Approval

A thesis titled (Multiclass Emotion Classification : A deep learning approach) submitted by Abid Saleh to the Department of Computer Science and Engineering, University of Chittagong, Bangladesh has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc.(Engg.) in Computer Science and Engineering and approval as to its style and contents board examiners.

——————

Supervisor

Dr. Hanif Seddiqui

Professor,

Department of Computer Science and Engineering

University of Chittagong, Bangladesh.

Acknowledgement

Bismillahir Rahmanir Rahim. To begin with, I'd like to thank the Almighty Allah for keeping me in good health and bestowing me with the ability to engage myself in this thesis work.

Then, I'd like to thank my supervisor Professor Dr. Md. Hanif Seddiqui. I'm very much grateful to my supervisor for being so patient, kind and helpful towards me whenever I needed his help.

Not to forget, my seniors whom I've consulted with time to time and they've inspired me with their support . Also my thesis partners were a great driving force for my work.

Finally, we are especially grateful to our family and friends for being there for me.

Multiclass Emotion Classification: A Deep Learning Approach

Abstract

The usage of internet is increasing exponentially, so is online communication as a result. People are adopting different ways of communication over internet to satisfy their need to communicate And each of these communication sessions include wide range and variety of emotions in them. Most of the communications available in public are in textual format rather than audio or video format. This is because people are discovering and adopting new ways to communicate their emotions properly with text. As text communication is getting popularity, the size of text data is also increasing exponentially. Thus, an automated tool is needed to extract emotion in text to reduce misunderstanding and promote expressive communication. Various rule based approach have been proposed to extract the emotion of text automatically. Those works used emotion intensity lexicon to extract the emotion of text. But creating emotion intensity lexicon is tedious and time consuming process. Moreover, absence of hard and fast rule for annotating words with intensity make this task more challenging. Various works based on Machine approach to sentiment classification are done. Those works, involved only finding positivity, negativity and or neutral property of text. In this work we analyzed knowledge based Machine Learning approach to extract the emotion of the text. The system depends on examples rather emotion intensity lexicon. We have experimented Naïve Bayesian classifier Recurrent Neural Network and Support Vector Machine with Bigram Thus an automated Learning modelling for finding emotion from text. In our setup, RNN-LSTM outperformed other classification algorithms with promising accuracy.

Chapter 1: Introduction

1.1 Emotion Detection – An introduction

Emotion is a solid inclination getting from one's conditions, state of mind, or relationships with others. Emotions assume imperative role in human intelligence, decision-making, social association, recognition, memory, learning, imagination, and that's just the beginning.
"During the 1970s, psychologist Paul Eckman identified six basic emotions that he suggested were universally experienced in all human cultures. The emotions he identified were happiness, sadness, disgust, fear, surprise, and anger. He later expanded his list of basic emotions to include such things as pride, shame, embarrassment, and excitement" (Source: Wikipedia) These days they have additionally pulled in the consideration of researchers of computer science, particularly in the field of human computer collaborations. The focal point of research in emotion analysis is processing the emotions to identify the opinionated information rather than mining and retrieving factual information which is the target of most of the research up until now in natural language processing and textual analysis. Emotion recognition is the fresher region of textual analysis and these lines don't have so many strong methods. While board theme of Emotion has been concentrated in psychology for years, next to no exertion has been spent on detecting emotion from text. In this work, we accept that Emotion response of an input sentence is spoken to by its word appearance.

1.2 Emotion Detection vs Sentiment analysis

Though "Sentiment" and "Emotion" are synonyms, they do not refer to the same thing.
If we look at the dictionary, a "sentiment" is defined as an opinion or view. And the term "emotion" refers to "a strong feeling deriving from one's mood" .

As we can see, sentiment and emotional analysis are two different evaluation methods of people's moods. They both aim to better perceive the writers, and provides insights about their emotional responses. Although, when it comes to how they accomplish it, each sector has its own ways.

1.2.1 Sentiment Analysis

The goal of sentiment analysis is to capture the overall feel or impression individuals get from consuming a product. It doesn't specialize in the precise articulate emotions.
It rather depends on a rather simple binary system of "positive" and "negative" responses. We have a tendency solely to look to know if the customer had a positive or negative feedback with the content.
The influence of emotions don't count in this case. A simplified analysis technique presents information that are easy to process and quantify.
This technique proved its efficiency in acquiring important information about both customers and content. On one hand, it assists businesses to know the preferences and inclinations of the customers. If one gets a 60% negative feedback on a product review, he'd know that his customers are not quite interested in that product.
On another hand, obtaining sentiment feedback helps us better evaluate our content. If one gets a positive response that person would realize that he is doing something right, and he should do more of it.
Overall, sentiment analysis help us find the proper content/audience match.

1.2.2 Emotional Analysis

Contrary to sentiment analysis, the emotional analysis depends on a rather complicated and subtle system.
Here the former one uses a simplified binary classification and the latter depends on a deeper analysis of human emotions and responses. This method focuses on the nuances between the various feelings customers express. It's a additionally meticulous, thorough checking into the degrees and intensities related with the expressions of each emotion.

Contradictory to sentiment analysis, emotional analysis is comprehensive and considerate of quite a few variations of human mental subjectivities. It's mostly based on a wide range of moods rather than a number of static categories. Emotions like happiness, satisfaction, or excitement – are associated with positive, depending on how it's configured.

Emotional analysis looks further into subjects motives and impulses. It offers valuable and precise insights that can easily be converted into actions.
If one gets a higher number of confusions as a response, that person should know that ne need to write simpler and clearer content. If the prominently detected emotion is boredom, he will need to add some storytelling or jokes. If the audience is angry, he knows that he should consider a different perspective.

Here in our work we have focused on the rather complex task of emotional analysis.

1.3.1 Emotion detection in Bangla text

Bangla (or Bengali) is an Indo-Iranian language spoken in the Indian subcontinent. With over 250 million speakers, bengali is the seventh most spoken native language in the world. It is the primary language in Bangladesh and second language in India. Over the internet world, a lot of people write or read bangla text, many e-commerce site are also in bangla text. A lot of people in many social media site, such as facebook, twitter etc also use bangali text.

People may joke that others spend too much time on the internet, but this intricate series of tubes has become an important part of everyday life so much that its become a human rights violation to take it away.

These social networking sites and other platforms lead to the generation of petabytes of data per week. Some of the popular and widely used social networking platforms with its brief overview.

Facebook- This statistic presents the number of daily active Facebook users as of the fourth quarter of 2018. During this period of time, it was found that 1.52 **billion** active users visited the social network on a daily basis. Overall, daily active users accounted for 66 percent of monthly active users.[1]

Twitter- “With more than 321 million monthly active users worldwide as of the fourth quarter of 2018, Twitter is one of the biggest social networks worldwide.[2]

Goegle+ It has about 2.2 billions users.

The statistics mentioned above gives an idea about the rate at which the web has been increasing. With such vast data generated regularly, it provides enormous business opportunities to handle this data safely and precisely.

1.3.2 Emotion detection from English text

Detecting emotion from text is useful in understanding users’ feelings towards particular discussion in intelligent learning system. To test our algorithm, we use ISEAR (International Survey on Emotion Antecedents and Reactions), dataset collected by Klaus R. Scherer and Harald Wallbott [3]. ISEAR dataset contains seven major emotions: joy, fear, anger, sadness, disgust, shame, and guilt.
The process to classify sentences in this work involves two main steps: representing. 95% of dataset to allow learning,

1.4 Main Challenges

Because of the relatively new field, there are many challenges to be faced. According to the current techniques are just identification emotional expression and Topic Identification Mainly these challenges are related to the authenticity of the extracted data and the methods used in it. Reference [4] also discusses some issues of opinion mining. A summary of challenges of opinion mining is as follows:

1.4.1 Unstructured Data: Unstructured Sentiments are an informal and free text format, the writer does not follow any constraints. The data available on the internet is very unstructured, there are different forms of the data talking about the same entities, persons, places things and events. The web contains data from different sources varying from books, journals, web documents, health records, Companies logs, internal files of an organization and even data from multimedia platforms comprising of texts, images, audios, videos etc. The diverse sources of the data makes the analysis more complex as the information is coming in different formats.

1.4.2 Noise (slangs, abbreviations):

The web content available is very noisy. In todays era of 140 characters texting, for their ease people use various abbreviations, slangs, emoticons in normal text which makes the analysis more complex and difficult. Now a day people write their text in small format. For example If a people want to write awesome the he write it asm or osam e.t.c.

1.4.3 Foreign Word:

There are many foreign word mixing in bangla text. For example -(I miss you) . The are no exact bangla meaning some time it's a challenging work identifying emotion or opinion from this type of text.

1.4.4 Contextual Information:

Identifying the context of the text becomes an important challenge to address. Based on the context the behavior/use of the word changes in a great aspect

Ex-1: I was sick the other day.

Ex-2: I am sick of working 6 days a week.

Ex-3: Their activities make me sick.

