Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

Performance Evaluation and Comparison using Deep Learning Techniques in Sentiment Analysis

A. Pasumpon Pandian,
Professor, Dean (R&D),
CARE College of Engineering,
Trichy, India.

Abstract: One of the most common applications of deep learning algorithms is sentiment analysis. This study delivers a better performing and efficient automated feature extraction technique when compared to previous approaches. Traditional methodologies like surface approach will use the complicated manual feature extraction process, which forms the fundamental aspect of feature driven advancements. These methodologies serve as a strong baseline to determine the predictability of the features, and it will also serve as the perfect platform for integrating the deep learning techniques. The proposed research work has introduced a deep learning technique, which can be incorporated with feature-extraction. Moreover, this research work includes three crucial parts. The first step is the development of sentiment classifiers with deep learning, which can be used as the baseline for comparing the performance. This is followed by the use of ensemble techniques and information merger to obtain the final set of sources. As the third step, a combination of ensembles is introduced to categorize various models along with the proposed model. Finally experimental analysis is carried out and the performance is recorded to determine the best model with respect to the deep learning baseline.

Keywords: Machine Learning, Sentiment analysis, deep learning, ensemble, performance

1. Introduction

The development of user generated content in social networks and websites like Trip advisor, Amazon, Facebook, Twitter and Instagram have resulted in making social networks

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

the ultimate platform to express opinions about events, products, services, etc. [1]. This ability, when integrated with the quick spreading nature of online content has increased the value of the opinions posted. To analyze this large amount of data, a number of natural language processing tasks are currently in use. Sentiment analysis is one of the key analyses that is currently used with the aim of classifying sentiments and opinions generated by human beings and in text. Machine learning techniques are predominantly used in sentiment analysis [2]. In the beginning, the Bag of Words (BoW) model was adapted which maps a particular feature vector with a document and then segments using machine learning techniques. This methodology was greatly appreciated for its efficiency and simple nature. However, the drawback with this method is that the original natural language is lost, synthetic structures are broken and word order is destroyed. Due to this, a number of unique methods like higher order n-grams were introduced [3-6].

As observed by the authors in [7] part of speech tagging is a type of methodology that is used in the syntactic analysis process. This is also called as surface forms since they are used in tactical [8] and lexical information [9] which depends on the text pattern [10-14] instead of its semantic aspect. Information on sentiment gathered is also used for analysis. This includes addition of word polarity with features described previously. This information gathered [15] is generally in the form of sentiment lexicons. This methodology has a number of advantages [16]. Consideration of negotiations and intensifiers along with sentiment valence shifting is taken into account in linguistic content. Moreover depending on the characteristics it is possible to differentiate the sentiment orientation of lexical entities. However, the number of drawbacks faced by lexicon based approaches is many such as requirement of a reliable [17] and consistent lexicon along with a range of options for opinion words with respect to languages contexts and domains [18-20]. This leads to difficulty in maintaining the independent lexicons. In a typical machine learning-driven methodology, selecting a classification algorithm determining the relevant features and extracting features from text are some of the fundamental questions that are addressed. In the beginning, manual feature engineering was used which proved to be time consuming. An alternative to this approach is deep learning techniques.

This methodology has shown positive performance in a number of Natural Language Processing (NLP) algorithms [21]. The most important concept involved in deep learning is

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

the use of deep neutral network to train the nodes for extracting complex features from the available information with limited contribution. These algorithms adapt easily and do not require manual input giving them the ability to learn new complex features automatically. However, the drawback with using deep learning approaches is that they require a vast collection of data for high efficiency output. In comparison to the other traditional machine learning approaches this techniques will give crucial importance to availability of resources and automatic feature extraction [22]. However there is no clarity in the capability of deep learning-based models and traditional methodologies when viewed with respect to specialization and generalization capacity. Hence, in this proposed work, we have used a combination of the two methodologies with the help of many ensemble models with emphasis on the different features required for sentiment analysis. To be more specific we have taken into consideration an ensemble of features and ensemble of classifiers [23]. Here, various features are combined to a servant in sentiment classified as and every feature combination is done at the feature level. In order to test the proposed system in real time we have used 6 public test data from two categories namely Twitter and Spicinemas.

