International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Study of Learning Classifiers Over Review Text Dataset for Aspect Level Sentiment Analysis

Subhadra Biswal[1], Prithviraj Mohanty[2], Ajit Kumar Nayak[3]

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: With the passage of time, public and consumer reviews on social media gained a lot of attention and have become the easiest way for quick judgment. Many studies have been conducted in this field of sentiment analysis. The need for textual mining or sentimental analysis was felt or increased suddenly due to the outbursts of the internet and various social media platforms being available for the public to express their views or opinions. Since the number of people using the internet is growing all the time, there are a lot of different points of view available online. Users can openly voice their opinions, provide star ratings, and write reviews for books or any products they have read or seen. The abundance of information available from unstructured data aids in many knowledge productions. In this case, the difficulty is to choose an effective categorization method where the imbalanced dataset can have a significant detrimental impact on the machine classifier model's performance. In this research work, a compressive study has been done for three classifiers like Support Vector Machine (SVM), Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) considering with imbalanced dataset. Various experiments are conducted to visualize different performance measures like accuracy, precision, recall, F-measure, specificity and G-mean over the above classifier models.

Keywords: Sentiment Analysis, Classification Algorithm, SVM, ANN, CNN

1. Introduction

The internet has become an integral aspect of modern life, particularly because it can be used in a variety of disciplines, including the submission of opinions. People can now get opinions from strangers as well as family and friends via the internet, which enables virtual spaces for people to share their experiences via word-of-mouth (WOM) which is a method for determining whether a text contains positive or negative ideas or feelings. Now a days both Machine Learning (ML) and Deep Learning (DL) has been used by companies all over the world to do sentiment analysis automatically in order to gain insights from clients' feedback [5]. The most popular algorithms for sentimental analysis are compared for the purpose of this study. In a system or model classifier's work is considered usually important as it analyses sentiments at the aspect level. Here, in this paper three fundamental classifier models: - SVM, ANN and CNN are used.

The process of collecting reviews has changed throughout time and sentiment analysis on review data sets are very prevalent. When a customer submits a review for a product, his/her possible expressions may vary. However, social media has given all the privilege and freedom to the users to post reviews in the form of open-text about the features of the products [15] or comment about a controversial opinion. These textual reviews are analysed and given output as positive or negative polarities. In order to correct them, it is helpful to identify the factors that are causing important changes.

In this research, we anticipated three machine learning models for this level of sentiment analysis. Here, dataset taken from "imdb review dataset" site [14] which has around total 50000 files of benchmark review text datasets. The strategy of this paper is as follows: section 2 reflects some predefined literature surveys or some of the related works. Section 3 and 4 contain the frame work for Aspect Level Sentiment Analysis and models used for classification purposes. Experimental consequences and analysis are discussed in section 5 and at last section 6 end with conclusion and future work.

2. Related Work

As the digital era is evolving, more and more online data are produced including reviews of products which are posted on different social media sites with different languages. Many businesses including internet service providers are greatly benefit from this analysed information. Sentiment Analysis

[1 1] Department of CSE, Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be) University, Bhubaneswar, India [1]
Department of CSIT, Institute of Technical Education and Research, Siksha 'O' Anusandhan (Deemed to be) University,

is the process of understanding and characterizing the feelings or thoughts behind the text that is review texts using text analysis tools. When a statement has both positive and negative polarity at once, it can occasionally result in conflict situations, which is difficult for the sentiment analysis to predict the polarities accurately. So, here comes the concept of Aspect Based Sentiment Analysis which identifies fine-grained opinion polarity towards a given aspect connected to a target. With aspect category we can comprehend the sentiment analysis issues.

Attention-based Bidirectional CNN-RNN Deep Model (ABCDM) has been proposed by the author Mohammad., 2021 [3] by taking the temporal information bidirectional flow. This model will extract both past and future contexts by using two independent bidirectional LSTM and GRU layers. Attention mechanism is used here and outputs are emphasized on some words. This model also uses convolution and pooling algorithms to lower the dimensionality of the features and extract position invariant local features.