In all the above 3 examples, the same word “sick” is used. In ex-1 ‘sick’ means being unwell, which expresses the emotion sadness. In ex-2, usage of the ‘sick’ expresses anger of the speaker. In ex-3, ‘sick’ associates to the emotion ‘disgust’. It’s clear that same word with same meaning can have multiple usage depending on the context. So, it becomes important to detect the context to find the subjective information in a text.

1.5 Motivation

Understanding emotions, analyzing situations and the sentiments associated with it is the natural ability of a human being. But how efficiently can we train a machine to exhibit the same phenomenon becomes an important and vital question to be explored and answered.

Emotional analysis provides an effective mechanism for understanding individual’s attitude, behavior, likes and dislikes of a user.

Its believed that when a child is small the mother knows very well what and when he/she is going to need, at what time the child drinks, eats or even the time of difficulty when it cries and the possible causes for that. She very well knows the difference in the cry for food and cry for getting the diapers changed, thus the mother can analyze and take the necessary action very well.

Analogous to this small story, how well our lives would be if what we want can be automatically analyzed, suggested and provided to us without putting much efforts? Emotional analysis provides us with the services and products we want of our taste at our ease. With e-commerce business spreading at a great speed the task of mining opinions on various products becomes an useful resource to guide and help people in making choices and decisions. Mining sentiments and subjective information helps to provide products and services in a personalized fashion and as per individuals taste and likings. With more emphasis laid on personalized information it becomes necessary and important to go about catering information to an individual, based on his likings and taste. The study of emotional analysis also provides enough information about how human beings perceive and express information in the form of text to express their feelings and emotions. This wide multi-dimension aspects discussed above, motivated me to take this problem as my Research Problem Today the Web has become an excellent source of opinions with the explosive growth of social media (e.g. reviews, forum discussions, blogs, micro-blogs, Twitter, comments, and postings in social network sites). Individuals and organizations are increasingly using the content in these media for decision making. Nowadays, if one wants to buy a consumer product, one is no longer limited to asking ones friends and family for opinions because there are many user reviews and discussions in public forums on the Web about the product. For an organization, it may no longer be necessary to conduct surveys, opinion polls, and focus groups in order to gather public opinions because there is an abundance of such information publicly available.

1.6 Application

The internet has become a rich platform for people to express their opinion, attitude, feeling and emotion. From this point of view, Web is an important source of product reviews, news reviews, blog reviews, movie reviews, stock market reviews, travel advice, social issue discussions, consumer complaints, etc. Nowadays, Bangla has been using widely in the Web. Automatic sentiment classification will become very useful in above applications Sentiment analysis is now a great interest to the social networking media such as Twitter, Facebook, Google+ as well.

Emotional analysis is widely used for understanding subjective nature of text or writer opinion. There are few area where emotional analysis are applied –

1.6.1 Decision making System

What other people think has always been an important piece of information for most of us during the decision making process. Long before awareness of the World Wide Web became widespread, many of us asked our friends to recommend, an auto mechanic or to explain whom they were planning to vote for local elections, requested reference letters regarding job applicants from colleagues, or consulted Consumer Reports to decide what dishwasher to buy. But, the Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well known professional critics - That is people we have never heard of. And conversely more and more people are making their opinion available to strangers via the internet.

The interest that individual users show in online opinions about products and services and the potential influence such opinions wield, is something that vendors of these items are paying more and more attention to. Thus, aside from individuals, an additional audience for systems capable of automatically analyzing consumer sentiment, as expressed in no small part in online venues, companies are anxious to understand how their products and services are perceived. **1.6.2 Recommendation Systems**

Most of the websites, we visit have a recommendation system built to assist us, online media, entertainment, music, film, Industry to other forms of art this system uses our personal information, previous history, likes and dislikes and our friend's information to make a suggestion.

1.7 Proposal

In this report we tried to focus on extracting emotion from the text. There are a lot of method for detecting emotion from text, We tried to classify data instances as one of the following emotion classes for Bengali dataset:

1) Happy

2) Sad

3) Disgust

4) Anger

5) Surprise

6) Fear

Before classifying, we have done some preprocessing on this data.

And for English dataset, classes are:

1) Joy

2) Anger

3) Fear

4) Disgust

5) Shame

6) Sadness

7) Guilt

1.8 Report organization

This report is organized in 5 chapters.

In Chapter 2 we discuss the related work done in the corresponding area.

In chapter 3 we depict the methodology of our proposed method

In chapter 4 we evaluate the performance of our system and compare that with other systems.

In chapter 5 we come up with a conclusion.

Chapter 2: Literature Review

2.1 Previous Work

Affective computing researchers are inspired by the challenging issues involved in detection, interception and representation of affect. Emotional expression can be expressed by different means such as written text, speech, facial expression, posture and different psychological activity. Different studies are done in all of the means by researchers[5] Different ways are also used to display the calculated emotion. Researchers attempted to display the emotion of text through still image, simplified text and avatars [6], [7],[8], [9]. The study in [10] found that using avatar give some feeling of presence of partner along with better understanding of emotion.

- The architecture of the proposed in [6] contains two stages: training stage, and classification stage. The training stage happens on the server side. They applied the dominant meaning methods [6] on the ISEAR dataset [11] to form the hierarchy tree. Based on the ISEAR, the tree consists of seven concepts: joy, fear, anger, sadness, disgust, shame, and guilt.

  The classifier unit receives two types of information. A hierarchy tree for dominant meaning for seven classes and ISEAR examples. The classifier in general uses a large amount of labeled training data for text classification, which is a labor- intensive and time-consuming task. In contrast, their approach is to construct the dominant meaning tree and then use this tree to classify incoming examples from Emotion Models unit. This unit contains two types of set of words. First, set coming from Emotion Agent, which extract some features from Chatting GUI unit during the chatting between users, remove stop words, and reformulate in the way Emotion unit can deal with it. Stop words are those that occur commonly but are too general—such as "the" , "an" , "a" , "to" , etc. The algorithm removed the stop words from the collection. Emotion agent use Emotion Algorithm to assign an emotion for each set of features based on the emotion models coming from emotion models unit. After determining the emotion,

Emotion Expression assigns a suitable expression for it and sends it to be shown in the Chatting GUI (see Figure 1)

- **Constructing Emotion Dominant Meaning Tree**

To represent the proposed approach to classify sentiment, suppose that the collectionconsists of m emotion. Given the limited set of examplesfor each emotion, we tried to represent the collection as a hierarchy of dominantmeanings.In this definition, each emotion is represented by a finite set of examples.

The question now is how can they use those examples toconstruct dominant meanings of the corresponding emotion? In other words, thoseexamples include some words that almost come with the corresponding emotion.The challenge is how to determine those words.Each example is represented by a fixed set of words.

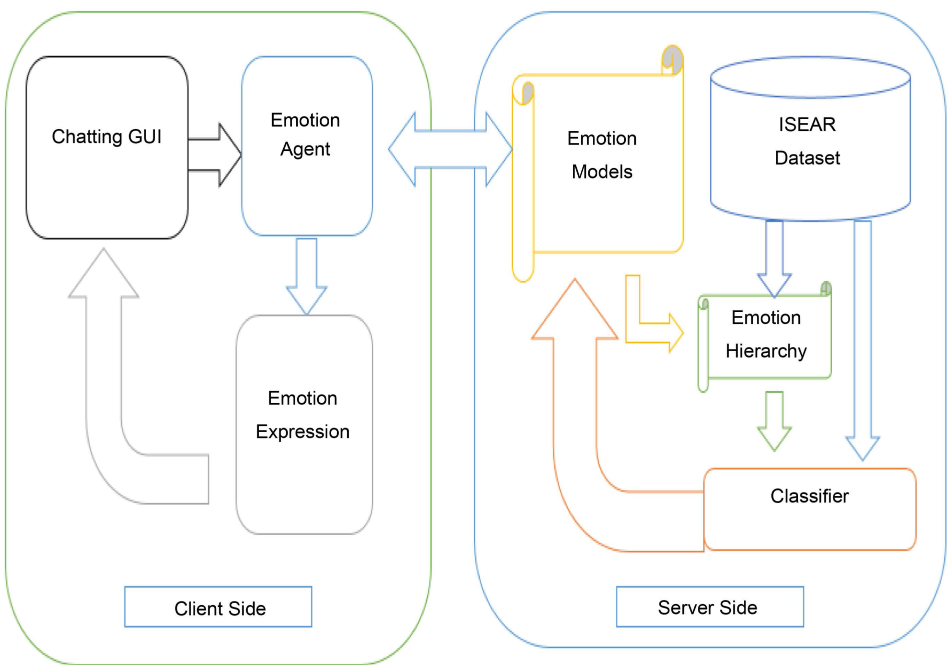


Figure 1: Architecture of the emotion detection system in [6]

- **Constructing Emotion Dominant Meaning Models**

The proposed system creates sevens models one for each emotion: joy, fear, anger, sadness, disgust, shame, and guilt. For each emotion ζ_k , we have a collection of N examples For each collection. After applying proper formulas, They get a set of dominant meanings each word in the set has P_{ki} value for a word w_i and in emotion ζ_k . Thry rank the terms of collection $\{ P_{k1}, P_{k2}, \ldots , P_{kn},\}$ in decreasing order.

