Research Article

Sentiment analysis of legal emails using Plutchik's Wheel of Emotions in quantified format

Vaishali Ganganwar[a], Nihal Babu[b], Pooja Kudale[c] , Rohit Singh[d], Sandesh Tanwar[e]

[a,b,c,d,e] Department of Computer Engineering, Army Institute of Technology, Pune, India

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: Sentiment analysis, which automatically extracts expressions from text, has gained a great deal of analysis attention within the past decade. Sentiment analysis for social networking sites has become an emerging field in text mining, however once we quote email that is wide employed in communication in our everyday tasks analysis into email sentiment analysis isn't to identical proportion. Earlier very less work has been done in extracting emotions from emails. The aim of this paper is to perform sentiment analysis and quantify the emotional intentions expressed in emails and highlighting the dominant one by using Machine learning models like Naïve Bayes, Support vector machine (SVM), RNN (Recurrent Neural Network), Convolution Neural Network (CNN), Word2Vec and comparing the performance of these model. We have classified the emotion into eight totally different classes of the Plutchik's emotion wheel: joy, trust, fear, surprise, sadness, expectation, anger, and disgust. we have used TF-IDF (Term Frequency Inverse Document Frequency) for feature extraction, to train Naïve-Bayes classifiers and SVM. We have trained all our models on Dens-Dataset and predict emotions from the document passed to that later, which resulted in achieving maximum accuracy using RNN. Dens-Dataset has 10,710 entries containing emotions, which are in Plutchik's wheel.

Keywords: Sentiment Analysis, Emotion Classification, NLP (Natural Language Processing), Plutchik Wheel, Emails, Dens-Dataset

1. Introduction

Email is one amongst the foremost reliable means that of on-line correspondence and has become an especially important means that of official communication for many organisations and people. Within the company world, folks use email as a proper technique of human activity with their customers or sharing their opinions. With the increasing use of email, prioritising associated organising emails is turning into an insurmountable task. A common user spends a big quantity of your time reading, understanding, and responding to emails. Filtering emails based on emotions will reduce efforts and save time.

We have focused on finding the emotions present in the email and quantifying them also highlighting the dominant one. A lso, to narrow down the emotions to a limit we are using Plutchik's wheel and have took 8 prominent emotions from it which are: anger, expectation, disgust, sadness, joy, trust, surprise, fear.

Robert Plutchik took eight emotions of basic categorised as basic and eight other emotions categorised as advanced, each one contains 2 basic emotions and created an emotion wheel. Aggression, contempt disappointment, submission, optimism, remorse, awe and love are the advanced unit of emotions formed by the following 8 fundamental unit of emotions anger, expectation, disgust, sadness, joy, trust, surprise, fear.

Varied approaches like reaction, visual communication, facial expressions, etc. can be used for detecting emotion. However, the utilization of text desires a lot of improvement. And email was not in focus while finding out emotions from text.

Various algorithms like NB, SVM, CNN, RNN, Word2Vec are matched mistreatment experimental results. The dataset employed in this paper is named Den's dataset. The analysis is considered as classification. The complete implementation is completed in Python.

Highlights of this paper are as follows:

- Using Plutchik's wheel of emotions for emotion analysis.
- Focused on emails and finding out the emotions present in email.
- Quantification of emotions from the emails and highlighting the dominant one.
- Used Naive Bayes, Word2Vec, SVM, RNN, CNN methods to find of emotions.
- Comparing the accuracy and f1 score of the above-mentioned methods.

The schema of this paper is delineated as follows. Second section provides an outline of previous analysis and work done on sentiment and feeling analysis. Third section defines Dataset and its statistics. Fourth section defines emotion classifications and Plutchik’s wheel. Fifth section defines Pre-processing and Methods used and their accuracies. Finally, Section six summarises the findings, limitations, and future work of this paper.

2. Related Work

In the recent years with the increasing importance of emotion or sentiment analysis/detection in text, audio, or videos feed (facial expressions) the field has been an area of research by data analysts. But here as we focus on the emotion detection on text, the road has been rather bumpy because of the nature of the problem itself. Walking towards the target the first task is to choose the framework considering the sentiment or emotion. Previously in this area **Rayan Salah Hag Ali** and **Neamat El Gayar** [1] made utilisation of the Enron email dataset to train the classifier. Then implemented TF-IDF for feature extraction, to train Naïve Bayes and Support vector machine (SVM) classifiers.

