Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 3
Open Access

Prediction Of The Growing Stock In Stock Market On Analysis Of The Opinions Using Sentiment Lexicon Extraction And Deep Learning Architectures

[1]Ashfauk Ahamed A.K., [2]Manogar E., [*3]Jeevitha S., [4]Sujithra L. R., [5]Boopathi Kumar E.,
[1]Assistant Professor (Sr.Gr), Department of Computer Applications,
B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai.
[2,3]Assistant Professor, Department of Computer Applications,
B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai.
[4]Assistant Professor, Department of Artificial Intelligence and Data Science,
Sri Eshwar College of Engineering, Coimbatore, Tamilnadu.
[5]Department of Information Technology, Bharathiar University, Coimbatore, Tamilnadu.

Cite this paper as: Ashfauk Ahamed A.K.,Manogar E.,Jeevitha S.,Sujithra L. R., Boopathi Kumar E, (2024) Prediction Of The Growing Stock In Stock Market On Analysis Of The Opinions Using Sentiment Lexicon Extraction And Deep Learning Architectures. *Frontiers in Health Informatics, 13* (3), 1382-1392.

Abstract

Extraction of the sentiment in the social networks and stock exchange is a growing research in the area of data mining. Clustering technique considered to be most essential in discovering the significant stock information about the stock market with respect to various opinion of the traders in the online social network. Conventional clustering technique considered to be inadequate as multiple opinions of the traders is non actionable and sparse in nature. The extracting the sentiment is alone not feasible in the opinion mining as numerous information is related to stock forecasting and stock prediction analysis on the social network data extraction. Subspace clustering technique will be a suitable solution to challenges exhibited in the clustering on inclusion of domain knowledge and parameter sensitive information's. Further thresholding mechanism is capable of predicting the data Sensitiveness. In spite of numerous benefits of the subspace clustering approach, extracting the correct dimension found to be inconsistent and challenging issue. In order to manage those issues, hybrid deep learning architecture named as BERT-CNN model with incorporation of sentiment and domain knowledge has been proposed in this article. Initially the opinion of lexicon is extracted using BERT which is a natural language processing model and extracted feature is represented as word embedding vector. Those embedding vector processed using convolution layers with inbuilt kernel and filter constraints to eliminate the inconsistent dimension. Resultant feature is down sampled in the max pooling layers to obtain high level features as feature map. ReLU Activation Function is applied to fully connected layer to explore the growing stock features on processing the weight of the feature on the particular domain knowledge. Extracted actionable feature representing the growing stocks is clustered and refined periodically in its subspaces on utilizing the domain knowledge. An experimental result on the twitter benchmark dataset containing stock news proves that proposed architecture outperforms other conventional technique on performance metrics such as accuracy and efficiency.

Keywords: *Sentiment analysis, BERT Model, Convolution Neural Network, Lexicon, Prediction, Stock Analysis*

Introduction :

Sentiment analysis [1] in the online opinion of the user in the online social networks can facilitate investors in making investment decision and to analyse the stock companies' risk perception. It is identified that stock opinions posted in the social network has significant impact on stock markets. Investment analysis and prediction tools[2] is functioning on basis of the information explored from the historical and present financial statement of the industries and stock information posted on the stock forums and stock message boards on stock exchange.

In Stock Data analysis, Investors spending more amount of time to explore the information updated on the stock forums to compute the stock price based on the information of changing factors in index values on the stock exchanges[3]. Investment Recommendation system is formulated with respect to multiple financial information of stock issuers and investors on the stock exchange. Many conventional model has been examined related to sentiment analysis on the opinion of the investors posted on the social network applications and financial forums.

Investment information regards to the financial and business related discussion. However, those model has incorporated efficient opinion mining analysis utilizing the sentiment analysis mechanism in order to achieve effective investment suggestions. Hence it is necessary to obtain the detailed insight to increase the accuracy of stock forecasting using time series data. Further intent of the stock prediction as investment decision is established with different level of data volatility and evidence under different financial conditions. Insider information and rumour propagation in the social networks and business forum were need to be considered as significant as it misleads the investment decision.

