**CHAPPTER TWO**

**LITTRATURE REVIEW**

According to the paper work of (Nazeer et al., 2020), titled Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis. The researchers stated that twitter is a platform where the users can express there feeling and opinion. It has also become a major respiratory where many sectors get their data and then later analyzed it for sentiment. According to the researchers, many machine learning algorithm are available for twitter sentiment analysis which can be used for predicting the sentiments of tweets automatically. However, there are disadvantage that affect the machine learning classifiers to achieve an outstanding performance in term of classification. In this paper the authors proposed a novel feature generating technique to produce desired features for tanning model. This novel ensemble classification system classifier, which make use of many commonly used statistical model like naïve Bayes, random forest, and logistical regression which are weighted according to their ability on historical data, where weights are chosen differently for each model. The researchers then concluded that in future, the proposed ensemble model will not participate in tanning the dataset for classification of sentiments. The model will then be extended by the use of optimization algorithm for further refining the future se for outstanding performance in term of classification in sentiment analysis.

Reviewing this article by (Wang et al., 2020) , which is been titled Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis. According to the authors a research is been done on dimensional sentiment analysis, in which dimensional sentiment analysis aim to identify continues numerical values in double dimensions such as the valence-arousal(VA) space. Matched to the categorical approach that aim on sentiment classification such as binary classification (positive and negative), the dimensional approach can give a better sentiment analysis. In this paper the research proposed a tree-structured regional CCN-LSTM model with two part: regional CNN and LSTM to predict the VA performance of texts. Unlike a conventional CNN. The proposed regional CNN make used of some text as region, by dividing an input into many region so that the useful information in each region can be extracted and weighted their advantage to the VA prediction. It’s been concluded that in other to further improve the performance, a region division approach is proposed to discover task-relevant phrases and clauses to combined structured information into VA prediction. The experimental result on different corpora shows that the proposed method outperform other methods.

In this paper (Raisa et al., 2021) title A Review on Twitter Sentiment Analysis Approaches. According to the researcher it is been stated that sentiment analysis solves the problem of interpreting data presented in different format and realm in term of feelings and opinions. Sentiment analysis is also needful to different sectors that need awareness of the view of people on their products or other cases. Twitter is among the most widely accepted social platform and widely used platform site where people can express their mind, view and opinions, which make this space an outshine tools for analyzing sentiment. The authors reviewed the work done by other researchers in which the reviewed work are separated according to their implementation classification algorithms. A contrastive analysis on different approach and method of sentiment analysis using Twitter data in the review article. Also result analysis, research gaps, and future scope of twitter sentiment are tackled.

In this article (Shitole, 2021) titled Twitter Sentiment Analysis Using Supervised Machine Learning. The researcher stated that sentiment analysis aim to figure out opinions, attitude, and also emotions from different social media site such as twitter. In the recent years twitter has become the most popular research area. The main focus of the conventional way of sentiment analysis is on textual data. According to the researcher twitter is the most popular smallblooging online networking site in which the users post updates about different topics in form of tweet. In this article the authors make use of label dataset which is publicly available on Kaggle, and also a well arrange pre-processing algorithm which make the tweets manageable to the normally used language plan is structured, in which each example in the dataset is a pair of tweet an sentiment. So the supervised machine learning is used to support the sentiment analysis models based on naïve Bayes, logistic regression, and support vector machine are proposed. In this paper the researcher main goal is to break down sentiments adequately. In the twitter sentiment analysis tweet are categorized into positive and negative sentiment. The researcher concluded that higher accuracy is obtained by using sentiments features instead of conventional text classification. In which the feature can be utilized by various sector.

Reviewing the work of (Machuca et al., 2021) titled Twitter Sentiment Analysis on Coronavirus: Machine Learning Approach. In this paper the author stated that the main challenges in machine learning is the analysis of data to identify feelings using algorithms that allow us to identify a bad feelings from good ones in a tweet. Social network is a vital source of getting information, which is been used to express personal point of view and ideas. Based on this article the researchers proposed a sentiment analysis of English tweet at the time of pandemic COVID-19 in 2020. According to them the tweet where categorized as good or bad by applying the Logistic Regression algorithm, using this method the authors got a classification accuracy of 78.5% . In this paper the researcher’s focus on the analyzing people’s behavior to the pandemic. The main goal is to find out whether the sentiment of the public opinion is positive or negative by using machine learning algorithm and natural language processing techniques. Even though that the analysis figure out variation of ideas. It seems that the users mostly remain positive about the pandemic. The month of January is the only month in which negative ideas is much than the positive ideas, then march is the month when the COVID-19 disease was announce as a pandemic and several countries started to apply percussion measures, due to the rise of positive thought. To make it short, 54% of the people showed positive feelings and 64% of the people showed negative feelings.

