Stock market analysis using sentiment analysis

[1]Anoop Pande, [2]Kanak Jaiswal, [3]Jasmin Kaur Batra, [4]Dr. Manjiri Ranjanikar
[1]Student,[2]Student,[3]Student,[4] Assistant Professor, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India.

ABSTRACT :- One of the major problem on twitter is hateful tweets that can beanalyzed and can be removed from Twitterin accordance with twitter policies. The Algorithms Exploration explores Unison models and feature extractors. Different learning models are compared, including convolutional and recurrent neural networks. Human decision-making is heavily influenced by emotions. Affective computing takes this fact into account with the intention of adapting decisionsupport to an individual's emotional state.Deep learning can improve performance, but as this study shows, the uniqueproperties of this task include: bidirectional processing, dropout layers asregularization techniques, weighted lossfunctions, Requires recurrent neural network adaptation. In this case, the network is first trained for another goal, inthis case sentiment analysis, and then the output layer is adapted to the emotion detection task.

Keywords :- Sentiment Analysis, Cyberbullying, Emotion, Emotion Recognition, Unison model, Feature extractor, Convolutional NeuralNetwork(CNNs), Recurrent Neural Network(RNNs).

1. INTRODUCTION :-

Sentiment analysis, also known as opinion mining, is the process of extracting subjective information from text and categorizing it as positive, negative, or neutral. In recent years, sentiment analysis has gained significant attention in the field of natural language processing due to its wide range of applications in variousdomains, including social media analysis, customer feedback analysis, and stockmarket prediction. Twitter, being one of themost popular social media platforms, has become a popular source of data for sentiment analysis. However, analyzing Twitter data presents unique challenges, such as the limited length of tweets and theinformal nature of the language used. To address these challenges, researchers have developed various algorithms and models for Twitter sentiment analysis. One popularapproach is to use machine learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to classify tweets as positive, negative, or neutral. CNNs areparticularly useful for analyzing text data because they can automatically extract features from the input data, while RNNs can capture the temporal dependencies in the data, which is useful for analyzing sequences of tweets. Another approach is touse lexicon-based methods, which involve the use of dictionaries or lists of words withpre-assigned sentiment scores. These methods can be effective for analyzing social media data, as they can capture the nuances of language used in tweets. In addition to these approaches, there are alsohybrid methods that combine multipletechniques, such as using both lexicon-based methods and machine learning algorithms. Overall, sentiment analysis of Twitter data is an important task that has wide-ranging applications in various domains. While there are many different approaches to sentiment analysis, selectingthe most appropriate method depends on the specific application and the characteristics of the data being analyzed.

2. LITERATURE REVIEW :-

Hashtags are exploited to get three differentclassification of emotions i.e. Ekman(six basic emotions), Plutchik(eight emotions) and Profile of Mood States(POMS)(sixty five adjectives for six emotions). Ekman defines six basic emotions i.e. anger, disgust, fear, joy, sadness and surprise. These tweets are then tested on Neural Network BagOfWords(BOW) and thenpossibility to build it on Unison Model waspredicted. In the Unison model, trained dataset of Ekman can be input to other twoclassifications.

