IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECT AND ANALYZE THE EMOTIONS FROM TEXTUAL MESSAGES

ARANGI SAHITHI [#1], MADUGULA MURALI KRISHNA[#2] ,

[#1] M.Tech Student, Department of Computer Science and Engineering,

[#2]Assistant Professor, Department of Computer Science and Engineering,

Sri Sivani Colleges of Engineering, Chilakapalem Junction, Srikakulam-532402.

ABSTRACT

Emotion may be displayed in a variety of visible ways, including voice, written language, gestures, and facial expression. Text document emotion detection is essentially a content-based classification issue that incorporates ideas from both the machine learning and natural language processing fields. Most users utilise a variety of social media sites to convey their present feelings, which they may do through text messages, photographs, audio, and video. The user will often attempt to read our material before attempting to determine the present emotional state of that specific shared OSN user. In this study, we create a model that automatically recognises emotions from textual input and classifies them depending on those emotions on textual data is identified automatically and the appropriate emotion is classified based on the keywords which are entered in that input text.

1. INTRODUCTION

Due to the fact that most textual expressions do not only result from the direct use of emotion words but also from the interpretation of the meaning of concepts and interactions of concepts which are described in the text document, determining an individual's emotional state through the analysis of a text document written by that person can be difficult but is frequently necessary. In the connection between humans and computers, emotion recognition is crucial [1]. Speech-based emotion, facial expression-based emotion, and text-based emotion are three ways that people can convey their emotions. While sufficient research has been done on voice and face emotion identification, more attention has to be paid to text-based emotion recognition systems [14]. In.

Joy, sadness, wrath, surprise, hatred, fear, and many more emotions can be communicated. Since there is no established hierarchy of emotions words, the related study on emotion in the field of cognitive psychology is the major focus. In his 2001 book "Emotions in Social Psychology," W. Gerrod Parrot[2] defined the human emotion system and formally divided human emotions into six major groups, including love, joy, anguish, sadness, fear, and surprise. Other terms can also be classified as secondary or tertiary. This study suggests ways to enhance the functionality of existing text-based emotion recognition techniques

The primary issue is that there is currently no framework for categorising all the many types of expressions and feelings from text messages. This inspired me to create a programme that can accurately predict the sort of emotion from the input text provided to the system

ML models are being used to assess the information found in online social networks and attempt to extract the emotions from the text messages. Administrators will particularly benefit from being able to determine how many people are experiencing which emotions

We attempt to utilise Nave Bayes and LSTM models in our system to extract the emotion from text messages collected from the KaggleWen website

2. LITERATURE SURVEY

1) "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2001, pp. 32–80, doi: 10.1109/79.911197, by R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, and S. Kollias

Two channels have been distinguished in human interaction: one transmits explicit messages, which may be about anything or nothing; the other transmits implicit messages about the speakers themselves. Both linguistics and technology have invested enormous efforts in understanding the first, explicit channel, but the second is not as well understood. Understanding the other party's emotions is one of the key tasks associated with the second, implicit channel. To tackle that task, signal processing and analysis techniques have to be developed, while, at the same time, consolidating psychological and linguistic analyses of emotion. This article examines basic issues in those areas. It is motivated by the PKYSTA project, in which we aim to develop a hybrid system capable of using information from faces and voices to recognize people's emotions

2) Emotions in Social Psychology by W.G. Parrott, Psychology Press, Philadelphia, 2001

There is no area of social psychology that does not involve emotions. Not only has social psychology contributed enormously to theory and research on the nature of emotions, it also has emotions at the heart of its basic subject matter, from attitudes and dissonance to altruism and aggression. This reader presents a collection of articles on the nature of emotions and their role in social psychological phenomena, along with recent work that reflects the current state of the art. Articles have been selected and edited for readability, succinctness, and interest. For the beginning student, this reader serves as an introduction to the social psychology of emotions, and makes a useful text for advanced undergraduate and graduate courses on emotions, social processes, and related topics. It may also serve as a supplement to a general text on social psychology.