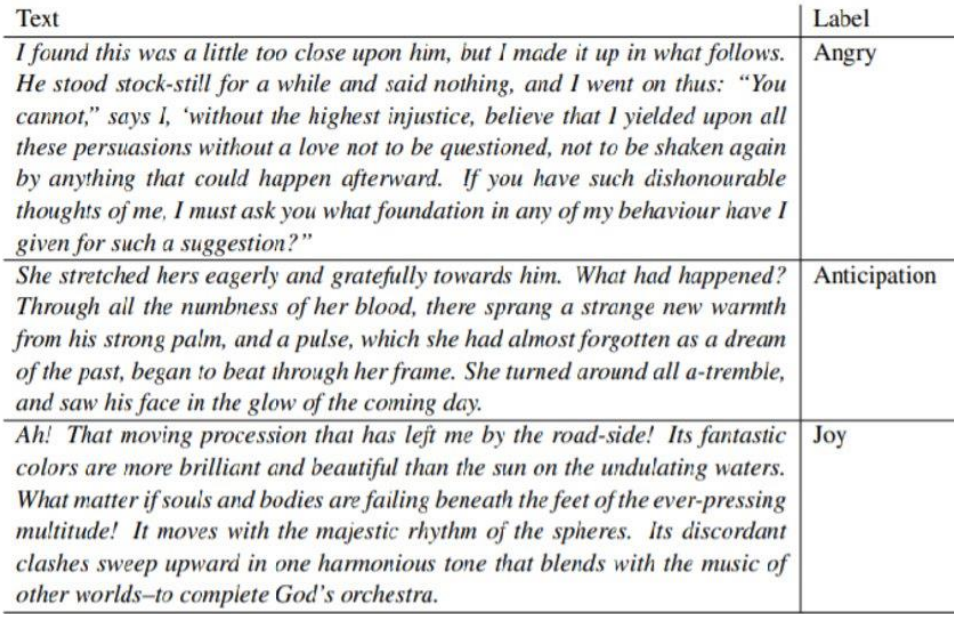
LSTM (Long Short-Term Memory) based approach on sentiment classification was proposed by **Dr. Gorti Satyanarayana Murty** and **Shanmukha Rao Allu** [2]. The distinguished feature provided by LSTM is that it generates output at every step taken or time and this output obtained is utilised to train the network using gradient descent. Looking at performance of LSTM we see a significant accuracy of 85% in emotion detection with the condition that more training data is provided to the model. Tool SENTA can be handed down to investigate the practicability of numeric form categorization by performing it on a dataset which consist of tweets distributed under 11 different sentiment classes by **Mondher Bouaziziand Tomoaki** [3]. The tweet dataset is manually annotated, and the result so obtained was compared against human annotations and the F1 score obtained after completion of experiment was 45.9%. **Muhammad Babar Abbas** and **Mukarram Khan [4]** performed sentiment detection using various algorithms, their main purpose was to choose a suitable algorithm to be utilized in automatic email response system. The performance of Naïve Bayes, SVM, FNN (Forward Neural Network) and RNN algorithms were compared with each other. The accuracy of RNN improved with each epoch to a final of 87% where it started with 26% only at the first epoch.

A different approach of creating a vast collection of tweets labelled wih Plutchik’s, Ekman’s and POM’s classification of sentiments was done by **Niko Colneric** and **Janez Demsar**[5]. RNN was able to perform better than the mark set by common bag of words model. Research suggests that it is best to train RNN on sequence of characters rather than on sequence of words. Using this approach, the model gives much accurate results and no pre-processing or tokenisation is required. A hybrid sentiment analysis model formed by combining K-means clustering and SVM was formed which took Email data as dataset was performed by **Sisi Liu** and **Ickjai Lee** [6]. This approach gave better results in terms of accuracy as compared to SVM, NB, LR and J48. Thereby making the combined K-means and SVM algorithm the suitable algorithm for our problem statement.

3. Dataset

The dataset that has been used in this paper is Dens-dataset [7]. DENS stands for Dataset for Emotions of Narrative Sequences. It was collected from each classic literature out there on Project pressman and fashionable on-line narratives out there on Wattpad, annotated victimization Amazon Mechanical Turki.

