Abstract

A Trump Tweet that is insulting and negative has the best chance of receiving the most favorites. This discovery reveals that, in terms of user engagement with digital political media, emotional content maximizes interaction. In our research, we aim to uncover the factors that contribute to the popularity of a Trump tweet. To analyze our dataset of 26,312 tweets we will utilize feature generation methods to create variables based on the text data and we will apply machine learning techniques like Random Forest and Gradient Boosting to explore the relationship between different factors of a tweet and its success.
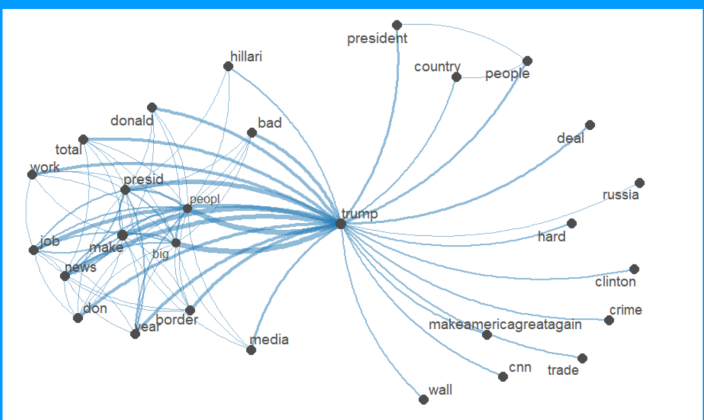
Features

- **ALL CAPS:** T/F for when the tweet is written in all caps. (True/False)
- **Time of Day:** When did Trump post the tweet? Morning, Afternoon, Evening, or Midnight. (Categorical)
- **Length of Tweet:** We created 4 buckets based on the word count of each tweet. (Categorical)
- **Political Words:** From our personally built list of 180+ political words Trump often uses, (i.e. "election", "presidency", "vote") here we measure the proportion of the words in each tweet that are political. (Numerical)
- **Insulting Words:** From our personally built list of 10+ insulting words Trump often uses, (i.e. "crooked", "weak", "sleepy") here we measure the proportion of the words in each tweet that are insulting. (Numerical)
- **Sentiment:** Using python libraries, we assessed a sentiment score for each tweet to see if the sentiment was positive or negative. (Numerical)

Text Mining

This word cloud shows the top one hundred most frequent words to appear in Trump tweets. The most commonly used words seem to either be referencing politicians (i.e. "trump", "hillary", "obama") or surrounding the topic of politics itself (i.e. "president", "vote", "democrats").

The text network displays connections between words, offering insights into how words generally relate to each other within tweets. The strongest word association occurs between the word "trump" to words like "president", "people", and "big".

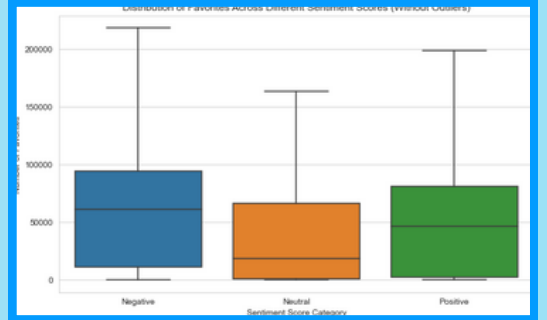**Donald J. Trump** ✓ @realDonaldTrump · 1 hr

A Trump Tweet that is **insulting** and **negative** has the best chance of receiving the most favorites.