The suggested model here has used a variety of Machine Learning (ML) and Deep Learning (DL) models for classification. Word2Vec and fast Text [2]. So, using these two techniques LSTM model beat all other techniques with accuracy rates of 57.93 and 52.32 % respectively. In the year 2020, Nazir and Ambreen [5] had done a survey on challenges and issues at Aspect Level Sentiment Analysis where they had highlighted the factors which are responsible for the respective analysis progress with the comparing the ongoing current solutions. The challenges were to find out the measures which will increase the accuracies. Klinger and Kim [6][7] did a survey on Emotional and Sentimental Analysis for Computational literature studies which mostly contain the machine learning and statics works. In 2018, Yue et al. did the categorization using new tools about the latest technologies and processes of Sentiment Analysis. Different tools have been used to overcome limitations of social media analysis. Root causes for the commonly occurring classification error has been stated by Zimbra et al, 2018. A survey on Sentiment Analysis using DL is done by Aspect Classification approaches and solutions related to Aspect Entity, Schouten et al, 2016. Elaborated discussion on tasks related to twitter Sentiment Analysis like detection of emotion, detection of sarcasm with change in Sentiment over a particular period of time by Giachanou et al. 2016 [11]. Around 2015, Ravi et al, focused on document level and sentence level sentiment analysis with their respective subtasks. Here, overly he discussed about lexicon relationship, extraction of opinion-word, subjectivity classification with open issues and future direction for this analysis.

3. Sentiment Analysis procedure at Aspect Level

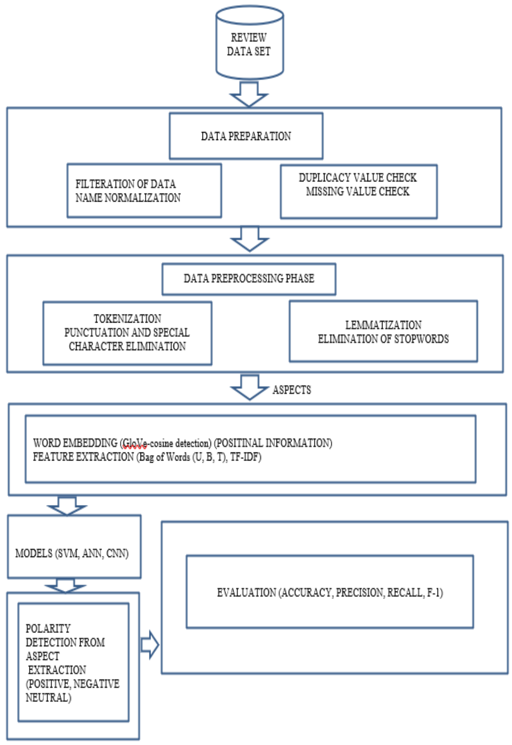

Fig 1. Frame Work for Aspect Level Sentiment Analysis

The above figure describes about the research studies using a number of strategies to arrive at a final classification result which is very useful for each classifier or we can say it briefly gives the idea about the proposed work. The processes are like data collecting, data labelling, data labelling which can also be referred as review of the data set, data cleaning, data splitting, resampling of the data using time series that produce distinct outputs, classification of data, and model evaluation. Collections of review text data are gathered for training and testing purposes. First of all, the data is collected using Pandas Python packages, which import data from a csv file into JupyterNotebook. Data Frame structure is used here. The label should be transformed from a string to an integer after the data has been read [22]. In this study the positive class is referred with digit '1' and negative class is referred with '-1'. The removal of punctuations, new lines, digits as well as URL removal, stop word removal are all part of the filtering strategy used in this study. The stemming process is then used to convert words to their root form. Data Split is done using train_test_split function with 90:10, 80:20 and 70:30 percent ratios. Random state parameters is to be defined when train_test_split() is used by the splitting data. Its main goal is to build an internal random number of generators that will determine how training and testing data will be divided. If the parameter part is found empty, the splitting step will produce different training and testing data each time the code is run. When it comes to data resampling, most algorithms assume balanced class distributions, as most of the common machine learning algorithms strive to maximize overall accuracy, which is overwhelmed by majority classes. So, they are biased towards the majority classes [16]. Over sampling of minority class are found by randomly duplicating minority data to modify the amount of data in majority classes are used for the resample. Unlike human, classifiers are unable to comprehend text data, they can only understand numbers. So, TF-IDF approach is used to convert the corpus to numbers. This is the major method for conversion of data. The TF-IDF formula is made up of two terms that is TF (Term Frequency) and IDF (Indicative Distribution Function) or it can be referred as Inverse Document Frequency.