Dataset contains 10710 passages extracted from on-line narratives on Wattpad and literature on project Johannes Gutenberg and categorised into 8 broad emotions that are anger, joy, sadness, anticipation, surprise, fear, disgust and trust. Fig 1 shows email passage samples from the dataset.



Text	Label
I found this was a little too close upon him, but I made it up in what follows. He stood stock-still for a while and said nothing, and I went on thus: "You cannot," says I, 'without the highest injustice, believe that I yielded upon all these persuasions without a love not to be questioned, not to be shaken again by anything that could happen afterward. If you have such dishonourable thoughts of me, I must ask you what foundation in any of my behaviour have I given for such a suggestion?"	Angry
She stretched hers eagerly and gratefully towards him. What had happened? Through all the numbness of her blood, there sprang a strange new warmth from his strong palm, and a pulse, which she had almost forgotten as a dream of the past, began to beat through her frame. She turned around all a-tremble, and saw his face in the glow of the coming day.	Anticipation
Ah! That moving procession that has left me by the road-side! Its fantastic colors are more brilliant and beautiful than the sun on the undulating waters. What matter if souls and bodies are failing beneath the feet of the ever-pressing multitude! It moves with the majestic rhythm of the spheres. Its discordant clashes sweep upward in one harmonious tone that blends with the music of other worlds–to complete God's orchestra.	Joy

Fig 1: sample email passages from Dens Dataset

Dataset for Emotions of Narrative Sequences (DENS). The DENS dataset contains a total of 10710 passages that are unit narratives, the common sentences per passage is half-dozen or half-dozen sentences per passage. Also, the common word gift in every sentence is sixteen i.e., sixteen words per sentences. and therefore, the average length of eighty-six words.

The size of this dataset is 4.5 mb. It has total 10,710 entries with totally different emotions. There are unit eight classes of emotions gift in this dataset. These are unit anger, joy, sadness, anticipation, surprise, fear, disgust and trust. Fig 2 shows number of samples per class in Dens dataset. The entries that are unit gift within the passages of multi-class feeling analysis are unit long-form recital in English. From classic literature on the market on Project Gutenberg and fashionable on-line narratives on the market on Wattpad, annotated mistreatment Amazon Mechanical Turki the dataset (DENS) for sentiments of Narrative Sequences was collected.

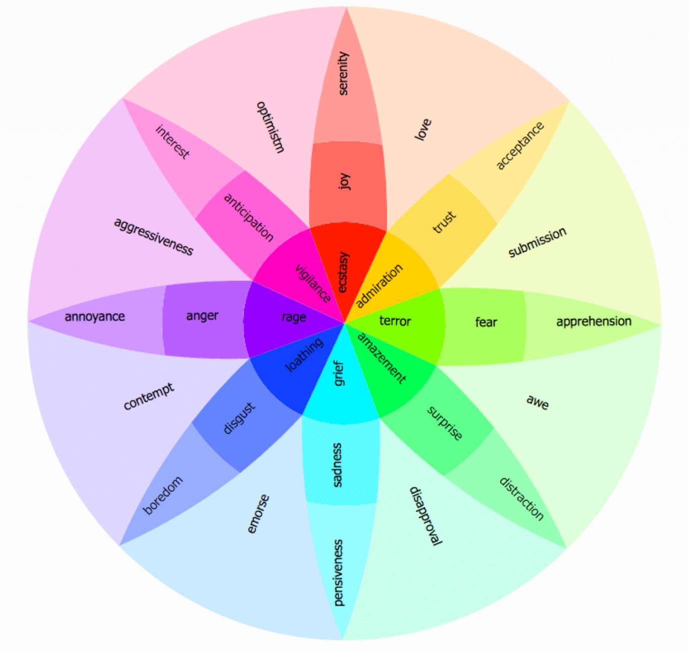
Emotions	Number of samples
Joy	3301
Trust	2130
Fear	1313
Surprise	1302
Sadness	880
Disgust	722
Anger	596
Anticipation	466

