"Emotion Analysis of Data"

Ms. RuchaPardeshi, Ms. BhagyashriPatil, Ms. ChetanaPatil, PranaliYenkar, Dr. S. D. Sawarkar

Datta Meghe College of Engineering Airoli, Navi Mumbai Maharashtra

ABSTRACT-Emotion detection is playing a very important role in our life. People express their emotions in different ways that is face expression, gestures, speech, and text. This research focuses on detecting emotions from the text. In this paper, we address the problem of detection, classification and quantification of emotions of text in any form. We consider English text collected from social media like Twitter, which can provide information having utility in a variety of ways, especially opinion mining. Social media like Twitter and Facebook is full of emotions, feelings and opinions of people all over the world. However, analyzing and classifying text on the basis of emotions is a big challenge. This paper proposes a method to classify text into six different Emotion-Categories: Happiness, Sadness, Fear, Anger, Surprise and Disgust. In our model, we use two different approaches and combine them to effectively extract these emotions from text. The first approach is based on Natural Language Processing, and uses several textual features like emoticons, degree words and negations, Parts Of Speech and other grammatical analysis. The second approach is based on Machine Learning classification algorithms. We have also successfully devised a method to automate the creation of the training-set itself, so as to eliminate the need of manual annotation of large datasets. Moreover, we have managed to create a large bag of emotional words, along with their emotion intensities. On testing, it is shown that our model provides significant accuracy in classifying tweets taken from Twitter.

Keywords – Emotion Analysis, Social Media, Machine Learning, Natural Language Processing, Emotion Category, Text-to-Speech Conversion.

I. INTRODUCTION

Emotion Recognition is the process of identifying human emotion from both Speech and verbal expressions. Emotions are described as intense feelings that are directed at something or someone in response to internal or external events having a particular significance for the individual. And the internet, today, has become a key medium through which people express their emotions, feelings and opinions. Every event, news or activity around the world, is shared, discussed, posted and commented on social media, by millions of people. Eg. "The Syria chemical attacks break my heart!! :' or "Delicious dinner at Copper Chimney! " or "OMG! That is so scary!". Capturing these emotions in text, especially those posted or circulated on social media, can be a source of precious information, which can be used to study how different people react to different situations and events.An emotion could be a combination of any of two or more basic emotions. But the algorithm analyses the given sentence and determines the emotion of the sentence based on the strength of the emotion in the given sentence. This is done by analyzing the sentence pattern and identifying the emotion which represents the given pattern of sentence.

II. PROBLEM STATEMENT

Emotion Analysis is Web Based Applications Focus on Single Tweet Only. With the rapid growth of the World Wide Web, people using social media such as Twitter which generates big volumes of opinion texts. The Problem in emotion analysis is classifying the polarity of a given text at the document, sentence or feature/aspect level. Whether the expressed opinion in a document, a sentence, or an entity feature/aspect is Fear, Sadness, Love, Joy, Anger, Surprise, etc. In any recognition task, the most common techniques are text-classification, Convolution Neural Network, and sequential model will be usedtheir use depends on factors such as availability of data, domain expertise, and domain specificity. In the case of emotion analysis, this task can be tackled using machine learning, or a concept-level approach. Here, we are exploring how we can achieve this task via a machine learning approach, One of the biggest challenges in determining emotion is the context-dependence of emotions within text. A phrase can have an element of anger without using the word "anger" or any of its synonyms. For example, the phrase "Shut up!".

III. EXISTING SYSTEM & ANALYSIS OF THE ISSUES

Text classification system in which they were using unstructured data for their classification which is very difficult to extract because of the large numbers of data which includes media file also so that it is not that easy to extract data from it but the system worked properly in this scenario also but it was more time consuming. Due to the unstructured data it was also not possible to separate the emotions or text according to their categories properly this was the another issue faced in this existing system. And the other issue was the user interface it was not that possible for a particular user to search for the respective categories of that particular text document for emotion analysis.

ANALYSIS OF ISSUES IN EXISTING SYSTEM

- Unstructured data.
- No proper Categories of articles
- No easy to use UI.

Unstructured Data:

Unstructured data in the form of text is everywhere: emails, chats, web pages, social media, support tickets, survey responses, and more. Text can be an extremely rich source of information, but extracting insights from it can be hard and time-consuming due to its unstructured nature. it's often very difficult to analyze unstructured data.

No proper categories of articles:

Due to the unstructured data on internet it is difficult and time consuming process to separate the articles according to the particular categories from that unstructured data(sports, electronics politics etc.) This is one of the issues faced on that system.

No easy to use UI (User Interface):

This is one of the issue faced by the user interface used but not that straight forward and quite easy to use classification was done through the user interface was done manually. Due to the manual process of classify the articles it was to time consuming and difficult and also sometimes expensive.
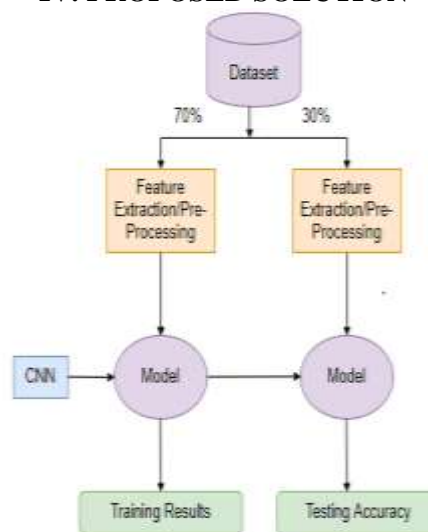
IV. PROPOSED SOLUTION

Fig.1 Proposed Solution Diagram

The dataset is divided into two parts. 70% of the dataset is used for training the model while 30% is used for testing purposes. The dataset is first preprocessed. The pre-processing of the dataset involves removing garbage value, null value or unwanted features from the dataset. After pre-processing necessary features are extracted from the dataset. This processed dataset is then used for training the model. The CNN algorithm used for training purposes.