💬 82k   🔁 45k   ❤ 91k   🔖 78k   📊 87K

"CROOKED"   "WEAK"   "DISHONEST"   "PUPPET"   "SLEEPYJOE"   "DOPEY"

"WITCH"   "CORRUPT"   "SLEEPY"   "PHONY"   "PATHETIC"   "FAT"   "CROOKEDHILLARY"

Team Mean Squares
Daphne, Ethan, Kevin, Nishant, and Mia
UCLA Statistics 140XP

EDA for Sentiment
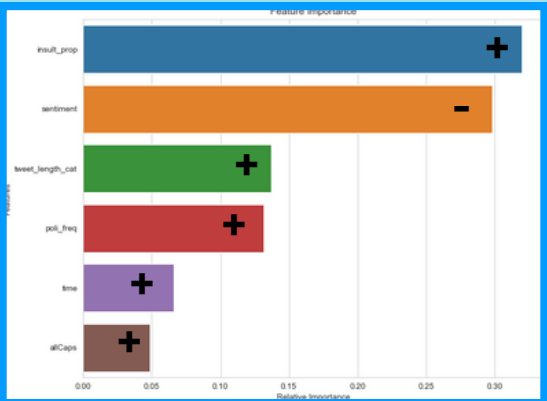

The side-by-side boxplot shows that tweets expressing negative sentiments have a higher lower quartile, median, and upper quartile in the number of favorites compared to the other tweets. We observe that tweets expressing either negative or positive sentiments have a higher median number of favorites compared to tweets that are neutral. Such observation is reasonable because tweets expressing negative or positive sentiments often contain more emotions, thus causing Trump's followers to react strongly and favor the tweet more.

Statistical Analysis