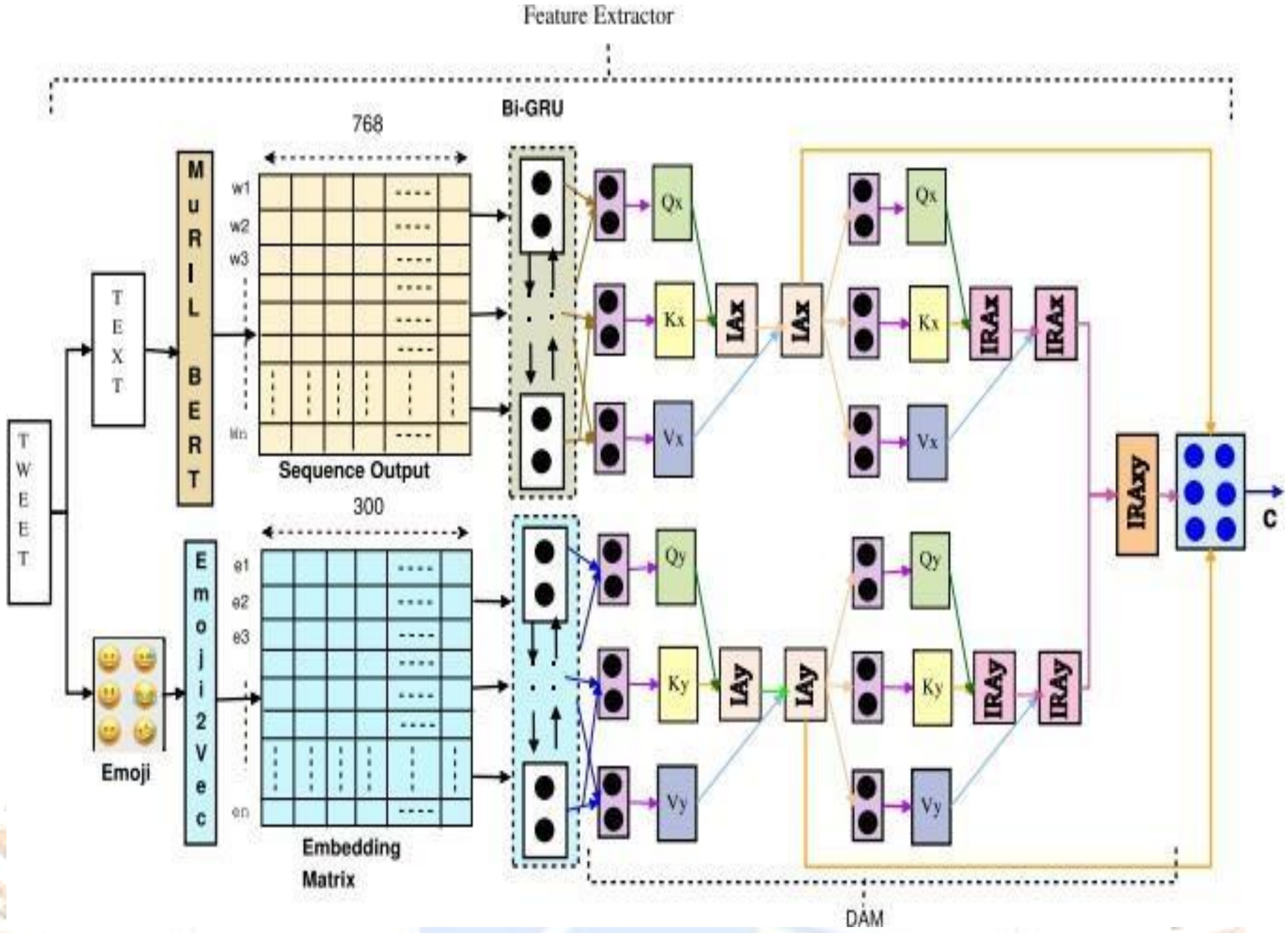

Figure 1 : Feature extraction module of proposed MM-CBD architecture.

Tweets were classified as textual tweets andemojis. Textual features from the input tweets were extracted using the BERT model. Emoji2vec was used to create 300- D vector representation from emoji. Now, the Bi-GRU layer is being sent across both word vectors. To capture long-term dependency, Bi-GRU encodes vectors inboth the forward and backward directions. The outputs of Bi- GRU are then passed to three fully connected layers(queries(Q), keys(k), values(V)). To calculate attention values triplets are passed to intramodel andintermodal attention. Intramodel is used to understand dependence between the currentword and previous portion of the tweet. With the help of these intramodel attention scores we calculate intermodal attention. All the calculated or obtained IAand IRA vectors are combined together known as attention fusion.

