Deep Learning Based NLP and LSTM Models for Sentiment Classification of Consumer Tweets

[1]**P.Rajyalakshmi**, Assistant Professor, Dept of CSE, TKRCET(A), Hyderabad
[2] **Koushik Reddy Chaganti,** Masters in Computer Science, Dept of Computer and Information Systems, University of the Cumberlands, KY USA
[3] **Sathvik Chaganti,** [3] Masters in Computer Science, Dept of Computer Science, Georgia State University, Atlanta
[4]**A Anusha,** Assistant Professor, Dept of CSE -Cyber Security, CMR Engineering College, Hyderabad

Abstract:

The expanded use of virtual entertainment and online business sites is continually creating a monstrous volume of information about picture/video, sound, text, and so on. The text among these is the main sort of unstructured information, requiring exceptional consideration from scientists to obtain significant data. To gain insight from these data, numerous methods have recently been proposed. However, dealing with the enormous text poses additional difficulties; As a result, accurately detecting polarity in customer reviews is a challenging but exciting challenge. Because of this, it is difficult to precisely interpret the textual information in consumer reviews, comments, tweets, posts, and other such materials. There has been some work done in the past to make it easier to get exact meanings from these data. When interpreting such data, proper data collection, preprocessing, and classification are crucial. Deep Learning Based NLP and LSTM Models were implemented with parameters F1 score, Precision, Recall and accuracy. All parameters were applied to two different datasets and our proposed method showed good performance, accuracy.

Keywords: sentiment analysis; consumer reviews; artificial intelligence; deep learning

1. Introduction

This article aims to explain the concepts of Natural Language Processing and how to build a model using LSTM (Long Short Term Memory), a deep learning algorithm for performing sentiment analysis.

Natural Language Processing:
Natural Language Processing (NLP) is a subfield of Artificial Intelligence that deals with understanding and deriving insights from human languages such as text and speech. Some of the common applications of NLP are Sentiment analysis, Chatbots, Language translation, voice assistance, speech recognition, etc. examples Google translator,Chatbots in Apps like Flipkart & Swiggy, Autocompletion feature in Gmail, Personal Assistance like Alexa, Siri & Google Assistance, Email spam detection, Document summarization

Importance of NLP: The reason for this is that in today's world, roughly 2.5 quintillion bytes of data are generated every day. And the majority of them are inherently unstructured. Examples: Text, audio, etc. To make use of the majority of these data and to derive meaning out of it, we need to have a technology that can handle the textual and voice data. NLP is one of technology which helps to extract meaning from these types of data.

Stages in Natural Language Processing:
There are five important stages in NLP:

[1] Lexical Analysis
[2] Syntactic Analysis
[3] Semantic Analysis
[4] Discourse Integration
[5] Pragmatic Analysis

Lexical Analysis:
It is the first step in the NLP process where we break the texts into series of tokens or words for easy analysis. It also involves removing unnecessary blanks/white spaces from the sentences

Syntactic Analysis
This step refers to the study of how the words are arranged in a sentence to identify whether the words are in the correct order to make sense. It also involves checking whether the sentence is grammatically correct or not and converting the words to root form.

Stemming / Lemmatization: It is the process of converting the words to their root form. **Example:** Converting the word 'Studying' to 'Study'. The main difference between stemming and lemmatization is stemming might not necessarily result in an actual meaningful word. But lemmatization would result in an actual meaningful word

Semantic Analysis:
This step involves looking out for the meaning of words from the dictionary and checking whether the words are meaningful.

Discourse Integration:
The meaning of a sentence in any paragraph depends on the context. Here we analyze how the presence of immediate sentences/words impacts the meaning of the next sentences/words in a paragraph.

Pragmatic Analysis:
This is the last phase of the NLP process which involves deriving insights from the textual data and understanding the context.

1.1 RNN:

RNN is a type of supervised deep learning algorithm. Here, the neurons are connected to themselves through time. The idea behind RNN is to remember what information was there in the previous neurons so that these neurons could pass information to themselves in the future for further analysis. It means that the information from a specific time instance (t1) is used as an input for the next time instance(t2). This is the idea behind RNN.

