message popularity
- High popularity messages tend to have their Mention Count keep around 1 and Noun Ratio between 5% to 10%

- High popularity messages tend to resemble real human language.
- Low popularity messages are more like coding language.