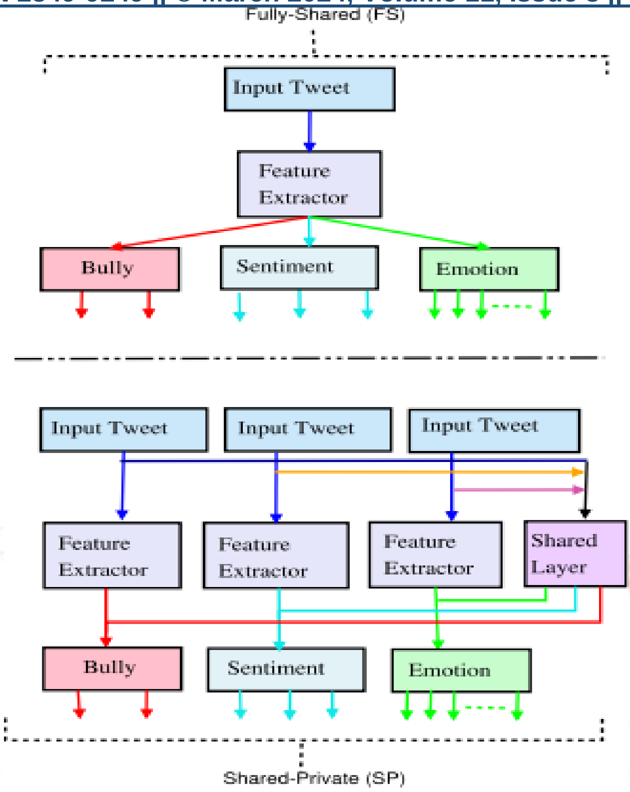

Figure 2 : FS and SP multitask modules of MM-CBD model.

The complexity and multidimensionality of human emotions and opinions, and the current requirement for in-depth analysis of the content at hand, make such solutions obsolete. Because of these requirements,research is currently focused on identifyingemotions as well as moods expressed in specific texts. However, given the complexand multi-layered nature of human emotions and opinions, and today's demands for in-depth study of the content at hand, such solutions are no longer appropriate. We are currently focused on identifying emotions, not just the moods depicted in certain texts. A consistent lack of social media posts communicating expectations was demonstrated. The hardest emotions to find in tweets are joy and surprise, but anger and trust are equallyelusive.

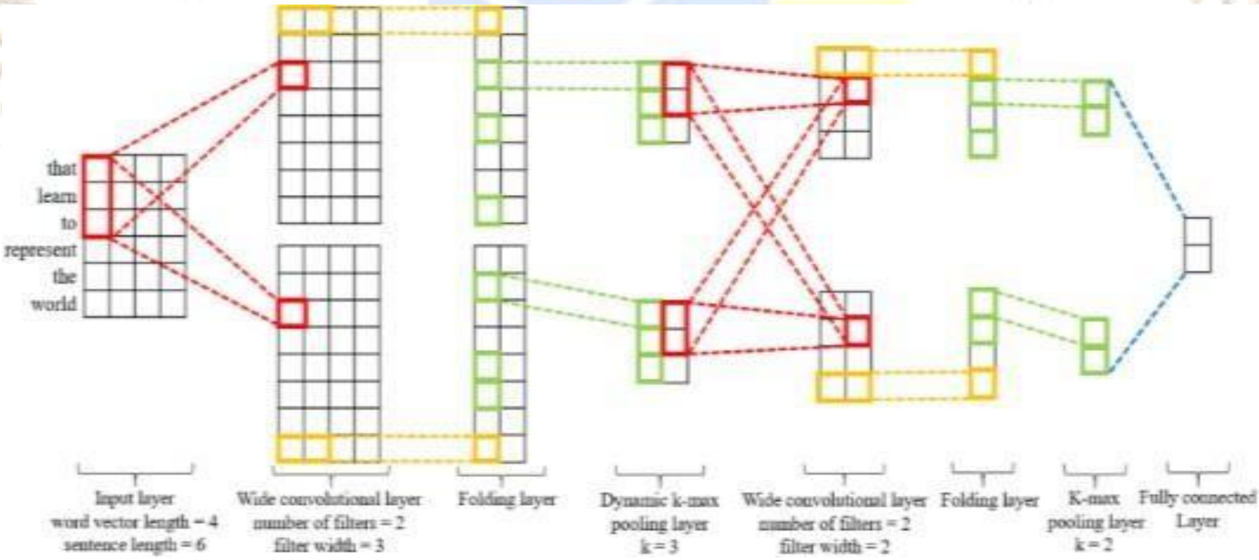

Figure 3: Dynamic Convolutional Neural Network

Sentiment analysis (SA) is crucial indetermining the overall contextual polarityof a document by inferring sentiment or emotion from text and visual components like photographs and videos. The domains of image identification and classification in machine learning are currently expanding quickly (ML). In order to provide a baselinefor customized SA algorithms, this research reviews open-source machine learning algorithms created using neural network- based frameworks like TensorFlow and Keras. This study also supports the use of open-source Scikit- Learn models tocategorize images and text tweets. For the purpose of enabling and assisting in thecorrelation of this study with existing, current, and future avant-garde and revolutionary methodologies, two notable, freely accessible, and manually annotated benchmark text and image datasets were used. Deep learning can increaseperformance, however as this research shows, according to the particulars of the task, recurrent neural networks must be customized with regard to bidirectional processing, dropout layers forregularization, and weighted loss functions.We also propose sent2affect, a special typeof transfer learning for affective computing.In this case, the network is first trained for another task (sentiment analysis). Python isthe programming language used in this project. Python has garbage collection and dynamic typing.