(Model et al., 2021)

(Model et al., 2021), tiled A Novel Word-embedding Method for Real-time Sentiment with Integrated LSTM-CNN Model. According to them, Artificial intelligence (AI) is a focused research area technology in which Natural language processing (NLP) is a vital technology in Artificial intelligent. It is also stated that sentimental analysis goal is to absorb and classify the people’s idea by Natural language processing (NLP). The machine learning and lexicon dictionaries have less power to efficiently analyze large live media data. In the last few years, deep learning method have successfully improve the accuracy of recent sentiment models. However the existing method provide the part that reduces each word accuracy if a sentence doesn’t follow the information in real-time. Therefore, the authors proposed a novel word embedding method for the real-time sentiment (WRS) for word representation. The word real-time sentiment novel is a word embedding method namely, word-to-word Graph (W2WG) embedding that make use of the word2vce approach. This WRS method combine different lexicon resources to employ the W2WG embedding method to achieve the word features water vector. Achieve the word feature vector. Robust neural networks leverage these features by integrating LSTM and CNN to improve sentiment classification performance. LSTM is utilized to store the word sequence information for the effective real-time SA, and CNN is applied to extract the leading text features for sentiment classification. The author concluded that the analysis are performed on Twitter and IMDB datasets. The results demonstrate our proposed method's effectiveness for real-time.

According to the research work of (Calleja-solanas et al., 2021) titled Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. The authors stated that the growth if technology in the last few years as become very high, this is known to us by different social media that have come up. Twitter is one of the popular social media, twitter is a platform which is created to preach activities, discuses and distribute stories among the users. It is also stated that twitter have become a place in which customers log complains about a product to the company, one of which is PT Telkon Indonesia. Some complainant prefer not to contact the customer service that have been provided by the company, but prefer to log complain via the twitter platform. In this paper the data gotten during a certain period, 3324 tweets were obtained, which include the keywords indihome, myindihome, useetv and wifi.id. The tweet data that have been taken, if processed well, will be a useful information for the company. According to the authors t tweet were classified where the keywords indihome, myindihome, useetv and wifi.id. furthermore, several data preprocessing techniques were carried out, sentiment analysis and visualization in the view of histograms, pie chart, and word clouds, from 3324 tweets that have been analyzed, the result shown that there are 34.4% positive tweet , 16.1% negative tweets, and 49.6% neutral tweets

(Hanif et al., 2021), in this paper work titled Sentiment Analysis of Tweets through Altmetrics: A Machine Learning Approach. It is been stated that the goal of this research is to impact an Altemetrics datasets among five studies, undertake sentiment analysis using different machine learning and natural language processing –based algorithm, and finding the best-performing model and provide a python library for sentiment analysis of an altmetrics datasets. In this paper the researcher gave a set of rules to two human annotators which is related with the job of tweet annotation of scientific literature. It is also stated that the sentiment is duly label, achieving an inter-annotators agreement of 0.80(Cohen’s kappa). And also the exact analysis were conducted on two different versions of the dataset: one with tweet in English and the other with tweet in 23 languages, including English. According to the authors 6388 tweet was use and about 300 article indexed in web of since. The impact of adopting machine learning and natural machine language processing models was measured by comparing with well know sentiment analysis model, that is sentiStrenght and sentiment140, as the baseline. The researcher concluded that support vector machine with un-gram outperformed all the other machine learning algorithm and the baseline methods employed, with an accuracy of over 85% followed by logistic regression at 83% accuracy and naïve Bayes at 80%. Then the precision, recall and F1 scores for support vector machine, logistic regression and naïve Bayes were (0.89, 0.86, 0.86), (0.86, 0.83, 0.80) and (0.85, 0.81, 0.76), respectively.