Study of the result analysis obtained on incorporating the deep learning methodology [24] is also analyzed and tabulated. The complexity involved in the proposed work is also presented and the previous approaches contributed by various researchers over the years have also been reviewed and analyzed to determine the best.

The highlights of this proposal are as follows:

1. Performance analysis of surface and deep ensembles.

2. Benefits of deep learning approaches when compared with surface approaches.

3. Characterization of existing methodologies in sentiment analysis with respect to ensemble of traditional [25] and deep learning methodologies.

The rest of the paper is organized such that, Section 2 describes the proposed methodology and its implementation in an experimental design setup. Section 3 shows a detailed discussion on various methodologies, their performance and draws a comparison on its results. Section 4 concludes the work and outlines the possible future scope.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

2. Methodology

2.1 Dataset

A total of six data sets are used in this methodology to analyze the data. In particular concern, three data sets will together form a total of 43 tweets. These tweets comprises of 728 words along with a vocabulary of 12 words each. Following that, the tweets are processed in order to teach the algorithm to perform well. As a result, during pre-processing, some of the characters are changed or deleted. This might involve the removal of emoticons, special characters, and case conversions.

2.2 Word Embedding

Word embedding is used such as GloVe and Word2Vec. GloVe is implemented along with pre-trained word vectors. It uses 25 dimensional vectors which have been formulated based on 3 billion tweets. This helps develop a large training data set from the available contact. The Word2Vec methodology also uses similar 25 dimensional word vectors. In this type of word embedding words that are less often used are removed automatically. It uses the calculation known as skip length which defines the least number of times a word has to be used before it gets discarded. The following equation is used to normalize the work vector:

$$n_i' = \frac{n_i - n_{min}}{n_{max} - n_{min}} \tag{1}$$

Here n_{max} and n_{min} are the maximum and minimum value of the vector while represents the normalized i-value

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

2.3 Ensemble of Features N-Model

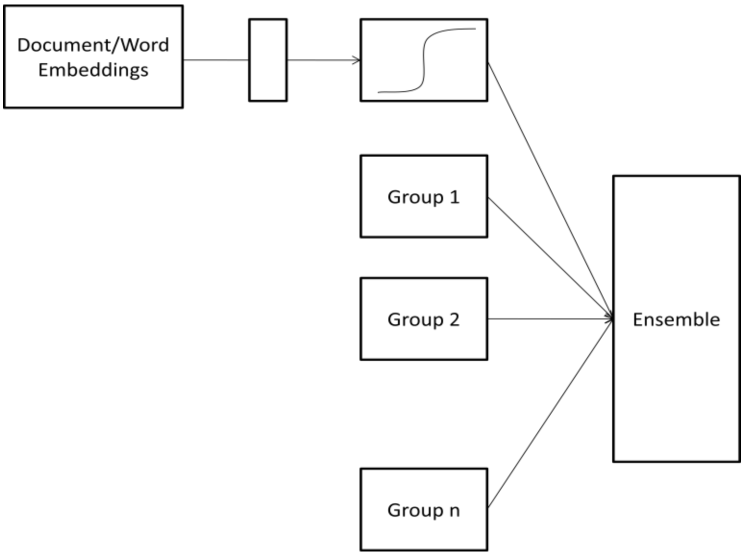

Fig.1. Ensemble Model with combination of the different features

The N_{SG} model combines a number of features in order to create a new set of features based on the individual information obtained from the features. The learning methodology that works on unified collection of data will be able to perform a better in comparison with that one that works by using a single feature. This is one manner of differentiating the two types of feature analysis. When both deep learning and surface features are combined and used, they belong to the first category as an ensemble of features. This model is commonly referred to as the surface model and can be represented as N_S. On the other hand, if the method uses only deep learning techniques then they are known as N_G model or deep learning model. Here both effect words as well as generic words N_G are combined as shown in Figure 2 and Figure 3. The resultant vector from training pre-trained collection of vectors is known as affect vector N_A. A combination of all the three features together is known as N_{SGA} model.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