3) C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human machine communication," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, September 2008, pp. 1210–1215.

The ability to recognize emotion is one of the hallmarks of emotion intelligence. This paper proposed to recognize emotion using physiological signals obtained from multiple subjects. IAPS images were used to elicit target emotions. Five physiological signals: Blood volume pulse (BVP), Electromyography (EMG), Skin Conductance (SC), Skin Temperature (SKT) and Respiration (RESP) were selected to extract 30 features for recognition. Two pattern classification methods, Fisher discriminant and SVM method are used and compared for emotional state classification. The experimental results indicate that the proposed method provides very stable and successful emotional classification performance as 92% over six emotional states.

4) "Towards Text-based Emotion Detection: A Survey and Possible Improvements," International Conference on Information Management and Engineering, 2009, Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, and Von-Wun Soo

This paper presents an overview of the emerging field of emotion detection from text and describes the current generation of detection methods that are usually divided into the following three main categories: keyword-based, learning-based, and hybrid recommendation approaches. Limitations of current detection methods are examined, and possible solutions are suggested to improve emotion detection capabilities in practical systems, which emphasize on human-computer interactions. These solutions include extracting keywords with semantic analysis, and ontology design with emotion theory of appraisal. Furthermore, a case-based reasoning architecture is proposed to combine these solutions

5) J.G. Taylor and N. Fragopanagos, "Emotion recognition in human-computer interaction," Department of Mathematics, King's College, Strand, London, WC2 R2LS, UK (2005).

we outline the approach we have developed to construct an emotion-recognising system. It is based on guidance from psychological studies of emotion, as well as from the nature of emotion in its interaction with attention. A neural network architecture is constructed to be able to handle the fusion of different modalities (facial features, prosody and lexical content in speech). Results from the network are given and their implications discussed, as are implications for future direction for the research.

3. EXISTING SYSTEM & ITS LIMITATIONS

There was no appropriate way to extract an emotion from text in the current system. Emotion keywords may be used to easily identify linked emotions, however their meanings can be ambiguous and varied as most words can have several meanings depending on usage and context. Furthermore, in certain extreme circumstances, such as sardonic or cynical statements, even the bare minimal set of emotion labels (without all of their equivalents) might have a distinct feeling. The set of emotion keywords is the sole foundation of the keyword-based method. As a result, statements without any keywords might be interpreted as having no emotion at all, which is plainly incorrect. An example would be, "Hooray! I passed my qualified test today." It should indicate the same that "I passed my qualified exam today." emotion (joy), but the former without “hooray” could remain undetected if “hooray” is the only keyword to detect this emotion.

LIMITATIONS OF THE EXISTING SYSTEM

1. The process of extracting emotion from text takes longer.
2. There is no system in place that can organically process text messages.
3. There is no system that can identify emotions using NLTK.

4. PROPOSED SYSTEM & ITS ADVANTAGES

In this project, the dataset used has 7 different categories of emotions; these emotions will have smaller margins between them, so if we add more related emotions, accuracy will decrease even more. Here, we preprocess the data before converting this text data into numerical vectors to apply machine learning; initially, we use Nave Bayes Algorithm, which we found to have an accuracy rate of 84.4% training accuracy and 64% validati Similar to accuracy, we put the LSTM architecture to work to see which model will work best for our issue statement. Using the LSTM model We are extremely near to architectures, reaching 84% training accuracy and 63% test accuracy for both models. Given that Nave Bayes is a straightforward model that produces results more quickly, we recommend that it be used as the optimal model. For further information, see the following descriptions of internal data cleansing and preparation techniques.

ADVANTAGES OF THE PROPOSED SYSTEM

1) We can quickly identify the sentiment in text messages by employing a variety of modalities.

2) In this study, we review several studies that employ one or more algorithms for emotion recognition.

