Product Opinion Mining for Competitive Market Analysis

Problem Statement

Micro-Blogging website Twitter is a social networking website that has a large and speedily rise in the number of user using it. This research proposes an approach that is based on Twitter based opinion mining forproduct opinion mining for competitive market analysis on a given set of Tweets containing varied opinion. The objective is to extract expressions of opinion describing the product and classifies them as positive or negative, also this system is going to encounter Hinglish language which is mash up of Hindi and English language and do the competitive market analysis. 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 minimize the maintaining cost in advertisement of product. 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 service of companies can be tracked. This process involves building a system which collect and categorize opinions about a product. Attitudes and feelings of public is 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. Following are the methods used for opinion mining:

Naïve Bayes **t**his 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 modelwhich performs probabilistic classification, making use of the exponential model. One major advantage of thisclassifier is that it makes no conditional independenceassumption on the features of the documentsto be classified, given a sentiment class. Hence, itis applicable to real-life scenarios, unlike in caseof Naive Bayes.

Methodology
Step 1: Data collection and dataset preparation
This will involve collection ofthe data from Twitter and The Opinion Mining, Sentiment Analysis, and Opinion Spam Detection dataset is used. Tweepy - client for Twitter Application Programming

Interface (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 Product Opinion Mining For Competitive Market Analysis

In this step a RNN model with LSTM integrated is designed for performing the opinion mining. Different hyperparameters are tuned for this model. This model will classify the opinions into 2 categories that is positive and negative opinion.

Step 3: Training and experimentation on datasets

The RNN based model that consist LSTM for product opinion mining for competitive market analysiswill be trained on the training dataset to do Opinion Mining accurately and notify.

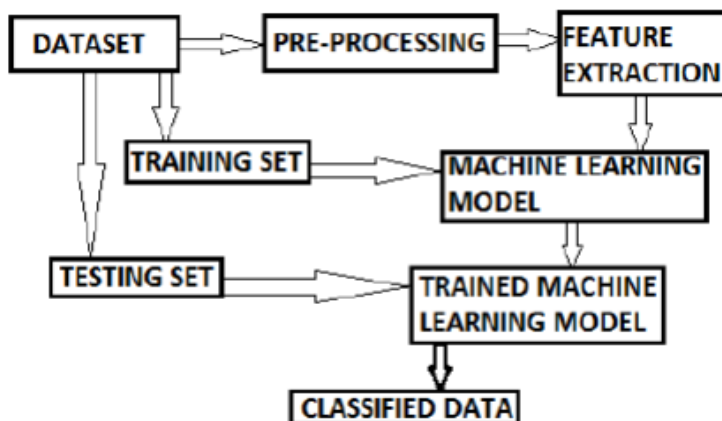**Step 4: Deployment and analysis on real life scenario**
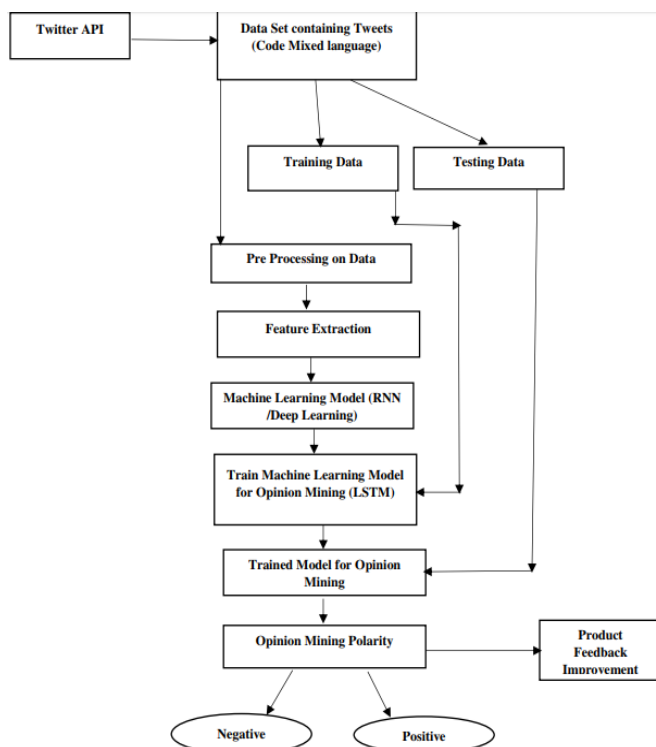

Figure.1 Basic Methodology of Opinion Mining

Figure.2 Shows The Flow Chart of Opinion Mining Based on Twitter

The trained and testedproduct opinion mining for competitive market analysismodel will be deployed in a real-life scenario to get the opinion of public on the product & will be leveraged for further improvement in the methodology and will follow the above architecture.

Experimental Design
Dataset
The Opinion Mining, Sentiment Analysis, and Opinion Spam Detection dataset (https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html),Twitter tweets from twitter APIwill be used for experimentation and evaluation.

Evaluation Measures
Measures such asconfusion matrix, accuracy and loss rate will be measured and evaluation of opinion mining for competitive market analysis 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 product opinion mining for competitive market analysis.