Political Opinion Mining for Popularity Prediction

Problem Statement

The proliferation of social media in the recent past has provided end users a powerful platform to voice their opinions. One such application is in the field of politics, where political entities need to understand public opinion and thus determine their campaigning strategy. Also by the help of this popularity of a person can be predicted in Politics also which will help the party to understand the sentiment and opinion of public about their party member which can help them in wining a election. Sentiment analysis on social media data has been seen by many as an effective tool to monitor user preferences and inclination. This research proposes an approach that is based on Twitter based Political opinion mining for predicting the popularity of a person on a given set of Tweets containing varied opinion. The objective is to extract expressions of opinion and predict the personality of political member by classifying them as positive or negative; also this system is going to encounter Hinglish language which is mash up of Hindi and English language. This approach applies deep learning techniques to the task of sentiment analysis and opinion mining. In order to this recurrent neural network (RNN) is used. In this research Long Short-Term Memory Units are used and a full Tensorflow based Opinion and Sentiment classifier is made at the end. This approach is going to use the concept of Sentiment Analysis i.e. tracking opinion of public, which uses the natural language processing and extract the information like either public's view is positive or negative which can be used further to predict popularity of political arty member. The data that is to be taken here is from twitter tweets.

Background

Opinion mining sometimes also referred as sentiment analysis, it can be used for natural language processing. By the help of opinion mining mood of public about any product and person can be tracked. This process involves building a system which collect and categorize opinions about a person's popularity. Attitudes and feelings of public are tracked in an opinionated document with classifying it as either positive or negative according to the sentiment expressed in it. Automated opinion mining uses machine learning to mine text for sentiment.

Besides the challenges traditional sentiment analysis systems face additional difficulties like Short Length of text, Spelling Mistakes ,Special tokens like URLs, emoticons, Diversity of content, Different style of Language, Multilingual content, Slang words. Some approaches of sentiment extraction are based on supervised Learning, & unsupervised methods as well. In recent survey following are the methods that were explained for political opinion mining based on popularity Naïve Bayes this algorithm is based on Bayes theorem which uses conditional probability by counting the frequency of values and combinations of them in a data set. Text categorization works well with this approach. Support Vector Machine transform text into the format which matches into input of machine learning algorithm input. So this process includes preprocessing and transformation on text documents. SVM has been proved one of the powerful learning algorithms for text categorization. Decision trees classify the data by hierarchically

sorting them based on feature values. Most commonly used models for feature extractions are entropy and information gain measure. Maximum Entropy classifier is another model which performs probabilistic classification, making use of the exponential model. One major advantage of this classifier is that it makes no conditional independence assumption on the features of the documents to be classified, given a sentiment class. Hence, it is applicable to real-life scenarios, unlike in case of Naive Bayes.

Methodology

Step 1: Data collection and dataset preparation

This will involve collection of the data from Twitter and Politics and political attitude dataset is used. Tweepy - (API) is used to collect the data from twitter. Then, preprocessing is done on the dataset and features are extracted from the data. Also the data is divided into 2 parts testing data and training data.

Step 2: Developing A RNN (LSTM) based Model for with CNN for Political Opinion Mining For Popularity Prediction

In this step a RNN model with LSTM integrated is designed for performing the political opinion mining and predicting the popularity. Different hyperparameters are tuned for this model. This model will classify the opinions into 2 categories that is positive and negative opinion and then popularity of that political party member is predicted according to the classified opinion of public from Twitter.

Step 3: Training and experimentation on datasets

The RNN based model that consist LSTM for political opinion mining for popularity prediction will be trained on the training dataset to do Opinion Mining accurately and notify.