Model	MSE	R-squared
Random Forest Regressor	0.3465	0.7283
Gradient Boosting Machine	0.2846	0.7768
Linear Regression	0.4074	0.6805
Decision Tree Regressor	0.5219	0.5907

Results/Conclusion

The top 4 variables for predicting Trump tweets' popularity are proportion of insulting words, proportion of political words, sentiment, and tweet length.

The combination of heightened emotional expression, provocative language, and a strong political stance creates a recipe for increased attention and interaction. However, the variables are highly correlated so we cannnot accurately isolate the effect of a single variable. Ultimately, looking at the importance of features and their R-squared values as a result of our model, we can see that a Trump tweet that would receive the most favorites is insulting and has negative sentiment.
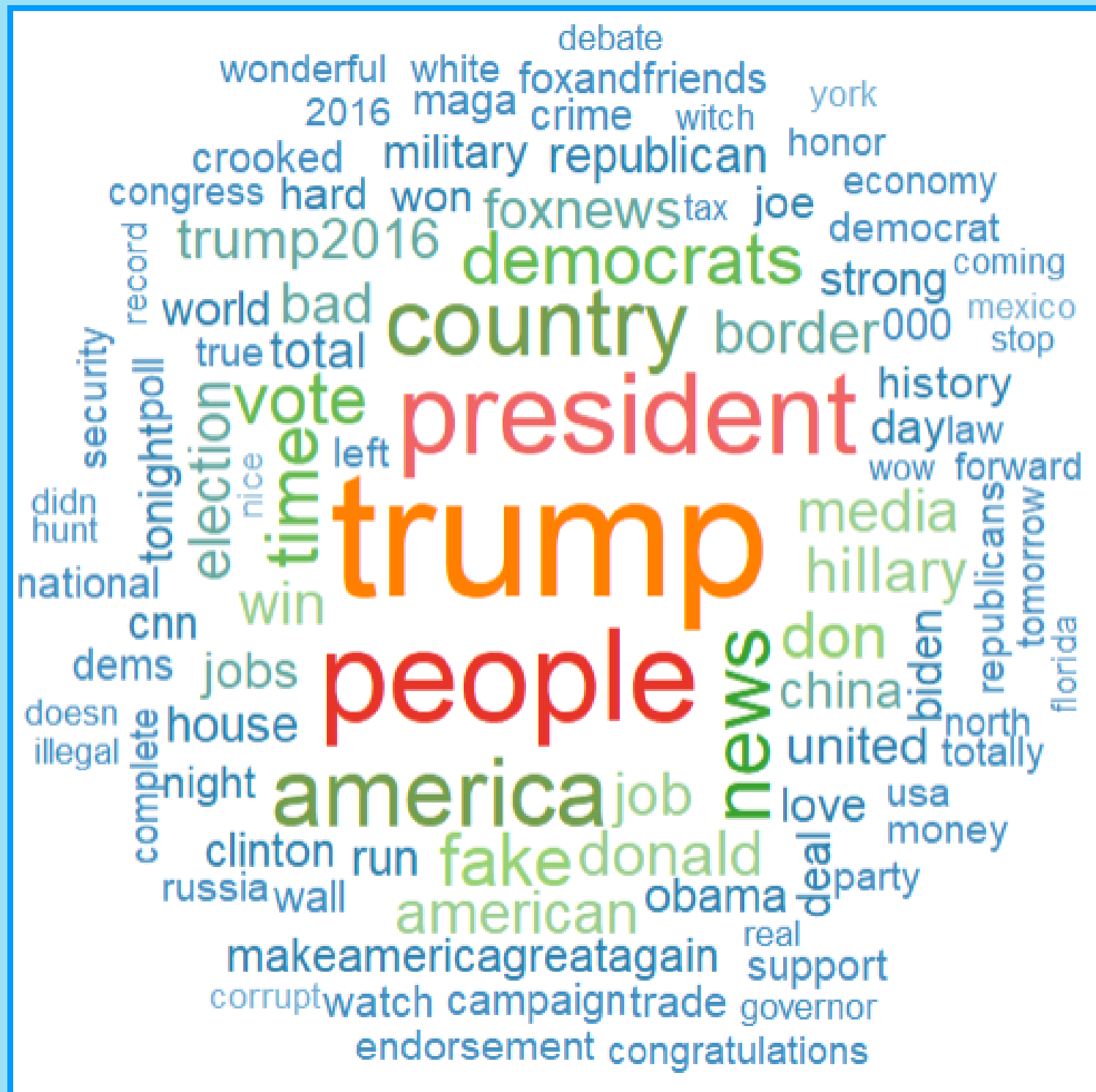
Abstract