Machine learning approaches[4] such as recommendation system implemented using principles of data classification and data clustering architectures which lead to more risk and failure of decision system. Hence to enhance the efficiency of recommendation system with risk analysis[5] in to stock exchange data and its various opinion by the investors in the social networks, deep learning architecture on incorporating the effective feature extraction technique has been developed as novel recommendation system using sentiment analysis to online opinion of the investor to various stocks in the stock exchanges on utilizing domain specific and context sensitive sentiment constraints to the model. Further methodology is designed with familiar sentiment analysis architecture for feature extraction using BERT model[6] and convolution neural Network[7[architecture for sentiment feature classification and prediction to stock exchange market.

The rest of the article is represented as follows, review of related literature are discussed in section 2, the proposed approach titled as deep sentiment learning architecture is represented in section 3. Implementation outcomes and performance outcome of the current architecture is illustrated in section 4 implementing twitter stock opinion on comparing its efficiency against conventional architectures with respect to accuracy measures has been highlighted. Finally article has been summarised with its achievements in section 5. Related Work

In this segment, Sentimental analysis approaches using deep learning models has been analysed in detail with respect to deep learning architectures along BERT model for feature extraction, selection and representations process. Further prediction performance of opinion mining is measured using similarity measures. Among those deep learning architecture which generates good outcomes with respect to accuracy on the assessment of the approach were illustrated in depth and few approaches which outcomes similar to the current technique is represented as follows

1.1. Improved Semantic Oriented Approach towards Sentiment Classification for Chinese Movie Reviews

In this literature, Sentimental classification aims at classifying the Chinese movie reviews into positive or negative reviews. Random forest technique considered as Machine learning architectures along semantic orientation approaches were employed for sentiment classification of the review. Initially part-of-speech tagging is used to determine phrases in the opinion that accumulative adjectives or adverbs and then semantic orientation is calculated using point mutual information algorithm [8]

1.2. Predicting opinion holder and topic on Extracting Opinions from the Online News Media

In this literature, we discuss a prediction technique for determining an opinion holder and opinion topic on extracting the opinion in the online new media. Detailed analysis of an model is to exploit the semantic information on the opinion sentence in terms of opinion holder and topic bearing the opinion. It is carried out on the phrase of the opinion containing the verb or adjective of the sentence. Prediction architectures employs semantic role labelling opinion holder and topic as an intermediate step to label to the data. Opinion analysis process containing three phases [9]is as follows

- Computing an opinion-bearing word
- Labelling semantic roles to the phrases in the opinion
- Finding the holder and the topic of the opinion word among the labelled semantic roles.

Proposed model

Design specification of proposed sentiment analysis approach with deep learning architecture for predicting the growing stock on analysis of the stock opinion from the online social network is provided in this section. Further hyper parameter tuning is included to the deep learning layers to enhance the performance of the predicting and forecasting the growing stocks from the investor opinions

1.3. Problem Definition

The important challenge of the opinion mining is to select the appropriate sentiment to the stock opinion of the investors using the rule- and learning-based approaches as those approaches relies on the data to more extent as it predict the sentiment on basis of the feature in the review. The following statements proves the problems clearly

- Lengths of the review is dynamic and it changes on basis of investors perception . Opinion of investor to stock vary from few words to thousands of words. Even Opinion might contains single word as opinion.
- Features of the opinion is diverse as web 2.0 platform provides equal platform to the entire user of the web and there is no regulation or norms on words used to information sharing or communication. Further different user express their opinion in multiple ways which leads to multiple expressions.
- Finally there are frequent emerging new Internet expressions. On basis of time changing, the same sentiment is expressed in multiple ways. In extreme cases, the similar word might carry a various sentiment polarity to online or offline events[10].

1.4. Data Pre-processing

The Analysis of Stock data for recommendation system is carried with extraction of the sentiment in the social networks and stock exchange is considered as open research in the data mining community. Prediction of growing stock on processing the opinion data of twitter data through deep learning architectures leads inconsistency as data contains more inadequate information's which considered as non actionable data. Only stock information which is considered to be sparse is considered for processing. However extraction the sentiment without pre-processing data will lead to misinterpretation.