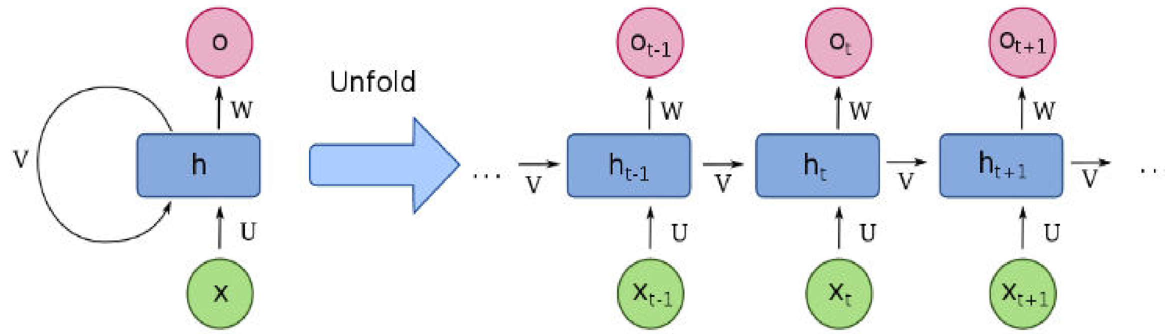

Fig.1 RNN Model

1.2 Long Short Term Memory:

LSTM is an updated version of Recurrent Neural Network to overcome the vanishing gradient problem. Below is the architecture of LSTM with an explanation.

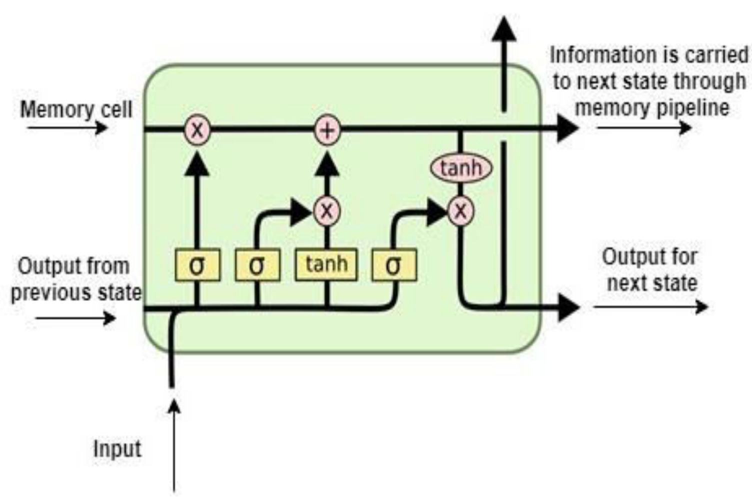

Fig.2 LSTM Architecture

It has a memory cell at the top which helps to carry the information from a particular time instance to the next time instance in an efficient manner. So, it can able to remember a lot of information from previous states when compared to RNN and overcomes the vanishing gradient problem. Information might be added or removed from the memory cell with the help of valves.
LSTM network is fed by input data from the current time instance and output of hidden layer from the previous time instance. These two data passes through various activation functions and valves in the network before reaching the output.

2. Background Survey

Sentiment Analysis is a process of extracting opinions that have different polarities. By polarities, we mean positive, negative or neutral. It is also known as opinion mining and polarity detection. With the help of sentiment analysis, you can find out the nature of opinion that is reflected in documents, websites, social media feed, etc. Sentiment Analysis is a type of classification where the data is classified into different classes. These classes can be binary in nature (positive or negative) or, they can have multiple classes (happy, sad, angry, etc.).
Sentiment analysis is a vital topic in the field of NLP. It has easily become one of the hottest topics in the field because of its relevance and the number of business problems it is solving and has been able to answer.

Fig.3 Need of Sentiment Analysis

After recognizing the significance of sentiment analysis, the system contributes to improving the process of sentiment analysis in web-based living. With the highest level of assurance and the least amount of computational complexity, the proposed method yields superior or, at the very least, identical results [8]. We have looked into how consumer reviews are affected by various preprocessing steps like data cleaning, normalizing, removing hashtags, punctuation, changing the text to lowercase, and tokenizing.