Fig 2: Dens dataset statistics

4. Emotion Classification

Paul Vagn Walfrid Ekman designed a stock of six salient emotions based on facial expressions: joy, disappointment, disgust, anger, surprise and fear. A wheel-like diagram was traced by Robert Plutchik with a stock of eight salient contrastive emotions pairs: trust - disgust, joy - sadness, surprise – anticipation and fear-anger. We tend to treat every of those emotions as a independent class also ignore the various degree of intensity identified by Plutchiks wheel of emotion. The Mood States Profile may be a psychological tool for assessing a human state of mood. Sixty-five adjectives are recognized and rated on a five-point-scale by the topic. Every adjective fall into one in every of six classes. Looking at instance, feeling irritated includes a positive impact on the anger class. The upper the score of associate degree adjective, a lot of it contributes to the general score for its class, except for relaxed and effective, whose contribution to the freelance classes is negative. we've removed the adjectives relaxed and economical, that have a negative contribution, because of the text containing them would represent counter examples to the relevant class.

From now on, we will refer to these classifications and go on with Plutchik's wheel.



Plutchiks Wheel of Emotions: The dataset is elucidated on transformed Plutchiks wheel of emotions. The original Plutchiks wheel includes eight primary emotions: Surprise, Joy, Disgust, Sadness, Anger, Trust, Fear, Anticipation, Trust. Also, a lot of complicated emotions can be shaped by fusing two salient emotions. Consider Love which can be outlined as a mixture of Joy and Trust. Anger will vary from the emotion of Annoyance (modest) to Rage (temper). Plutchik's wheel also represents the intensity of a feeling.

5. Methodology

The modeling process of our product is divided into two prominent parts.

1. Data Preprocessing part

2. Machine Learning models

5.1 Data Preprocessing

We have applied following text preprocessing steps.

Lowercasing: Lowercasing all of your text information in text preprocessing is useful and applicable to many texts mining and informatics issues and may facilitate in situations wherever your dataset is not terribly giant and considerably helps with consistency and uniformity to expected output.

Tokenization: Tokenization is a technique of breaking a given text into tiny chunks that may be words or sentences known as tokens. These tokens create some type of context or develop some informatics models. We tend to separate the sentences into words as we want to perform stemming and stop word removal techniques which might create the natural language processing economical.

For eg: "Shirt with dogs and cats" ----Apply Tokenization----> ["Shirt", "with", "dogs", "and" "cats"]

Stop Word Removal: Words that give no meaning to the data but are used to join couple of words are known as stop words like "and", "the" etc. that don't highlight a exact meaning and therefore they must be removed.

For eg: ["Shirt", "with", "dogs", "and" "cats"] ---Apply Stop Word Removal----> ["Shirt", "dogs", "cats"]

Notice that "and", "with" words are removed within the output.

Stemming: In order to understand the meaning of the word distinctly the suffix part of it is removed and the word is reduced to its root word, the process is known as stemming.

For eg: [“Shirt”, “dogs”, “cats”] ---Apply Stemming---> [“Shirt”, “dog”, “cat”]

Notice that the “s” prefix of “cats” and “dogs” are removed within the output.

Normalization:

Text normalization is a highly look-over pre-processing technique. It is the method of mutating the text into standard form. Take an example, the word “gooood” and “gud” is remodeled to “good”, its standard type. One more example is mapping of close to similar words like “stopwords”, “stop-words” and “stop words” to merely “stopwords”.

5.2 Machine Learning models

In this part we have implemented some of the most prominent machine learning model after doing intensive literary review in the field of sentiment analysis and emotion analysis. The results produced and the working mechanism encourage us to select all these models out of all present in the emotion analysis realm. Our task over here is to test the model with our database in hand and check the accuracy/performance of all the model. Our final decision comes on output but also the form in which the output is obtained, thereby finalizing one model to be used in the final project stage.

Method 1: Naïve Bayes’ Classifier

A machine learning algorithm Naive Bayes is used to solve problems based on classification. It uses Bayes Theorem assuming all independent predictors. Naive Bayes is time efficient and is appropriate for multiclass classification. It makes assumption of feature independence which if holds true then it performs a cut above other model. Training data required for this model is less. It works more efficiently for variables with categorial input than numerical. The beginning points of the Bayes theorem of conditional probability, states for a given data point q and D class:

$$P\ (D\ /\ q) = P(q/D)\ /P(q)$$

For a data point we assume that

q= {q1, q2 ,...qj}, The probability can be estimated for x by considering the probability of each of its traits that occurs in the provided class as independent, consider:

$$P(D\ /\ q) = P(\ D\ ).\ \prod P\ (qi/D)$$

After dividing the dataset into 8:2 ratio as training and testing data we get an F1 score of 0.32.