In order to compute accurate sentiment to opinion, stop words removal and stemming process is carried out in the pre-processing to eliminate the inconsistent phrases and to incorporate the domain knowledge and parameter sensitive information for further processing of the data to predict the sentiment. Data Sensitiveness is computed through thresholding mechanism. Further identifying the correct phrases is challenging issue in subspace clustering [12].

Finally the preprocessing of the data along inconsistent information correlating to phrases has been managed on implementing singular value decomposition and normalization process. Principle Component Analysis is employed to identify or extract the actionable phrases in the opinion text. In addition ,utilization of the domain knowledge is to refine and validate the optimal phrases dynamically. Figure 1 represents the work flow representation of the current growing stock prediction architecture.

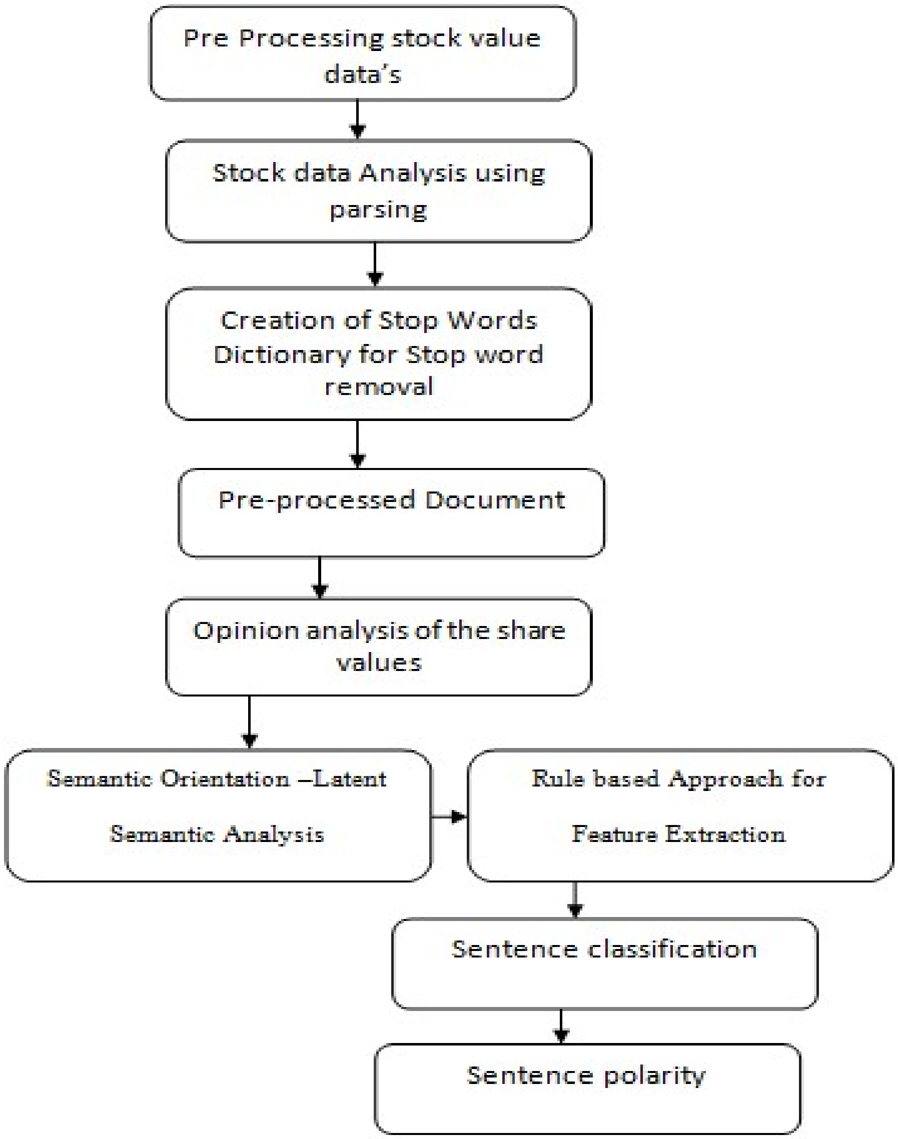

Figure 1: Architecture of the proposed Sentiment analysis

1.5. Opinion Analysis of the stock values

In this module, we obtain the stock review information from the online social network especially from twitter stock review containing various categories of user like investors, issuers brokers, companies etc. The opinion of the stock ensures the conclusions on personnel or public opinions. The sentiment polarity to those user category is easy to distinguish on true and meaningful contents whereas it is highly complex in distinguishing the rumours and insider information. Twitter stock opinion data is collected by the application programming interface or correspondent crawler with numerous categories of information. This data wrappers allows the collection of large accurate data on a emphasised topic [13].Further stock context is incorporated to obtain the polarity of the stock.

