**Election Result Prediction Using Twitter sentiment Analysis**

Abstract - The proliferation of social media in the recent past has provided end users a powerful platform to voice their opinions. Businesses (or similar entities) need to identify the polarity of these opinions in order to understand user orientation and thereby make smarter decisions. One such application is in the field of politics, where political entities need to understand public opinion and thus determine their campaigning strategy. Sentiment analysis on social media data has been seen by many as an effective tool to monitor user preferences and inclination. Popular text classification algorithms like Naive Bayes and SVM are Supervised Learning Algorithms which require a training data set to perform Sentiment analysis. The accuracy of these algorithms is contingent upon the quantity as well as the quality (features and contextual relevance) of the labeled training data. Since most applications suffer from lack of training data, they resort to cross domain sentiment analysis which misses out on features relevant to the target data. This, in turn, takes a toll on the overall accuracy of text classification. In this paper, we propose a two stage framework which can be used to create a training data from the mined Twitter data without compromising on features and contextual relevance. Finally, we propose a scalable machine learning model to predict the election results using our two stage framework.

Keywords— sentiment analysis, text classification, training data, labeling, Vader, Twitter

Existing System: An **election exit poll** is a poll of voters taken immediately after they have exited the polling stations. A similar poll conducted before actual voters have voted is called an entrance poll.

Disadvantages:

1. Data is less
2. Time consuming
3. More Efforts
4. We can not cover total population

Proposed System:

we have a dataset,which is not labelled. In order to assian label to dataset we require polarity of the tweet. Added polarity to each tweet and candidate names are asssigned as a classes to dependent variable y which we ahve to predict.Now, dataset is ready with independent variable x and dependent variable y. it can be used to train a supervised machine learning model to perform public sentiment analysis and predict election outcome. We split the dataset in 80:20 ratios to prepare the training and testing sets.

Built machine learning model using decision tree, logistic regression, naïve bayes, random forest

Built machine learning model using decision tree, logistic regression, naïve bayes, random forest

Steps of proposed system

* Data collection
* Data preprocessing
* Data splitting
* Model Building
* Model Evaluations

Advantages:

More Accuracy

We can reach more people within less time.

Modules:

Python

Pandas

Numpy

Matlpotlib

Textblob

Seaborn