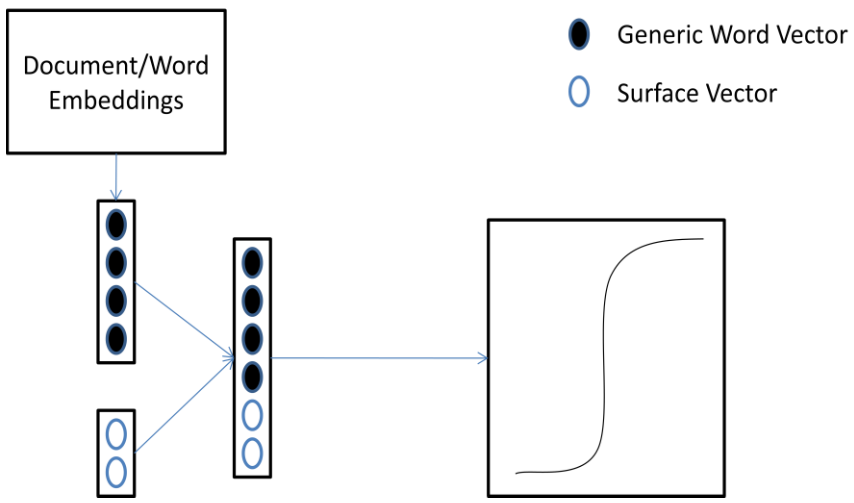

Fig.2. Ensemble Model with Surface vector

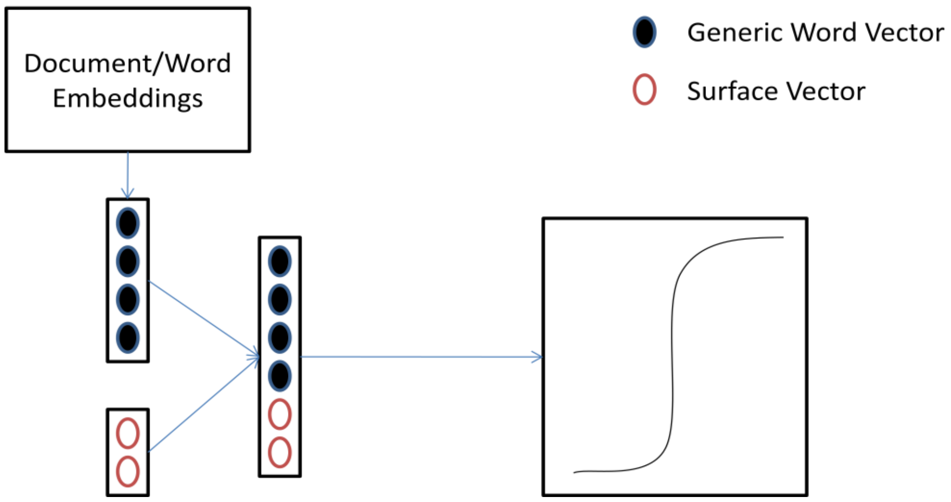

Fig.3. Ensemble Model with Generic and Surface vector

3. Results and Discussion

To evaluate the performance of this work, real datasets are used and applied to record the performance and further compare it with other methodologies. Six different datasets are used

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

for Sentiment analysis and the parameters measured are Recall, precision, accuracy and FI-score. The experiments indicated are as shown below:

3.1 Statistical Analysis

In order to analyze the various methodologies that are proposed in this work, based on experimental analysis, a statistical calculation is required with the help of post-hoc test and Friedman test. When many data sets are used, these tests enable a proper channel of comparison. According to the post hoc test, the F-distribution can be expressed as shown below:

$$F_F = \frac{(N-1)\gamma_F^2}{N(k-1)-\gamma_F^2}$$

where the degree of freedom is (k-1)(N-1). On the other hand Friedman's test uses the rank of an algorithm to determine the states of existence with respect to null-hypothesis such that:

$$\gamma_F^2 = \frac{12N}{k(kk+1)}(\sum_i R_i^2 - \frac{k(k+1)}{4})^2$$

where the degree of freedom is k-1. When a tie occurs, the issue is solved using a proper mechanism of average rank calculation.