The objective of former is to count the number of times a word appears in a document, sentences or a paragraph. However, knowing more unique words that the most common ones will be more beneficial in analyzing the sentiments [9] which is the purpose of IDF. This vectorizer function is used from the Scikit-Learn for the conversion process. The last method is the data classification approach where three algorithms like SVM, ANN and CNN have been employed.

4. Machine Learning Models Used

The commonly used machine learning models for sentiment analysis are SVM, ANN and CNN.

4.1. SVM

It is a machine learning model that chooses the optimal distinction between vectors that fall under a certain group and those that do not fall under the category. Therefore, it can be used with any type of vector that encodes any type of data. Here, text is converted into vectors [18] in order to fully utilize the potential of text categorization. It is a supervised method that may be applied to both regression and classification problems. Vectors and list of integers are used to represent a set of coordinates. SVM determines where to place the best line which is also known as “optimal hyperplane”. This separates the space into two subspaces: one for the vectors that fall within the specified category and other one is for the vectors that do not fall under the required plane. So, feature vectors are made to make it possible to represent the text in an n-dimensional plane [19]. As, categorization becomes simpler, it is particularly helpful when the test data is small. Here, we used 70% training and 30% test data as the starting point of our experiment with cost of constraint violation C=1.0, halting criteria e as 0.01 and maximum iteration as 1000.

4.2. ANN

It is mostly applied in scenarios where we want to mimic the behaviour of biological things or find patterns in data. A good application case for an ANN is to make medical diagnoses, speech recognition, data visualization and handwriting digit prediction [24]. This network is utilized when it is necessary to comprehend relation between inputs and outputs. It could be challenging to understand to correlations between the variables and the instances. This network is used as a frequent approach to retain the information of the data.

In Fig 2., one hidden layer has been used with “Bag-of-Words” approach to our training data followed by two hidden layers. This implementation has been used to convert an input review text sentence into an array of 0s and 1s. Sigmoid function was employed to normalize the data and its derivation is used for the calculation of Error Rate. The learning rate decay coefficient, batch size as 12 and alpha in the Leaky Rectified Linear Unit are just few hyperparameters that are tuned by experimenting with various values on the networks.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 30)	45030
leaky_re_lu (LeakyReLU)	(None, 30)	0
dropout (Dropout)	(None, 30)	0
dense_1 (Dense)	(None, 10)	310
leaky_re_lu_1 (LeakyReLU)	(None, 10)	0
dropout_1 (Dropout)	(None, 10)	0
dense_2 (Dense)	(None, 3)	33

Fig 2. ANN Architecture

4.3. CNN

The number of hidden layers used between the input and output layers defines it’s as a feed forward network. A collection of features is extracted at each layer with collection of features at each layer. In the below specified CNN architecture i.e., (Fig 3.) a series of filters are applied to the input to produce feature maps. The input values are multiplied by the weights of each filter as it travels through the fully connected layer. Then this is followed by passing the output to an activation function like ReLU, sigmoid or tanh [10]. The set of weight is evaluated using a loss function. The filters produce feature maps that emphasize various components of the input. Despite of being frequently utilized in images and video processing [20], CNN has recently been applied for text mining which is a technique of Natural Language Processing (NLP).

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 1500, 50)	250000
spatial_dropout1d (SpatialDr	(None, 1500, 50)	0
conv1d (Conv1D)	(None, 1498, 250)	37750
global_max_pooling1d (Global	(None, 250)	0
dense (Dense)	(None, 250)	62750
dropout (Dropout)	(None, 250)	0
activation (Activation)	(None, 250)	0
dense_1 (Dense)	(None, 1)	251
activation_1 (Activation)	(None, 1)	0

Fig 3. CNN Architecture

In this method the preprocessing step converts the text input to a matrix representation. Characters of sentences are used as rows and alphabetic characters are used as columns in the matrix form. In the processing technique, filter moves through the matrix's words. Convolution, ReLU, pooling and fully connected layers are the four keys used here. The total number weights of the convolutional layer are equal to the sum of weights and biases. Global Average Pooling is a pooling operation designed to replace fully connected layers in classical CNNs. But here Max pooling is used as it is better than average pooling.

