

Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms

K. Amulya¹

M.Tech, Data science
Kakatiya Institute of Technology & Science
Warangal, India
amulyakompallie@kit.ac.in

Dr. P. Kamakshi³

Professor & Head, Dept. of IT
Kakatiya Institute of Technology & Science
Warangal, India
pk.it@kit.ac.in

S. B. Swathi²

Assistant professor, Dept. of IT
Kakatiya Institute of Technology & Science
Warangal, India
sbs.it@kit.ac.in

Dr. Y. Bhavani^{4*}

Associate professor, Dept. of IT
Kakatiya Institute of Technology & Science
Warangal, India
yerrambh@gmail.com
ORCID: 0000-0002-8578-7126

Abstract—Sentiment analysis is the study, to classify the text based on customer reviews which can provide valuable information to improve business. Previously the analysis was carried out based on the information provided by the customers using natural language processing and machine learning. In this paper, sentiment analysis on IMDB movie reviews dataset is implemented using Machine Learning (ML) and Deep Learning (DL) approaches to measure the accuracy of the model. ML algorithms are the traditional algorithms that work in a single layer while deep learning algorithms work on multilayers and gives better output. This paper helps the researchers to identify the best algorithm for sentiment analysis. The comparison of the machine learning and deep learning approaches shows that DL algorithms provide accurate and efficient results.

Keywords— machine learning, NLP, deep learning, comparison

I. INTRODUCTION

Sentiment analysis is a perspective, thought, or judgment of a specific feeling. Most of the analysis is done based on online reviews. It is troublesome to analyze using forum discussions because most of the information provided is irrelevant. To overcome this, aspect-based sentiment analysis can be used as it is a text analysis technique used to categorize data and identify the sentiment attributed to specific review. Aspect-based sentiment analysis can be used to analyze customer feedback through sentiments with different aspects of a product or service. By using this analysis, the sentiment of the reviews can be identified easily.

Natural language explanations have become challenging due to the complexity of human languages. The main reason is that textual data doesn't have a proper structure. There is a need to parse the data as it helps the machine to understand and utilize the data. Sentiment analysis helps companies to analyze large amounts of data within less time. Sentiment analysis helps to analyze and prepare strategies based on

customer reviews. Reviews are considered as short text by which the opinion of a person can be identified. As everything is digitalized, people check blogs for reviews before they watch a movie or purchase a product. These reviews drive people to theatres and customers towards products. This analysis of sentiment helps to implement a strategic plan and attract more customers.

As per the previous work, sentiment analysis implementation was through NLP (Natural Language Processing) and machine learning algorithm. Sentiment analysis of movie reviews includes pre-processing and implementation of modelling techniques. These modelling techniques are used to identify the reviewer sentiment that is categorized as either positive or negative. Model accuracy can also be known by machine learning algorithm which provides the accuracy percentage of the sentiment identified on a particular review dataset.

In this paper, sentiment analysis is done using ML and Deep learning approach on the considered IMDB review dataset. The text reviews categorised as negative and positive reviews.

Deep learning is used to reduce human intervention and the issues of big data are dealt effectively to provide a better model for handling the data efficiently [17].

Deep learning can also be used in NLP to analyze the customer's point of view [18]. In this paper, deep learning and machine learning approaches are used to analyze the sentiment of the movie reviews and to identify whether the reviews are negative or positive. These reviews can be utilized as a tool for recommending movies. Filmmakers can also utilize this information to make marketing decisions and attract customers.

II. LITERATURE SURVEY

Sentiment Analysis is based on feature extraction and sentiment classification. It is achieved by applying the statistical approach using machine learning [2]. IMDB reviews were analyzed using sentiment analysis and show reviews by n-gram approaches.

Different classifiers are used to train the model and the unigram approach is performed well in comparison to others. The n-gram approach of experimentation was finished and tried to produce the simplest results. Similar work was done by Tripathy et al. [1], where TF, TF-IDF was used for the conversion of the text file to a numerical vector. Experimentation was done with n-gram approaches and its combination are tried to get the best results.

Not only the word features, but special symbols present with words can also be considered as features [3]. The emotions with the word features can be utilized. Classifier ensemble is used to classify the results obtained by different ML algorithms and produce a good result.

Research is done to extract features that include the parts of the speech by using a tagger [4]. The unigram model is used to extract adjectives that describe the positive or negative sentence. Emotion recognition had always been attracted attention in multiple fields that include NL processing, psychology [12].