As a result, the dominant meanings of the emotion ζ_k can be represented by the set of words that is corresponds to the set $\{ P_{k1}, P_{k2}, \ldots , P_{kn},\}$; i.e . $\zeta_k^n = \{ w_{k1}, w_{k2}, \ldots . w_{kn},\}$

Therefore, they select the top-N values of P_{ki} to form motion dominant meaning tree (EDMT). EDMT represents seven emotions suggested by (Klaus, 1994) as a tree. Each emotion is joined with a slave word. This slave is represent a dominant meaning and associated with the dominant meaning probability of that emotion as shown in Figure 2. In their paper they put the top-N of P_{ki} values as an arbitrary value.

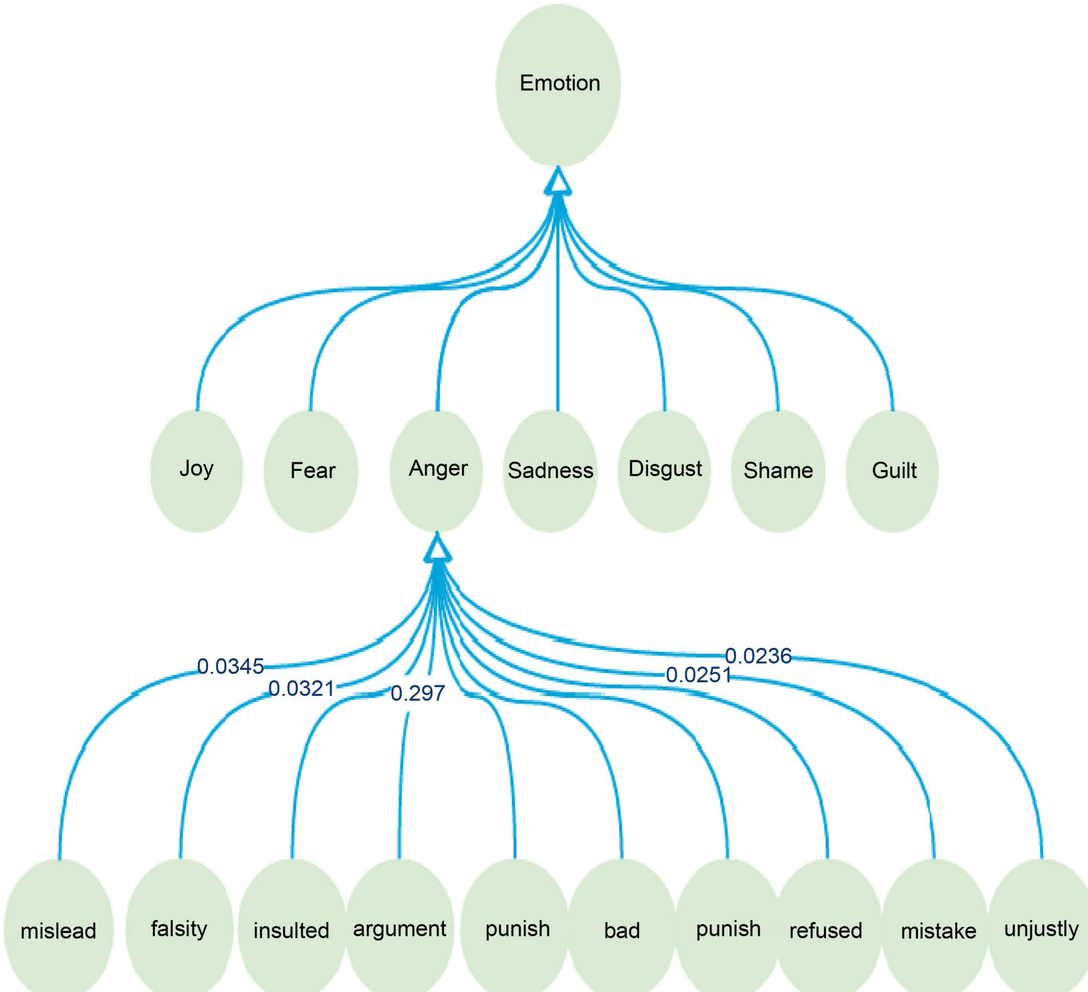


Figure 2: Emotion dominant meaning tree in [6]

Accordingly, they created seven models to represent the emotion. Each model is a set called emotion dominant meaning models. Each model contains the top-N dominant meaning probability.

The emotion detection algorithm returns the emotion ζ_i that represents a set of words. The algorithm uses appropriate equation to compute the model value for each emotion for the example ψ . Therefore, it calculates the highest value and then returns the index of this value. This index is used to determine
the emotion.

○ Experiments and Results

This section presents two purposes. First purpose is used to build Emotion Dominant Meaning Tree. The second purpose is to test the accuracy of using this tree for detecting the emotion

- **Data Sets**

The dataset uses ISEAR dataset [11] that contains emotional statements. ISEAR contains 7666 sentences (as shown in Table 1). The dataset is collected from 1096 participants with different cultural background who completed questionnaires about seven emotions: anger, disgust, fear, sadness, shame, joy, and guilt.

Table 1. Characteristics of the ISEAR Dataset.

Emotion	**No. of Examples**
Anger	1096
Disgust	1096
Fear	1095
Sadness	1096
Shame	1096
Joy	1094
Guilt	1093
Total examples	**7666**

Table 2. Characteristics of dataset used to build tree.

Emotion	**No. of Examples**
Anger	658
Disgust	658
Fear	657
Sadness	658
Shame	658
Joy	656
Guilt	656
Total examples	4601

- **Building Emotion Dominant Meaning Tree**

  Most of text classification methods use keyword-based methods with thesaurus.

  In contrast, they use the dominant meaning methods as features to improve accuracy and refine the categories. To build the dominant meaning tree, they use 60% of ISEAR dataset for seven emotion categories (as shown in Table 2): anger, disgust, fear, sadness, shame, joy, and guilt.

  Stop words were removed in all examples for examples: for, an, the, a, an, another, but, or, yet, so, towards, before, etc. Based on given equations they built the dominant meaning tree of seven emotion categories, as shown in Figure 2.

- **Detecting Algorithm Accuracy**

  The goal of the experiments is to measure the accuracy of the proposed algorithm to predict a single emotional label given an input sentence. They followed Cohen's Kappa [13]to measure the accuracies of the experiment. They use average precision, recall, and F-measure to measure the classification accuracy. In their experiment, they use ISEAR dataset to figure out the performance of their proposed mechanism.

  The precision and recall of their proposed approach shows a considerable performance comparing to those in related works.

  In his classification he found that using SVM produced better results for sadness (F1 = 0.733) which is better than their approach for sadness (F1 = 0.67). In contrast, their approach has better results in others classes such as anger (F 1 = 0.66), disgust (F 1 = 0.47), fear (F1 = 0.56), shame (F1 = 0.55), joy (F 1 = 0.58), and guilt (F1 = 0.50). Where Balahur results were for anger (F1 = 0.38), disgust, (F1 = 0.264), fear (F1 = 0.49), shame (F1 = 0.43), joy (F1 = 0.46), and guilt (F1 =0.42).

  Danisman and Alpkocak [14] used the ISEAR collection and used vector space models (VSM) to categorize 801 examples. Their approach showed a significant results anger (F1 = 0.38), joy (F1 = 0.46), and sadness (F1 = 0.67) compared to Danisman and Alpkocak (2008) for anger (F1 = 0.242), joy (F1 = 0.496) and sadness (F1 = 0.371).

On the other hand, in order to test the performance of their proposed approach with alternative methods for emotion detection, they chose the work done by Balahur et al . [12], as shown in Table 3.

Table 3 Comparison of precision and recall between Razek-Frasson[6] and Balahur[12]

Emotion	Balahur (Precision)	Razek & Frasson (Precision)	Balahur (Recall)	Razek& Frasson (Recall)
Anger	0.353	0.202	0.414	0.521
Disgust	0.292	0.224	0.241	0.469
Fear	0.482	0.262	0.491	0.557
Guilt	0.462	0.203	0.386	0.519
Joy	0.439	0.266	0.474	0.506
Sadness	0.707	0.272	0.76	0.602
Shame	0.441	0.202	0.412	0.482

Chapter 3: Methodology

In this work, we followed a supervised machine learning based approach to identify the 6 emotion category type proposed by Paul Eckman .In supervised machine learning system, the system is trained with example with known class label. After training, the system is tested with a data set with unknown class labels. Test data set is disjoint from training data set. 3.1

3.1 Architecture of the proposed diagram

Architecture of the Proposed System As like other machine learning systems, our system also performs data preprocessing and classification model creation on prediction upon the created model. The architecture of our proposed system is shown below.