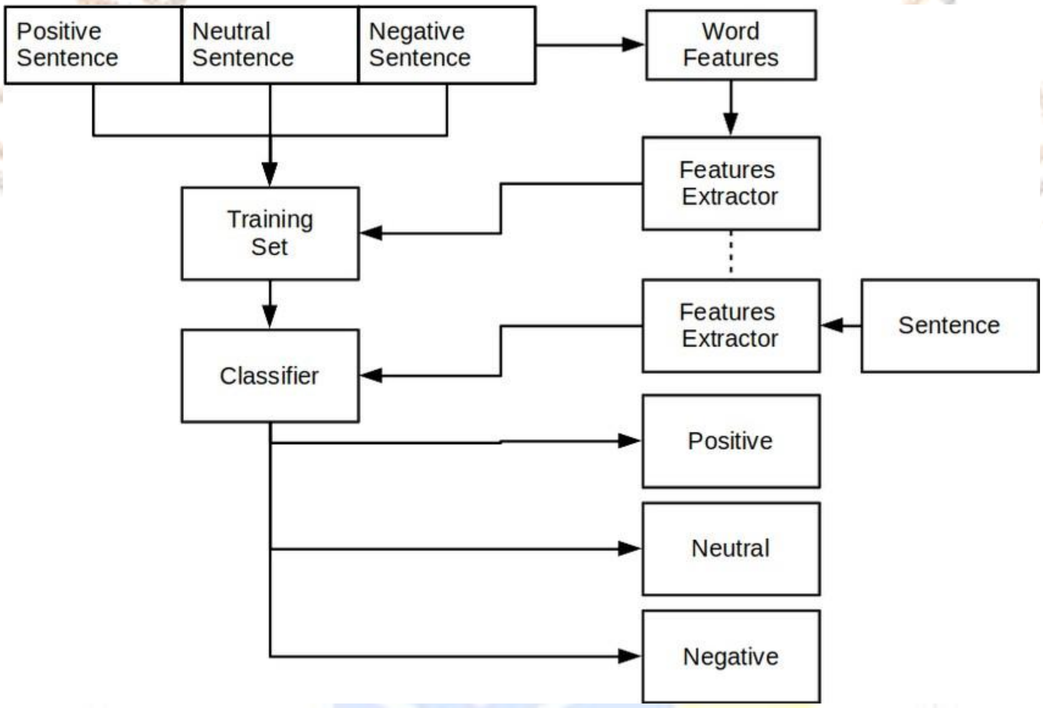

Figure 4 : Sentiment analysis architecture

The paper talks about how the widespread informal communities have made itpossible to access an unprecedented amountof publicly available client-created information, which may also be used togauge people's opinions and emotions. Bidirectional Encoder Portrayals from Transformers (BERT) designs are used for Twitter information feeling awareness and each sensation analysis. There is a vast amount of publicly available user- generated content as a result of movingvirtual entertainment. The method for robotic extraction of feelings is called feeling assessment. Extreme views of the author and how they are portrayed in thoseview: impartial, neutral, and favorable However, feeling like investigation has thetext's communication of the point about recognizing the inclination. we engaged in work-area learning on how to handleframework usage designs that aggregate,examine, and group. A crucial component of developing frameworks for emotionalintelligence is awareness of the feelings of those who communicate with others. Feeling is a complex region of the idea thatis influenced by external events, physiological changes, or relationships with other people. According to test results,our device can categorize user emotions into seven categories with a remarkably high degree of accuracy, including happiness, shock, anger, contempt, sorrow,dread, and neutrality. Physiological changes, or relationships with other people. According to test results,our device can categorize user emotionsinto seven categories with a remarkably high degree of accuracy, including happiness, shock, anger, contempt, sorrow, dread, and neutrality.

