IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 8, Issue.11, November 2020 E-ISSN: 2347-2693

Multi-Class Sentiment Classification using Machine Learning and Deep Learning Techniques

Saurav Singla[1*], Vikash Kumar[2]

[1]Independent Researcher, Gurgaon-122011, India
[2]Independent Researcher, Patna-800030, India

Corresponding Author: sauravsingla08@gmail.com, Tel.: +91-99587-93952

DOI: https://doi.org/10.26438/ijcse/v8i11.1420 | Available online at: www.ijcseonline.org

Received: 08/Nov/2020, Accepted: 19/Nov/2020, Published: 30/Nov/2020

Abstract **-** With the rapid growth of usage of online social media platforms in daily life there has also been an increase in opinion mining or sentiment analysis to extract the user's sentiments or views towards any topic. Twitter's data or tweets has been the focus point among the researchers as it provides abundant data and in a wide variety of fields. While most of the study in this field has been in the extraction of polarity scores of the sentiments namely positive negative and neutral in a tweet, this paper focuses on extracting the real sentiments such as love, hate, worry, sadness and more out of the tweets. This paper proposes different machine learning and deep learning techniques such as Random Forest, Bi-directional LSTM, BERT and more to present a comparative analysis of the performance of different techniques and extract the sentiments with high accuracy. Tweets have been collected from the Crowdflower dataset and experimental findings reveal that the methodology comprising BERT produces the maximum accuracy followed by the methodology that comprises bi-directional LSTM and then the rest of the model follows.

Keywords— Sentiment Analysis, BERT, Bi-directional LSTM, Multi-class Classification, Random Forest, GloVe

I. INTRODUCTION

Today we are living in a world where people share everything on the social media whether they are happy, sad or anything else. Twitter, Facebook & Instagram are some of the most renowned social media platforms that are being used by millions of users on daily basis. As per the stats almost 50 crores tweets are recorded each day and 6K tweets are recorded almost every second. This being the stats of twitter only. Facebook also records a breathtaking having 500 plus terabytes of data daily. Thus, enabling the extraction of people's opinion regarding any topic whether it is related to politics, education, society etc. On twitter, people share their opinion regarding various topic whether it being trendy or not in the form of text data, emojis or in other ways.

Now a days, companies use these opinions as a real time feedback against any product they possess and based on these feedbacks they find out the sentiment or opinion and make decisions. So, sentiment analysis is a technique in which users find the sentiment polarity i.e. (negative, neutral and positive) about a text message or the comments which contain opinion or the objective.

Tweets generally consist of noisy, incomplete, unstructured data with irregular expressions and slangs with meaningless and sarcastic words. Hence, a lot of preprocessing is required to structure the data before extracting the features for sentiment analysis. This paper presents a comparative analysis of different machine learning and deep learning approaches for twitter sentiment classification.

In this paper, we apply random forest, Long Short-Term Memory (LSTM), Bi-directional Long Short-Term Memory (Bi-LSTM), Bi-directional Gated Recurrent Units (GRUs), and Bidirectional Encoder Representations from Transformers (BERT) along with different vectorization techniques such as tf-idf, word2vec, and glove. First, we go through the preprocessing of the text and then these preprocessed texts are vectorized using above mentioned techniques and then the vectors obtained are fed into these machine learning and deep learning models.

The remainder of this paper is structured as follows: in section II we discuss about our motivation behind the work, section III presents the some of the works related to our work. In section IV we describe the methodology followed for our work and section V consists of the experimental results and analysis. Finally, section VI concludes this paper and proposes future work that can be carried out.

II. MOTIVATION

In previous works [1], [2], it was observed that the sentiment analysis has been performed as positive, negative or neutral. However, just distinguishing the sentences as positive, negative or neutral does not provide enough information about the reaction provided by people.

In this work, we aim to classify the statements in various sentiments such as sadness, boredom, neutral, worry, surprise, love, fun, hate, happiness, anger and relief. This would help in extracting the real feeling of the person. Importance of such classification can be represented as an example:

- “This ABC car I bought proved to be just an awful experience for me”.
- “Another guy gave a great bump on my ABC car while I was returning home”.