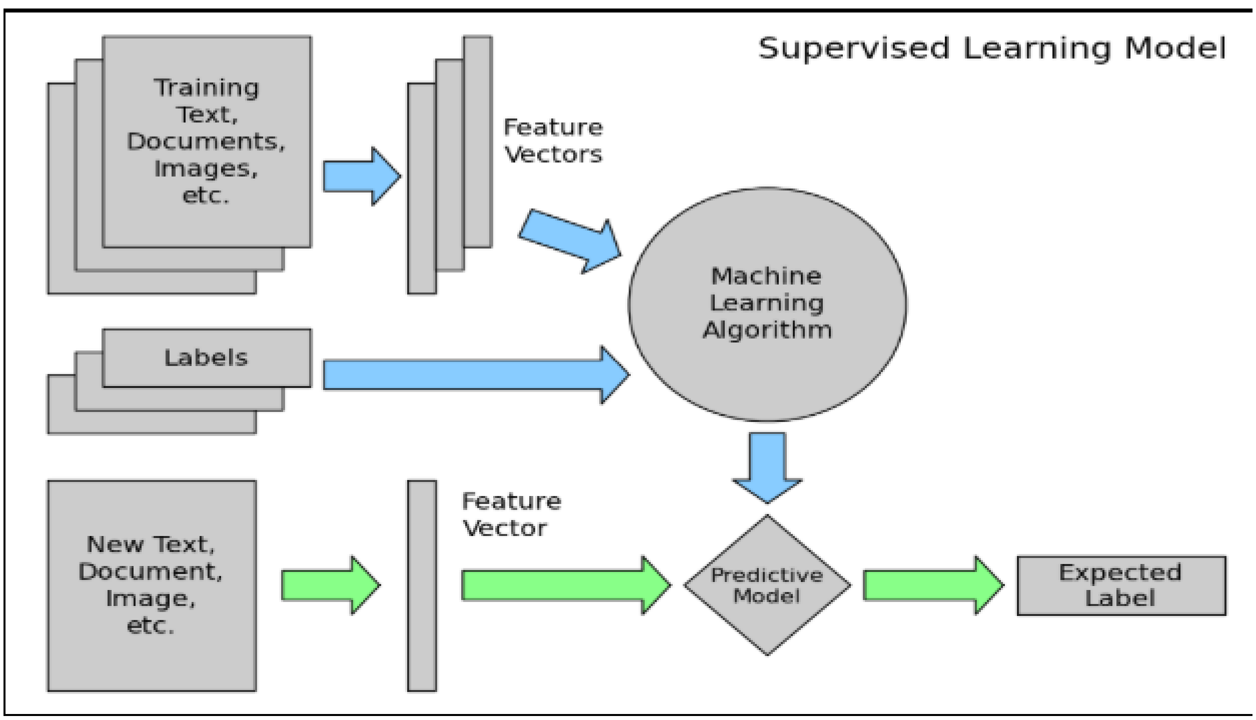


Figure 3: Architecture of our proposed system

Our system consists of two modules. They are:

1) Data Preparation Module

2) Emotion Extraction Module

In the following section, each module is described briefly.

3.2 Data Preparation Module

In data preparation module, data is processed so that data can be fed into classifier for model creation. In this module, processing of data until the model creation is done. The steps of data preparation are briefly described below.

3.2.1 Data collection & Labelling

In this step data is collected for training and testing the system. We collected data from social networking site because they are great source of emotional text. People express their emotion through different ways provided by the social networking site. There are various networking sites. One of the popular sites is Facebook. People express their emotion and opinion as a short form of text in this site via posts and comments. Facebook may contain hash tag, emoticons, link to other news, multimedia content etc. One post can be shared many times. We collected data from facebook public posts and comments. Collected data contained different types of noise, so we performed cleaning operation on the data. As the emotion of posts containing link to other content may depend on the content referred by the link, we eliminated such posts form our dataset. After performing the cleaning operation, we manually labelled the tweets to one of the emotion category we are working with and defined by Paul Ekman. We removed the data containing no emotional content. After completion of labelling, data are stored for preprocessing step.

3.2.2 Data Preprocessing

Only labelling of data is not sufficient for using the data to train the system. Data may not be suitable for processing by the classification algorithm. Moreover all data are not important for our task. For this reason preprocessing is done. In preprocessing we eliminate uninteresting and trivial part of the data. Preprocessing is very critical step because the accuracy of a natural language processing system greatly depends upon the quality of preprocessing. It also reduces the dimensionality of data by eliminating unnecessary detail from the data. To make system more efficient, we performed preprocessing of labelled data. Preprocessing is done in following steps.

3.2.2.a Tokenization

Tokenization is the first step of preprocessing. In tokenization, the text is chopped into stream of words, phrases, symbols or other meaningful units called tokens. Along with decomposing to tokens, tokenization may eliminate some symbols such as punctuations. Each token contains two elements, token name and token attribute. Token name is the content of the token and token attribute may be the index of symbol table entry. The collection of all token is called dictionary. Index of symbol table entry is important because they are stored in the computer memory Tokenizing the posts "I am sick today" and "I am angry" yields the following symbol table.

Table 4: Example of symbol table yield by tokenization

	0	1	2	3	4
Post	I	am	Sick	Today	Angry
T1	1	1	1	1	0
T2	1	1	1	0	1

Regular expression is used for tokenizing. A program named tokenizer tokenizes the text according to the given regular expression. Different challenge may arise during tokenization. These challenges are very much language specific. In many languages the words can't be decomposed into morphemes. Challenge also arises in application of delimiter. Decomposing text using different delimiter may yield different sequences of tokens

3.2.2.b Stopword Elimination

Stopwords are the token or term having zero or very less importance. They occur frequently and are essentialy meaningless in the sense that they contribute very less to the context and content of the document. Thus they are not important for classification of text and hence eliminated. We can find the importance of a token by finding the inverse document frequency (IDF) of the token. IDF of a token t is defined as, IDF(t) = log $\frac{|D|}{|dt|}$

Where,

| D| = Total umber of documents.

| d t| Number of documents containing the term t

In our context posts are analogous to document. Terms having large IDF value are more important. The terms having zero or very small IDF value can be marked as stopwords.

3.2.2.c Normalization

Normalization is text transformation process that transforms the words of text into one ideal form. Words of abbreviation may remain the same after using pronunciation. USA and U.S.A. both indicate the same country and are the same to humans. But these are different to machine. In this case, normalization comes into play. Normalization converts the both term into one ideal form so that both are identified as a same term. This is helpful for reducing dictionary size and finding similarity of documents.

3.2.2.d Lemmatization

Different words are formed by adding suffix and prefix. These words may convey the almost same meaning and hence can be considered as the same word. Lemmatization is the process of converting variants of word into a root representation by eliminating suffix and prefix so that machine can identify all of them as the same word. There are two ways of lemmatization,

1. Morphological analysis
2. Stemming

In morphological analysis, we start from the root word and remove suffix and prefix from the word by comparing them with the content of dictionary. Morphological analysis is more accurate but very much language dependent, complex and expensive to perform by machine. That's why another less accurate but simple and cheap approach named stemming is used for lemmatization.

In stemming we remove the suffix from the inflected form of a word step by step and convert it into a stem. Stem is the uninflected form of a word that conveys the main meaning. Affixes are added to stem for generating inflected form of the words. For example, words "accommodation", "accommodated" and "accommodating" can be converted into them stem accommodate" by applying stemming.

Stemming is also a challenging task because over-stemming and under-stemming may introduce errors in preprocessing. Over-stemming converts words of different stems into one stem. On the other hand, under-stemming converts words of same stem into different stems

3.3 Emotion extraction module

The main processing of the emotion extraction is done by this module. In this module model is created based on the training data and emotion of the test data are extracted using the model. This module performs its work in three steps. The steps are described below.

3.3.1 Model creation

In this step, model is created which is used for classification. Train dataset is used to create the model. Extracted feature of the train data are fed into the classifier for creation of model. Model is created by classification algorithm We trained our system with three classification algorithms. They are,

i. Recurrent neural network (LSTM)

ii. Support Vector Machine

3.3.1.a Recurrent neural network (LSTM)

Artificial neural network is based on the modelling of neuron of human brain. Neuron is the smallest working unit of human brain. It has input and output connection and is acti when the input exceeds some threshold value. We can consider a neural network as a weighted directed graph in which neurons represent the nodes and weighted directed edges can be considered as input and output connection of neuron . McCulloch and Pitts devise a simple mathematical model of neuron in 1943. The model is shown below.