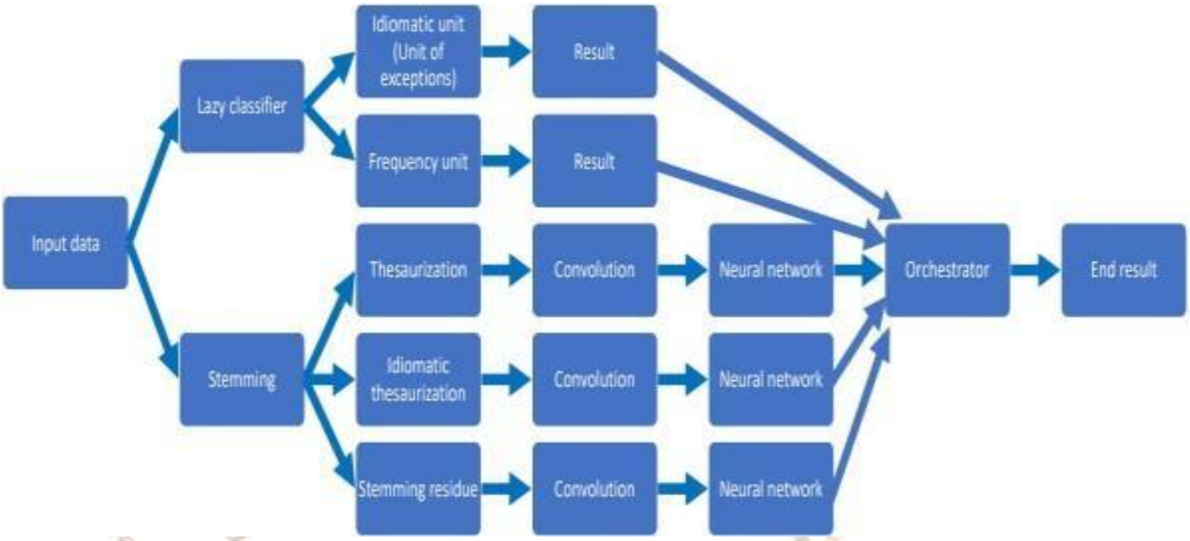

Figure 5 : System architecture

In this paper, Cyberbully Detection(CBD) of tweet is done. Three parameters were taken to determine tweet, namely, Sentimentclass (positive/neutral/negative), Emotion Recognition(Eckman's six emotions), and on basis of these two parameters third parameter is decided(Bully or Not Bully).Datasets were passed to Unimodel (Text only) and Multimodel (Text+Emoji). Result showed that Multimodel outperform Unimodel in both accuracy and FI score in all three parameters.

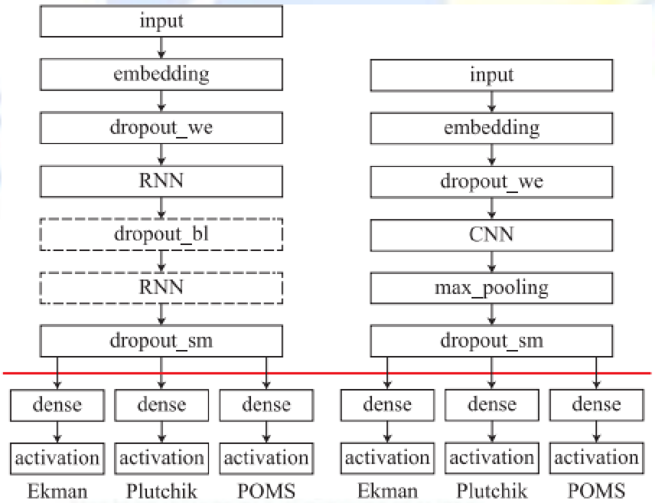

Figure 6 : The architecture of RNN (left) and CNN(right) unison models.

Table 1. Comparisons of state-of-the-art models andthe proposed model for CBD in terms of accuracy and F1-Score.