Reviewing the paper work of (Priyadarshini & Cotton, 2021), titled A novel LSTM–CNN–grid search‑based deep neural Network for Sentiment Analysis. According to them the amount of people getting attached to the use of the internet is increasing at a high rate, there is more user-generated content on the web, understanding hidden opinions, sentiments, and emotions in emails, tweets, reviews and comments is problem and equally vital for social media monitoring, brand monitoring, customer services, and market research. Sentiment analysis provide the, emotional tone behind a series of words may be importantly be used to comprehend the attitude, opinion, and emotions users. In this article the researchers proposed a novel long short-term memory (LSTM)-convolution neural network (CNN)-grid search –based deep neural network model for sentiment analysis. The researcher considered, baseline algorithms like convolutional neural networks, K-nearest neighbor, LSTM, neural networks, LSTM-CNN, AND CNN-LSTM which have been analyzed using accuracy, precision, sensitivity, specificity, and F-1 score, on multiple datasets. According to the result shown by the researcher, the proposed model base on hyper parameter optimization outperforms other baseline models with an overall accuracy greater that 96%.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| S/N | AUTHOR | TITLE | PROBLEM STATEMENT | METHODOLOGY | RESULT | LIMITATION |
| 1 | (Nazeer et al., 2020) | Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis. | Disadvantages that affect the machine learning classifiers to achieve an outstanding performance in term of classification. | A novel feature generating technique to produce desired features for tanning model. This novel ensemble classification system classifier | The proposed ensemble model will not participate in tanning the dataset for classification of sentiments. The model will then be extended by the use of optimization algorithm for further refining the future se for outstanding performance in term of classification in sentiment analysis. | Other machine learning algorithm are not considered such as decision tree and support vector machine. |
| 2 | (Wang et al., 2020) | Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis. | The categorical approach that aim on sentiment classification such as binary classification (positive and negative), the dimensional approach can give a better sentiment analysis. | Proposed a tree-structured regional CCN-LSTM model with two part: regional CNN and LSTM to predict the VA performance of texts. | The experimental result on different corpora shows that the proposed method outperform other methods. | In this paper the experiment is carried on a structured data  In which unstructured data is not considered. |
| 3 | (Raisa et al., 2021) | A Review on Twitter Sentiment Analysis Approaches. | Interpreting data presented in different format and realm in term of feelings and opinions. Also result analysis, research gaps, and future scope of twitter sentiment are tackled. | A contrastive analysis on different approach and method of sentiment analysis using Twitter data in the review article. | Result analysis, research gaps, and future scope of twitter sentiment are tackled. | Review was carried on recent papers in which old papers are not considered. |
| 4 | (Shitole, 2021) | Twitter Sentiment Analysis Using Supervised Machine Learning. | The conventional way of sentiment analysis is on textual data. | Supervised machine learning is used to support the sentiment analysis models based on naïve Bayes, logistic regression, and support vector machine. | Higher accuracy is obtained by using sentiments features instead of conventional text classification. In which the feature can be utilized by various sector. | The research make use of supervised machine learning algorithm. As unsupervised machine learning algorithm are not considered |
| 5 | (Machuca et al., 2021) | Twitter Sentiment Analysis on Coronavirus: Machine Learning Approach | Analysis of data to identify feelings using algorithms that allow us to identify a bad feelings from good ones in a tweet. | A sentiment analysis of English tweet at the time of pandemic COVID-19 in 2020. According to them the tweet where categorized as good or bad by applying the Logistic Regression algorithm. | That the users mostly remain positive about the pandemic. The month of January is the only month in which negative ideas is much than the positive ideas, then march is the month when the COVID-19 disease was announce as a pandemic and several countries started to apply percussion measures, due to the rise of positive thought. To make it short, 54% of the people showed positive feelings and 64% of the people showed negative feelings. | The analysis was conducted of COVID-19 other diseases are not considered. |
| 6 | (Calleja-solanas et al., 2021) | Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob. | Twitter have become a place in which customers log complains about a product to the company. | Data preprocessing techniques were carried out, sentiment analysis and visualization in the view of histograms, pie chart, and word clouds, | The result shown that there are 34.4% positive tweet, 16.1% negative tweets, and 49.6% neutral tweets. | The analysis carried was on twitter social media, in which other social media data was not considered. |
| 7 | (Hanif et al., 2021), | Sentiment Analysis of Tweets through Altmetrics: A Machine Learning Approach. | Finding the best-performing model and provide a python library for sentiment analysis of an altmetrics datasets. | A set of rules to two human annotators which is related with the job of tweet annotation of scientific literature. It is also stated that the sentiment is duly label, achieving an inter-annotators agreement of 0.80(Cohen’s kappa). And also the exact analysis were conducted on two different versions of the dataset: one with tweet in English and the other with tweet in 23 languages, including English. | Support vector machine with un-gram outperformed all the other machine learning algorithm and the baseline methods employed, with an accuracy of over 85% followed by logistic regression at 83% accuracy and naïve Bayes at 80%. Then the precision, recall and F1 scores for support vector machine, logistic regression and naïve Bayes were (0.89, 0.86, 0.86), (0.86, 0.83, 0.80) and (0.85, 0.81, 0.76), respectively. | Other machine learning algorithm was not considered like decision tree. |
| 8 | (Priyadarshini & Cotton, 2021). | A novel LSTM–CNN–grid search‑based deep neural Network for Sentiment Analysis. | Understanding the hidden opinions, sentiments, and emotions in emails, tweets, reviews and comments is problem and equally vital for social media monitoring, brand monitoring, customer services, and market research. | Article the researchers proposed a novel long short-term memory (LSTM)-convolution neural network (CNN)-grid search –based deep neural network model for sentiment analysis. The researcher considered, baseline algorithms like convolutional neural networks, K-nearest neighbor, LSTM, neural networks, LSTM-CNN, AND CNN-LSTM which have been analyzed using accuracy, precision, sensitivity, specificity, and F-1 score, on multiple datasets. | The result shown by the researcher, the proposed model base on hyper parameter optimization outperforms other baseline models with an overall accuracy greater that 96%. | In this article, The researcher considered, baseline algorithms like convolutional neural networks, K-nearest neighbor, LSTM, neural networks, LSTM-CNN, AND CNN-LSTM. Which other machine learning algorithm are not considered. |
| 9 | (Model et al., 2021). | A Novel Word-embedding Method for Real-time Sentiment with Integrated LSTM-CNN Model | Machine learning and lexicon dictionaries have less power to efficiently analyze large live media data. | The authors proposed a novel word embedding method for the real-time sentiment (WRS) for word representation. The word real-time sentiment novel is a word embedding method namely, word-to-word Graph (W2WG) embedding that make use of the word2vce approach. This WRS method combine different lexicon resources to employ the W2WG embedding method to achieve the word features water vector. Achieve the word feature vector | The results demonstrate our proposed method's effectiveness for real-time. | the analysis are performed on Twitter and IMDB datasets. |
| 10 |  |  |  |  |  |  |