A Trump Tweet that is insulting and negative has the best chance of receiving the most favorites. This discovery reveals that, in terms of user engagement with digital political media, emotional content maximizes interaction. In our research, we aim to uncover the factors that contribute to the popularity of a Trump tweet. To analyze our dataset of 26,312 tweets we will utilize feature generation methods to create variables based on the text data and we will apply machine learning techniques like Random Forest and Gradient Boosting to explore the relationship between different factors of a tweet and its success.
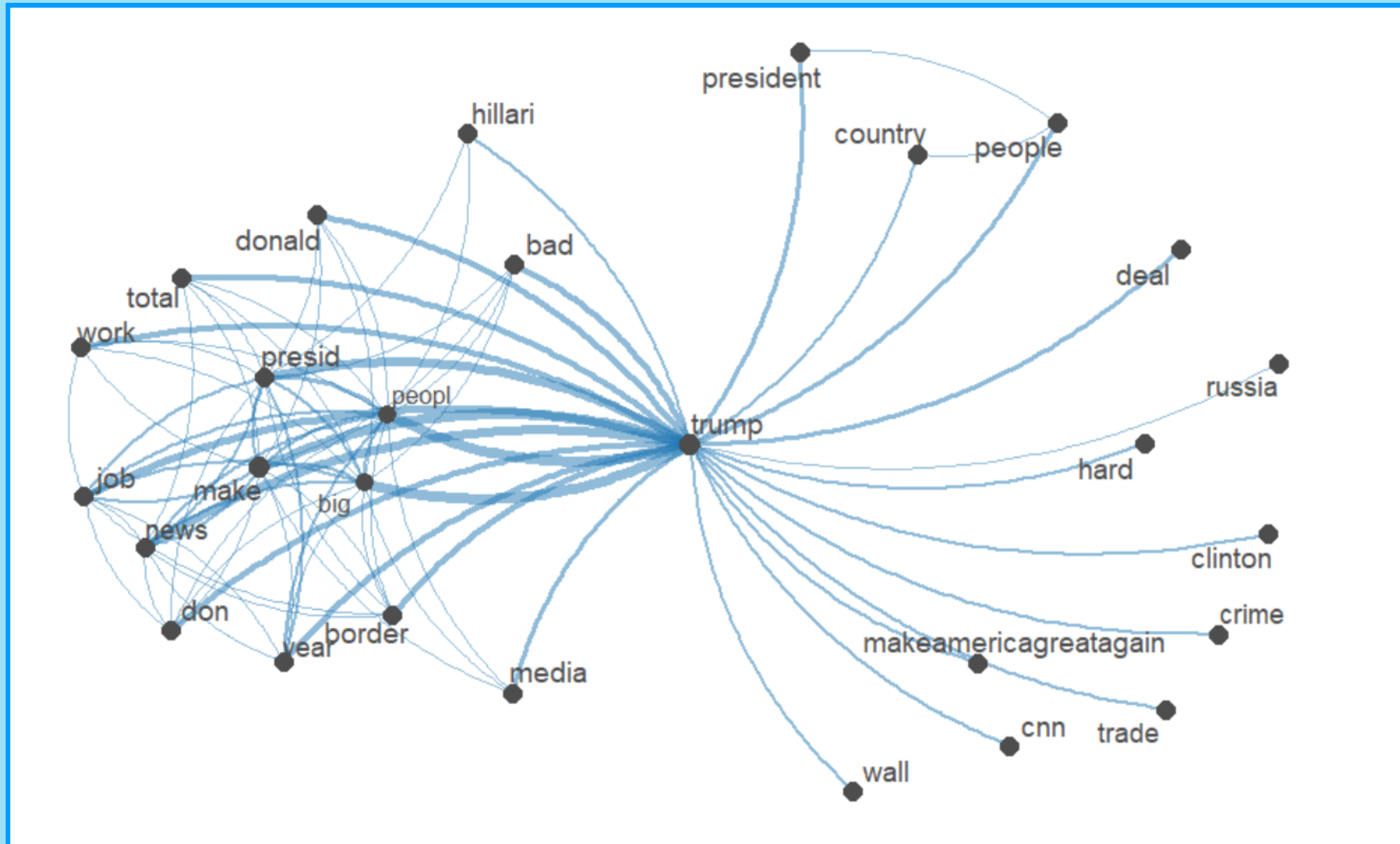
Features