3) When utilising Nave Bayes to discover emotions, the suggested system performs the best.

5. PROPOSED MODELS

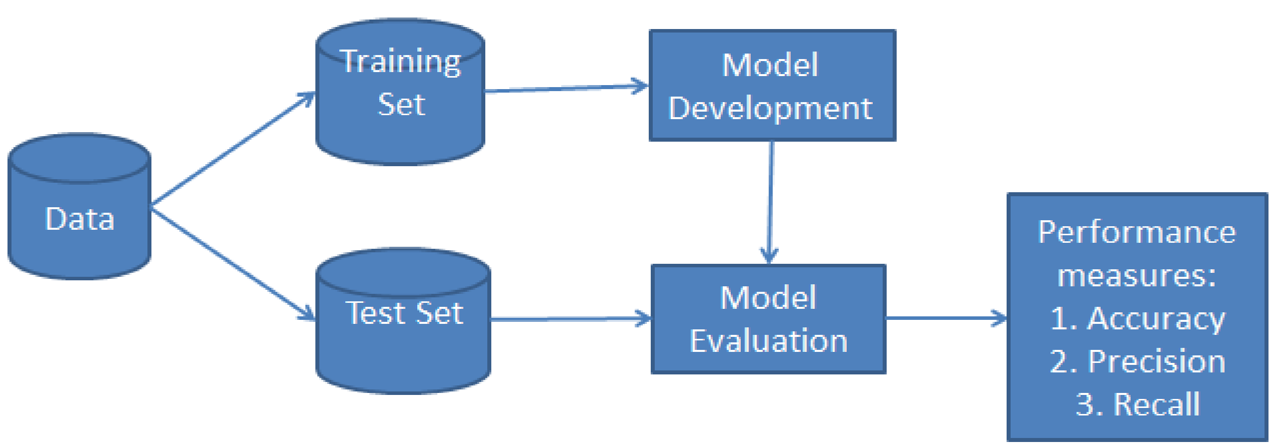
Machine Learning

A branch of artificial intelligence is called machine learning (AI). The main aim of machine learning is to comprehend the structure of data and fit it into models that people can comprehend and use. Despite being a branch of computer science, machine learning is distinct from conventional computational methods. Algorithms are collections of deliberately coded instructions that computers employ to do calculations or solve problems in conventional computing. Instead, machine learning techniques enable computers to train on data inputs and make use of statistical analysis to produce results that fall inside a particular range..

Classification Workflow

When doing classification, the initial stage is to comprehend the issue, find potential features, and assign a label to them. Features are those traits or qualities that have an impact on the label's outcomes. For instance, bank managers identify a customer's employment, income, age, location, past loan history, transaction history, and credit score in the event of a loan distribution. These traits are referred to as features that assist the model in categorising clients..

A learning phase and an evaluation phase make up the categorization process. A given dataset is used to train the classifier's model during the learning phase, and performance is assessed during the evaluation phase. Performance is assessed based on a number of factors, including memory, accuracy, and mistake..



Naive Bayes Classifier

A statistical classification method based on the Bayes Theorem is called naive Bayes. One of the easiest supervised learning methods is this one. The quick, accurate, and dependable approach is the naive Bayes classifier. On huge datasets, Naive Bayes classifiers perform quickly and accurately..

Naive The Bayes classifier makes the assumption that an individual feature's impact on a class is unrelated to the effects of other characteristics. For instance, a loan applicant's suitability depends on factors including their income, history of loans and transactions, age, and geography. These traits are nonetheless taken into account separately even though they are interconnected. This assumption is regarded as naïve since it makes calculation easier. It is known as class conditional independence.e.