3.2 Performance of Base Classifiers

Based on the F-Score values recorded in Table 1, it is observed that TextBlob was able to attain a higher performance when compared with the other classifiers. Moreover, higher performance is observed in the pattern.en. Similarly, the classifier that shows minimal performance is Sent140. The performance of the base classifiers are highly influence by the nature of the domain.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

Table 1. F-Score of Base Classifiers

Dataset	TB	Pattern.en	WSD	Sent140
1	83	83	77	79
2	70	72	74	61
3	86	86	75	79
4	67	81	72	74
5	74	74	88	71
6	71	69	80	70

3.3 Computational Complexity

The biggest drawback in the ensemble approach is the cost at which the resources are available. In order to fully understand the performance of the proposed model, a study on the computational cost is also included. As the actual cost of the text cannot be determined, we have used the training model in order to determine the cost of the system. The table below summarizes the computational costs that are associated with the proposed work.

Table 2. Computational Complexity of Word Embeddings

Word Embeddings	Approach	IMDB	Sentiment140
Word2vec	Compute Model	98.5 s	110 s
	Train Model	82 s	154 s
GloVe	Compute Model	2h	6.5 h
	Train Model	7 s	9 s
SSWE	Compute Model	27 d	-
	Train Model	180 s	89 s

Table 3. Computational Complexity in N-Models with Training Time

Model	IMDB	Sentiment140
N_G	1 s	13 s
N_A	1 s	14 s
N_S	38 s	13 s
N_{SGA}	39 s	980 s

Table 3 shows an analysis of various sets of features that show the time taken for training in two different environments. Here, it is observed that, N_{SGA} requires higher time of about 980

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

s in order to train the data sets. Similarly, in a meta-data learning environment, the time required to process data will be 1.5 s, which shows no noticeable change between the training data.

4. Conclusion

This paper presents a number of models that use a combination of automatic extraction and handcrafted separation of features along with a group of analyses it had trained accordingly. Moreover, a deep learning methodology is proposed to evaluate the performance of the combined work. A total of six data sets are being used with respect to two social websites of different domains. Statistical analysis is performed to determine the information combined through different analyses and features that are sufficient to outperform the performance of sentiment classification. The proposed work addresses the basic framework required to characterize the already available sentiment analysis based on the traditional research methodologies with respect to deep learning techniques. Analysis also indicates that, the proposed work shows significant improvement over the already existing techniques in terms of performance. This shows the ensemble of data obtained from different sources like affect word vectors, generic and surface features, which will result in a positive improvement with sentiment analysis tasks. Finally, this research work will also conclude on the test methodology that can be used to enhance the performance of deep sentiment analysis. As a possible future scope, this methodology can also be incorporated evenly in other languages. Research is underway to incorporate the proposed work with respect to emotion analysis.

References

[1] Adam, Edriss Eisa Babikir. "Evaluation of Fingerprint Liveness Detection by Machine Learning Approach-A Systematic View." Journal of ISMAC 3, no. 01 (2021): 16-30.

[2] Dashtipour, K., Ieracitano, C., Morabito, F. C., Raza, A., & Hussain, A. (2021). An Ensemble Based Classification Approach for Persian Sentiment Analysis.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

In *Progresses in Artificial Intelligence and Neural Systems* (pp. 207-215). Springer, Singapore.

[3] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation–A Systematic Approach." Journal of Artificial Intelligence 3, no. 01 (2021): 1-16.

[4] Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data*, *5*(1), 1-25.

[5] Chakrabarty, Navoneel, and Sanket Biswas. "Navo Minority Over-sampling Technique (NMOTe): A Consistent Performance Booster on Imbalanced Datasets." *Journal of Electronics* 2, no. 02 (2020): 96-136.