- **ALL CAPS:** T/F for when the tweet is written in all caps. (True/False)
- **Time of Day:** When did Trump post the tweet? Morning, Afternoon, Evening, or Midnight. (Categorical)
- **Length of Tweet:** We created 4 buckets based on the word count of each tweet. (Categorical)
- **Political Words**: From our personally built list of 180+ political words Trump often uses, (i.e. "election", "presidency", "vote") here we measure the proportion of the words in each tweet that are political. (Numerical)
- **Insulting Words:** From our personally built list of 10+ insulting words Trump often uses, (i.e. "crooked", "weak", "sleepy") here we measure the proportion of the words in each tweet that are insulting. (Numerical)
- **Sentiment**: Using python libraries, we assessed a sentiment score for each tweet to see if the sentiment was positive or negative. (Numerical)
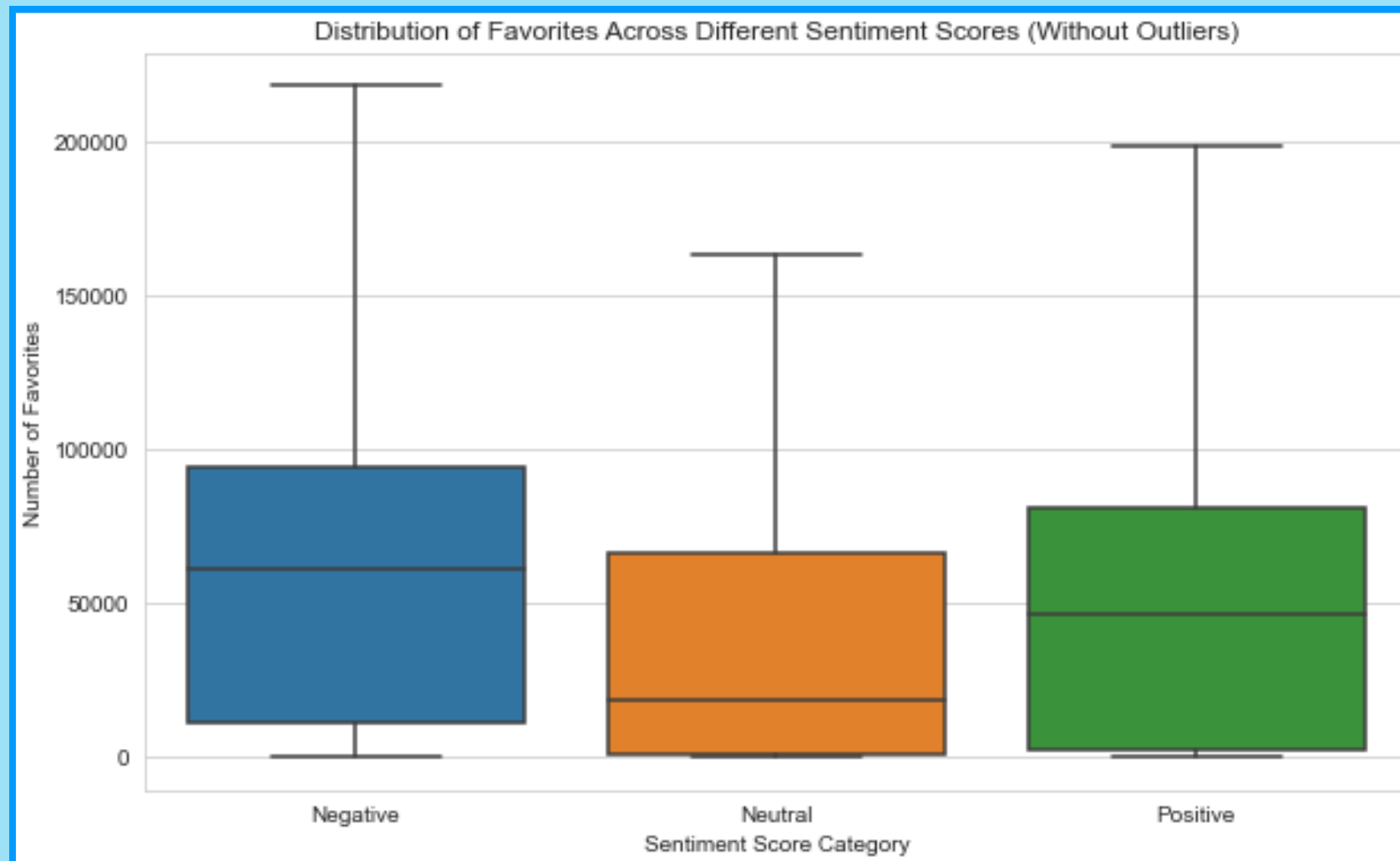
Text Mining