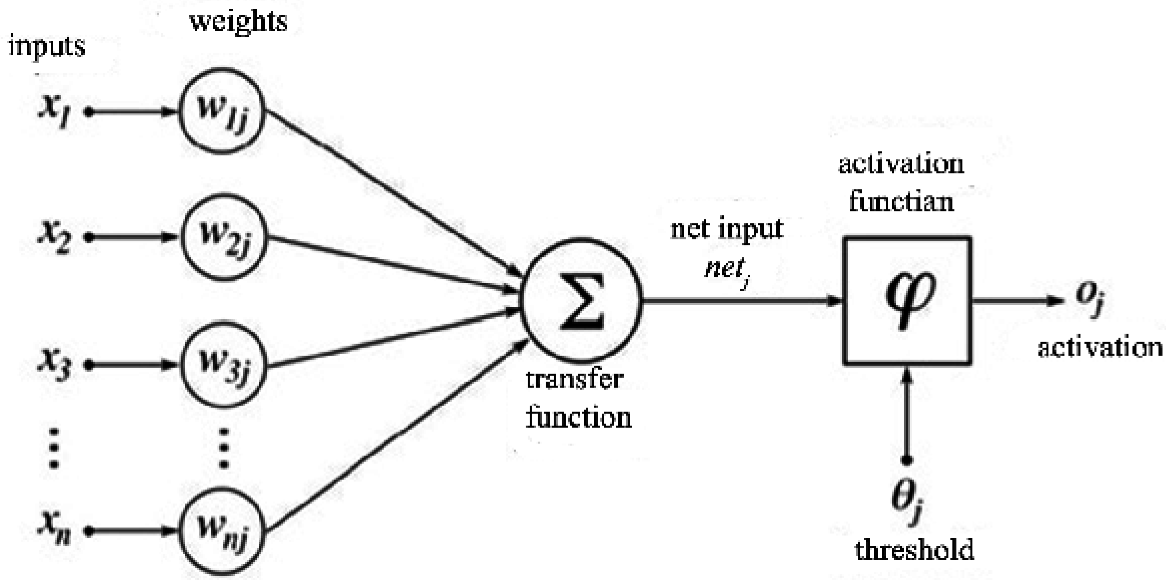


Figure 4: A simple mathematical model of a neuron

Neural network is necessarily connected collection of simple units, Links are used for propagation. Each link has a weight w_{ij} which defines the properties of the link. Mathematically neural network is nothing but linear regression along with sigmoid function. As like linear regression it also has a bias input $a_0 = 1$ with weight w_{oj} Each unit j computes the weighted sum of the inputs and then applies the sigmoid function to generate the output This type of unit is called a perceptron.

A neural network has one input layer, one output layer and one or more hidden layer ca consists of perceptron. A simple neural network is shown below

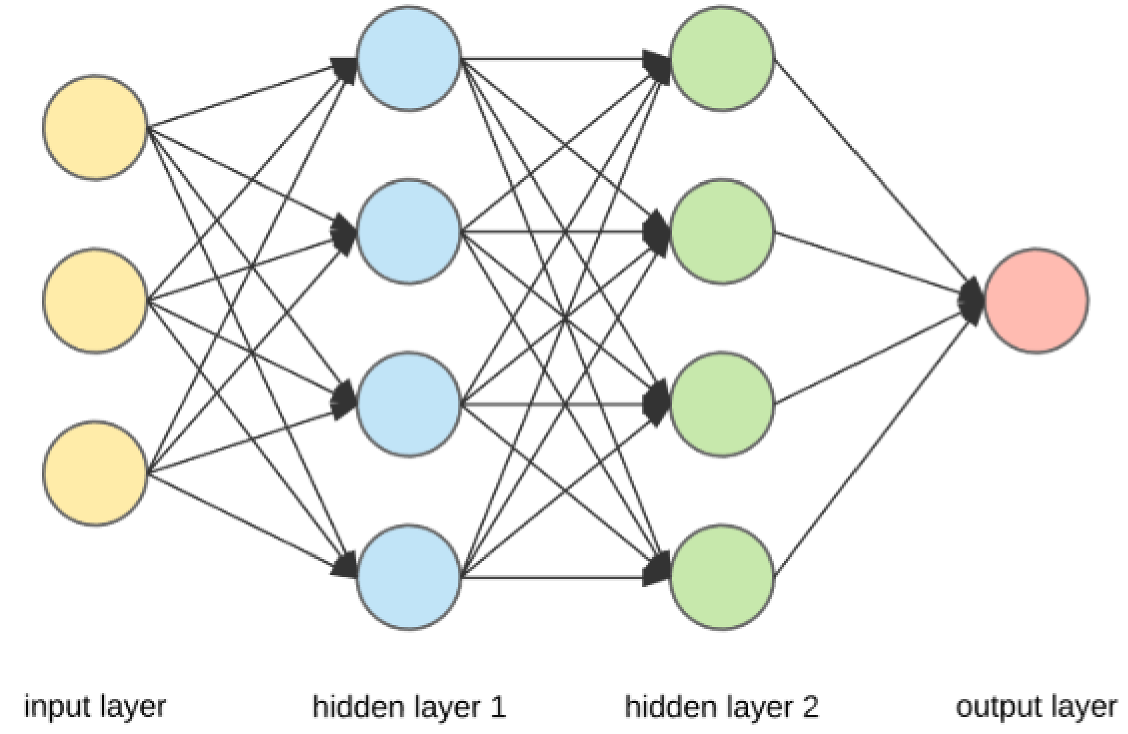


Figure 5: A multilayer feed forward deep neural network

Neural network learns by iteratively processing each tuple of training set. It calculates the error by comparing the networks output with given target value. It than back propagates the error so that weight of connection links are adjusted to reduce the mean-squared error between the network output and actual target value.

Long Short Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people in following work..[16] They work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn!

LSTM was invented specifically to avoid the vanishing gradient problem. It is supposed to do that with the Constant Error Carousel (CEC), which on the diagram below (from [15].)

correspond to the loop around cell.

Here's a figure that depicts how LSTM works

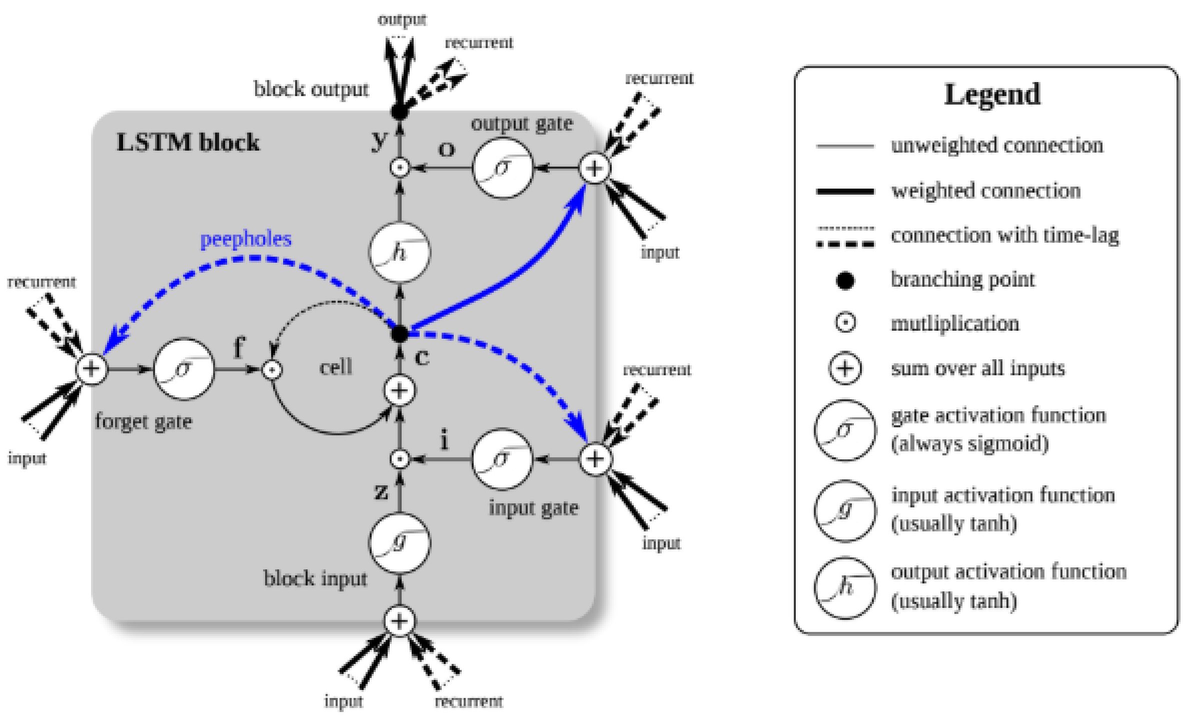


Figure 6: How LSTM prevent the problem of vanishing gradient

3.3.1.b Support Vector Machine

Support vector machine is a supervised classification technique in which we can classify linear and non-lincar data. It maps the original data into a higher dimension space using nonlinear mapping. It then searches the higher dimension for a decision boundary, which separates the tuple of one class from another, which is called optimal separating hyperplane. SVM finds this hyperplane using support vectors. Support vector are also training tuple that are closest to the decision boundary and pushed up by hyperplane The following figure shows the decision boundary and marginal hyper plane for a binary classification problem.

