Is Twitter an accurate predictor of election results and is Proposition 30 likely to pass?

**Intro Information:**

To determine if Twitter is an accurate predictor of election results, we used web scrapping and sentiment analysis.

We scrapped 2,612 Tweets related to Proposition 30 which were posted between July 1, 2022 and November 8, 2022. We then used two sentiment analysis models (RoBERTa and VADER) to determine whether a Tweet was in support of Prop. 30 or not, as well as comparing the effectiveness of each model.

<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest>

The RoBERTa model we used is a deep learning model that was trained on “~124 million tweets from January 2018 to December 2021” and is finetuned for Twitter sentiment analysis. This model is much more powerful than the VADER model, so we expected it to perform much better. Thus, the sentiments determined by the RoBERTa model are what we used to answer our primary question.

<https://towardsdatascience.com/social-media-sentiment-analysis-in-python-with-vader-no-training-required-4bc6a21e87b8>

The VADER model is from the NLTK package and is a “lexicon and simple rule-based model for sentiment analysis”. It uses a dictionary to determine the sentiment scores of individual words and phrases and then calculates a compound score to rate the sentiment of the entire text. Since this model is not trained and only follows relatively basic rules, its performance is not as good as the RoBERTa model.

After determining the sentiments of the Tweets, we obtained the counts of each sentiment and compared those with the actual poll results using two-proportion z-tests.

**Conclusions:**

\*Plot of all counts\* \*Plot of counts w/o neutral\*

Based on the plot above we see that the VADER model resulted in primarily Positive sentiment, followed by Negative, and then Neutral. However, the RoBERTa model resulted in primarily Neutral sentiment, followed by Neagtive, and then Positive. If we remove the Neutral sentiments as in (other table w/o neutral), we see that the RoBERTa model resulted in slightly more Negative than Positive sentiments, whereas the VADER model resulted in over twice as many Positive sentiments than negative. Overall, we see that the two models resulted in opposite majorities and that the VADER model had a much larger majority.

We will now compare the results of the two models with the actual poll results.

\*Plot of percentages w/ poll\*

Based on the plot above we see that the proportion of each sentiment from the RoBERTa model are very similar to the proportions of the actual poll results. The proportion of each sentiment from the VADER model, however, are very different from the actual poll results.

To determine the statistical significance of these visible similarities and differences, we will perform several two-proportion z-tests. Our null hypothesis will be that the proportion of a specific sentiment from a model is equal to the proportion of that same sentiment from the poll. Our null hypothesis will be that that proportions are different.

The results of these z-test are shown in the table below.

\*table of p-values\*

From the table, we see that we failed to reject the null hypothesis for both tests involving the RoBERTa model. Thus, there was not significant evidence to conclude that the sentiment proportions from the RoBERTa model were different from the sentiment proportions from the poll.

We also see that we rejected the null hypothesis for both tests involving the VADER model. Thus, there was significant evidence to conclude that the sentiment proportions from the VADER model were different from the sentiment proportions from the poll.

From these results, we can conclude that RoBERTa model is far more accurate than the VADER model, and that the proportion of opinions on Twitter can accurately predict the results of an election.