5. Experimental Analysis and Result Discussion

It is quite tough to construct a data set from various sources. It is sometimes considered as critical stages for the researchers [25]. This dataset collected from Kaggle site [14] for binary sentiment classification. The sample format of data that we took is shown below (i.e, Fig 4.). Each sentence's identification number appears in the first column and the review-based sentence in the second column.

In the following sentence, the given aspect term is also considered as opinion target to predict the sentiment label for the aspect term.

id_number	text_data
1	One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked.
2	A wonderful little production. The filming technique is very unassuming- very old-time-
3	I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air
4	Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents
5	Petter Mattei's "Love in the Time of Money" is a visually stunning film to watch. Mr. Mattei offers

Fig 4. Sample of the Review Text Dataset

There are different models' evaluation methods, like accuracy, confusion matrix, precision, recall, F-measure, ROC Curve and G-mean. We can say these are the global measures or these are the most typical methods for the evaluation purposes. When the data is imbalance, predicted accuracy may not be appropriate [17]. So, we use confusion matrix to have a breakdown of the mistakes in the predictions for an unknown dataset. Precision, Recall, F-Measure [8] and G-mean are used to determine the exactness, completeness and balance. Scikit-Learn Functions are used to describe the entire model assessment process. Algorithms are primarily compared based on their accuracies. This performance evaluation compares the algorithms in the broadest way, rather focused on each class. As, a result this does not differentiate between the number of correct labels assigned to various classes whereas, confusion matrix is used to assess the performance of classification models [16]. This technique aids in the detection of confusion and provides information not only about the errors but also about the errors made. G-mean is considered as central value in a set of numbers that is calculated by determining the nth degree root of the set with n number of products so, it is only applied to set of positive numbers.

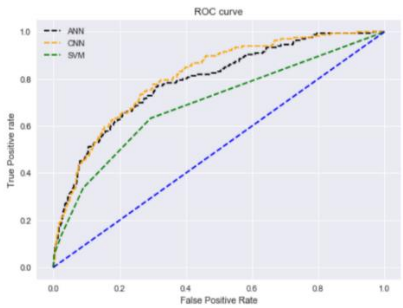

Fig 5. ROC Curve Graph for ANN, SVM, CNN Classifiers

Figures (fig 6. and fig 7.) shows the curve graph between accuracy and loss at Iteration 700. The solid line depicts training part and dotted line represent the validation part.

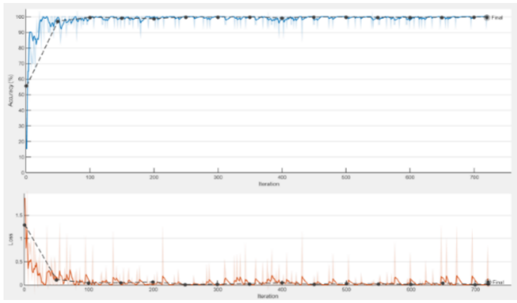

Fig 6. (Accuracy and Loss curve) Simulation graph for ANN Classifier at 700 Iteration

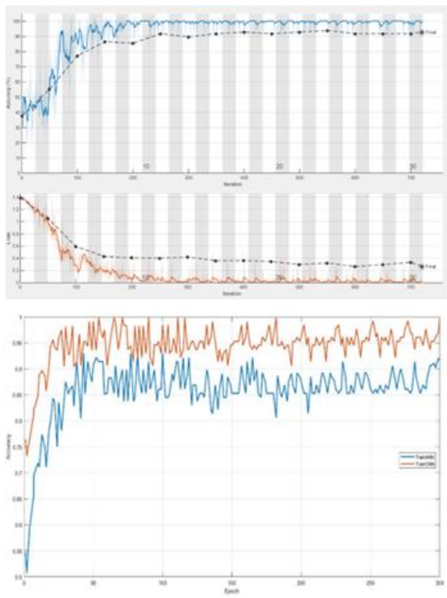

Fig 7. (Accuracy and Loss curve) Simulation graph for CNN Classifier at 700 Iteration