Method 2: Support Vector Machine

Support Vector Machine is supervised machine learning algorithmic program which will be used for each regression or classification challenges. SVM performs classification by finding the hyper-plane that differentiate the categories we tend to aforethought in n-dimensional house. SVM attracts that hyperplane by reworking our knowledge with the assistance of mathematical equations known as “Kernels”. SVM try and highlight words that square measure additional attention-grabbing, e.g., recurrence during a document however not across documents, thus help us to encode as much as information as possible from text. We have created a Radial basis perform kernel (RBF)/ mathematician Kernel SVM classifier that may be a standard Kernel technique it is a perform whose worth depends on the space from the origin or from some purpose.

Following is format of Gaussian Kernel:

$$K(X_1, X_2) = exponent(-\gamma \| X_1 - X_2 \|^2)$$

$| |X_1 – X_2 | |$ = Euclidean distance between X_1 and X_2

Using the Euclidean distance in the original space, we obtained the dot product that is similarity of X_1 and X_2. After implementation we got an F1 score of 0.33

Method 3: Convolutional Neural Network

CNN is the part of Supervised Deep Learning Algorithm. CNNs the first supervised learning algorithm to successfully get trained on multilayer network structure. CNN use spatial relation to decrease the number of parameters and to improve the training performance and accuracy. Its schema is multilayer convolution.

Advantages of CNN

CNN are the regularized version of multilayer perceptron. It means fully connected network, and each neuron of one layer is connected to all the neurons present in the next layer. This full connectivity helps this neural network to prone overfitting of data. It requires very less preprocessing as compared to other classification algorithms.

Because of little preprocessing it reduces the human efforts in developing and building its functionality

CNN have a very different approach to do regularization. It takes hierarchical pattern advantage and aggregate patterns of rising complexity using kernel that pulvinate in its filters.

The Pooling layer of CNN are responsible in reducing the spatial size of the convolved feature. It helps to decrease the computational power and helps to extract dominant features, which is rotational and invariantly to the position. There are two types of pooling: 1. Average Pooling 2. Max Pooling. We are using Max Pooling in our model. Max Pooling also helps in noise suppression, so it is better than average pooling.

Also, CNN has highest accuracy among all the algorithms that are used in prediction of image classification.

Our CNN Model

We have designed a CNN model which will train on our Dens-Dataset and will predict the emotions from the document given to it later. We have used Keras-Encoder to encode the textual data, and Keras -Tokenzier to divide a text in tokens. The dataset has been distributed in 2:8, 80% for training and 20% Testing. Epochs was set to 100.

CNN Model consists of

1Embedding Layers

2 Convo 1D

3 Global max Pooling

4 Dropout

5 Dense

The architecture is shown in fig 3. We got the accuracy of 24.59 % using this model.