Both the statements can be classified as having negative sentiment. However, if seen from the company’s point of view that produces car, the first statement could be classified as hate or anger towards the car the customer has bought and can be judged as more important for the company, whereas, the second statement can be viewed more of as sadness or bad luck by the customer and can be judged of as less importance for the company. This puts focus on the fact that a positive tweet can have multiple interpretations such as surprise, happiness, love, fun etc. Typecasting all the different sentiments in a single class can be observed as sort of dropping part of information.

This provides a highlight on the importance of multi-class classification of the sentiments. As we go further forward in the coming sections, we will observe that a single tweet does not contain just one sentiment to classify it. Nevertheless, some sentiments are highly correlated with each other like tweet showing fun can also show happiness.

III. RELATED WORK

Twitter, being one of the most used social media platforms, has become a large source of data and has attracted a large number of researchers to mine data and apply various data analytics such as sentiment analysis.

In this work author [3] proposed a hybrid method by combining lexicon-based approach with a fuzzy classification approach to deal with language ambiguity and analyze tweets into seven sentiment classes. Social Network Analysis (SNA) and UCINET tool were used to obtain different measures of users and Artificial Neural Network (ANN) helps to rank the users based on influence level. The final model gave a sentiment result which expose certainty and uphold decision makers from misguided decisions based on inaccurate sentiment results.

Bibi et al. [4] investigate the feasibility of three hierarchical clustering approaches named as Single Linkage (SL), Complete Linkage (CL) and Average Linkage (AG) for twitter sentiment analysis. An interdependent framework of all the approaches were built to find out the best cluster for tweets and was done using majority voting and comparative analysis was done with k-means and two state-of-art classifiers. CL gives better results in terms of quality of clusters than SL, AL and k-means. K-means takes smallest time than any other approaches.

An ordinal regression-based sentiment analysis using machine learning techniques was performed by Saad and Yang [5]. After preprocessing of tweets data, feature extraction approach is applied to obtain efficient features. Machine learning algorithms like Multinomial logistic regression (SoftMax), Support Vector Regression (SVR), Decision Tree (DT) and Random Forest (RF) were used to find out the sentiment classification results. The best result obtained by DT model with accuracy score of 91.81%.

Rehioui and Idrissi [6] made use of two clustering methods i.e. k-means and Density based clustering (DENCLUE) with its different variants for analyzing sentiments. They find out the optimum number of clusters. K-DENCLUE-IM outperforms better than any variant of DENCLUE.

Author aimed attention at combining text data and sentiment diffusion pattern for opinion mining [7]. They proposed a new iterative algorithm called as SentiDiff which find out sentiment scores in term of polarity values of twitter text data. They study about the sentiment diffusion by examining the sentiment reversal to find out some good attributes based on repost cascade trees and diffusion network.

Author [8] proposed a word embedding technique acquired from unsupervised learning which uses latent contextual semantic association with statistical features between the words. N-grams features along with word sentiment polarity values were used to form sentiment features set data and these feature set data are inputted to deep convolution neural network for training and testing the sentiment classification labels.

A metaheuristic technique based on combination of k-means and cuckoo search was proposed [9]. This technique finds the best cluster head of the sentiment content. The efficiency of the proposed technique outperforms the existing state-of-art method. K-means helps to find out the k clusters and these k-clusters are inputted to cuckoo search algorithm to find out the best cluster head.

In this section a transformer-based technique which transforms the representation by the help of transformer and used deep intelligent contextual embedding to increase the quality of tweets by eliminating the outliers while quantifying sentiments, syntax, semantic knowledge and polysemy to analysis the sentiment called as DICET was proposed by Naseem et al. [10]. Instead of this technique, they applied bidirectional long and short-term memory network which determine the opinion of a tweet.

In another study, Amrani et al. [11] proposed a hybrid approach of random forest and support vector machines for analyzing sentiments for product's review. The random forest is based on agglomeration of many decision trees which are used for classification as well as regression. It is one of the ensembles learning techniques after bagging and boosting. Constructs number of decision trees at training time and predicts the class with maximum probability in case of multiclass classification. In order to minimize error rate, correlation between any two trees should be reduced by randomly selecting trees. Hence, the trees will be independent of each other. Disadvantage of this classifier is that it easily overfits its class. To reduce the overfits, reduce the number of trees and also decrease the ambiguous links that exist.