On digital platforms, the text-based consumer review data has grown dramatically. Text reviews have been analyzed using a variety of methods by marketing researchers.
The empirical trade-off between diagnostic and predictive skills was examined by the authors in [10]. They found that AI methods in view of brain networks give the most exact expectations. Topic models, on the other hand, aren't very good at making predictions, and neural network models aren't good for diagnosing problems.

Sentiment analysis looks at the in-depth customer reviews of any product. To propose decisive ideas, the area of viewpoint based feeling examination (ABSA) investigations also, arranges the perspectives expressed on the numerous angles remembered for these sentiments. This essay examines the development of the ABSA model in Hindi reviews [11] in order to broaden an interpretation of the submitted Hindi text.

Natural language processing (NLP) and machine learning were used by the authors of [12] to determine how the reviews in our dataset felt. They also used business intelligence (BI), specifically Microsoft PowerBI, to help businesses that sell these products improve customer satisfaction and streamline their operations. Due to reviews left by customers who have previously purchased the product, the two claims above connect. Potential customers and businesses making the products need to know how to analyze and learn from such active feedback. This paper explains how businesses and customers can benefit from sentiment analysis and business intelligence. It provides an overview of how their products or services perform in the market, various use cases for producers and customers, and customer satisfaction.

3. Sentiment Classification using NLP and LSTM

By altering the parameters and architecture of the network, three distinct models were created. Model 1, Model 2, and Model 3 are the names given to these models. The training dataset for our research consisted of a corpus of 1,194,704 reviews; the remaining 512,016 reviews were utilized to evaluate the efficacy of selected classifiers. The data for testing and training were divided into two distinct groups. For each analysis, the classification algorithms were tested twice: once during the training phase and once during the testing phase. Before any classification algorithms dealt with the review text, it was encoded into a numerical feature vector in the following step. The word embedding vector model was used in this process. The training of LSTM-based classifiers was the third step. In the last step, we applied the prepared classifiers to the test dataset to think about their grouping execution in contrast to the anticipated and real marks that were not seen before the characterization calculations.

The initial step in the process is to obtain review data from the source of benchmark data. The information is taken from postings, comments, reviews, and tweets. Before extracting data from the needed media, search parameters were established for themes and customer evaluations. Twitter tweets, movie reviews, news feeds, product reviews, and Facebook postings are some of the sources of frequently utilized datasets. The extracted data are fed into the system at this step, which is used for data mining and analysis. This stage serves as the central component of the sentiment analysis process.

a) Load in and visualize the data
b) Data Processing — convert to lower case
c) Data Processing — Remove punctuation
d) Data Processing — Create list of reviews
e) Tokenize — Create Vocab to Int mapping dictionary
f) Tokenize — Encode the words
g) Tokenize — Encode the labels
h) Analyze Reviews Length
i) Removing Outliers — Getting rid of extremely long or short reviews
j) Padding / Truncating the remaining data
k) Training, Validation, Test Dataset Split
l) Dataloaders and Batching
m) Define the LSTM Network Architecture
n) Define the Model Class
o) Training the Network
p) Testing (on Test data and User- generated data)