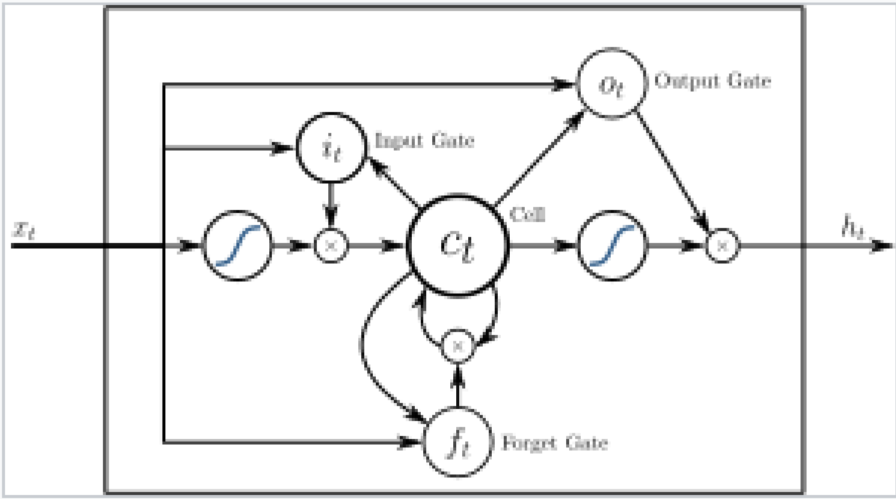
$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- P(h): the probability of hypothesis h being true (regardless of the data). This is known as the prior probability of h.
- P(D): the probability of the data (regardless of the hypothesis). This is known as the prior probability.
- P(h|D): the probability of hypothesis h given the data D. This is known as posterior probability.
- P(D|h): the probability of data d given that the hypothesis h was true. This is known as posterior probability.

LSTM (Long Short term Memory)

An artificial neural network called Long Short-Term Memory (LSTM)[1] is used in deep learning and artificial intelligence. LSTM features feedback connections as opposed to typical feedforward neural networks. Such a recurrent neural network may analyse complete data sequences in addition to single data points (such as photos) (such as speech or video). For instance, LSTM may be used for applications like networked, unsegmented handwriting identification, speech recognition, machine translation, robot control, video games, healthcare, and more. [11] The most frequently used neural network of the 20th century is LSTM. [12]

A cell, an input gate, an output gate, and a forget gate make up a typical LSTM unit[13]. [14] The three gates control the flow of information into and out of the cell, and the cell remembers values across arbitrary time periods. Since there may be delays of uncertain length between significant occurrences in a time series, LSTM networks are well-suited to categorising, processing, and generating predictions based on time series data. To solve the vanishing gradient problem[15] that might occur when training conventional RNNs, LSTMs were created. In many cases, LSTM has an advantage over RNNs, hidden Markov models, and other sequence learning techniques due to its relative insensitivity to gap length..



A peephole LSTM unit with input (i.e. i), output (i.e. o), and forget (i.e. f) gates.