In this section Wang et al. [12] presented dimensional sentiment analysis using a regional CNN-LSTM model. It predicts valence arousal (VA) ratings of texts, by acquiring local information within sentences as well as long-distance dependency along sentences. LSTM is able to combat long-distance dependency problem as well as it retains information for either short or long time.

In SemEval-2018 competition, Zhang et al. [13] implemented Bi-LSTM with Attention based sentiment analysis for affect in Tweets. Bi-LSTM enables to capture textual information from both directions while attention mechanism emphasizes the most relevant words and helps to improve scores.

Sharfuddin et al. [14] proposed deep RNN with Bi-LSTM based end-to-end model for Bengali texts sentiment classification.

A target-dependent Sentiment analysis of Tweets using Bi-GRU is proposed by Jabreel and Moreno [15]. It outperforms target independent LSTM, Bi-GRU.

In another study Munikar et al. [16] proposed an approach used the pretrained BERT model and tuned it by applying dropout regularization along with SoftMax activation function in dense layer. They used this tuned model for fine-grained sentiment classification problem.

Aggarwal et al. [17] discussed about the sentiment analysis, various methodology to find out the sentiment along with its applications and challenges. This paper helps new comer to have basic knowledge regarding the sentiment analysis, various techniques to find out the sentiments, application and challenges.

Huddar et al. [18] discussed approaches and challenges in Multimodal sentimental analysis. Data is in various modalities such as text, image, audio, and video. Sentiments are extracted from transcribed content, visual and vocal features. Multimodal approach outperforms unimodal approaches.

IV. METHODOLOGY

Although multi-class classification of tweets among various sentiments proves to be useful for extracting the correct sentiment rather than just the positive, negative or neutral sentiment. But classification of the tweets among multiple categories has its own limitations. A single tweet can contain more than one sentiment and just extracting one most dominant sentiment out can lead to misclassification.

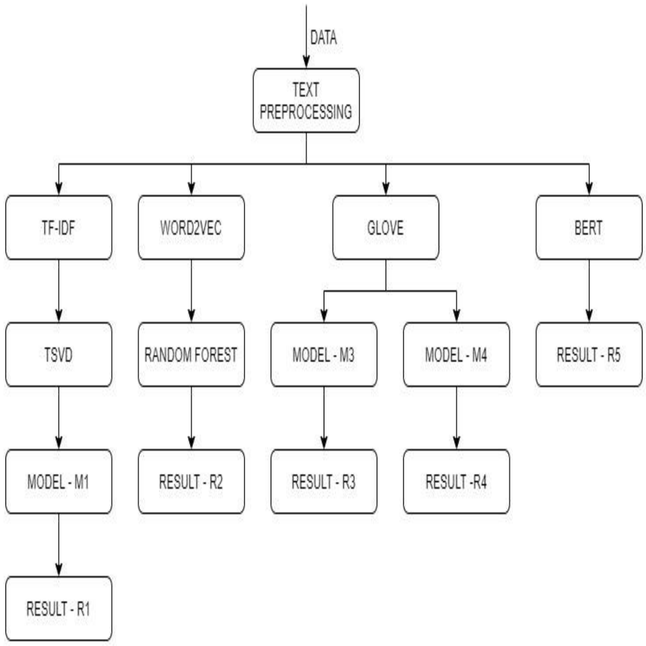

Figure 1. Systen Diagram of Proposed Methodology

Crowd flower dataset has been used by us for our work which contains tweets labelled with 11 different sentiments namely sadness, boredom, neutral, worry, surprise, love, fun, hate, happiness, anger and relief for classification. We have used five different approaches for the classification of sentiments with the preprocessing phase being the same for all the models.

In preprocessing, we have removed the abbreviations and slangs followed by the lemmatization of the tweets. In the end we have removed the stop words and then these tweets are fed to the model.

In the first proposed methodology M1 we have use tf-idf vectorizer for transforming the tweets followed by truncated SVD for dimensionality reduction. These vectors are used by model M1 shown in figure 2 comprising of bidirectional GRU layers, dropout layers, global max pooling layer and dense layers.