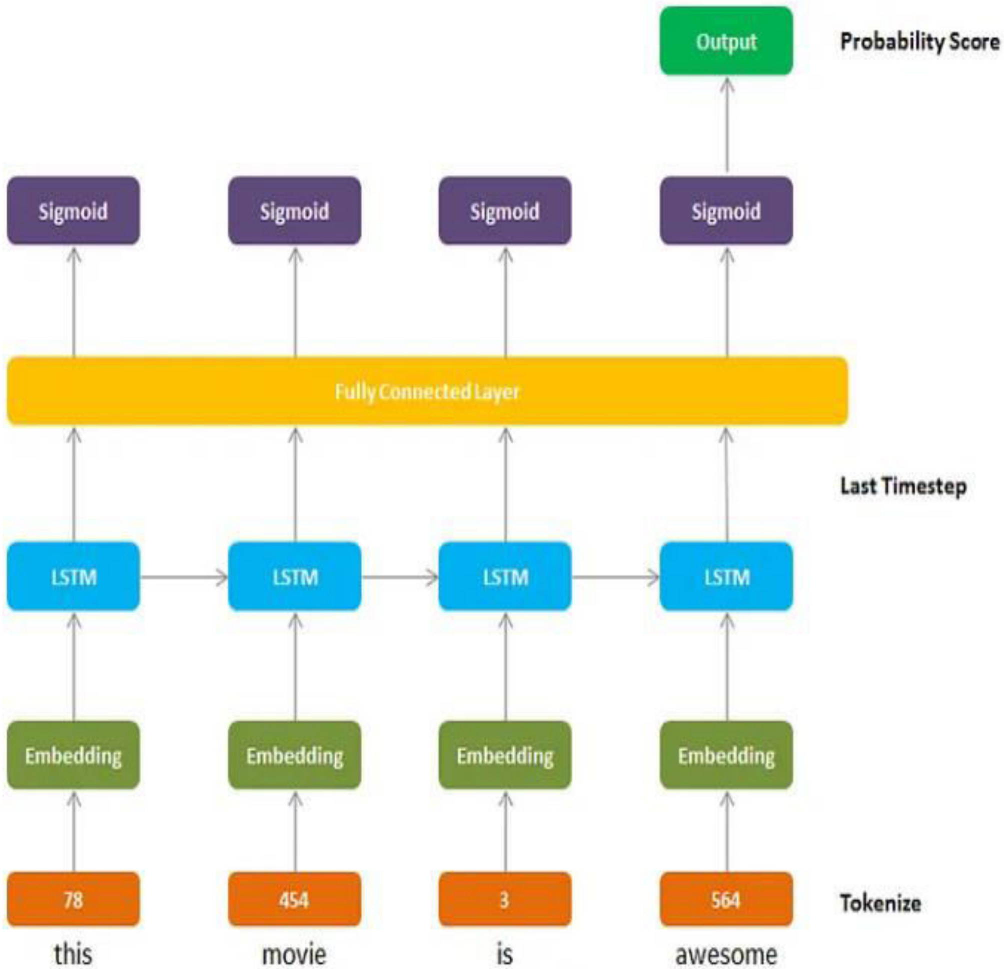

Fig.4 Proposed LSTM with RNN Model

1. Embedding Layer: that converts our word tokens (integers) into embedding of specific size
2. LSTM Layer: defined by hidden state dims and number of layers
3. Fully Connected Layer: that maps output of LSTM layer to a desired output size
4. Sigmoid Activation Layer: that turns all output values in a value between 0 and 1
5. Output: Sigmoid output from the last timestep is considered as the final output of this network

Presently, a massive amount of personal data appear in consumer reviews; classification is becoming popular in sentiment analysis and evaluation . During this phase, the recurrent neural network-based LSTM and Deep LSTM classifiers were used for the task of classification. The LSTM network consists of LSTM units next to the input and output layer. An LSTM framework allows long and short recording values and uses no device in several parts of its action . A three-layer LSTM stack has been developed to build a deep RNN . Moreover, peephole connections in a similar cell between its internal partitions and the entrances can also be used for the cutting-edge LSTM Design to evaluate precise performance. Deep LSTM RNNs (DNN) have been commonly used for the more resounding speech recognition architecture. Using Deep LSTM RNN in the regular LSTM, parameters can be best optimized by spreading them across many layers . This study uses a Deep LSTM model with one input layer, two LSTM layers in a row, two dense layers, and either one output layer or two dense layers. Three models were developed by changing the LSTM network architecture and parameters. These models were considered Model 1, Model 2, and Model 3, respectively. The following subsections provide a summary of each model. One LSTM layer, an embedding layer with vocabulary size, embedding vector length, and maximum review length, and a dense layer with a fully coupled sigmoid activation function make up the architecture of Model 1. A binary cross-entropy loss is used in the model's construction and training according to the nature of our challenge. A better optimization tool is Adam (faster and more reliably reaching a global minimum when minimizing the cost function in training neural nets).