Models	Accuracy	F1-Score
BERT+CNN+Capsule	77.70	76.58
BERT+LSTM+Capsule	78.18	78.48
BERT+GRU+Capsule	78.33	77.19
BERT+CNN+GR U+ Capsule	79.28	80.30
SP+DAM+Adv	82.87	82.86

Shared-Private(SP) + Dyadic AttentionMechanism(DAM) + Adv has better F1score than other learning models. Several techniques for determining emotional states from narrative content arepresented in this section. Due to thefundamental nature of affective computing,we must first summarize our baselines from classical machine learning and deep learning while simultaneously developing several improvements for the networkdesign. Finally, we describe our innovativemethod of transfer learning, called sent2affect, which applies understanding from the related task of sentiment analysis to emotion recognition. Figure shows how this pipeline works.

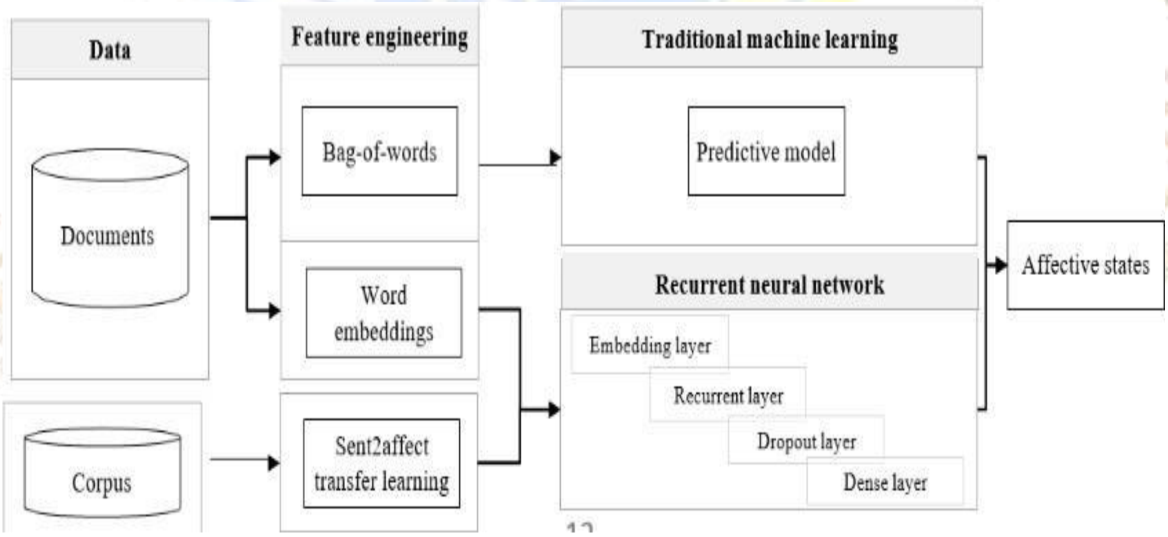

Figure 7 : Illustrative pipeline for inferring affective states from narrative material

Table 2 : Holistic comparison of traditional machine learning and recurrent neural networks

Dataset	Traditional Machine Learning	Traditional Machine Learning	Deep Learning (RNNs)	Pretrained Word Embedding
Text Classification	Accuracy: 88%	Accuracy: 91%	Accuracy: 93%	Accuracy: 94%
Time Series Forecasting	MAE: 12.5	MAE: 10.2	MAE: 9.6	NA
Image Recognition	Accuracy: 92%	Accuracy: 93%	Accuracy: 96%	NA
Speech Recognition	WER: 12%	WER: 10%	WER: 8%	NA
Sentiment Analysis	F1 Score: 0.85	F1 Score: 0.88	F1 Score: 0.90	NA

In this case, a single label is represented bythe result variable using a predetermined categorical emotion model. As a result, theperformance is evaluated using the F1- score, with a higher score being better. All models that perform better than the topbaseline model is bolded. The relative improvement over the best baseline from conventional machine learning is indicated by the percentage changes.

NRC Lexicon There are eight basic emotions: anger, fear, joy, sadness, and trust. Disgust, eagerness, and astonishment)and two emotions are on the NRC Emotion Lexicon (negative and positive).We used crowdsourcing to manually annotate the pictures. The Support Vector Machine (SVM) algorithm is thought to perform sentiment analysis well. SVM definespreference, limits and applies evaluationtechniques, and assesses records that are obtained inside the index region. The vectorarrangements of magnitude includeimportant information. To do this, the data has been sorted by kind and represented as a vector. After that, in two training sessions, the border is stratagem-classified.Regression and classification applications typically use the supervised ML strategy known as the decision tree algorithm. Selecting the tree's root node at each level is the main challenge, sometimes referred toas attribute selection. The Gini index and knowledge gain are the two most widely used methods for attribute selection. In thisstudy, the attribute value total is squared, then one is subtracted to determine the likelihood of the root node using the Gini index.