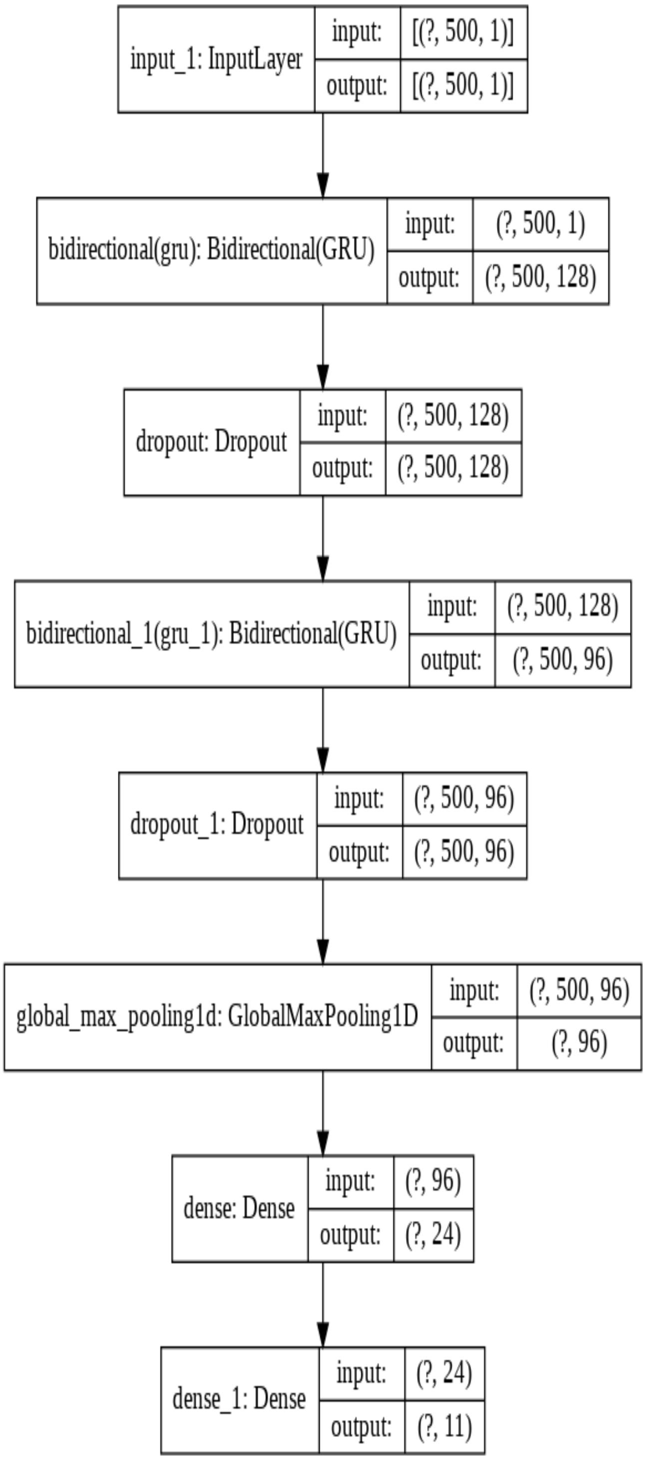

Figure 2. Model Architecture of M1

For the second methodology M2 word2vec is used for vectorizing the tweets and random forest is used for classification.

We have used GloVe for vectorization of the tweets for proposed methodologies M3 and M4. Figure 3 and figure 4 represents the models proposed for the classification for proposed approaches M3 and M4 respectively.