4. Results & Discussion

The IMDB movie reviews dataset is a collection of **review -> rating** pairs that ofen look like the following (this is an imitation, not pulled from IMDB):
The entire dataset consists of around 50,000 of these pairs, where the input reviews are usually a few sentences and the output ratings are between 1 and 5 stars. This dataset consists of with over 25,000 reviews for training and 25,000 for the testing set. Reviews have been preprocessed, and each review is encoded as a sequence of word indexes (integers). For convenience, words are indexed by overall frequency in the dataset, so that for instance the integer "3" encodes the 3rd most frequent word in the data. This allows for quick filtering operations such as: "only consider the top 10,000 most common words, but eliminate the top 20 most common words". Rating have been olso preprocessed, and each rating is encoded as 1 or 0 (positive or negative). As a convention, "0" does not stand for a specific word, but instead is used to encode any unknown word.

```
Train on 25000 samples, validate on 25000 samples
Epoch 1/15
25000/25000 [==============================] - 79s 3ms/step - loss: 0.45
65 - accuracy: 0.7874 - val_loss: 0.3688 - val_accuracy: 0.8396
Epoch 2/15
25000/25000 [==============================] - 76s 3ms/step - loss: 0.30
37 - accuracy: 0.8770 - val_loss: 0.3740 - val_accuracy: 0.8404
Epoch 3/15
25000/25000 [==============================] - 76s 3ms/step - loss: 0.21
74 - accuracy: 0.9168 - val_loss: 0.4306 - val_accuracy: 0.8291
Epoch 4/15
25000/25000 [==============================] - 74s 3ms/step - loss: 0.15
41 - accuracy: 0.9432 - val_loss: 0.4645 - val_accuracy: 0.8250
Epoch 5/15
25000/25000 [==============================] - 76s 3ms/step - loss: 0.11
44 - accuracy: 0.9587 - val_loss: 0.5888 - val_accuracy: 0.8228
Epoch 6/15
25000/25000 [==============================] - 76s 3ms/step - loss: 0.08
60 - accuracy: 0.9688 - val_loss: 0.6636 - val_accuracy: 0.8215
```

Fig.5 Sample output

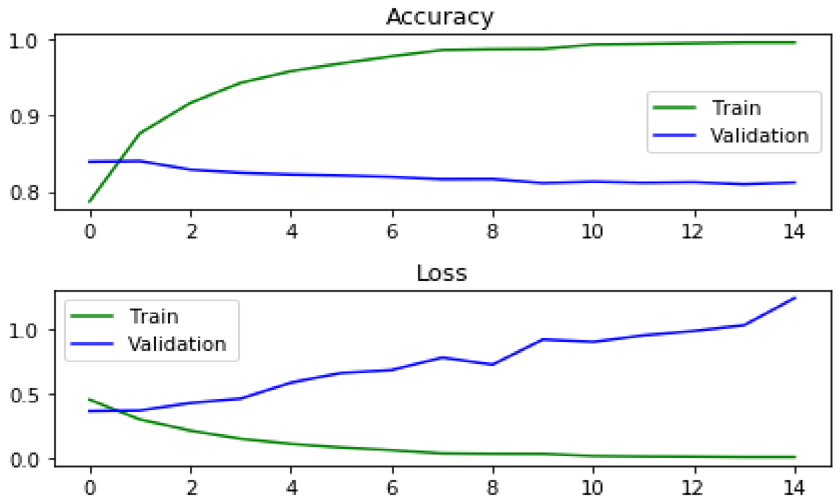

Fig.6 Accuracy and Loss graph

Accuracy: The most well-known performance metric is accuracy. It is convenient and straightforward to compute and identify. Accuracy assesses a predictor's capacity to correctly identify all samples, regardless of their effectiveness or unfavourability.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

where TN = True negative, FN = False negative, FP = False positive, TP = True positive, P = Positive, and N = Negative.

Sensitivity/Recall: The true positive rate or recall can be used to define sensitivity. The true positive percentage may be quickly identified by following a few simple methods.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{P}}$$

Precision: The precision demonstrates the accuracy of the classifier. Low accuracy and high precision both result in lower accuracy and fewer false positives. The improvement in precision is the cause of reduced sensitivity and is inversely proportional to the sensitivity.