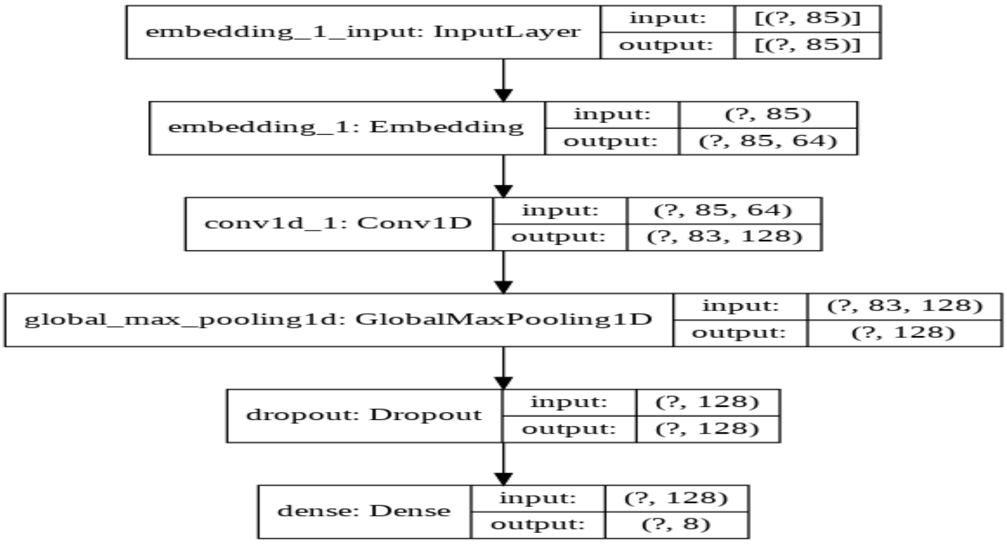


Fig. 3: CNN Model architecture

Method 4: Recurrent Neural Network

Recurrent (RNN) was chosen since it will naturally handle texts of variable length, as a result it has already shown its efficiency for text classification. It tends to experiment with 2 levels of graininess. Within the 1st approach, we tokenise the text and so introduce a concatenation of tokens into the RNN. For predicting emotions, a suitable representation of text is obtained by combining words by RNN. Thus, the task of Neural Network is to mix the characters into an appropriate illustration and to predict sentiments. Note that the RNN itself must understand that character sequences forms words, since a space isn't handled different from the other character. One advantage of the symbolic approach is that it does not need enough pre-processing and normalization.

When operating with words, Firstly, we have used tokenizer to separate the text into tokens. Next, we have got to remove the normalization drawback. Morphological variations that is internal structure of words area unit similar enough that we can use an identical token to represent them for instance, because the same token within the character settings, of these selections area unit left to the RNN's discretion. Concatenation of words or characters area unit 1st mapped into vectors, that is typically referred to as embedding. RNN is good with sequential data, because layers it uses it equipped with short term memory. And it uses these layers for accurate prediction. In the last is SoftMax layer for multinomial output.

Method 5: Word2Vec and LSTM

We applied a integrated method implementing Word2vec and LSTM model. The wiki-news pre-trained vector model is used as word embedding. Large-scale corpora can be trained by word2vec and produces word vectors of low dimension. It contains Continuous Bag of word and Skip-gram. Then LSTM is fed by trained vectors for further classification.

Architecture:

(X) Text -> Embedding (W2V pretrained on Wikipedia articles) -> Deep Network (LSTM/GRU) -> Fully connected (Dense) -> Output Layer (SoftMax) -> Emotion class (Y)

Embedding Layer

Word Embedding is giving homogenous representation to the texts with relatable interpretation. We have used word vectors pre-trained on Wikipedia articles having 300 dimensions. We could have used our dataset to train w2v model but due the small size of the dataset the efficiency won't match that of the pretrained w2v.

Deep Network

The progression of embedding vectors is set as input to Deep network and is then converted into its compact representation. All the particulars present in course of words in the text are functionally recorded in compact representation. Deep Network section is conventionally an RNN or some forms of it like LSTM/GRU.

Fully Connected Layer

The deep representation deriving out of RNN/LSTM/GRU is grabbed by fully connected layer to convert it into class scores or concluding output classes. This module includes fully connected layers in addition to batch normalization spontaneously dropout layers for normalization.

Output Layer

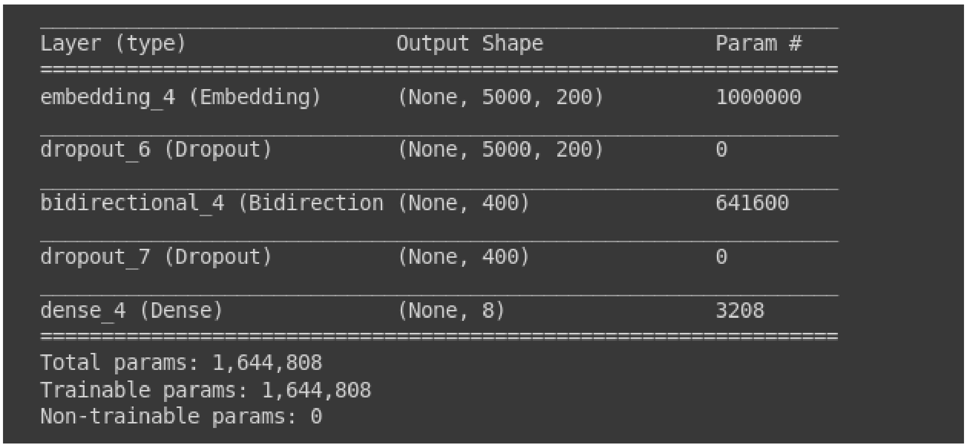
Output layers have SoftMax for both binary and multiclass classification results.