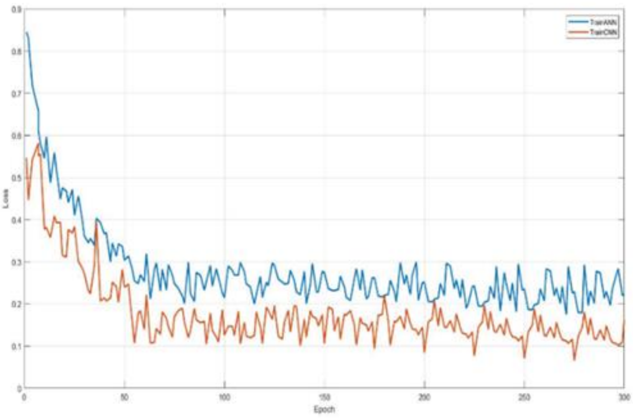

Fig 8. Loss and Accuracy curve comparision graph for ANN and CNN Classifier at an Epoch 300

The above figure (Fig:8) shows the accuracy and loss plot graph at epoch 300. CNN accuracy increases as training time progresses as compared to ANN accuracy curve and CNN loss curve decreases as training time progresses.
Below mentioned tables (Table 1. and Table 2.) shows the value comparision between the performance meausre parameters of the three used classifiers. Mainly the correlation is shown between two classifers "ANN" and "CNN" at different epochs like 100, 200 and 300 with different training and testing data split ratios. It is found that CNN outperforms better results.

Table 1. Performance Metrics Comparision of Classifiers

PERFORMANCE MEASURES OF CLASSIFIERS	SVM	ANN	CNN
ACCURACY	0.863	0.889	0.911
RECALL	0.681	0.720	0.796
SPECIFICITY	0.940	0.961	0.960
PRECISION	0.827	0.886	0.894
F-MEASURE	0.747	0.794	0.842
G-MEAN	0.800	0.832	0.874

Table 2. Performance Metrics Comparision of ANN and CNN Classifiers with different Epochs and Split Ratios

EPOCHS	SPLIT RATIO (%)	ACCURACY		PRECISION		RECALL	
		ANN	CNN	ANN	CNN	ANN	CNN
100	70-30	0.867	0.892	0.764	0.882	0.602	0.731
200	80-20	0.880	0.901	0.851	0.840	0.714	0.780
300	90-10	0.889	0.911	0.886	0.894	0.720	0.796

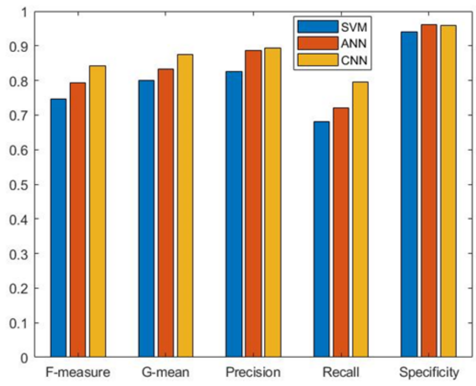

Fig 9. Bar Graph Representation of Performance Metrics with [0-1] value

The above figure (Fig:-9) depicts the compairision between the performance measures between SVM, ANN and CNN Classifiers.

6. Conclusion and Future Work

In this study, we discussed our strategies for tackling the issue of ABSA for a binary class review text data set, one for positive and another for negative. With the help of SVM, ANN and CNN classifiers, we have created the Aspect Based Sentiment Analysis system employing supervised methods for identifying aspects and their polarities. These approaches used various performance metrics including accuracy, precision, recall, F-measure and G-mean. In comparison to ANN, at various epochs CNN technique scored the highest on this benchmark dataset with 91.1% accuracy.
Furthermore, this work will be extended by working on other Indian language text review data sets and different review text datasets on products or restaurant or places with different classes like neutral, average positive, average negative etc. This process may require manual labelling under an experienced guidance. We can make this work more efficient by ensemble methods or hybrid approaches between different classifiers with different labels in which sentiments to be expressed effectively.

7. References

[1]. Dash, Shiyona, et al. "Deep learning–based decision support system for multicerebral disease classification and identification." *Brain Tumor MRI Image Segmentation Using Deep Learning Techniques*. Academic Press, 2022. 91-122.

[2]. Aparna, T. Sai, et al. "Aspect-Based Sentiment Analysis in Hindi: Comparison of Machine/Deep Learning Algorithms." *Inventive Computation and Information Technologies*. Springer, Singapore, 2021. 81-91.

[3]. Sally Fouad Shady. (2021). Approaches to Teaching a Biomaterials Laboratory Course Online. Journal of Online Engineering Education, 12(1), 01–05. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/43

[4]. Basiri, Mohammad Ehsan, et al. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." *Future Generation Computer Systems* 115 (2021): 279-294.