Aspect-based sentiment analysis requires two conceptual tasks, syntactical information to explore the grammatical methods and address this problem by utilizing the effective encoding of the syntax.

Analyzing the sentiment of the customers is essential as a customer plays an important role to improve business. The work of sentiment recognition mainly focused on semantic processing to understand the customer better and analyze through the reviews. This approach, shows the graphs, similarity measures, algorithms using graph theory to make the process simplified and easy [13] by considering the texts that could promote to understand the language better so that the sentiment analysis tasks can be done easily.

Preethi et.al. [5] introduced a new application (RNN) with deep learning system for sentiment analysis using pic reviews. Qian et.al.[6] projected a model to train with sentence-level annotation, they conjointly did an effort to come up with linguistic coherent representations of the model using regularizers.

Nguyen et.al. [7] Designed a model that predicts stock worth movement by mistreatment of the sentiment collected from social media. It shows the analysis of stock prediction task and the authors adopted the neural network methodology to demonstrate that ML methodology is beneficial for finding out the variations and commonalities of various quantizing strategies of quantum correlation [8].

The study of Recursive Neural Tensor Network was accustomed to determine sentences based on the user sentiment, and they used a dataset that contains 11,855 movie reviews. The accuracy achieved by RNTN is 80.7%.

Socher et.al. [9] projected that a text cannot be analyzed in isolation which states that sentences are closely associated with the words.

The invention [10] expressed that (RNN) model will subsume the short dependence during a sequence of information however, includes a drawback of gradient explosion as they will have a major impact on overall polarity to unravel this semipermanent dependence drawback, the LSTM model is proposed by Vu et al. [11].

Remote sensing applications are not fully engaged in the use of CNN. In order to address this issue, they introduced a novel CNN so that there is an increase in the performance of detectors [14]. The research study performs an analysis by considering a dataset obtained from online social media, where the detection is based on a ML algorithm [17].

III. METHODOLOGY

In this paper, two approaches are compared on the data set of 50000 IMDB movie reviews, the reviews are in the text format the sample format of the used dataset is shown in Fig. 1.

The first implementation is performed on this dataset by applying ML algorithms for the prediction of accuracy on the model. The second implementation is performed using deep learning techniques which resulted in better accuracy for sentiment analysis.

The input data set is applied as a set of movie reviews and the expected output is the accuracy of the model. In this paper, both the ML and DL approaches are implemented and a comparison of these approaches is shown. A short note on the techniques used is given below.

Natural language processing: NLP is taken into account as a field of applied science and is additionally concerned with the interactions between machines and human languages. It helps to spot the sentiment of the reviewer and consists of many pre-processing techniques to convert the information into simpler text. So, that it will be easily understood.

Machine Learning Algorithms: ML is the study of systems related to field of computer science and it can learn from data. It mostly focuses on predictions supported on known properties and learns from the training data. It requires training data set to be considered and therefore the classifier has to be trained on some labelled training data, before it is applied to the particular classification Task.

Some of the machine learning algorithms is given below.

- Logistic Regression
- Naive Bayes
- SVM
- XGBOOST

Deep Learning Approaches

- **CNN:** CNN often used for identifying objects inside images and for text classification by using word embeddings. It's been found effective for text in search query retrieval, sentence modeling, and NLP tasks.
- **RNN:** RNN takes a sequence of information as input; recursive process is performed in the evolution direction of the sequence. It is the study on nonlinear characteristics of the sequence and has many advantages.

RNN is applied on NLP, like speech recognition, language modelling, and other fields and it is a deep learning approach that may be used for sentiment analysis. It produces the output supported by previous computation and taking sequential information.

- **LSTM (Long Short-Term Memory):** It is used to overcome the RNN problem for memorizing data for a longer time. LSTM works as a part of long-term dependence. This can be used for text classification and it produces long-term memorizing of the data compared to RNN. Thus, LSTM is also used for the implementation and analysis of the sentiment based on reviews.

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive
5	Probably my all-time favorite movie, a story o...	positive
6	I sure would like to see a resurrection of a u...	positive
7	This show was an amazing, fresh & innovative i...	negative
8	Encouraged by the positive comments about this...	negative
9	If you like original gut wrenching laughter yo...	positive
10	Phil the Alien is one of those quirky films wh...	negative

Fig. 1. Sample format of the text reviews of IMDB movie reviews dataset

IV. EXPERIMENTATION

In this paper, sentiment analysis is done using Machine learning Algorithms and an advanced implementation based on the Deep Learning RNN method to identify the better model for sentiment analysis on movie reviews. The experiment was done by using python code on the google collab platform.