This word cloud shows the top one hundred most frequent words to appear in Trump tweets. The most commonly used words seem to either be referencing politicians (i.e. "trump", "hillary", "obama") or surrounding the topic of politics itself (i.e. "president", "vote", "democrats").

The text network displays connections between words, offering insights into how words generally relate to each other within tweets. The strongest word association occurs between the word "trump" to words like "president", "people", and "big".

EDA for Sentiment

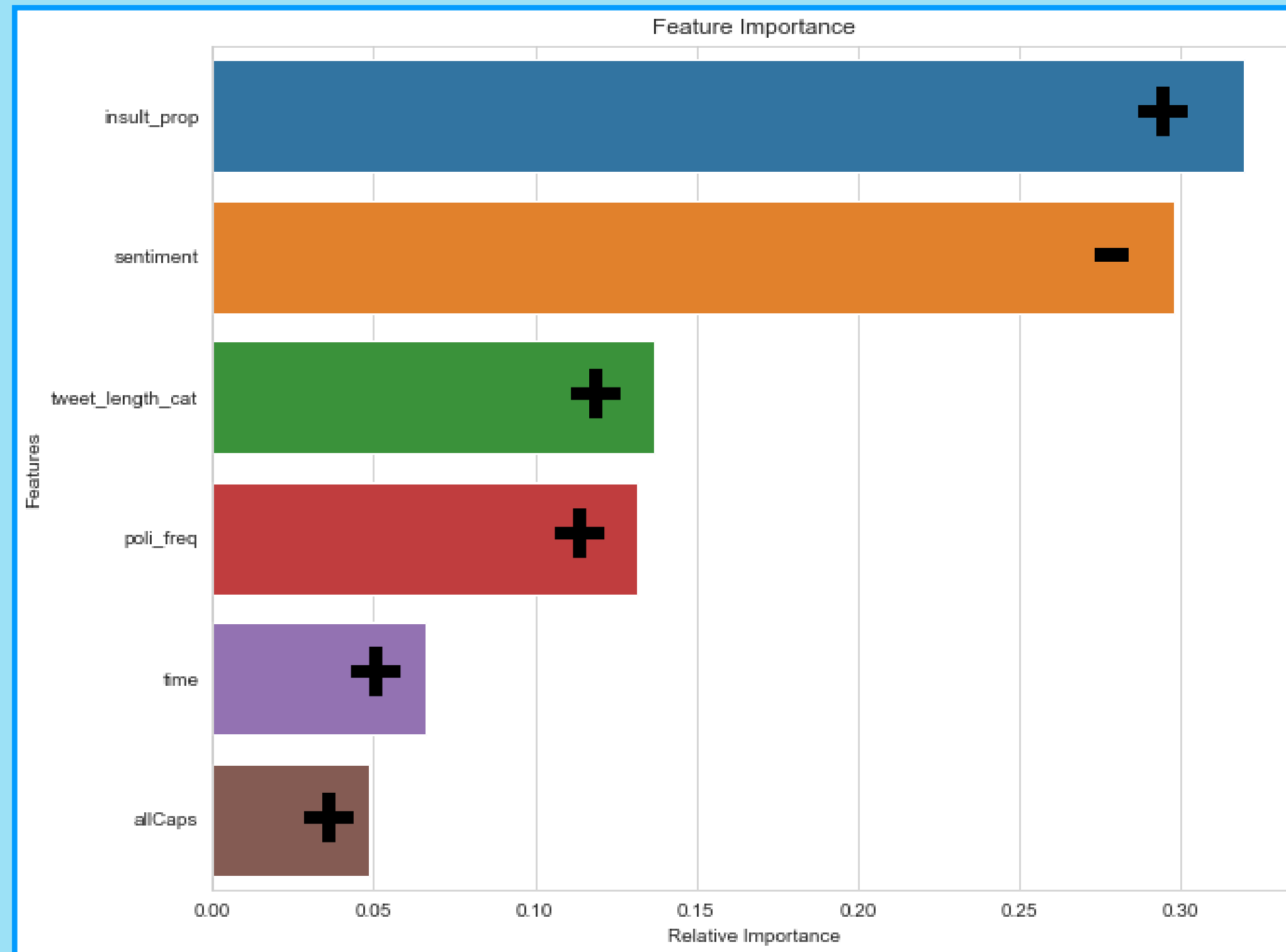Distribution of Favorites Across Different Sentiment Scores (Without Outliers)

The side-by-side boxplot shows that tweets expressing negative sentiments have a higher lower quartile, median, and upper quartile in the number of favorites compared to the other tweets. We observe that tweets expressing either negative or positive sentiments have a higher median number of favorites compared to tweets that are neutral. Such observation is reasonable because tweets expressing negative or positive sentiments often contain more emotions, thus causing Trump's followers to react strongly and favor the tweet more.