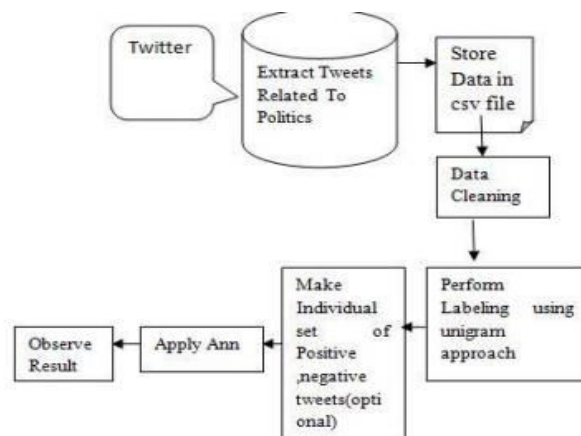**Step 4: Deployment and analysis on real life scenario**

Figure.1 Basic Methodology of Opinion Mining [Akash, Brahmbhatt, and Risha Tiwari. "Opinion Mining to Predict Election Results."]
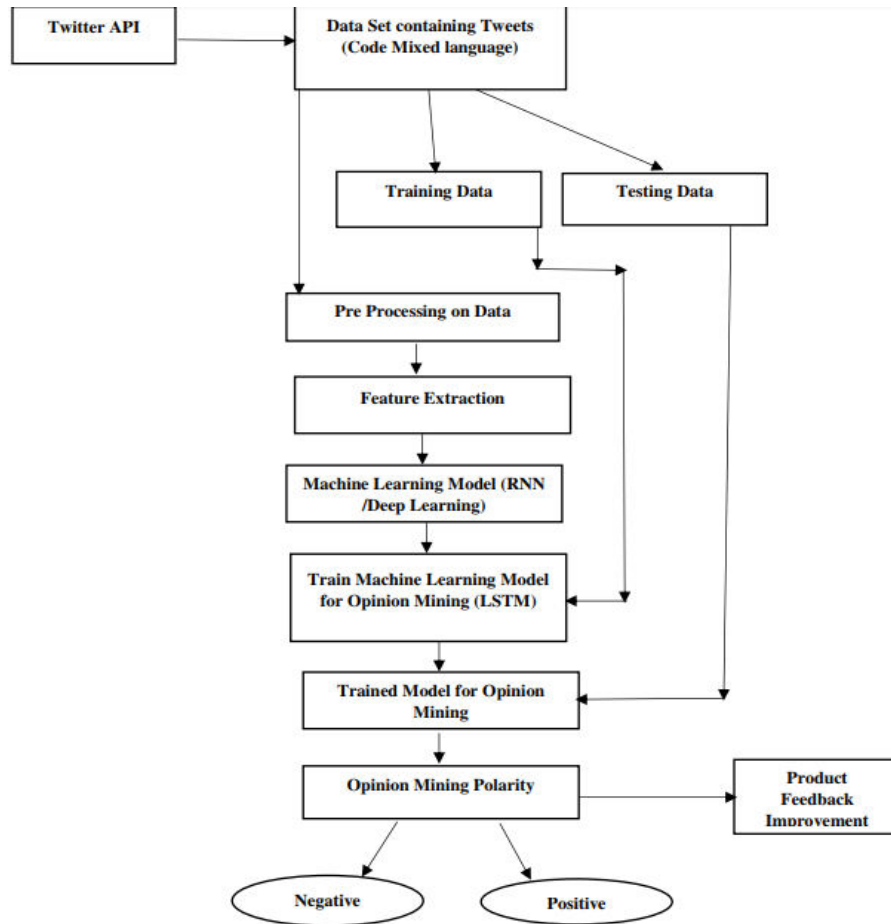
Figure.2 The Flow Chart of political Opinion Mining for Popularity Prediction.

The trained and tested popularity prediction based on political opinion mining model will be deployed in a real-life scenario to get the opinion of public and predict the popularity of political party member & will be leveraged for further improvement in the methodology and will follow the above architecture.

Experimental Design

Dataset

The Roper Center archive consists of over 23,000 datasets from public opinion surveys (https://ropercenter.cornell.edu/polls/dataset-collections/). Politics and political attitude dataset(http://libguides.library.qut.edu.au/c.php?g=428115&p=2923239) which will be used for experimentation and evaluation.

Evaluation Measures

Measures such as confusion matrix, accuracy and loss rate will be measured and evaluation of political opinion mining for popularity prediction will be done.

Software and Hardware Requirements

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python and libraries such as Tensorflow will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the above proposed system that contains a deep learning based approach for political opinion mining for popularity prediction.