1.6. Data Pre-processing for Stop Word Removal and extraction of context information

In the opinion pre-processing, the following processes are incorporated: 1) text segmentation, 2) POS tagging and phrasing ,3) labelling of phrase, and 4) the replacement of synonymous on the phrase to avoid word ambiguity. In order eliminate the irrelevant data distributions and to enhance the accuracy of the prediction task, pre-processing is considered as vital task using stop removal, stemming and tokenization. However the "sentiment base" to the stock prediction application is carried out using term frequency and inverse document frequency computations.

$$\mathrm{Tf}_{u,\,wi} = \frac{TF'}{\sum \mathrm{wi} \,\xi\, \mathrm{W}\,(TF')} \quad \text{........Eq.1}$$

Where TF' is the term frequency value of phrases in opinion .
W is the set of phrases

- **Semantic Orientation**

Wordnet tool is employed to identify the Subjects and objects of the phrase to establish a context. Context mining is carried out on context generated to provide the domain specific information and context sensitiveness for the prediction task. Context mining helps to determine the semantic orientation of opinion. The semantic extraction of context contains the sentiment, modifier, and rule base.
Maximum likelihood of the phrase is computed as

$$\hat{P}(S_j) = \frac{N(S = S_j)}{N} \quad \text{....Eq.2}$$

1.7. Designing and incorporating the Sentimental analysis based on the data polarity

In order to generate the sentiment, modifier, object, and condition base with polarity of the data using BERT which is a rule based Approach as follows.

1.7.1. Semantic Orientation –BERT Analysis

BERT Model uses the following process to analyse the sentiment attached to the stock data and it provide valuable suggestion to future investors.

- ***Sentiment base:***

The sentiment analysis composed of two strongly correlated functions such as sentiment lexicon and its sentiment polarity. Lexicon is considered as subjectivity clues along its strength annotations(strong or weak) and polarity(i.e., positive, negative, or neutral) on the occurrence of phrases as verbs, adjectives, and nouns. However, the lexicon is capable to compute original and actual phrase polarity as it is subject to change due to its context in the review.

Numerous approaches which considering the context of phrase is implemented to compute phrase sentiment orientation. Assume that the sentiment polarity of a phrase is computed by its morphemes. Occurrence of phrase morphemes determines the sentiment such that phrase morphemes appears more frequently the positive lexicon, the phrase is positive else it is considered as negative. The important correlations among the sentiment phrases and its modifiers depend on their locations of the phrases and classes in the opinion [14].

1.8. Prediction of the Stock value using Convolution Neural Network

The significant opinion containing the thematic (stock growth) phrases on the most prominent position either in the beginning , middle or ending of the opinion for emphasis. Therefore, in computing the overall polarity of a opinion, the location of the sentiment phrase should be examined. In processing, the significance of a phrase to a opinion can be illustrated by the weight in the overall polarity computation. The weight of thematic (stock growth)phrase should be greater than those of other sentences in a opinion.

Convolution Neural Network [15] considered as deep learning architecture is implemented in this work for prediction of the growing stock in stock data analysis. Hyper parametric tuning of the architecture is carried out in the activation function to predict the stock growth. Further activation functions of each layers handles the features effectively. Convolution Neural Network architecture for stock prediction is as follows

- **Convolution Layer**

Convolution layer composed of kernel filter and stride to generate feature map to principle feature. Convolution layer computes optimal feature for feature map on latent opinions. Activation function of Convolution Neural Network computes the pair-wise features representation to investor opinion to the stocks on the investor matrix[16].

