Chapter 10
Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis

Ishrat Nazeer
School of Computer Science and Engineering, Lovely Professional University, Jalandhar, India

Mamoon Rashid
https://orcid.org/0000-0002-8302-4571
School of Computer Science and Engineering, Lovely Professional University, Jalandhar, India

Sachin Kumar Gupta
School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Jammu, India

Abhishek Kumar
School of Computer Science and IT, Jain University, Bangalore, India

ABSTRACT

Twitter is a platform where people express their opinions and come with regular updates. At present, it has become a source for many organizations where data will be extracted and then later analyzed for sentiments. Many machine learning algorithms are available for twitter sentiment analysis which are used for automatically predicting the sentiment of tweets. However, there are challenges that hinder machine learning classifiers to achieve better results in terms of classification. In this chapter, the authors are proposing a novel feature generation technique to provide desired features for training model. Next, the novel ensemble classification system is proposed for identifying sentiment in tweets through weighted majority rule ensemble classifier, which utilizes several commonly used statistical models like naive Bayes, random forest, logistic regression, which are weighted according to their performance on historical data, where weights are chosen separately for each model.

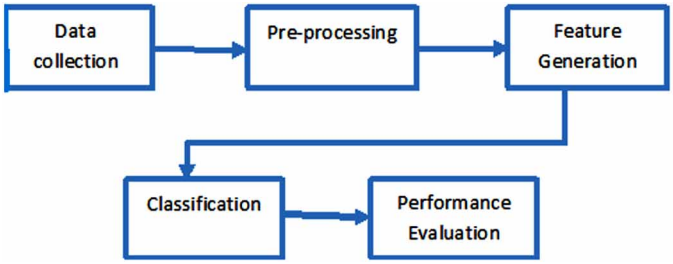
DOI: 10.4018/978-1-6684-6303-1.ch010

INTRODUCTION TO SENTIMENT ANALYSIS

In the current world of technology everyone is expressive in one or other way. People want to express their opinions about various issues be it social, political, economic or business. In this process social media is helping people in a great way. Social networking sites like Facebook, twitter, WhatsApp and many others thus become a common tool for people to express themselves. Analyzing the opinions expressed by the people on different social networking sites to get useful insights from them is called social media analytics. The insights gained can then be used to make important decisions. Among all the networking sites twitter is becoming most powerful wherein people express their opinions in short textual messages called tweets. Analyzing the tweets to retrieve insight information is called twitter sentiment analysis (SA) or opinion mining. Sentiment analysis classifies the sentiment of a tweet into three classes of positive negative and neutral (Ahuja, Ret al. 2019). Twitter sentiment analysis is helping the modern world in a great way as an example SA can help a company in knowing the customer reviews about a particular product and will help customers to select the best product based on opinion of people.

Figure 1 shows five main steps required in Sentiment Analysis.

Figure 1. General steps in Twitter sentiment analysis process



1. **Data Collection**: Process of SA begins by collecting the tweets from twitter using Application Programming Interface (API). API will allow us to interact with the twitter and extract the tweets in a programmatic way. The extracted tweets are then used for further processing,
2. **Pre-Processing**: Data preprocessing is done to remove extra features from the tweets. It decreases the size of tweets and makes them suitable for classification (Rane, A et al. 2018). The feature that are removed include following:
   a. The user name which is preceded by @ symbol.
   b. The retweets which are preceded by RT.
   c. Hashtags denoted by #.
   d. Slang words are replaced with words of equivalent meanings.
3. **Feature Extraction**: Feature extraction steps are responsible for extracting the features from the tweets. Different types of features are there like twitter specific features (includes features like hashtags, retweets, user names, URL), textual features (includes feature like length of tweet and length of words, emoticons, number of question marks), Parts Of Speech (features like nouns, verbs, adverbs, adjectives etc.), Lexicon Based features (comparison of positive and negative word percentages)(Permatasari, R. Iet al. 2018).