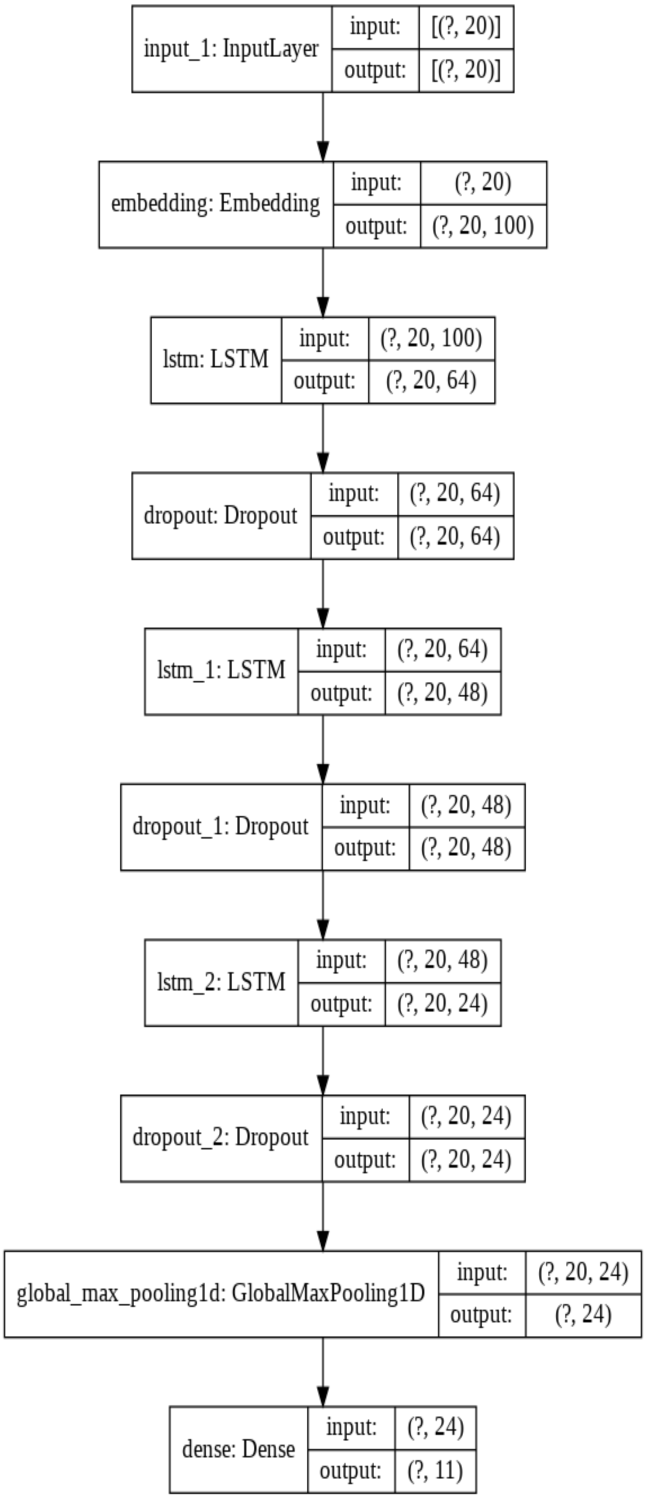

Figure 3. Model Architecture of M3

Both the models contain embedding layer, dropout layer, dense layer and golbal max pooling layer. However, the model in proposed methodology M3 contains LSTM layer where as the model in proposed methodology M4 contains Bidirectional LSTM.

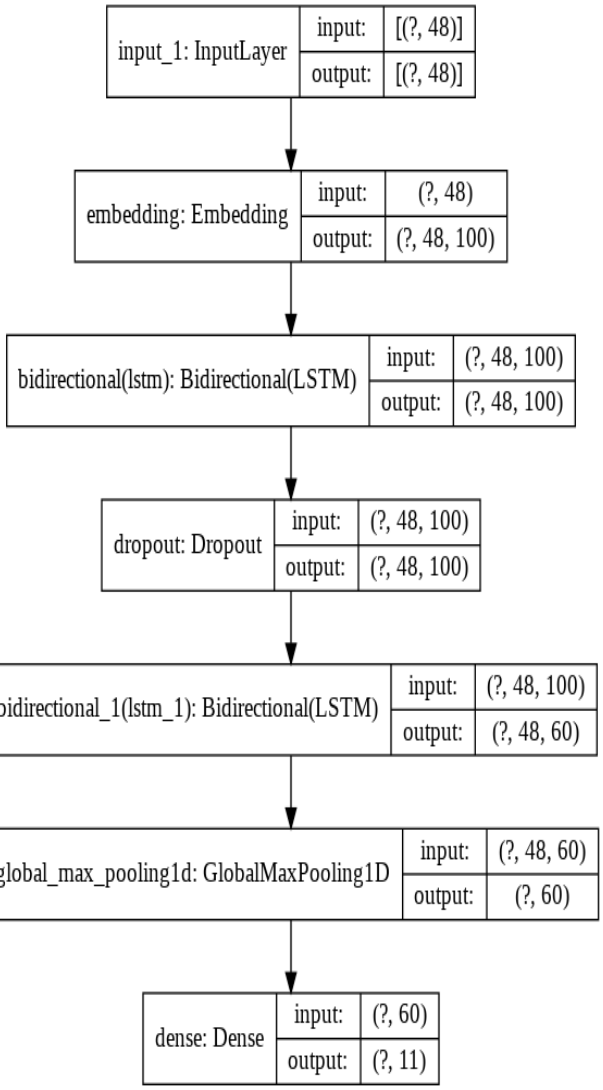

Figure 4. Model Architecture of M4

For the last approach we have fine tuned BERT for the classification of the pre-processed tweets.