[6] Chakraborty, K., Bhattacharyya, S., Bag, R., & Hassanien, A. E. (2018, February). Comparative sentiment analysis on a set of movie reviews using deep learning approach. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 311-318). Springer, Cham.

[7] Hariharakrishnan, Jayaram, and N. Bhalaji. "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things." Journal of ISMAC 3, no. 02 (2021): 69-81.

[8] Hanafy, M., Khalil, M. I., & Abbas, H. M. (2018, September). Combining classical and deep learning methods for twitter sentiment analysis. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition* (pp. 281-292). Springer, Cham.

[9] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.

[10] Pasupa, K., & Ayutthaya, T. S. N. (2021). Hybrid deep learning models for thai sentiment analysis. *Cognitive Computation*, 1-27.

[11] Haoxiang, Wang, and S. Smys. "Overview of Configuring Adaptive Activation Functions for Deep Neural Networks-A Comparative Study." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 01 (2021): 10-22.

[12] Ranganathan, G. "A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms." Journal of Innovative Image Processing (JIIP) 3, no. 01 (2021): 66-74.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

[13] Do, H. H., Prasad, P. W. C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, *118*, 272-299.

[14] 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 (TCSST) 3, no. 01 (2021): 24-39.

[15] Adam, Edriss Eisa Babikir. "Survey on Medical Imaging of Electrical Impedance Tomography (EIT) by Variable Current Pattern Methods." Journal of ISMAC 3, no. 02 (2021): 82-95.

[16] Joe, Mr C. Vijesh, and Jennifer S. Raj. "Location-based Orientation Context Dependent Recommender System for Users." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 14-23.

[17] Shirley D.R.A., Sundari V.K., Sheeba T.B., Rani S.S. (2021) Analysis of IoT-Enabled Intelligent Detection and Prevention System for Drunken and Juvenile Drive Classification. In: Kathiresh M., Neelaveni R. (eds) Automotive Embedded Systems. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-59897-6_10

[18] Chourasia, Mayank, Shriya Haral, Srushti Bhatkar, and Smita Kulkarni. "Emotion recognition from speech signal using deep learning." Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020 (2021): 471-481.

[19] Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, *97*, 106754.

[20] Janoria, Honey, Jasmine Minj, and Pooja Patre. "Classification of Skin Disease Using Traditional Machine Learning and Deep Learning Approach: A Review." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 643-651. Springer Singapore, 2021.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021

Journal of Soft Computing Paradigm (JSCP) (2021)
Vol.03/ No.02
Pages: 123-134
http://irojournals.com/jscp/
DOI: https://doi.org/10.36548/jscp.2021.2.006

[21] López, M., Valdivia, A., Martínez-Cámara, E., Luzón, M. V., & Herrera, F. (2019). E2SAM: evolutionary ensemble of sentiment analysis methods for domain adaptation. *Information Sciences*, *480*, 273-286.

[22] Jain, Sarika, Ekansh Tiwari, and Prasanjit Sardar. "Soccer Result Prediction Using Deep Learning and Neural Networks." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 697-707. Springer Singapore, 2021.

[23] Murugesan, G., G. Preethi, and S. Yamini. "A Deep Learning Approach for Detecting and Classifying Cancer Types." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 135-141. Springer Singapore, 2021.

[24] Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Information Processing & Management*, *57*(1), 102141.

[25] Gautam, K. S., Vishnu Kumar Kaliappan, and M. Akila. "Strategies for Boosted Learning Using VGG 3 and Deep Neural Network as Baseline Models." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 151-168. Springer Singapore, 2021.

Author's Biography

A Pasumpon Pandian has completed his UG and PG degree from the reputed colleges and works as Professor and Dean (R&D) in CARE College of Engineering, Trichy, India. His areas of research include Social Networks, Wireless Networks, Internet of Things, Computer Networks, Mobile Communication, Robotics and Electrical Infrastructure, Mobile APIs, Data Analysis and Visualization.

ISSN: 2582-2640 (online)
Submitted: 17.05.2021
Revised: 07.06.2021
Accepted: 26.06.2021
Published: 03.07.2021