Statistical Analysis

Model	MSE	R-squared
Random Forest Regressor	0.3465	0.7283
Gradient Boosting Machine	0.2846	0.7768
Linear Regression	0.4074	0.6805
Decision Tree Regressor	0.5219	0.5907

Statistical Analysis

Feature Importance chart showing Relative Importance of Features: insult_prop (+), sentiment (−), tweet_length_cat (+), poli_freq (+), time (+), allCaps (+)

Results/Conclusion

<u>The top 4 variables for predicting Trump tweets' popularity are proportion of insulting words, proportion of political words, sentiment, and tweet length.</u>

The combination of heightened emotional expression, provocative language, and a strong political stance creates a recipe for increased attention and interaction. However, the variables are highly correlated so it is difficult to accurately isolate the effect of a single variable.

Ultimately, looking at the importance of features and their R-squared values as a result of our model, we can see that a Trump tweet that would receive the most favorites is **insulting** and has **negative sentiment**.
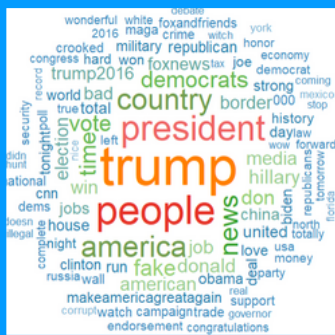
Abstract

A Trump Tweet that is insulting and negative has the best chance of receiving the most favorites. This discovery reveals that, in terms of user engagement with digital political media, emotional content maximizes interaction. In our research, we aim to uncover the factors that contribute to the popularity of a Trump tweet. To analyze our dataset of 26,312 tweets we will utilize feature generation methods to create variables based on the text data and we will apply machine learning techniques like Random Forest and Gradient Boosting to explore the relationship between different factors of a tweet and its success.
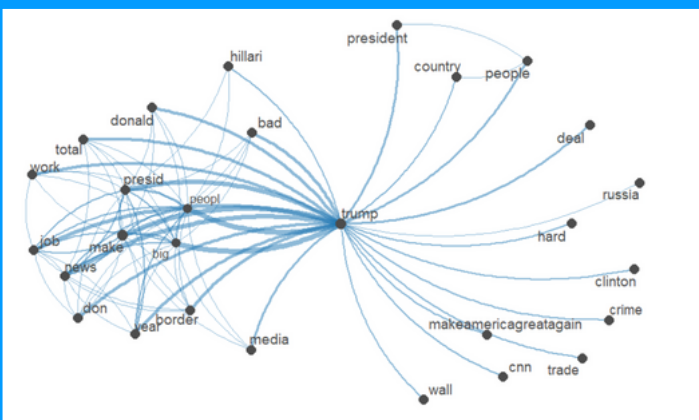
Features

- **ALL CAPS:** T/F for when the tweet is written in all caps. (True/False)
- **Time of Day:** When did Trump post the tweet? Morning, Afternoon, Evening, or Midnight. (Categorical)
- **Length of Tweet:** We created 4 buckets based on the word count of each tweet. (Categorical)
- **Political Words**: From our personally built list of 180+ political words Trump often uses, (i.e. "election", "presidency", "vote") here we measure the proportion of the words in each tweet that are political. (Numerical)
- **Insulting Words:** From our personally built list of 10+ insulting words Trump often uses, (i.e. "crooked", "weak", "sleepy") here we measure the proportion of the words in each tweet that are insulting. (Numerical)
- **Sentiment**: Using python libraries, we assessed a sentiment score for each tweet to see if the sentiment was positive or negative. (Numerical)

Text Mining

This word cloud shows the top one hundred most frequent words to appear in Trump tweets. The most commonly used words seem to either be referencing politicians (i.e. "trump", "hillary", "obama") or surrounding the topic of politics itself (i.e. "president", "vote", "democrats").