Calleja-solanas, V., Pigani, E., Hall, G., Bialek, W., Alash, H. M., & Al-sultany, G. A. (2021). *Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob Twitter Sentiment Analysis as an Evaluation and Service Base On Python Textblob*. https://doi.org/10.1088/1757-899X/1125/1/012034

Hanif, S., Radi, N., Hassan, S., Saleem, A., & Jamil, S. (2021). *Downloaded from : https://e-space.mmu.ac.uk/626042/ Version : Accepted Version Publisher : SAGE Publications Usage rights : Creative Commons : Attribution-Noncommercial-No Deriva- Sentiment analysis of tweets through Altmetrics : A machine learning approach*. *0052*, 712–726.

Machuca, C. R., Gallardo, C., & Toasa, R. M. (2021). *Twitter Sentiment Analysis on Coronavirus : Machine Learning Approach Twitter Sentiment Analysis on Coronavirus : Machine Learning Approach*. https://doi.org/10.1088/1742-6596/1828/1/012104

Model, I. L., Rasool, A., Jiang, Q., Qu, Q., & Ji, C. (2021). *WRS : A Novel Word-embedding Method for Real-time Sentiment with WRS : A Novel Word-embedding Method for Real-time Sentiment with Integrated LSTM-CNN Model*. *September*, 2–8. https://doi.org/10.1109/RCAR52367.2021.9517671

Nazeer, I., Rashid, M., & Kumar, A. (2020). *Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis*. *January 2021*. https://doi.org/10.4018/978-1-7998-4718-2.ch002

Priyadarshini, I., & Cotton, C. (2021). A novel LSTM – CNN – grid search ‑ based deep neural network for sentiment analysis. *The Journal of Supercomputing*, *0123456789*. https://doi.org/10.1007/s11227-021-03838-w

Raisa, F., Ulfat, M., Mueed, A. Al, & Reza, S. M. S. (2021). *A Review on Twitter Sentiment Analysis Approaches*. 27–28.

Shitole, A. (2021). *Twitter Sentiment Analysis Using Supervised Machine Learning*. *January*. https://doi.org/10.1007/978-981-15-9509-7

Wang, J., Yu, L., Lai, K. R., & Zhang, X. (2020). *Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis*. *28*, 581–591.