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

F1-Measure: F1-Measure is the blend of accuracy and sensitivity. This is the weighted harmonic way to be sensitive and accuracy/precision. The F1 measurement is proven as effective as precision.

$$F1 - Score = \frac{TP + TN}{TP + TN + FP + FN}$$

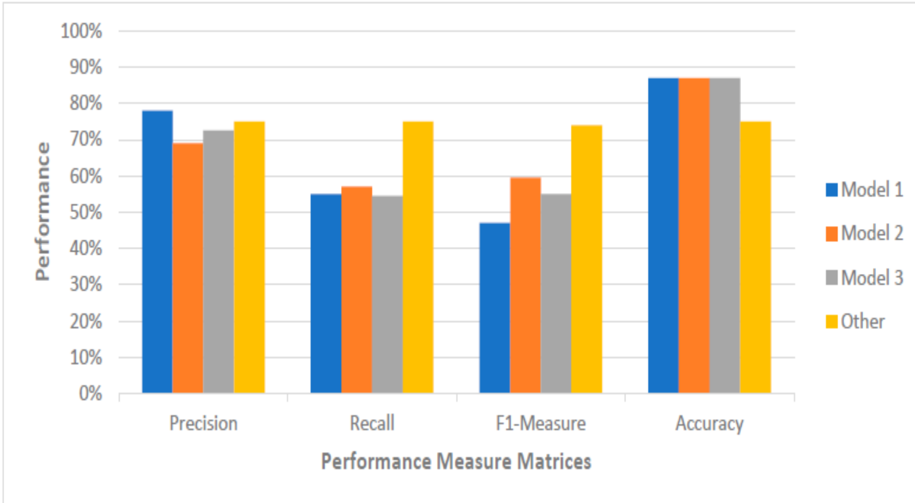

Fig.7. Comparison of performance measure matrices on the product Reviews dataset

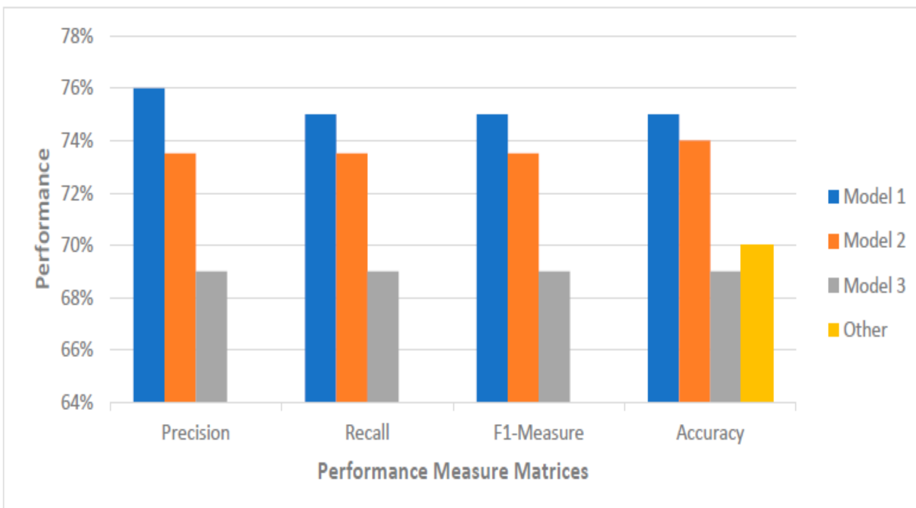

Fig.8 Comparison of performance measure matrices on the Movies dataset

5. Conclusion

Sentiment classification and analysis, this study used recurrent neural network-based models and long-short-term memory inspired by deep learning. These architecture- and parameter-tuning-based deep learning approaches were the foundation for the creation of three models. The accuracy, precision, recall, and F1-score of these models were used to evaluate their efficacy on the aforementioned datasets. The final results were compared to those from previous methods. The proposed results were better compared to or equivalent to past methodologies. On both datasets, the deep learning method outperformed Logistic Regression in terms of accuracy. The paper also shed light on the fact that traditional machine learning algorithms were more accurate with raw data than with processed data for both datasets, whereas deep learning algorithms were more accurate with processed data. It should also be noted that when subclasses share similar characteristics, the deep learning method for hierarchical classes produces better accuracy. For future work, we wish to stretch out this work to remember emoticon for our texts. The usage of emoji in user-generated content has increased. All emoji are removed from the texts during preprocessing. However, the predictions' accuracy could have been enhanced if emoji could have been processed and converted.