V. RESULTS AND DISCUSSION

In this section, we present the results obtained for multi-class classification of sentiments using different methods proposed. The models are evaluated using precision, recall and F1 score for each class and accuracy overall:

- Recall (Rec) or True Positive Rate measures the rate of the tweets classified correctly as being of that class with respect to the total number of tweets of that class:
  Recall = TP / (TP + FN)
- Precision (Prec) measures the rate of total tweets classified correctly as being of that class with respect to the total number of tweets classified of being in that class:
  Precision = TP / (TP + FP)
- F_1 score is a combination of precision and recall:
  F_1 score = (2 * Prec * Rec) / (Prec + Rec)
- Accuracy is the fraction of predictions model got right:
  Accuracy = (TP + TN) / (TP + TN + FP + FN)

TP, FP, TN, FN terms in the context of classification are defined as:

- TP (True Positive): Tweets that belongs to a class C and are classified as being in class C
- FP (False Positive): Tweets that does not belong to class C but are classified as being in class C
- TN (true Negative): Tweets that does not belong to class C and are classified as not belonging to class C
- FN (False Negative): Tweets that belongs to class C but are classified as not being in class C

Table 1 represents different sentiment classes and number of samples corresponding to each class in the dataset.

Table 1. Class Distribution

Class	Class Name	No. of Samples
0	Sadness	5165
1	Boredom	179
2	Neutral	8638
3	Worry	8459
4	Surprise	2187
5	Love	3842
6	Fun	1776
7	Hate	1323
8	Happiness	5209
9	Anger	110
10	Relief	1526

The dataset is divided into 70:30 ratio for the purpose of training and testing the models. Tweets are classified into different classes of sentiments using different methods.

Table 2. Comparative Analysis of Precision of Models

Class	M1	M2	M3	M4	M5
0	0.20	0.36	0.37	0.38	0.36
1	0.00	0.00	0.00	0.00	0.00
2	0.31	0.35	0.37	0.39	0.38
3	0.24	0.32	0.35	0.34	0.46
4	0.00	0.06	0.23	0.15	0.34
5	0.42	0.51	0.44	0.46	0.52
6	0.00	0.00	0.12	0.16	0.22
7	0.00	0.33	0.32	0.26	0.35
8	0.19	0.33	0.30	0.34	0.37
9	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.16	0.30

Table 2 represents the comparative analysis of precision of models for different classes. All the proposed approaches record their lowest precision value 0 for classes 1 and 9 and obtain their highest precision value for class 5. This shows that class 5 has the best rate of correct predictions among all classes with approach M5 having the greatest precision value among all approaches.

Table 3. Comparative Analysis of Recall of Models

Class	M1	M2	M3	M4	M5
0	0.03	0.08	0.13	0.21	0.35
1	0.00	0.00	0.00	0.00	0.00
2	0.20	0.59	0.60	0.52	0.65
3	0.79	0.61	0.45	0.49	0.34
4	0.00	0.00	0.02	0.05	0.03
5	0.18	0.26	0.42	0.39	0.40
6	0.00	0.00	0.01	0.11	0.04
7	0.00	0.01	0.14	0.28	0.30
8	0.11	0.26	0.45	0.38	0.52
9	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.03	0.04

Table 3 represents the comparative analysis of recall of models of different classes. Methodology M1 records its highest recall value of 0.79 for class 3 which is the highest value for any class obtained by any model. Methodology M2 also records its highest recall value for class 3. M3, M4 and M5 obtains their highest recall values for class 2.