• Data set

To identify the better model that can be used for sentiment analysis, a public IMDB dataset that contains 50,000 reviews is considered, of which 35,000 are used for training and the remaining 15,000 are used for testing which is shown in Table I and Fig. 2. is the graphical representation of two review categories.

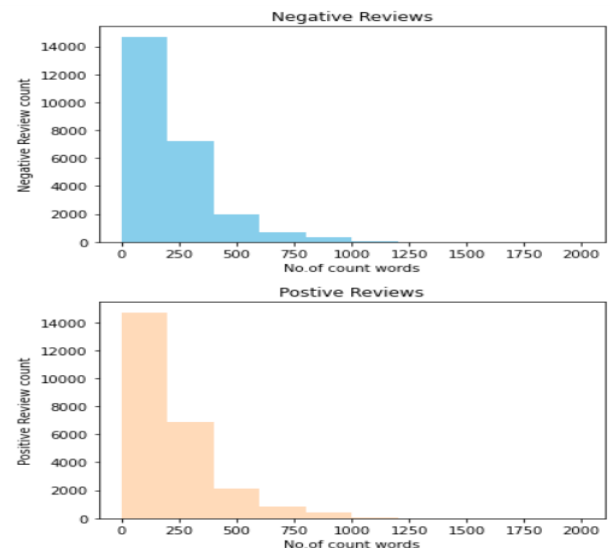


Fig. 2. Shows the graphical representation of two review categories

TABLE I. Shows the IMDB movie review dataset categories

Table 1: IMDB reviews			
Train data		Test data	
Negative	17500	Negative	7500
Positive	17500	Positive	7500

• Process of Implementation

In this paper, different ML and DL approaches are considered and the implementation of both Approaches is listed below briefly.

• NLP approach using Machine learning algorithms.

In Fig 3. The process is shown, firstly a Movie reviews data set is applied and then pre-processing of data that includes text normalization, removing noisy text, special characters, text stemming, removing stop words, and word embedding. Then modelling techniques in NLP are applied that include count vectorizer and TF-IDF. The data here was split into train and test set after that ML algorithms were applied to check the accuracy of the model.

The first algorithm considered by us is logistic regression which is a classification technique that serves to solve the binary classification problem. Movie reviews were considered as positive and negative thus logistic regression can be used and was applied on both count vectorizer and TF-IDF to find out the accuracy score. The accuracy score obtained is around 87% for TF-IDF and 86% for the count vectorizer.

The second algorithm considered is the support vector machines and was applied on TF-IDF and count vectorizer to find out the accuracy score and the accuracy score obtained is 89% for TF-IDF and 86% for count vectorizer.

The third algorithm considered is Multinomial Naive baye's algorithm that can be used for text analysis and this technique is accustomed to find the possibilities of classes assigned to texts by considering the joint probabilities of the words and classes. This algorithm was also applied on both TF-IDF and count vectorizer to understand the accuracy score. The obtained accuracy is 86% in both cases.

The fourth algorithm considered is XGBoost algorithm which boosts the speed and performance of the model. Here the accuracy obtained by applying XGBoost on TF-IDF and count vectorizer is 80%.

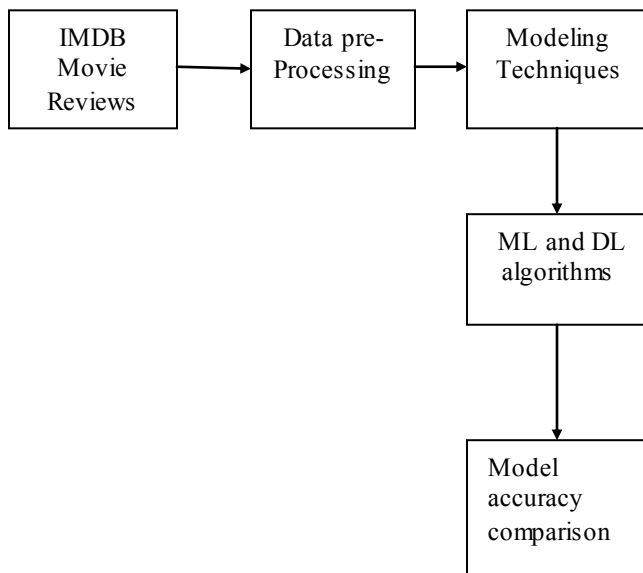


Fig. 3. Shows the block diagram of algorithms applied on IMDB movie reviews.