References:

[1] Study of Machine Learning Techniques in Automatic Product Categorisation", In Proceeding of the 14th International Symposium on Neural Networks (ISNN 2017), 21-23 June 2017, Hokkaido, Japan (Fengyu Cong, Andrew Leung, Qinglai Wei, eds.), vol. 10261, pp. 10-17, 2017.

[2] A Comparative Study of Machine Learning Methods for Verbal Autopsy Text Classification. / Danso, Samuel; Atwell, Eric; Johnson, Owen. In: International Journal of Computer Science, 18.02.2014.

[3] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," in International Educational Data Mining Society, 2015, pp.392-395.

[4] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), 1746-1751.

[5] Zhang, Ye and Byron C. Wallace. "A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification." CoRR abs/1510.03820 (2015):

[6] Aaron Defazio, Francis Bach, Simon Lacoste-Julien. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives. Advances In Neural Information Processing Systems, Nov 2014, Montreal, Canada.

[7] Saqib, M.S., Kundi, F.M., Ahmad, S. Unsupervised Learning Method for Sorting Positive and Negative Reviews Using LSI (Latent Semantic Indexing) with Automatic Generated Queries. IJCSNS International Journal of Computer Science and Network Security, Vol.18 No. 1, 2018 4. Cambria E, Das D, Bandyopadhyay S, Feraco A. A practical guide to sentiment analysis. Switzerland: Springer, Cham; 2017.

[8] P.Naresh,et.al., "Implementation of Map Reduce Based Clustering for Large Database in Cloud", Journal For Innovative Development in Pharmaceutical and Technical Science,vol.1,pp 1-4,2018.

[9] B.M.G. Prasad, P. Naresh, V. Veeresh, "Frequent Temporal Patterns Mining With Relative Intervals", International Refereed Journal of Engineering and Science ,Volume 4, Issue 6 (June 2015), PP.153-156.

[10] T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.

[11] V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

[12] P, N., & R Suguna. (2022). Enhancing the Performance of Association Rule Generation over Dynamic Data using Incremental Tree Structures. International Journal of Next-Generation Computing, 13(3). https://doi.org/10.47164/ijngc.v13i3.806.

[13] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206.

[14] Naresh, P., & Suguna, R. (2021). IPOC: An efficient approach for dynamic association rule generation using incremental data with updating supports. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 1084. https://doi.org/10.11591/ijeecs.v24.i2.pp1084-1090.

[15] Elmurngi, E., Gherbi, A., An empirical study on detecting fake reviews using machine learning techniques, In: Proceedings of the 2017 Seventh International Conference on Innovative Computing Technology (INTECH). (2017).

[16] W. P. Ramadhan, S. T. M. T. Astri Novianty, and S. T. M. T. Casi Setianingsih, “Sentiment analysis using multinomial logistic regression,” 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), 2017.

[17] Martineau, J. and Finin, T. Delta TFIDF - an Improved Feature Space for Sentiment Analysis. Third AAAI International Conference on Weblogs and Social Media, San Jose CA (2009).

[18] Bhavitha, B.K., Rodrigues, A.P., Chiplunkar, N.N. Comparative Study of Machine Learning Techniques in Sentimental Analysis. In: Proceedings of International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE (2017) 216-221.

[19] Mishra, Prakhar. Natural Language Preprocessing – Hacker Noon. Hacker Noon. November 27, 2017. Available at: https://hackernoon.com/natural-language-preprocessing- 630c28832fd1. Rain, Callen. Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning. Swarthmore College, Department of Computer Science. 2016.

[20] Dang, S.;Wen, M.; Mumtaz, S.; Li, J.; Li, C. Enabling Multi-Carrier Relay Selection by Sensing Fusion and Cascaded ANN for Intelligent Vehicular Communications. IEEE Sens. J. **2020**, 21, 15614–15625. [CrossRef]

[21] Cambria, E.; Poria, S.; Bajpai, R.; Schuller, B. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 2666–2677.