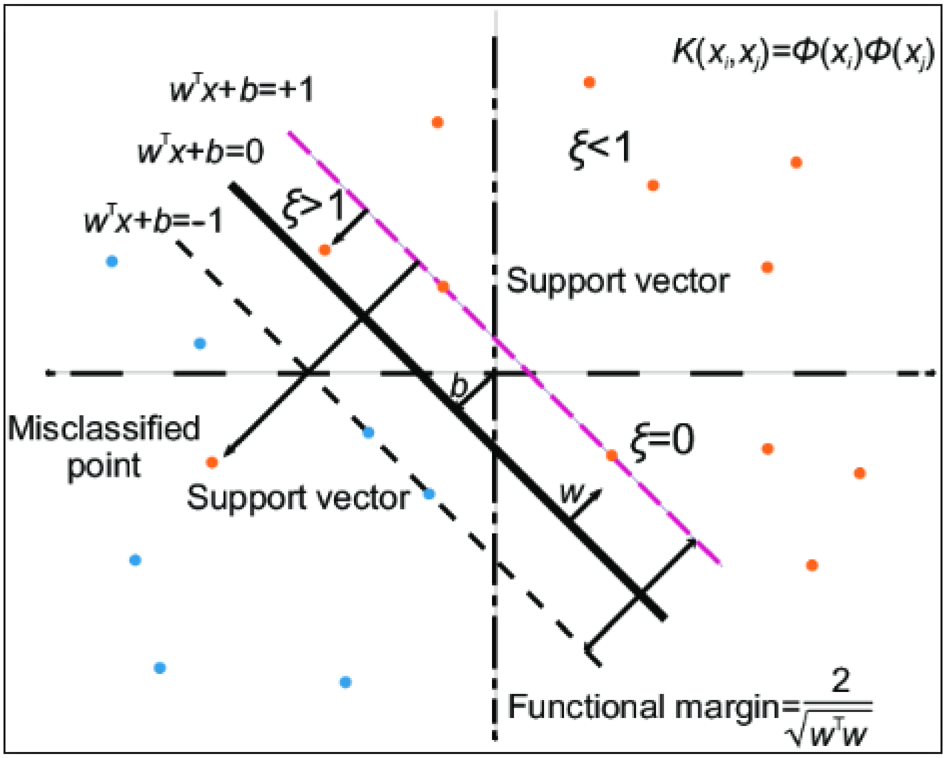


Figure 5: Marginal hyperplane and SVM for binary classification

To understand how Support vector machine works, consider a binary classification problem. We want to find a hyperplane that has the maximum margin because such type of hyperplane will be able to classify future data accurately. It is called maximum marginal hyperplane (MMH) because it maximizes the margin between classes. The main function of SVM is to find a maximum marginal hyperplane separating the classes.

A separating hyperplane can be written as,

$$W.X + b = 0$$

Where W is weight vector of the form is, W = { W_1, W_2, W_3 ....Wn} ;

n is the number of attributes and b is a scalar quantity referred to as bias.

Let's consider training tuples has two attributes A_1 and A_2. So training tuples are of the form X-(x_1,x_2) where x_1 is value of attribute A_1 and x_2 is the value of attribute A_2 respectively for the tuple X. We denote b_0 as additional weight w_0.

So, any point lies above the separating hyperplane satisfies.

$W0 + w1x1 + w2x2 > 0$

So, any point lies below the separating hyperplane satisfies,

$W0 + w1x1 + w2x2 < 0$

The hyperplanes defining the sides of the margin can be written as,

$H1 : w0 + W1x1 + w2x2 \geq W2x2$ for $yi = +1$

$H1 : w0 + W1x1 + w2x2 \leq W2x2$ for $yi =$ -1

Combining these above equations we get, $y(wo+$ Wi it $W2x2) \geq 1$ for all i.

The distance of any point on HI from the separating hyperplane is $1 / || w||$,

Where, w is called the Euclidean norm of W.

As the distance of both hyperplane are same from the separating hyperplane, The maximum margin is $2/|| w||$ The equation of the MMH can be written as a constrained convex optimization problem which can be solvable using Lagrangian formulation. After solving we find the following equation for MMH as decision boundary.

$$d\left(x^{T}\right) = \sum_{i=1}^{l} y_{i}\, a_i X_i X^T + b_0$$

Where, y_i is the class label of the support vector X_i ; X^T is a test tuple: a_i and b_0 are e parameters that are determined automatically by the SVM algorithm and l is the number of support vector.

Given a test tuple X^T we put in it the equation above. If the sign of the result is positive then the tuple is of the positive class and if negative then negative class.

Multiclass Classification: SVM is basically a binary classifier All classification problems can be represented as a binary classification problem. We can take one class and put all other classed together. In this way we can convert a multiclass classification problem into a binary classification problem. We can apply the same procedure as the number of classes we have. The class for which we for which we find the desired result, will be the output of classification task.

3.3.2 Emotion Extraction

As stated earlier, we created three classification models using Naive Bayesian classifier Artificial Neural Net and Support vector Machine. After creation of models we used each of them for testing on the test data. Test data are disjoint from training set. Using the features of test data and based on classification model, system showed one of the emotion categories for the corresponding test.

3.4 Experimental Setup

To perform the experiment according to the methodology described, we need some hardware and software. In the following section, hardware requirement and software requirements and described briefly.

3.4.1 Hardware Requirement

To perform the experiment we used a notebook pc as hardware. The pc was of 3rd generation Intel core i3 processor with 8GB RAM. The pc was running on 64-bit windows8 OS. The programs run within reasonable time in this setup.

3.4.2 Software requirement

We used different software to perform the experiment. We used Python language for our experiment. Python is widely used high level general purpose programming language. Python is popular language because use of its easy syntax, dynamic type system and support for multiple programming paradigms. Because of availability of well rich library, python is the first choice for data analysis system development. Python has two versions, python 2.x and python 3.x. As python 3.x is more popular and structured, we used python 3.6 for our experiment. Different compatible with different python versions. Anaconda is the most popular of them. We used Anaconda3 as our python interpreter. interpreters provide a bundle of commonly used python modules that are Different modules of python are used in our experiment. These modules are noted in the following sections in accordance with our system architecture. Python modules used in data preparation: We collected data from popular social networking site facebook. We collected 2470 posts and manually labelled them. The facebook are of 6 emotion category as mentioned above. We saved them in a excel file for preprocessing.

Data preprocessing is the first step of emotion extraction. Python provides rich module for natural language preprocessing name Natural Language Toolkit(NLTK). It provides implementation of common methods and a rich set of corpora for analysis task. To perform tokenization, we used the tokenizer available in NLTK. NLTK also has a list of English stopwords. We filtered out the stopwords from data using that list. To perfom stemming. we used Porter stemmer available in NLTK Python modules used in emotion extraction Python has a rich machine learning library named sci-kit learn. Machine learning algorithms and feature extraction approaches along with other performance improvement tools are implemented there. We used Count Vectorizer to create language model for our dataset and Tf-Idf Transformer for feature extraction.

Multinomial Naïve Bayesian classifier is suitable for discrete features. As we used TF-IDF as text feature which has discrete values, we used this classifier for Naive Bayesian classification. We used MLP Classifier for ANN classification and Stochastic Gradient Descent, SGD Classifier for SVM classification. Moreover we used numpy module of python for array processing and evaluation task The implementation code of our system is provided in the Appendix section

CHAPTER 4: EVALUATION

After building a machine leaning system, the performance of the system is evaluated. There are various ways of validation and measuring performance of machine learning system. The approaches used in our experiment are described below

4.1 Performance Metrics

There are various measures of performance used in machine learning literature for evaluation These measures are given below

Positive Tuples (P): Tuples of the main class of interest.

Negative Tuples (N): Tuples of all other class.

True Positive (TP): These are positive tuples correctly classified by the classifier.

True Negative (TN): These are negative tuples correctly classified by the classifier.

False Positive (FP): These are negative tuples incorrectly classified by the classifier.

False Negative (FN): These are positive tuples incorrectly classified by the classifier.

Accuracy $= \frac{TP+TN}{P+N}$

Precision $= \frac{TP}{TP+FP}$

Recall $= \frac{TP}{TP+FN}$

F1 Score $= 2*$ (Recall $*$ Precision) / (Recall $+$ Precision)

4.2 Holdout Method

In this method whole data is partitioned into two sets, train set and test set. Typically two third of the data are used for training and one third is used for testing

We performed this type of validation of our system. We used one fifth of total data for testing and rest for training. We performed the test for three times and find the average of various measures our system. The result is shown below.

Classifier	Accuracy	Avg. Precision	Avg. Recall	Avg. F-Measure
SVM	48.625%	48%	49%	47%
Naïve Bayes	50.7499%	50%	51%	50%
RNN (LSTM)	53.65%	55%	54%	53%