For the training Phase, the dataset consists of tweets and a dependent variable which indicates whether the tweet is hate speech or not. As the machine learning approach for this project follows a supervised form of learning, the dependent variable is used for training the model.

The data type of this dependent variable is changed to an integer using the 'StringIndexer' function which takes in a column name and changes the datatype of that column.

After the first step, the tweets are tokenized, that is the tweets are split based on white spaces into words.

'StopWordsRemover' library is used to remove any stop word in the tokenized sets of words obtained from the above step. Stop words are words that do not make any contribution in identifying the sentiment of the tweet. For example 'a', 'an', 'but', 'house' etc does not have any sentimental value, so these words are removed from the tokenizing set of words.

'CountVectorizer' function will be used to count each word and store them in a dictionary format with the word and its count.

'IDF' function is used to calculate the tf-

idf value for a tweet.

After this pre-processing for both the training and testing phase, the data is fed to the machine learning model for further evaluation and result generation.The testing part of the dataset is inputted to the model for obtaining the final accuracy.

V. RESULTS

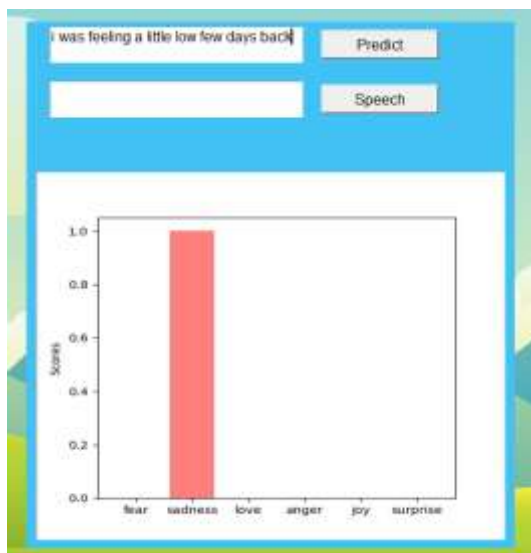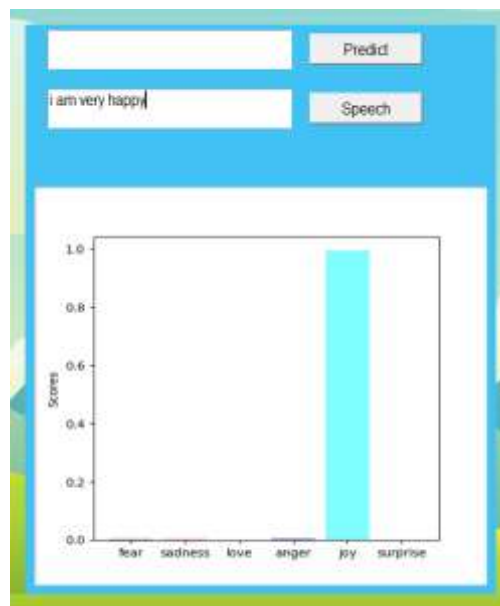**Fig.2**

Fig.3 Text Prediction

Fig.**4 Speech Conversion**

In this study, Keras and TensorFlow libraries were used for training CNN architecture and prediction of emotion states with proposed deep learning model. Intel I7 8300 CPU was used for all experiments and training custom dataset. Proposed CNN model was set with mentioned parameters. Fig.4 shows performance metrics (Score) of proposed architecture during training and testing. According to experiment results, accuracy is 85%.

VI. CONCLUSION

Emotion Detection can be seen as an important field of research in human-computer interaction. A sufficient amount of work has been done by researchers to detect emotion from text and audio information whereas recognizing emotions from textual data is still a fresh research area. The problems are faced by the emotion recognition system while processing raw text which contain both plain text and short messaging language. This paper addresses the existing different approaches for resolving processing of raw textual data which contain combination of both plain text and short messaging language. These types of system are applicable for different e-contents like Chats, Blogs, E-learning systems etc. This is successfully performed a comparative study of LSTM and CNN algorithm for text based emotion detection. From the results obtained after implementing this thesis we successfully conclude that CNN accuracy is much better when compared to LSTM algorithm.

REFERENCES

[1]. Alm, C.O., Roth, D., Sproat, R.: Emotions from text: machine learning for text-based emotion prediction. In: Proc. of the Joint Conf. on Human Language Technology/Empirical Methods in Natural Language Processing, 579–586 (2005).

[2]. Ekman, P.: An Argument for Basic Emotions. 169–200 (1992).

[3]. C. O. Alm, D. Roth, and R. Sproat. Emotions from text: machine learning for text-based emotion prediction. In Proceedings of the Human Language Technology Conference Conference on Empirical Methods in Natural Language Processing(HLT/EMNLP 2005), 2005.

[4]. SaimaAman and Stan Szpakowicz.Identifying expressions of emotion in text. In TSD, pages 196– 205, 2007.

[5]. M.M Bradley and P.J. Lang. Affective norms for english words (anew). University of Florida, 1999.

[6]. SM Mohammad, S Kiritchenko-Using hash tags to capture fine emotion categories from tweets - Computational Intelligence, (2015).

[7]. W Wang, L Chen, K Thirunaraya, AP Sheth: Harnessing twitter" big data" for automatic emotion identification: (2012)- IEEE.