Table 4. Comparative Analysis of F-1 Scores of Models

Class	M1	M2	M3	M4	M5
0	0.06	0.13	0.19	0.27	0.36
1	0.00	0.00	0.00	0.00	0.00
2	0.24	0.44	0.45	0.45	0.48
3	0.37	0.42	0.40	0.40	0.39
4	0.00	0.00	0.03	0.07	0.05
5	0.26	0.35	0.43	0.42	0.45
6	0.00	0.00	0.02	0.13	0.06
7	0.00	0.01	0.20	0.27	0.32
8	0.14	0.29	0.36	0.36	0.43
9	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.05	0.07

Table 4 represents the comparative analysis of F-1 score of models for different classes. All the methodologies except for M1 obtains their highest F-1 score for class 2 where as M1 records its best F-1 score for class 3. All the methodologies obtain their lowest value of 0 for classes 1 and 9.

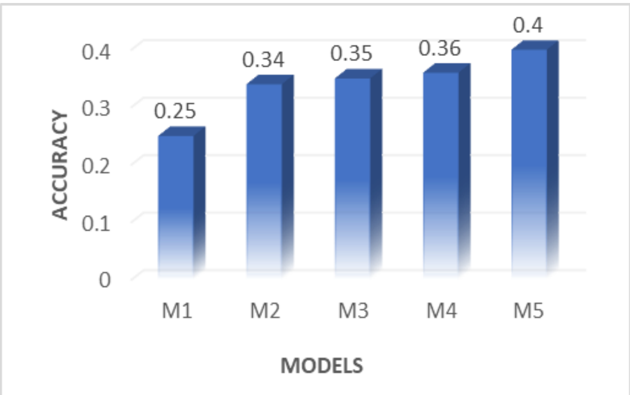

Figure 5. Comparative Analysis of Models

Figure 5 represents a comparison graph of accuracies of the different models proposed. The highest accuracy 40% is obtained by the model M5 which uses BERT for sentiment classification and lowest accuracy is obtained by model M1 which uses tf-idf for vectorization of the tweets and model comprises of layers of Bidirectional GRUs, Global Maxpool1D and flatten. However, none of the proposed succeeds in classifying the tweets as boredom or anger.

VI. CONCLUSION AND FUTURE SCOPE

In this paper we have introduced sentiment classification of twitter data for 11 different sentiments conveyed through the tweets. We have tried to extract the exact sentiments out of the tweets instead of the polarity scores of the tweets. We have followed a number of different vectorization techniques, machine learning and deep learning models to extract the sentiments out of the tweets. We obtained the maximum accuracy of 40% of classification with BERT with majority of the other models reaching the maximum accuracy of around 35%. However, none of the models were able to classify tweets as boredom or anger. One of the potential reasons for misclassification of the sentiments can be the presence of more than one sentiment in the same tweet. Presence of comparatively small number of samples for some classes can also be pointed as a reason for their misclassification as portrayed in Table 1.

In our future work, we will increase try to extract the sentiments more accurately by using different featurization techniques and deep learning techniques such as convolutional neural networks. We also aim to extract multiple sentiments, if present in a tweet and then mark out the most dominant one of them while keeping the rest of the sentiments extracted in consideration too.

REFERENCES

[1] M. Bouazizi, T. Ohtsuki, “A pattern-based approach for multi-class sentiment analysis in Twitter”, IEEE Access, **vol. 5, pp. 20617-20639, 2017.**

[2] M. Bouazizi, T. Ohtsuki, “Sentiment analysis: From binary to multiclass classification: A pattern-based approach for multi-class sentiment analysis in Twitter”, In the Proceedings of 2016 IEEE International Conference on Communications (ICC), Malaysia, **pp. 1-6, 2016.**

[3] M.M. Madbouly, S.M. Darwish, R. Essameldin, “Modified fuzzy sentiment analysis approach based on user ranking suitable for online social networks”, IET software, **Vol. 14, Issue 3, pp. 300-307, 2020.**

[4] M. Bibi, W. Aziz, M. Almaraashi, I.H. Khan, M.S.A. Nadeem, N. Habib, “A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis”, IEEE Access, **Vol. 8, pp. 68580-68592, 2020.**

[5] S.E. Saad, J. Yang, “Twitter sentiment analysis based on ordinal regression”, IEEE Access, **Vol. 7, pp. 163677-163685, 2019.**