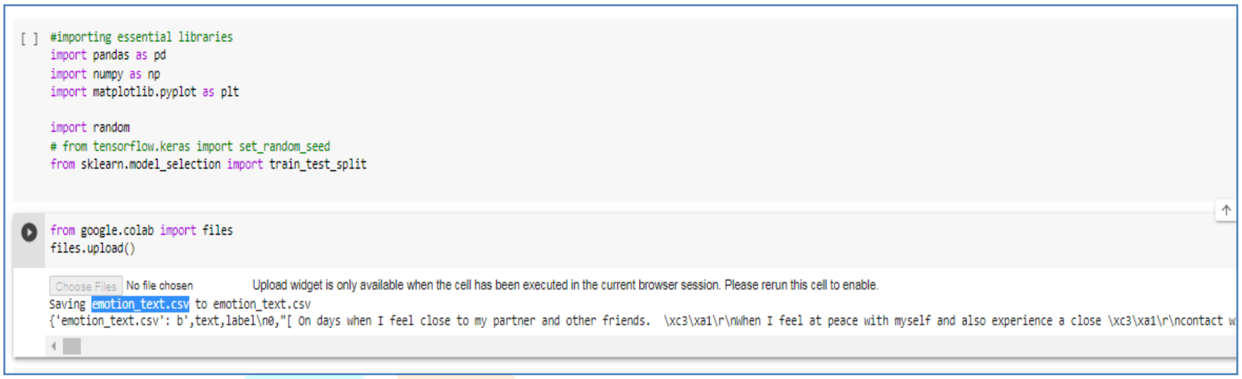
6. IMPLEMENTATION PHASE

The step of implementation is when the theoretical design is translated into a programmatically-based approach. The application will be divided into a number of components at this point and then programmed for deployment. The application's front end uses Google Collaboratory, while for the back end database, we used the dataset emotion text. Python is being used in this case to implement the existing programme, which is primarily separated into the following 6 modules. They are listed below:

1. Import Necessary Libraries
2. Load the dataset module.
3. Data preparation
4. List the terms for emotions
5. Extract the Text's Emotion
6. Comparative Analysis

1) IMPORT NECESSARY LIBRARIES

We must first import all the relevant libraries into this module in order to create the model. Here, we make an effort to employ every library available for converting data in a useful way. We try to import the numpy module since the data is separated into numerical values that the system can quickly identify, and we use the matplot library to plot the data in graphs and charts..



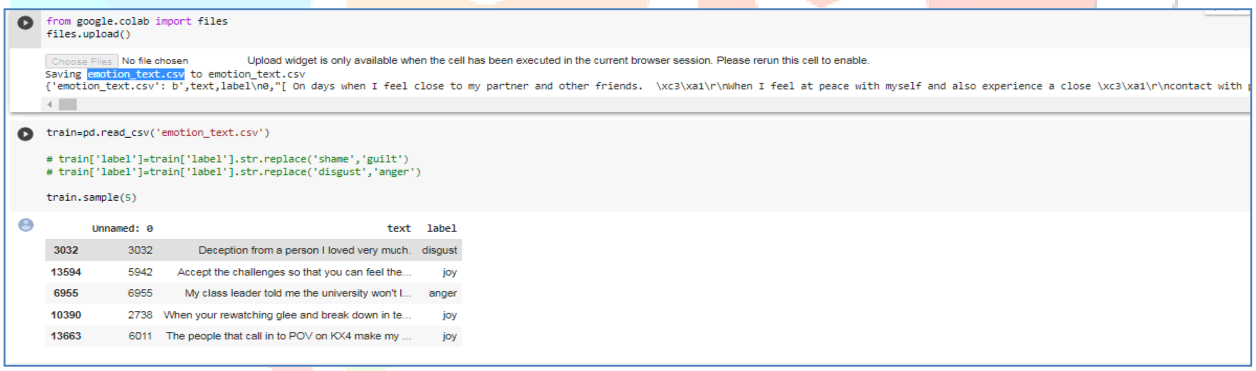
```
#importing essential libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

import random
# from tensorflow.keras import set_random_seed
from sklearn.model_selection import train_test_split
```

```
from google.colab import files
files.upload()
```

```
Choose Files No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving emotion_text.csv to emotion_text.csv
{'emotion_text.csv': b',text,label\n0,"[ On days when I feel close to my partner and other friends.  \xc3\xa1\r\nWhen I feel at peace with myself and also experience a close \xc3\xa1\r\ncontact w
```

2) LOAD DATASET MODULE

We attempt to load the dataset that was downloaded or gathered from the UCI repository in this module. The dataset in question is called "emotion text" and it includes the following data, including :



```
from google.colab import files
files.upload()
```

```
Choose Files No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving emotion_text.csv to emotion_text.csv
{'emotion_text.csv': b',text,label\n0,"[ On days when I feel close to my partner and other friends.  \xc3\xa1\r\nWhen I feel at peace with myself and also experience a close \xc3\xa1\r\ncontact with
```

```
train=pd.read_csv('emotion_text.csv')

# train['label']=train['label'].str.replace('shame','guilt')
# train['label']=train['label'].str.replace('disgust','anger')

train.sample(5)
```

	Unnamed: 0	text	label
3032	3032	Deception from a person I loved very much.	disgust
13594	5942	Accept the challenges so that you can feel the...	joy
6955	6955	My class leader told me the university won't l...	anger
10390	2738	When your rewatching glee and break down in te...	joy
13663	6011	The people that call in to POV on KX4 make my ...	joy

Every property comprises data that has been evaluated and gathered based on conversations with various users..