The text network displays connections between words, offering insights into how words generally relate to each other within tweets. The strongest word association occurs between the word "trump" to words like "president", "people", and "big".

Donald J. Trump ✔ @realDonaldTrump · 1hr

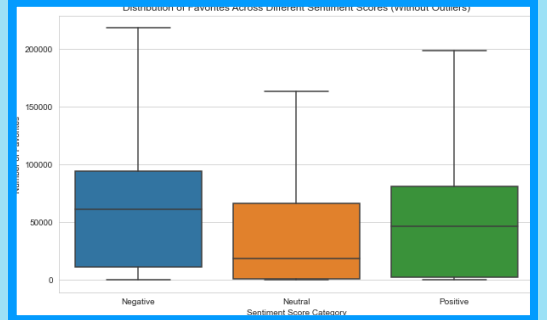# A Trump Tweet that is **insulting** and **negative** has the best chance of receiving the most favorites.

💬 82k   🔁 45k   ❤️ 91k   🔖 78k   📊 87K   ⬆️

"CROOKED"   "WEAK"   "DISHONEST"   "PUPPET"   "SLEEPYJOE"   "DOPEY"

"WITCH"   "CORRUPT"   "SLEEPY"   "PHONY"   "PATHETIC"   "FAT"   "CROOKEDHILLARY"

Team Mean Squares
Daphne, Ethan, Kevin, Nishant, and Mia
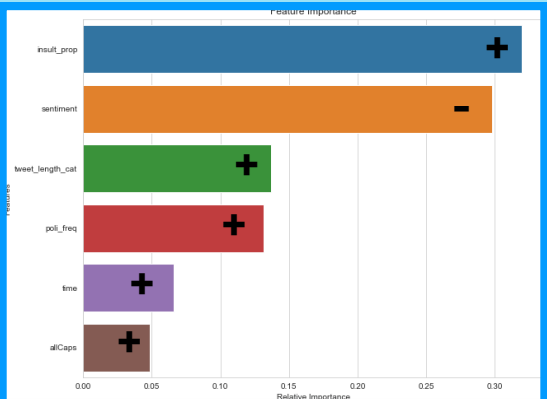*UCLA Statistics 140XP*

EDA for Sentiment

The side-by-side boxplot shows that tweets expressing negative sentiments have a higher lower quartile, median, and upper quartile in the number of favorites compared to the other tweets. We observe that tweets expressing either negative or positive sentiments have a higher median number of favorites compared to tweets that are neutral. Such observation is reasonable because tweets expressing negative or positive sentiments often contain more emotions, thus causing Trump's followers to react strongly and favor the tweet more.

Statistical Analysis

Model	MSE	R-squared
Random Forest Regressor	0.3465	0.7283
Gradient Boosting Machine	0.2846	0.7768
Linear Regression	0.4074	0.6805
Decision Tree Regressor	0.5219	0.5907

Results/Conclusion

The top 4 variables for predicting Trump tweets' popularity are proportion of insulting words, proportion of political words, sentiment, and tweet length.

The combination of heightened emotional expression, provocative language, and a strong political stance creates a recipe for increased attention and interaction. However, the variables are highly correlated so we cannnot accurately isolate the effect of a single variable. Ultimately, looking at the importance of features and their R-squared values as a result of our model, we can see that a Trump tweet that would receive the most favorites is insulting and has negative sentiment.