Table 3 : Comparison of Deep learning models

Model	Accuracy	Precision	Recall	F1 score
Stacked LSTM Model	0.824	0.80	0.16	0.27
Stacked LSTM with 1D convolution	0.81	0.82	0.95	0.88
Bert based Model	0.832	0.861	0.783	0.816
CNN with pre-trained word embeddings	0.841	0.839	0.84	0.839

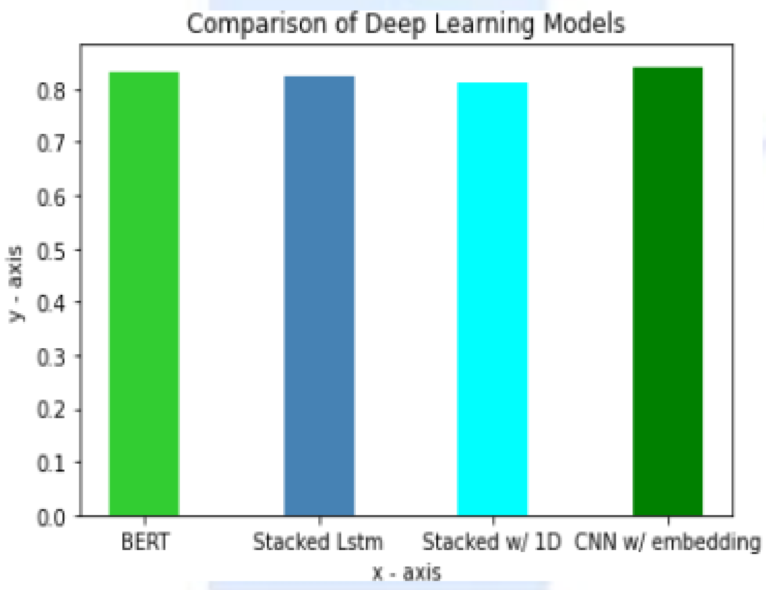

Figure 8 : Comparison of Deep learning models

A constructive Rule-Based algorithm is used with detailed execution steps in the following:

Step 1: (Data Collection) Data is obtained from datasets that consist of various attributes. The purpose is to extract the sentiment to predict whether the tweet is negative or positive.

Step 2: (Tokenization) Process of tokenization is carried out. Sentence tokenizer divides the text into a list of sentences and word tokenizer for dividing a sentence into words. The whole sentence is made into tokens.

Step 3: (Pre-Processing) Data that has not been carefully screened can produce misleading results. If there is much irrelevant and redundant information or noisy and unreliable data reduces the performance ,hence removing such noisy

data is preprocessing.

Step 4: (Tagging) In tagging, tag is applied for the word based on its previous and next (following) word. Tagging is based on context and each word is tagged.

Step 5: (Feature Selection) All specified tags may not be useful, to know the most specific emotional words we have taken the more frequently occurring words in the dataset as features. As a result, feature space does no longer include all the words, but instead it only contains the emotional words from the defined Knowledge base.

Step 6: (Emotion/Mood Detection) The results are classified into various categories depending on results obtained.

Table 4. SOTA – evaluation

Language specific algorithm	SOTA average accuracy	Dataset
M-BERT BaseFiT	0.874	RuTweetCorp (full version)
Dual-trained Lazy CNN	0.843	RuTweetCorp (clean version)
nb-blinov	0.816	ROMIP-2012
Naive-Bayes + Thesaurus	0.697	RuTweetCorp (full version)

Table 5. Table of emotion classes

Class	Training sample elements	sting sample elements
0 - Disgust/Hate	2288	2288
1 - Sad	2297	2297
2 - Happy	2283	2283
3 - Surprise/Fear	2374	2374

3. CONLUSION :

We explored different techniques or modelsfor detecting sentiment in tweets. We learned that a dataset trained with one model can be used as an input to another model of the unison model. We compared different learning models to determine the best learning model for detecting sentimentin tweets. There is a lot of research on usingmachine learning techniques to predict epidemics and outbreaks. However, there appear to be some drawbacks that needattention in order to get better results from machine learning-based sentiment analysis.Each of the various machine learning techniques described above has its strengthsand weaknesses. Deep learning has significantly improved the performance of a variety of natural language processingapplications.