3) DATA PRE-PROCESSING MODULE

In this part, we try to perform a pre-processing operation on the incoming dataset to identify any missing values or incomplete data. In the event that such information is included in the dataset, the programme will disregard it and only load rows that contain all valid inputs..

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
train['label_encoded'] = le.fit_transform(train['label'].astype(str))
train.head()
```

	Unnamed: 0	text	label	label_encoded
0	0	[On days when I feel close to my partner and ...	joy	4
1	1	Every time I imagine that someone I love or I ...	fear	2
2	2	When I had been obviously unjustly treated and...	anger	0
3	3	When I think about the short time that we live...	sadness	5
4	4	At a gathering I found myself involuntarily si...	disgust	1

```
[ ] from sklearn.feature_extraction.text import CountVectorizer
    from nltk.tokenize import RegexpTokenizer

    # def naive_bayes_preprocessing(df):
    token = RegexpTokenizer(r'[a-zA-Z0-9]+')
    cv = CountVectorizer(stop_words='english',ngram_range = (1,1),tokenizer = token.tokenize)
    text_counts = cv.fit_transform(train['text'])
        # return text_counts
```

4) TRAIN THE MODEL USING ML ALGORITHMS

Here, we attempt to train the present model on a particular dataset using a variety of ML classification methods in order to determine which algorithms work best for properly and effectively identifying and classifying the input dataset. Here, we try to apply the subsequent algorithms on the supplied dataset, including:

1. LSTM
2. Naïve Bayes

Using LSTM

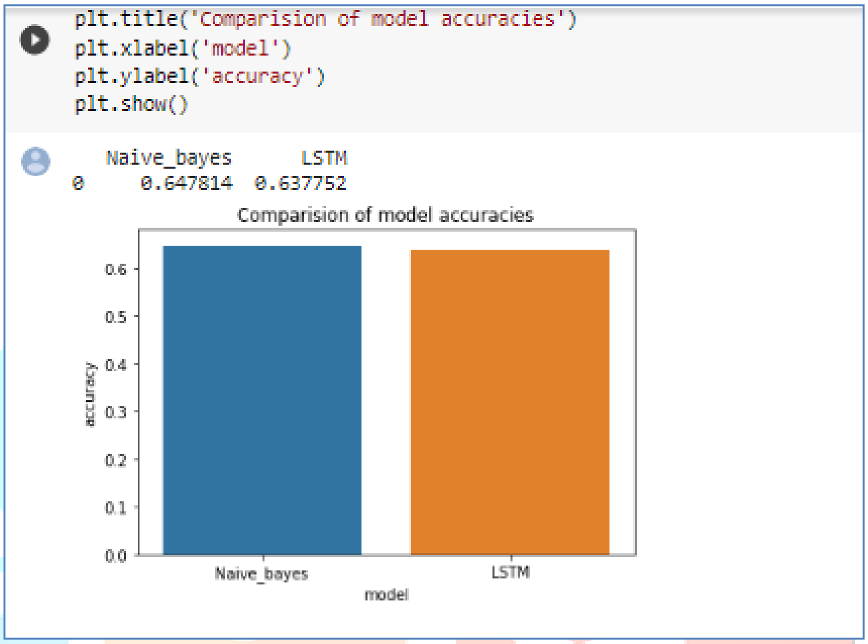
```
[ ] from tensorflow.keras.preprocessing import sequence
    from tensorflow.keras.preprocessing.text import Tokenizer
    from tensorflow.keras.layers import Dense,Dropout,Embedding,LSTM,Bidirectional
    from tensorflow.keras.callbacks import EarlyStopping
    from tensorflow.keras.losses import categorical_crossentropy
    from tensorflow.keras.optimizers import Adam
    from tensorflow.keras.models import Sequential
    from tensorflow.keras.utils import to_categorical
```