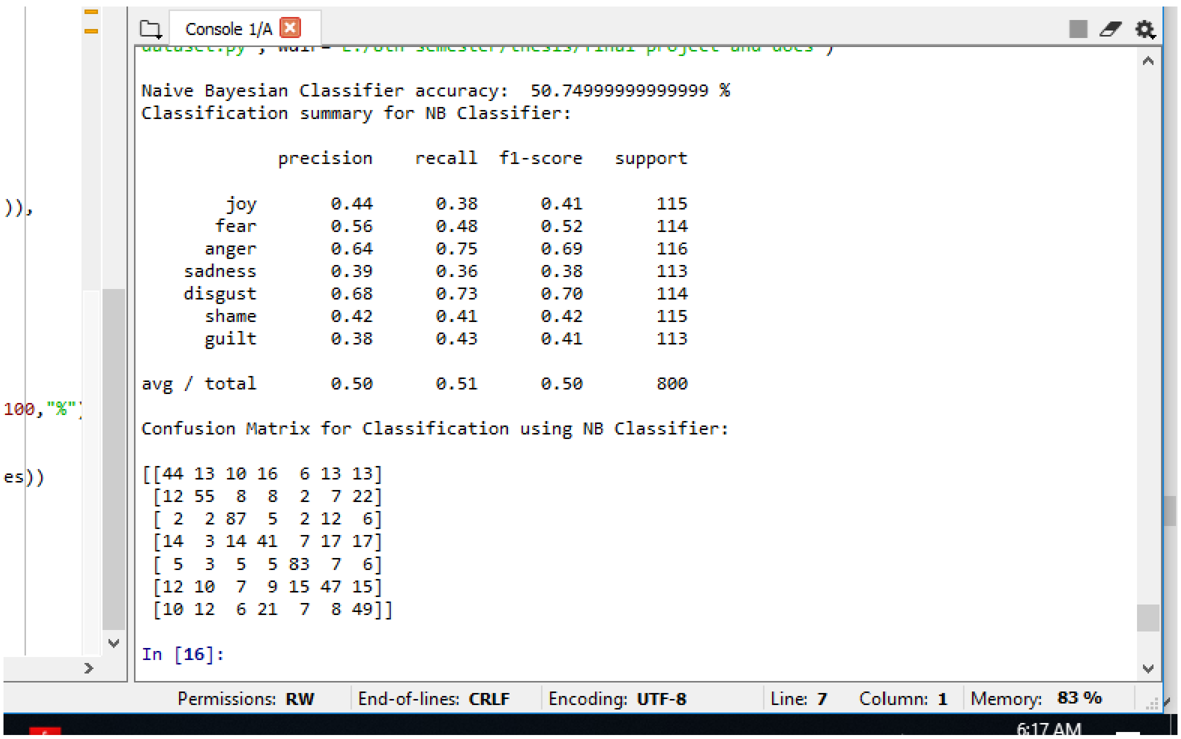
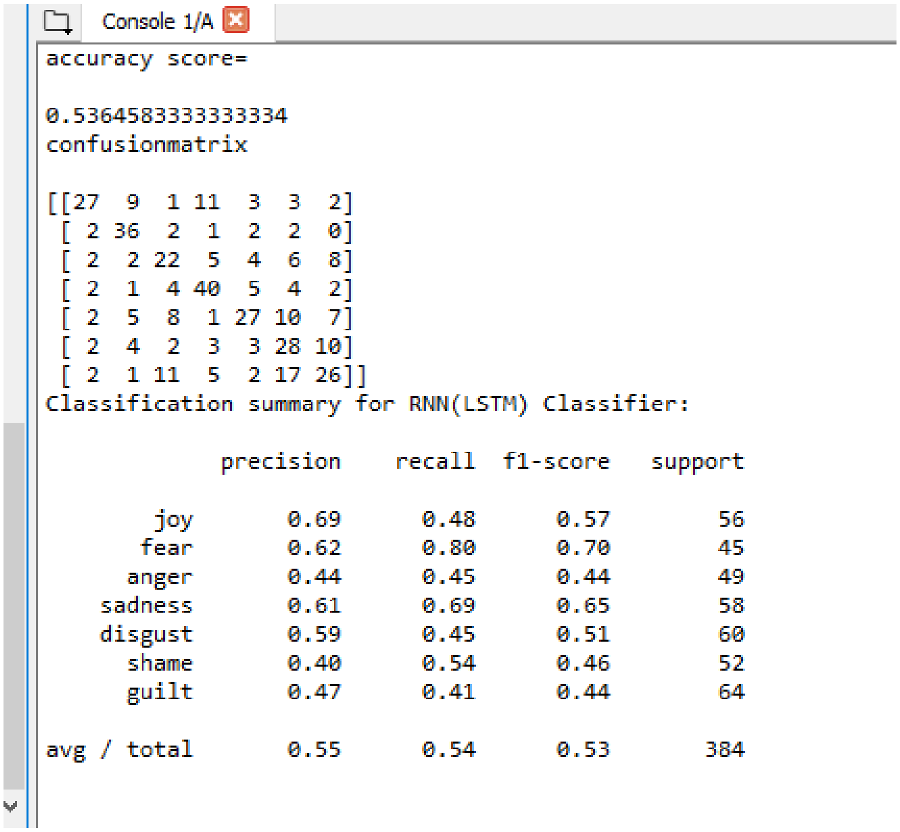


Figure 6: Classification summary of NB

```
max_iter and tol instead.
  DeprecationWarning)

SVM Accuracy:  48.625 %
Classification summary for SVM:

             precision    recall  f1-score   support

        joy       0.46      0.36      0.40       115
       fear       0.56      0.56      0.56       114
      anger       0.60      0.74      0.66       116
    sadness       0.43      0.31      0.36       113
    disgust       0.45      0.76      0.56       114
      shame       0.40      0.38      0.39       115
      guilt       0.48      0.28      0.36       113

avg / total       0.48      0.49      0.47       800

Confusion Matrix for Classification using SVM:

[[41 16 10 13 16 13  6]
 [ 9 64 10  5 14  6  6]
 [ 1  4 86  3  4 14  4]
 [14  4 12 35 21 19  8]
 [ 4  2  4  3 87 10  4]
 [13 10  8  7 27 44  6]
 [ 8 15 13 15 26  4 32]]
```

Permissions: **RW** | End-of-lines: **CRLF** | Encoding: **UTF-8** | Line: **47** | Column: **37** | Memory: **87 %**

Figure 7: Classification summary of SVM



```
accuracy score=

0.5364583333333334
confusionmatrix

[[27  9  1 11  3  3  2]
 [ 2 36  2  1  2  2  0]
 [ 2  2 22  5  4  6  8]
 [ 2  1  4 40  5  4  2]
 [ 2  5  8  1 27 10  7]
 [ 2  4  2  3  3 28 10]
 [ 2  1 11  5  2 17 26]]
Classification summary for RNN(LSTM) Classifier:

             precision    recall  f1-score   support

        joy       0.69      0.48      0.57        56
       fear       0.62      0.80      0.70        45
      anger       0.44      0.45      0.44        49
    sadness       0.61      0.69      0.65        58
    disgust       0.59      0.45      0.51        60
      shame       0.40      0.54      0.46        52
      guilt       0.47      0.41      0.44        64

avg / total       0.55      0.54      0.53       384
```

Figure 8: Classification summary of RNN(LSTM)

4.3 Comparison

Comparing the result given in the previous section it is clear that Artificial Neural Network algorithm outperforms the other two Support Vector Machine and Naive Bayesian. It is as expected. This is because in deep learning we don't have to select feature to train the model. Deep learning method is such that it searches for relevant feature by itself and trains on them. Moreover. LSTM performs even better than the conventional Deep neural networks specially in terms of text classification.

Now we see the comparison among our approach and the approach of others cited in literature review.

