**Twitter Emotion Analysis**

By

**Hari Krishna 2015103017**

**ABSTRACT:**

These days, Social networking sites like twitter, facebook, etc. are the great source of communication for internet users. So these become an important source for understanding the opinions, views or emotions of people. In this paper, we use data mining techniques for the purpose of classification to perform emotion analysis on the views people have shared on Twitter, which is one of the most used social networking sites nowadays.

We collect dataset, i.e. tweets from Twitter and apply text mining techniques – transformation, tokenization, stemming etc. to convert them into a useful form and then use it for building emotion classifier. Here, we are using different classifiers on the data and then compare the results to find which one gives better accuracy and better results.

**Table of Contents**

**1 Introduction……………………………………….. 4**

**2 Related Works…………………………………….. 6**

**3 Requirement Analysis…………………………. 8**

**CHAPTER 1**

**INTRODUCTION**

**Problem domain:**

In the current world of business analytics, analysts are constantly trying to identify the information about their users that can help them with providing better services. But to collect this information, which has to be credible and reliable, it’s not a very easy task. Today with microblogging websites like Twitter which provides developers to collect the information of the users, we can easily collect the information of the users and perform emotion analysis and figure what the general audience feel about any particular problem they face or any product in the market.

**Problem Description:**

Given any tweet crawled from the internet this model should be able to analyze the given tweet and determine what emotion is associated with the tweet or what the user wants to convey through his tweet, be it a happy emotion or a sad emotion.

**Scope:**

There various micro blogging sites in this fast paced world where information is constantly exchanged over these sites and a lot of data about the people and their needs are available to all the mega companies. By analyzing this data that is freely available to us we can easily help the people of this world and make their lives easier in whatever way suits them.

**Contribution:**

With this project, I am adding a different look at text mining by using CNN models which are usually preferred for image classification. By learning and understanding the user data at a fundamental level I am trying to one up the already existing models.

**CHAPTER 2**

**RELATED WORKS**

This section contains a review of the work previously done in the field of sentiment analysis for the live data. A lot of work has been carried out till date in this field for the data from the users on social media in order to extract the sentiments of people towards any topic, products, trend etc. The studies focus on extracting useful information from the natural language of users and process it to get the real sentiments from the language. Osaimi and Badruddin have done a lot of work on the sentiment analysis of the tweets on the twitter in the Arabic language. In this, they build different classifiers by training them with a proper dataset and then analyzed the accuracy and result of these classifiers in order to predict the correct sentiments. Pragya Tripathi, Santosh Kr Vishwakarma, Ajay Lala have proposed the work on the sentiment analysis of English tweets using rapid minor. They collect the dataset from the twitter that is in natural language and applies the techniques of text mining and use it to build the sentiment classifier. O’Keefe et al. have proposed a technique to select the features attributes weight and applied two classifiers on it i.e. Naïve Bayes and SVM. In this work, the author obtained classification accuracy of 87.15% by using only 29% of the selected attributes. Pak and Paroubek have also worked in this field. The author used the data of Twitter to perform linguistic analysis and then build a classifier that is highly efficient. Pang and Lee presented the broad overview of the existing work done by Pak and Paroubek. The authors describe the existing techniques and approaches for an information retrieval, in their survey.

K.Bhuvaneswari and R. Parimala have proposed in their work, a method for sentiment classification using correlation-based feature selection. They applied different data pre-processing techniques, then used a correlation attribute method for feature selection, and then finally two classifiers namely Naïve Bayes and Support Vector Machine are implemented and results were evaluated. Farhan Laeeq, Md. Tabrez Nafis and Mirza Rahil Beg have proposed a work on sentiment classification of social media. In the work, they used three classifiers namely K-NN, Naive Bayes and Decision Tree Classifier for sentiment classification and obtained a result that shows the accuracy of K-NN, Naive Bayes, and Decision Tree classifier is 77.50%, 80%, and 78% respectively. Mangal Singh, Md.Tabrez Nafis and Neel Mani have worked on Similarity Evaluation and Sentiment Analysis on the Reviews for Heterogeneous-Domain product. They demonstrated scaling and sentiment classification with similarity evaluation among the reviews on the product. And the Review data is pre-processed and cleaned for the data preprocessing. Mnahel Ahmed Ibrahim and Naomie Salim have worked on the sentiment analysis of Arabic tweets extracted through Twitter, and then various classifiers like Naive Bayes, SVM, K-Nearest Neighbour are applied on data to find the best result. Eibe Frank and Albert Bifet proposed challenges that Twitter data define, focused on classification problems, and consider for sentiment analysis. Isabella et al., have proposed movie reviews for sentiment analysis and evaluated feature selectors to improve the performance of the classifiers.

**CHAPTER 3**

**REQUIREMENT ANALYSIS**

**Functional Requirements:**

The system should be able to correctly predict what emotion is associated with any input tweet fed into it. It should do the following things:

* Stream the twitter data live using the available API
* Pre-process the tweet removing all unnecessary characters
* Build a training model based off the tweets collected
* Correctly predict any incoming tweet and associate an emotion to it.
* Use emotion to identify to understand users emotion towards anything

**Non Functional Requirements:**

* Hardware:

No particular additional hardware is required for this project. A simple 64-bit processor laptop or desktop is enough to complete this project.

* Software:

Operating system: Windows

Programming Language: Python

Twitter API

Tools: Anaconda Navigator