```
import nltk
nltk.download("popular")
nltk.download('stopwords')
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
lemmatizer = WordNetLemmatizer()
from bs4 import BeautifulSoup
import re
from nltk.corpus import stopwords # Import the stop word list
# stopwords.words("english")
```

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/cmudict.zip.
[nltk_data]    | Downloading package gazetteers to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/gazetteers.zip.
[nltk_data]    | Downloading package genesis to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/genesis.zip.
[nltk_data]    | Downloading package gutenberg to /root/nltk_data...
[nltk_data]    |   Unzipping corpora/gutenberg.zip.
```

5) PERFORMANCE ANALYSIS MODULE

On a given input dataset, we compare each classification method in this module in an effort to determine which one is most effective at producing correct results. Finally, we will choose the method that produces the most accurate results in the shortest amount of time. As we can see, compared to other algorithms, Nave Bayes provides more accurate results..



7. CONCLUSION

Human-computer interaction research in the area of emotion detection is crucial. While enough research has been done to identify emotions from visual and auditory data, textual data emotion recognition is still a new and active study topic. In this research, existing techniques for text emotion analysis are evaluated, along with their drawbacks, and a novel system design that would work effectively is offered

8. REFERENCES

[1] "Emotion recognition in human-computer interaction," in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2001, pp. 32–80, doi: 10.1109/79.911197, by R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, and S. Kollias

[2] Emotions in Social Psychology by W.G. Parrott, Psychology Press, Philadelphia, 2001

[3] C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human machine communication," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, September 2008, pp. 1210–1215.

[4] "Towards Text-based Emotion Detection: A Survey and Possible Improvements," International Conference on Information Management and Engineering, 2009, Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, and Von-Wun Soo

[5] J.G. Taylor and N. Fragopanagos, "Emotion recognition in human-computer interaction," Department of Mathematics, King's College, Strand, London, WC2 R2LS, UK (2005).

[6] C. Elliott's doctoral dissertation, "The Affective Reasoner: A Process Model of Emotions in a Multiagent System," May 1992, Northwestern University, Evanston, IL.

[7] "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, number 2, June 2006, pp. 165–183, doi:10.1145/1165255.1165259

[8] Z. Teng, F. Ren, and S. Kuroiwa, "Recognition of Emotion using SVMs," Lecture Notes in Artificial Intelligence, Volume 4114, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer, Berlin Heidelberg, 2006, pp. 701–710, doi: 10.1007/11816171 87.

[9] Using online blog corpora, C. Yang, K. H.-Y. Lin, and H.-H. Chen classified emotions. IEEE Computer Society, Nov. 2007, pp. Proc. IEEE/WIC/ACM International Conference on Web Intelligence..

[10] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining Acoustic and Language Information for Emotion Recognition," in Proceedings of the Seventh International Conference on Spoken Language Processing (ICSLP02), 2002, pp. 873–876.

[11] "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, number 2, June 2006, pp. 165–183, doi:10.1145/1165255.1165259

[12] C. Elliott's doctoral thesis at Northwestern University in Evanston, Illinois, published in May 1992, "The Affective Reasoner: A Process Model of Emotions in a Multiagent System,"

[13] Protégé tool, available at www.protege.stanford.edu ,

[14] Multimodal Approaches for Emotion Recognition: A Survey by Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S. Huang, USA.

[15] http://sail.usc.edu/kazemzad/emotion in text cgi/DAL app/index.php?overall=bad&submit evaluation=Submit+Query.

[16] http://www.wikipedia.org /