4. REFERENCES :-

1. Goyal, Shilpi, and Nirupama Tiwari. "Emotion recognition: A literature survey." Int. J. Technol. Res. Eng 4, no. 9 (2017).

2. Bandhakavi, N. Wiratunga, D.Padmanabhan and S. Massie. "Lexicon-based feature extraction for emotion text classification." Pattern Recognit. Lett., vol. 93, pp.133-142, Jul. 2017.

3. J. Capdevila, J. Cerquides, J. Nin and J. Torres. "Tweet-SCAN: An event discovery technique for geo-located tweets." Pattern Recognit. Lett., vol. 93, pp. 58-68, Jul.2017.

4. T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu and J. Planes. "An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships." Pattern Recognit. Lett., vol. 105, pp. 191-199, Apr. 2018.

5. Jamal, N.; Xianqiao, C.; Aldabbas, H. "Deep Learning-Based Sentimental Analysis for Large-Scale Imbalanced Twitter Data." Future Internet 2019, 11, 190.

6. Colnerič, Niko, and Janez Demšar. "Emotion recognition on twitter: Comparative study and training a unison model." IEEE transactions on affective computing 11, no. 3 (2018): 433-446.

7. Ahluwalia H., Abdul-Mageed M., Ungar L. (2018). "Enabling deep learning of emotion with first- person seed expressions." In Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in social media, 25–35. 10.18653/v1/W18-1104.

8. M.Abdullah, M. Hadzikadicy, S. Shaikhz Sedat. "Sentiment and emotion detection in Arabic text using CNN-LSTM deep learning." 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE (2018), pp. 835-840.

9. Barbieri, Francesco, Jose Camacho Collados, Leonardo Neves, and Luis Espinosa-Anke. "Tweeteval: Unified benchmark and comparative evaluation for tweet classification." arXiv preprint arXiv:2010.12421 (2020).

10. AlZoubi, Omar, Saja Khaled Tawalbeh, and AL-Smadi Mohammad. "Affect detection from Arabic tweets using ensemble and deep learning techniques." Journal of King Saud University-Computer and Information Sciences (2020).

11. C. Kariya and P. Khodke. "Twitter sentiment analysis." Proc. Int. Conf. Emerg. Technol. (INCET), pp. 212-216, Jun. 2020.

12. Wadhawan, Anshul, and Akshita Aggarwal. "Towards Emotion Recognition in Hindi-English Code- Mixed Data: A Transformer Based Approach." arXiv preprint arXiv:2102.09943 (2021).

13. Alhuzali, Hassan, and Sophia Ananiadou. "Improving Textual Emotion Recognition Based on Intra- and Inter-Class Variation." IEEE Transactions on Affective Computing (2021).

14. Keelawat, Panayu, Nattapong Thammasan, Masayuki Numao, and Boonserm Kijsirikul. "A comparative study of window size and channel arrangement on EEG-emotion recognition using deep CNN." Sensors 21, no. 5 (2021): 1678.

15. Abdelali A., Hassan S., Mubarak H., Darwish K., Samih Y. (2021). "Pre-training BERT on Arabic tweets: practical considerations." arXiv Prepr. arXiv2102.10684. 10.48550/arXiv.2102.10684.

16. Chiorrini, Andrea, Claudia Diamantini, Alex Mircoli, and Domenico Potena. "Emotion and sentiment analysis of tweets using BERT." In EDBT/ICDT Workshops. 2021.

17. Ivanyuk, Vera, and Ekaterina Tsapina. "Creating emotion recognition algorithms based on a convolutional neural network for sentiment analysis." In International Workshop on Reproducible Research in Pattern Recognition, pp. 66-79. Springer, Cham, 2021.

18. Mehta, Ansh, Shubham Patil, Soham Dave, and Prachi Tawde. "Emotion Detection using Social Media Data." 2021.

19. Sydor, Jakub, and Szymon Cwynar. "Emotion Recognition from Tweets." (2021).

20. Maity, Krishanu, Sriparna Saha, and Pushpak Bhattacharyya. "Emoji, Sentiment and Emotion Aided Cyberbullying Detection in Hinglish." IEEE Transactions on Computational Social Systems (2022).