[6] H. Rehioui, A. Idrissi, "New Clustering Algorithms for Twitter Sentiment Analysis", IEEE Systems Journal, **Vol. 14, Issue 1, pp. 530-537, 2019.**

[7] L. Wang, J. Niu, S. Yu, "SentiDiff: Combining textual information and sentiment diffusion patterns for Twitter sentiment analysis", IEEE Transactions on Knowledge and Data Engineering, **Vol. 32, Issue 10, pp. 2026-2039, 2020.**

[8] Z. Jianqiang, G. Xiaolin, Z. Xuejun, “Deep convolution neural networks for twitter sentiment analysis”, IEEE Access, **Vol. 6, pp. 23253 - 23260, 2018.**

[9] A.C. Pandey, D.S. Rajpoot, M. Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method”, Information Processing & Management, **Vol. 53, Issue 4, pp.764-779, 2017.**

[10] U. Naseem, I. Razzak, K. Musial, M. Imran, "Transformer based deep intelligent contextual embedding for twitter sentiment analysis", Future Generation Computer Systems, **Vol. 113, pp. 58-69, 2020.**

[11] Y. AI. Amrani, M. Lazaar, KE. EI. Kadiri, “Random forest and support vector machine based hybrid approach to sentiment analysis", Procedia Computer Science, **Vol. 127, pp. 511-520.**

[12] J. Wang, L.C. Yu, K.R. Lai, X. Zhang, “Dimensional sentiment analysis using a regional CNN-LSTM model”, In the Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, **Vol. 2: short papers, pp. 225-230, 2016.**

[13] Y. Zhang, J. Wang, X. Zhang, “YNU-HPCC at SemEval-2018 Task 1: BiLSTM with attention based sentiment analysis for affect in tweets", In the Proceedings of the 12th International Workshop on Semantic Evaluation, China, **pp. 273-278, 2018.**

[14] A.A. Sharfuddin, M.N. Tihami, M.S. Islam, “A deep recurrent neural network with bilstm model for sentiment classification”, In the Proceedings of the 2018 IEEE International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, **pp. 1-4, 2018.**

[15] M. Jabreel, A. Moreno, “Target-dependent Sentiment Analysis of Tweets using a Bi-directional Gated Recurrent Unit”, In the 13th International Conference on Web Information Systems and Technologies (WEBIST), **pp. 80-87, 2017.**

[16] M. Munikar, S. Shakya, and A. Shrestha. "Fine-grained sentiment classification using bert." In the Proceedings of the 2019 IEEE International Conference on Artificial Intelligence for Transforming Business and Society (AITB), Kathmandu, Nepal, **pp. 1-5, Vol. 1, 2019.**

[17] D. G. Aggarwal, “Sentiment Analysis: An insight into Techniques, Application and Challenges”, International Journal of Computer Sciences and Engineering, **Vol.6, Issue.5, pp. 697-703, 2018.**

[18] M.G. Huddar, S.S. Sannakki, V.S. Rajpurohit, “A Survey of Computational Approaches and Challenges in Multimodal Sentiment Analysis”, International Journal of Computer Sciences and Engineering, **Vol.7, Issue.1, pp. 876-883, 2019.**

AUTHORS PROFILE

Saurav Singla is a Senior Data Scientist and a Machine Learning Expert. He has fifteen years of comprehensive experience in statistical modeling, machine learning, natural language processing, deep learning, and data analytics. He has a Master of Science from the University of Westminster. He has been recognized for maximizing performance by implementing appropriate project management tools through analysis of details to ensure quality control and understanding of emerging technology. Outside work, Saurav volunteers his spare time for helping, coaching, and mentoring young people in taking up careers in the data science domain. He has created two courses on data science, with over twenty thousand students enrolled in it. He regularly authors articles on data science.

Vikash Kumar is machine learning engineer. He did Bachelor of Engineering from Sant Longowal Insitute of Engineering and Technology in year 2018 and Master of Technology from National Institute of Technology, Patna in year 2020. His main research work focuses on appliedAI, Big Data Analytics, machine learning, deep learning, with specialization in Natural Language Processing, Recommendation system and Image Processing. He has been selected in top 250 entries in Hackathon against COVID-19 held in April 2020. Outside work, he spare his time in helping and guiding newcomers in taking up the courses which can uplift the career in the field of data science and analytics.