• Deep Learning approach

In the Deep learning approach, CNN, RNN, LSTM are used and implemented on the considered IMBD movie review datasets. In Deep learning, feature extraction is done automatically and that makes solving complex problems easier. The First Deep learning approach is CNN, by applying this model the accuracy achieved is 87%, the second approach is the RNN model and the accuracy achieved is 88%. Third and final approach is the LSTM model and the accuracy achieved by this model is 72%.

- **Pre- Processing Stage:** As the dataset is required to be clean to apply and create and train model. In this stage, it includes removal of attribute missing values, Standard Scalar, Min-Max Scalar has been applied to the dataset to clean the data and obtain required clean data.

- **Predictive Models:** The predictive models used in this paper are Logistic regression, SVM, XGboost, Multinomial Naive Baye's and Deep Learning CNN, RNN, and LSTM model.

- **Performance Metrics and its Evaluation:** Evaluation metrics are used for classifiers to know the performance.

Metrics for a model on a binary classification problem

are listed with the equations:

- Recall metrics: $tp / (tp + fn)$ (1)

- F1 Score metrics: $2 tp / (2 tp + fp + fn)$ (2)

- Accuracy metrics = $(TP + TN) / (TP + TN + FP + FN) \times 100$ (3)

- Precision metrics: $tp / (tp + fp)$ (4)

The overall flow process of the implementation on IMDB movie reviews by using ML and DL approach is shown in Fig. 4.

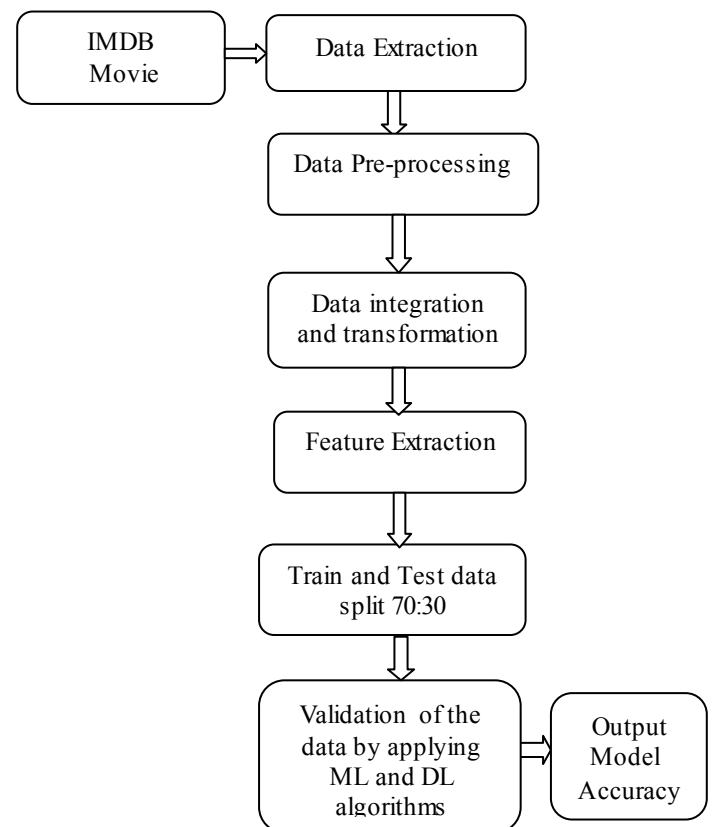


Fig. 4. Shows the flow diagram of the implementation procedure of the ML and DL algorithms

V. RESULTS

In this approach ML and DL algorithms are applied to the IMDB movie reviews dataset to analyze positive and negative sentiment by detecting the emotion of the reviewer through text that includes some emotional key words which determine the emotion of the reviewer. Some positive emotions include “good”, “like”, “best”, “great” and negative emotions include “worst”, “sadly”, “disappointed”, “uncomfortable”, “bad”. To analyze these positive and negative emotions machine learning algorithms (Logistic Regression, SVM, Multinomial Navies baye’s and XGBoost) and Deep learning (CNN, RNN, and LSTM) models are applied. The different approaches are compared with Recall metrics (eq. 1), F1 Score metrics (eq. 2), Accuracy metrics (eq. 3) and Precision metrics (eq. 4) as shown in Table II. The Accuracy comparison of ML and DL algorithms is represented as Bar graph in Fig. 5. The accuracy and loss of CNN and RNN are shown in graphical representation in Fig. 6, Fig. 7

TABLE II. The comparison scores of two approaches implemented on IMDB dataset.