[5]. Pérez, Juan Manuel, Juan Carlos Giudici, and Franco Luque. "pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks." *arXiv preprint arXiv:2106.09462* (2021).

[6]. Kastrati, Zenun, et al. "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study." *Applied Sciences* 11.9 (2021): 3986.

[7]. Nazir, Ambreen, et al. "Issues and challenges of aspect-based sentiment analysis: a comprehensive survey." *IEEE Transactions on Affective Computing* (2020).

[8]. Mitra, Ayushi. "Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset)." *Journal of Ubiquitous Computing and Communication Technologies (UCCT)* 2.03 (2020): 145-152.

[9]. Roy, R., and D. A. . Kalotra. "Vehicle Tracking System Using Technological Support for Effective Management in Public Transportation". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 2, Mar. 2022, pp. 11-20, doi:10.17762/ijritcc.v10i2.5515.

[10]. Prabha, M. Indhraom, and G. Umarani Srikanth. "Survey of sentiment analysis using deep learning techniques." *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE, 2019.

[11]. Zhang, Lei, Shuai Wang, and Bing Liu. "Deep learning for sentiment analysis: A survey." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4 (2018): e1253.

[12]. Jain, Kruttika, and Shivani Kaushal. "A comparative study of machine learning and deep learning techniques for sentiment analysis." *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. IEEE, 2018.

[13]. Zia, S. S., et al. "A survey on sentiment analysis, classification and applications." *Int J Pure Appl Math* 119.10 (2018): 1203-1211.

[14]. https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews

[15]. Zia, S., et al. "A survey on sentiment analysis, classification and applications." *Int J Pure Appl Math* 119.10 (2018): 1203-1211.

[16]. Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064

[17]. Narendra, B., et al. "Sentiment analysis on movie reviews: a comparative study of machine learning algorithms and open-source technologies." *International Journal of Intelligent Systems and Applications* 8.8 (2016): 66.

[18]. Madhoushi, Zohreh, Abdul Razak Hamdan, and Suhaila Zainudin. "Sentiment analysis techniques in recent works." *2015 Science and Information Conference (SAI)*. IEEE, 2015.

[19]. Dursun, M., & Goker, N. (2022). Evaluation of Project Management Methodologies Success Factors Using Fuzzy Cognitive Map Method: Waterfall, Agile, And Lean Six Sigma Cases. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 35–43. https://doi.org/10.18201/ijisae.2022.265

[20]. Schouten, Kim, and Flavius Frasincar. "Survey on aspect-level sentiment analysis." *IEEE Transactions on Knowledge and Data Engineering* 28.3 (2015): 813-830.

[21]. Rezaeinia, Seyed Mahdi, Ali Ghodsi, and Rouhollah Rahmani. "Improving the accuracy of pre-trained word embeddings for sentiment analysis." *arXiv preprint arXiv:1711.08609* (2017).

[22]. Joulin, Armand, et al. "Bag of tricks for efficient text classification." *arXiv preprint arXiv:1607.01759* (2016).

[23]. Cliche, Mathieu. "BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs." *arXiv preprint arXiv:1704.06125* (2017).

[24]. Kamilaris, Andreas, and Francesc X. Prenafeta-Boldú. "A review of the use of convolutional neural networks in agriculture." *The Journal of Agricultural Science* 156.3 (2018): 312-322.

[25]. Prabha, M. Indhraom, and G. Umarani Srikanth. "Survey of sentiment analysis using deep learning techniques." *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE, 2019.

[26]. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. *The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).*

[27]. Zuheros, Cristina, et al. "Sentiment analysis based multi-person multi-criteria decision-making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews." *Information Fusion* 68 (2021): 22-36.

[28]. Deepak Mathur, N. K. V. . (2022). Analysis & Prediction of Road Accident Data for NH-19/44. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(2), 13–33. https://doi.org/10.17762/ijrmee.v9i2.366

[29]. Zunic, Anastazia, Padraig Corcoran, and Irena Spasic. "Sentiment analysis in health and well-being: systematic review." *JMIR medical informatics* 8.1 (2020): e16023.

[30]. Balaji, Penubaka, O. Nagaraju, and D. Haritha. "Levels of sentiment analysis and its challenges: A literature review." *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. IEEE, 2017.