- **Max pooling layer**

In this pooling layer, Latent Feature processed in form of Subset is reduced further on operation of pair wise similarity estimation along various preference pairs of stock on its bias coefficients. Investor preference matrix is generated using matrix factorization to provide the weighted embedding features. Weighted embedding features is represented as high level features on the preference matrix. It contains crucial feature to determine the sentiment for the specific stock. The max pooling layer of the convolution neural network to generate weighted embedded feature for fully connected layer are follows

$$\mathbf{x} = \sum_{i=1}^{N} x_i v_i = x_1 v_1 + x_2 v_2 + \dots + x_N v_N \quad \text{....Eq.3}$$

The weighted feature representations which improves discrimination of the stock during the Pearson correlation similarity calculation is considered as investor preference to the stock. The prediction components of the Fully Connected Neural Network for hyper parameter tuning are given in table 1.

Table 1: Parameter setting of Proposed architecture

S. No	Hyper Parameter of Scheduling Component	Parameter Values
1	Batch Size of the embedding vector	118
2	Learning Rate	0.08
3	Size of Dimensions	25

4	Number of Epoch	30
5	Error function	Cross entropy

- **Fully Connected Layer**

The Fully Connected Layers utilizes the weighted embedding feature subset and discriminate the feature on softmax layer. Softmax layer acts classifier to classify the feature to the fully connected layer. Feature Flatten is a inherent mechanism which hierarchically extracts the abstract features and learn the discriminative features with few hyper parameters of fully connected network. Result of the feature space is latent distribution of feature of the evolving stock characteristics. In this layer, large evolving characteristics of the preference context have embedded to activation function to compute the value of the stock.

- **Activation Function**

The current architecture employs the rectified linear units (ReLU) as activation function[17] for growing stock forecasting and recommendation to evolving investor opinion to various stock. opinion is represented as the preference vector and pair wise relation is computed in fully connected layer to identify the growth of various stocks data. Activation function produces the appropriate stock on each class in the fully connected layer. weighted feature vector has been processed with parameterized values to produce the accurate prediction to the selected stock vector and to each epoch on parameter updating.

- **Output Layer**

The output layer of the convolution neural network contains the prediction result containing stock suggestion to investors. Further parametric flattening is applied to the prediction model using similarity measures for stock recommendation. Soft max optimization[18] is employed to the resultant set with cross entropy approach to access the effectiveness of the stock growth prediction.

- **Loss Layer.**

This layer is to evaluate the prediction accuracy on the model and it fine tuning with refine parameter of different layers of convolution neural network to reduce the reconstruction error along the feature in the classes. Further cross entropy loss function is implemented to manage prediction outcomes [19].

Algorithm 1: CNN learning for Stock Prediction using sentiment analysis
Input: Discriminative Stock data
Output: Growing Stock Prediction
Process
BERT sentiment analysis ()
Estimate latent feature Set F_s
Convolution Neural Network ()
Convolution Layer ()
Feature map generation
Max Pooling ()
Principle Stock characteristics
Fully Connected Layer ()
Activation Layer ()
ReLu Function

Output Layer()
Softmax() ---Recommendation of Growing

This model encourages the embedded latent feature on feature map to form growing stock list to recommendation.

2. Result and Discussion

The online Stock data is extracted from the Twitter stock data using API as data wrapper on the opinion of different issuer and investor to the stock containing 2.7lakh records. Those collected data is organized as dataset. Investor stock opinions extracted contains the information such as stock prize, company, opinion of the stocks etc. Initially these data has been pre-processed with removal of stop words and stemming process to transform it into the significant information[20].

2.1. Performance Evaluation

Performance of the model is evaluated against the following metrics

- **Precision** (also called positive predictive value) is the ratio of retrieved instances to the relevant instance in the stock classes.
- **Recall** (also known as sensitivity) is the ratio of relevant instances that are retrieved. Both precision and recall is measure of relevance.