Performance metrics	Predictive model	Precision	Recall	F1-Score	Accuracy
Performance metrics of tf-idf features	Logistic Regression	0.89	0.85	0.87	0.86
	SVM	0.89	0.86	0.86	0.87
	Multinomial Naïve Baye’s	0.87	0.85	0.86	0.86
	XGBoost	0.84	0.73	0.86	0.86
Performance metrics of Count-Vectoriser	Logistic Regression	0.87	0.86	0.87	0.86
	SVM	0.86	0.86	0.86	0.86
	Multinomial Naïve Baye’s	0.87	0.85	0.86	0.86
	XGBoost	0.84	0.74	0.79	0.81
Performance metrics of Deep learning Algorithms	CNN	0.94	0.85	0.87	0.87
	RNN	0.95	0.86	0.88	0.88
	LSTM	0.72	0.70	0.71	0.71

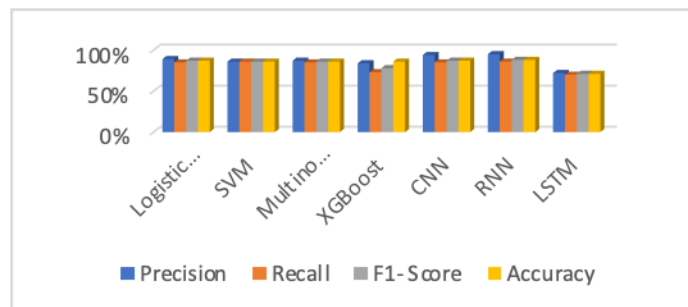


Fig. 5. The Bar graph of the ML and DL approaches model accuracy comparison.

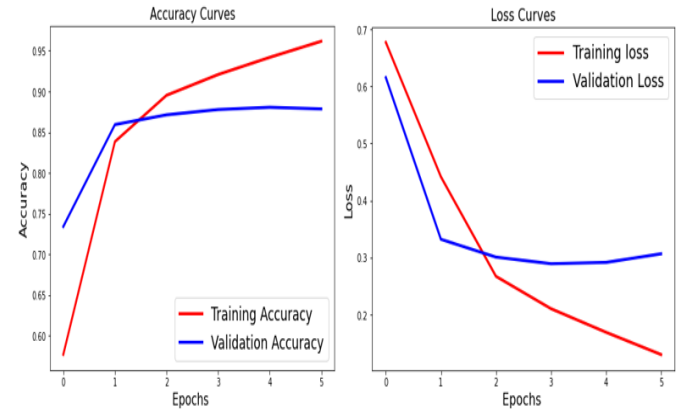


Fig. 6. The graphical representation of CNN validation accuracy and loss of trained model.

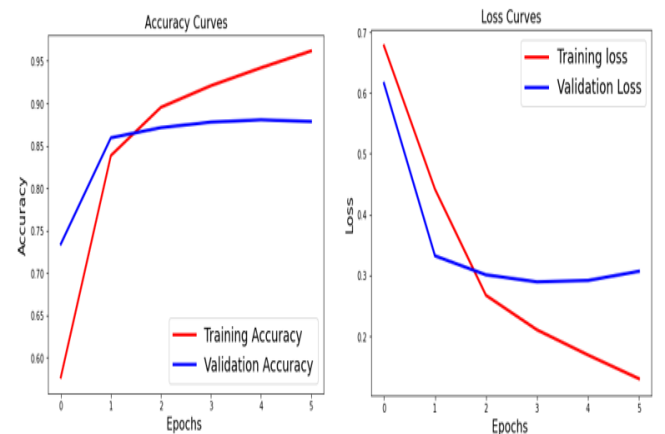


Fig. 7. The graphical representation of RNN validation accuracy and loss of trained model

VII. CONCLUSION AND FUTURE WORK

In this paper, the NLP approach using Machine Learning algorithms and the Deep Learning methods is used to classify reviews of the data set taken into positive and negative categories. Comparison of ML and DL approaches is done by considering IMDB movie reviews. From the observations it is found that DL approaches provided accurate results than ML algorithms. Among the DL algorithms (CNN, RNN, LSTM), RNN gives more accuracy of 88%. When ML algorithms is used the feature extraction should be done manually whereas in DL approach there is no need of human intervention and feature extraction is done by machine automatically. It is concluded that deep

learning algorithms are more accurate and efficient than machine learning algorithms.