F1 Score was 30.80%

6. Results & Discussion

We have performed emotion classification using Naive Bayes, SVM, CNN, RNN, and word2vec + LSTM comparing F1 score of each it is concluded that RNN is performing best among all when trained on DENS dataset.

We used these parameters configurations for RNN Model:



```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_4 (Embedding)      (None, 5000, 200)         1000000
_________________________________________________________________
dropout_6 (Dropout)          (None, 5000, 200)         0
_________________________________________________________________
bidirectional_4 (Bidirection (None, 400)               641600
_________________________________________________________________
dropout_7 (Dropout)          (None, 400)               0
_________________________________________________________________
dense_4 (Dense)              (None, 8)                 3208
=================================================================
Total params: 1,644,808
Trainable params: 1,644,808
Non-trainable params: 0
```

• Embedding dimensions: 5000, 200

•dropout for embedding: 0.2

•RNN layer kind: Bidirectional LSTM

•RNN neurons: 200 (hidden layer)

•RNN layer bi-directional: Yes

•RNN dropout of layers: 0.2

•Dense layer: 8

•Activation layer: SoftMax.

The optimizer we have used is RMSProp for RNNs, a batch size of 128 and epoch of 10. Our system performed best result and achieved an accuracy of 54.205%.The results of all the models are listed in Table 1.

Model	**Micro-f1 Score**
Naive Bayes	32.8
SVM	33.5
CNN	24.59
Word2vec+ LSTM	33.02
RNN	**54.20**

Table 1: F1-score of all classifier models

7. Conclusion

In this work we performed sentiment analysis of Email data. We have classified emails from dens dataset eight different classes of the Plutchik’s emotion wheel: joy, trust, fear, surprise, sadness, expectation, anger, and disgust. machine learning and deep learning techniques like naïve Bayes, SVM, CNN, LSTM and RNN are used for classifying email data and we observed RNN gives higher accuracy than all other models. In future we will try to improve accuracy by using attention mechanism over RNN.

References

1. Rayan Salah Hag Ali and Neamat El Gayar. 2019. Sentiment Analysis using Unlabelled Email data.
2. Dr.Gorti Satyanarayana Murty and Shanmukha Rao Allu. ISSN: 22780181 http://www.ijert.orgIJERTV9IS050290 Published by: www.ijert.orgVol. 9 Issue 05, May-2020. [22:36] 7422 Rohit Singh. Text based Sentiment Analysis using LSTM.
3. MONDHER BOUAZIZIAND TOMOAKI OHTSUKI.2018. Multi-Class Sentiment Analysis in Twitter: What If Classification Is Not the Answer.
4. Muhammad Babar Abbas1 and Mukarram Khan. 2019 sentiment analysis for automated email response system
5. Niko Colneric and Janez Demsar.2018. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model.
6. Sisi Liu and Ickjai Lee. 2017. A Hybrid Sentiment Analysis Framework for Large Email Data.
7. Chen Liu and Muhammad Osama and Anderson de Andrade. 2019. DENS: A Dataset for Multi-class Emotion Analysis.
8. N Kalchbrenner, E Grefenstette and P Blunsom, "A convolutional neural network for modelling sentences".
9. K Chatfield, K Simonyan, A Vedaldi et al., "Return of the devil in the details: Delving deep into convolutional nets".
10. Meylan Wongkar ,Apriandy Angdresey , “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter”.

11. Jishnusri Ojaswy Akella , LN Yashaswy Akella, “Sentiment Analysis Using Naïve Bayes Algorithm: With Case Study”.
12. Nemanja Mili, ćević . “Comparing Sentiment Analysis and Document Representation Methods of Amazon Reviews”
13. Boy Utomo Manalu; Tulus; Syahril Efendi. “Deep Learning Performance In Sentiment Analysis”
14. Manish Munikar, Sushil Shakya, Aakash Shrestha, “Fine-grained Sentiment Classification using BERT”.
15. Rohit Kumar Kaliyar, “A Multi-layer Bidirectional Transformer Encoder for Pre-trained Word Embedding: A Survey of BERT”