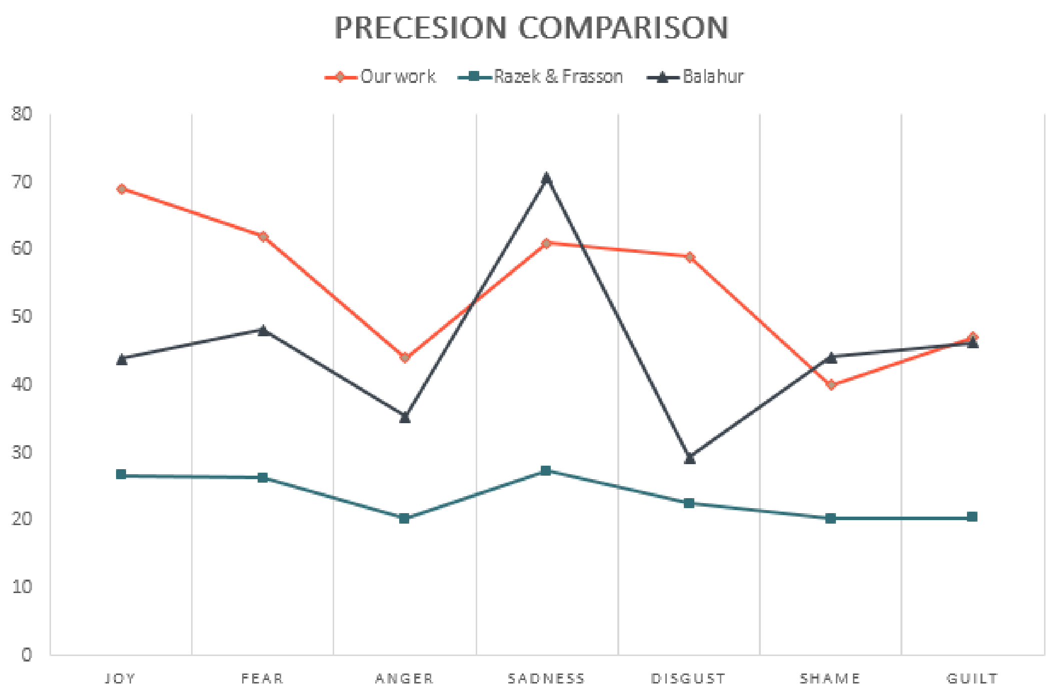


Figure 9: Precision comparison

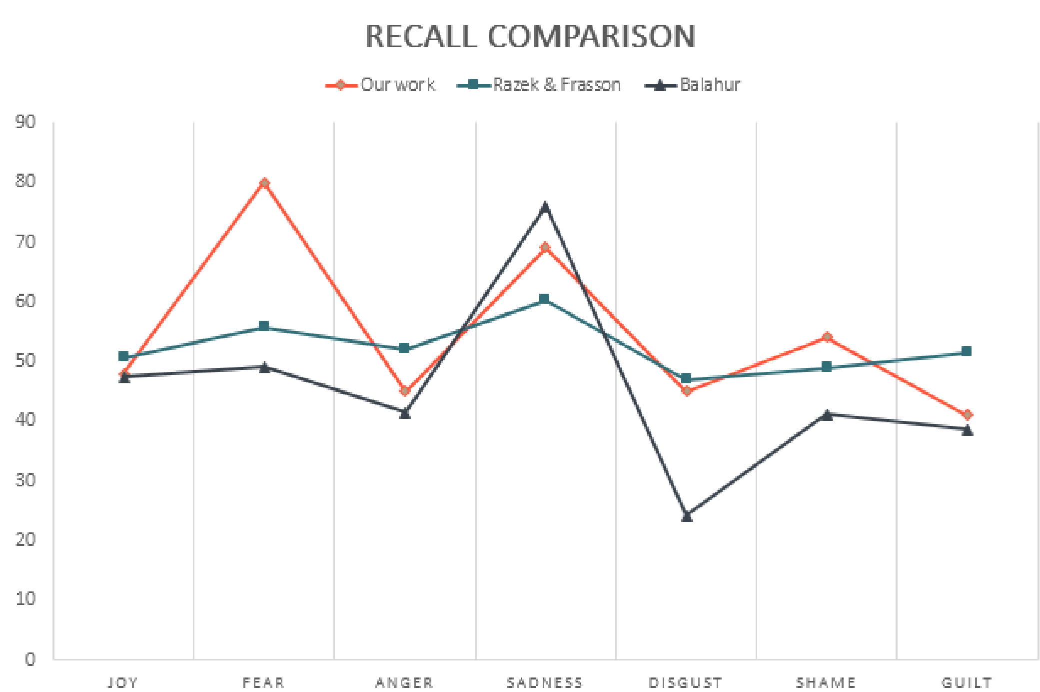


Figure 10: Recall comparison

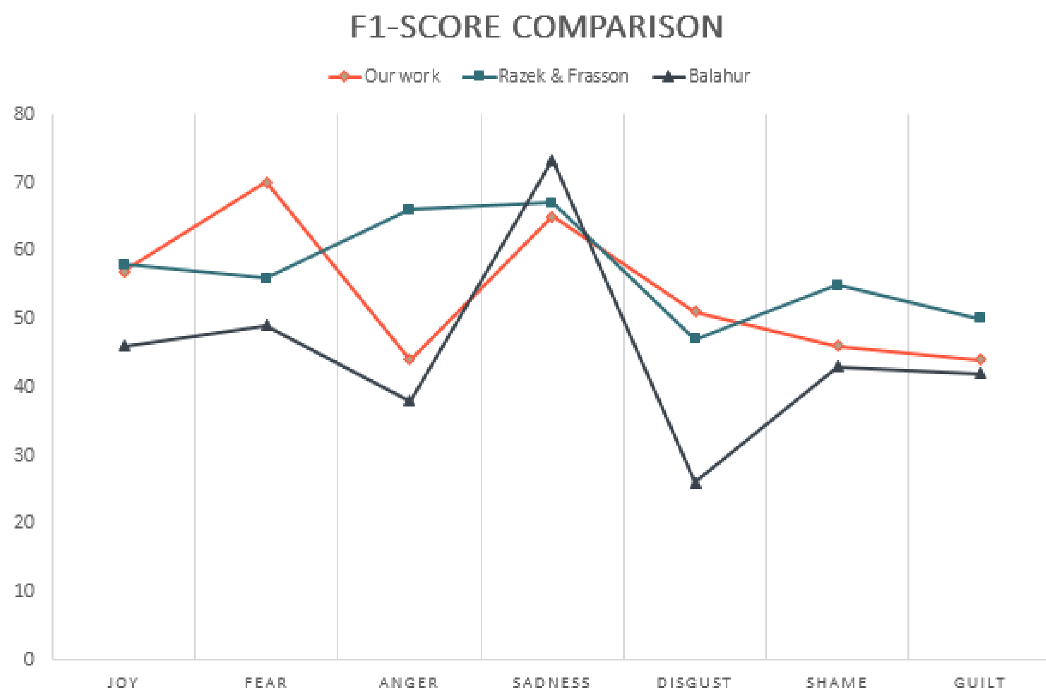


Figure 11: F-1 Score comparison

4.4 Discussion

As mentioned earlier, LSTM provides better accuracy than other two algorithms. Various studies have shown that, SVM provides a very good accuracy in the task of text categorization. But as number of data sample increases, accuracy of SVM declines and accuracy of LSTM increases. Our text to emotion extraction problem can also be considered as a problem of sentimental text categorization.

4.5 Future Works

We will complete the work we had originally started with the bangla dataset and then run our proposed algorithm as well as conventional baseline algorithms.

Will experiment on the LSTM model changing number of dense layers and batch sizes.

Will work on assigning weight values to specific emotional key words and try to build a model around it.

Chapter 5: Conclusion

Extracting emotion from text corpus has a lot of important applications, such as e-co product review analysis, personal or group level communication, automated motive detection for robotics etc. Extracting emotion from text can be done by a lot of methods. A lot of rule based methods are established there. The main problem of these methods is creating emotion intensity lexicon. We analysed supervised machine based approach to solve this problem. We experimented Naïve Bayesian classifier, Support Vector Machine and Recurrent Neural Network to solve this problem. In our setup, LSTM outperformed other two classifiers and provided good accuracy for the task. This work will be helpful to eliminate currently available tedious systems of creating emotion intensity lexicon.

REFERENCES

[1] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

[2] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[3] Scherer, K.R. and Wallbotth, G. (1994) Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning. Journal of Personality and Social Psychology, 66, 310-328. https://doi.org/10.1037/0022-3514.66.2.310

[3] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau. Sentiment Analysis of twitter data. In proceedings of the workshop on languages in social media, LSM 11, pages 3038 , Stroudsburg, PA, USA, 2011. Association for computational linguistics.

[4] Pichard, Rosalind W. Affective compting . vol 252, Cambridge: MIT press , 1997

[5] Izard Carol E. , Human Emotions. Springer science and business media, 2013.

[6] Razek, M.A. and Frasson, C. (2017) Text-Based Intelligent Learning Emotion System. Journal of Intelligent Learning Systems and Applications 9, 17-26.

[6] Improving Multi-class Text Classification with Naive Bayes
- Jason D.M. Rennie , AI Technical Report 2001-004 September 2001

[7] Gajjala, Abhiteja, "Multi Faceted Text Classification using Supervised Machine Learning Models" (2016).*Master's Projects*. 482.
http://scholarworks.sjsu.edu/etd_projects/482

[8] Text Classification Using Machine Learning Techniques, by M. IKONOMAKIS, S. KOTSIANTIS and V. TAMPAKAS .
WSEAS TRANSACTIONS on COMPUTERS, Issue 8, Volume 4, August 2005, pp. 966-974

[9] Kaur, Amandeep & Khaneja, Deepesh & Vyas, Khushboo & Saini, Ranjit. (2017). PROJECT REPORT SENTIMENT ANALYSIS ON TWITTER USING APACHE SPARK. 10.13140/RG.2.2.10737.79200.

[10] 2015 International Conference on Soft-Computing and Network Security (*ICSNS* -2015), Feb. 25 – 27, 2015, Coimbatore, INDIA
Overview on Emotion Recognition System by Ashwini Ann Varghese, Jacob P Cherian, Dr. Jubilant J Kizhakkethottam.

[11] Scherer, K.R. and Wallbotth, G. (1994) Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning. Journal of Personality and Social Psychology, 66, 310-328. https://doi.org/10.1037/0022-3514.66.2.310.

[12] Balahur, A., Hermida, J.M. and Montoyo, A. (2011) Building and Exploiting Emotinet, a Knowledge Base for Emotion Detection Based on the Appraisal Theory Model. IEEE Transactions on Affective Computing , 3, 88-101.

[13] Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement , 20, 37-46. https://doi.org/10.1177/001316446002000104

[14] Danisman, T. and Alpkocak, A. (2008) Feeler: Emotion Classification of Text Using Vector Space Model. Proceedings of the AISB 2008 Symposium on Affective Language in Human and Machine, University of Aberdeen, UK, 1-4 April 2008, 53-59.

[15] LSTM: A Search Space Odyssey Klaus Greff, Rupesh Kumar Srivastava, Jan Koutník, Bas R. Steunebrink, Jürgen Schmidhuber. https://arxiv.org/abs/1503.04069

[16] In addition to the original authors, a lot of people contributed to the modern LSTM. A non-comprehensive list is: Felix Gers, Fred Cummins, Santiago Fernandez, Justin Bayer, Daan Wierstra, Julian Togelius, Faustino Gomez, Matteo Gagliolo, and Alex Graves