In future work, better models are hoped to be identified using deep learning to achieve better accuracy and to improve the effect of movie reviews by using sentiment analysis. Data pre-processing plays an important role in such large data sets. The aim is to identify better data pre-processing methods to achieve improved accuracy for movie review sentiment analysis.

REFERENCES

- [1] A. Tripathy, A. Agrawal, and S.K. Rath. "Classification of sentiment reviews using n-gram machine learning approach." *Expert Systems with Applications*, Vol. 57, pp. 117-126. 2016.
- [2] B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up? sentiment classification using machine learning techniques." In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* Volume 10, Association for Computational Linguistics, pp. 79-86. 2002.
- [3] M. S. Mubarak, Adiwijaya, and M. D. Aldhi. "Aspect-based sentiment analysis to review products using Naïve Bayes." In *AIP Conference Proceedings*, vol. 1867, AIP Publishing, no. 1, pp 1-8. 2017.
- [4] G. Gautam, and D. Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." In *Contemporary computing (IC3)*, 2014 seventh international conference on, pp. 437-442. IEEE, 2014.
- [5] G. Preethi; Krishna, P. V.; Obaidat, M. S.; Saritha, V.; Yenduri, S. (2017). "Application of deep learning to sentiment analysis for recommender system on cloud". *International Conference on Computer, Information and Telecommunication Systems*, pp. 93-97.
- [6] Q. Qian.; M. Huang.; J. Lei; X. Zhu (2016): "Linguistically regularized LSTMs for sentiment classification. arXiv preprint arXiv:1611.03949. Sak, H.; Senior, A.; Beaufays, F. (2014): "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition". arXiv preprint arXiv:1402.1128.
- [7] T. H. Nguyen; K. Shirai; J. Velcin (2015): "Sentiment analysis on social media for stock movement prediction". *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603-9611.
- [8] Q. Li, X.; Zhu; Q. Meng; You, C.; M. Zhu (2019): "Researching the link between the geometric and r nyi discord for special canonical initial states based on neural network method". *Computers, Materials Continua* 1, vol. 60, no. 3, pp. 1087-1095.
- [9] R. Socher; A. Perelygin; J. Wu; J. Chuang; C. D. Manning (2013): Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631-1642.
- [10] T. Lin; B. G. Horne; P. Tino; C. L. Giles (1996): Learning long-term dependencies in nax recurrent neural networks. *IEEE Transactions on Neural Networks*, vol. 7, no. 6, pp. 1329-1338.
- [11] N. T. Vu; H. Adel; P. Gupta.; H. Sch tze, (2016): "Combining recurrent and convolutional neural networks for relation classification". arXiv preprint arXiv:1605.07333.
- [12] J. B. Delbrouck, N. Tits, M. Brousmiche, and S. Dupont, "A transformer based joint-encoding for emotion recognition and sentiment analysis", in *Proc. 2nd Grand-Challenge Workshop Multimodal Lang. (ChallengeHML)*, 2020, pp. 1-7.
- [13] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 6975-6988.
- [14] P. Karuppusamy "Building Detection using Two-Layered Novel Convolutional Neural Networks" *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 29-37.
- [15] Smys, S., and Haoxiang Wang "Security Enhancement in Smart Vehicle Using Blockchain- based Architectural Framework" *Journal of Artificial Intelligence* 3, no. 02 (2021): 90-100.
- [16] S.R. Mugunthan, "Soft computing based autonomous low rate DDOS attack detection and security for cloud computing" *J. Soft Comput. Paradig. (JSCP)* 1, no. 02 (2019): 80-90.
- [17] Smys, S., and Jennifer S. Raj "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study" *Journal of trends in Computer Science and Smart technology (TCSSST)* 3, no. 01 (2021): 24-39.
- [18] Kumar, T. Senthil "Construction of Hybrid Deep Learning Model for Predicting Children Behavior based on their Emotional Reaction" *Journal of Information Technology* 3, no. 01 (2021): 29-43.