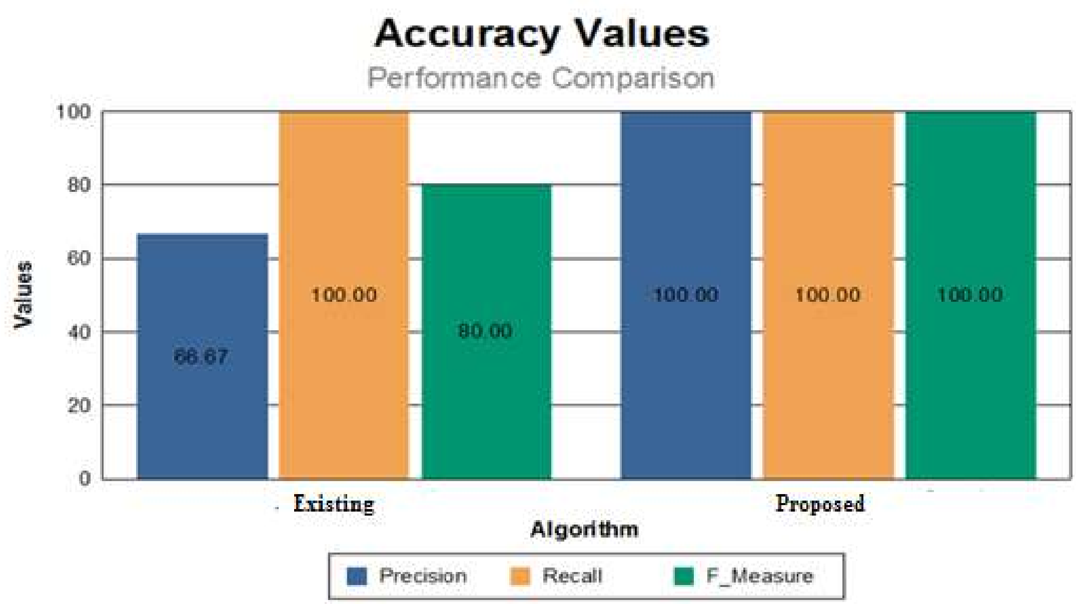

Figure 2: Performance evaluations of the Prediction results

Proposed model is evaluated on cross fold validation to measure the performance of sentiment analysis in predicting the growing stock with respect to the precision, recall and accuracy. The prediction accuracy is the degree of closeness of calculated outcomes to actual (true) value. Figure 2 illustrates the performance evaluation of the proposed architecture in terms of accuracy. Table 2 illustrates the performance evaluation of the sentiment analysis architectures.

Table 2: Performance Evaluation

S.No	Technique	Precision	Recall	Fmeasure
1	BERT+CNN –Proposed	100	100	100

2	Ensemble Scheme	66	100	80

As the proposed approach attains higher accuracy on compared with the conventional approach, hence it is recommend to compute growing stock on analysis of the opinion of the investor on social network and business forums.

Conclusion

The Sentiment analysis using deep learning approach has been carried out for growing stock data analysis in this work. Convolution Neural Network with BERT model which considered as rule based approach for sentimental analysis is designed and implemented. Investor opinion on the Stock market data is collected as dataset from twitter applications and its architecture is modelled as Stock prediction analytics to the new investors. The proposed architecture is modelled to analyse the opinion of the stock data to the identify the polarity and predict the value of the stock as outcome. Further opinion of the investor about stock data is processed using latent semantic analysis algorithm as it provide the latent information of the stock in terms of scaling rate and accuracy. Sentiment analysis calculate the feature polarity on the scores as Positive, Negative and neutral. Finally polarity of the sentiment is provided to fully connected layer of the deep learning model to provide the stock value as growing stock and diminishing stock. Experimental analysis and performance analysis proves that proposed architecture outperforms the conventional architectures.

References

[1] H. Li, L. Dombrowski, and E. Brady, "Working toward empowering a community: How immigrant-focused nonprofit organizations use twitter during political conflicts," in Proceedings of the 2018 ACM Conference on Supporting Groupwork. ACM, 2018, pp. 335–346.

[2] D. Tang, B. Qin, F. Wei, L. Dong, T. Liu, and M. Zhou, "A joint segmentation and classification framework for sentence level sentiment classification," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 11, pp. 1750–1761, 2015.

[3] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," in Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, 2012, pp. 115–120.

[4] D. Paul, F. Li, M. K. Teja, X. Yu, and R. Frost, "Compass: Spatio temporal sentiment analysis of us election what twitter says!" in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017, pp. 1585–1594.

[5] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "Annotate-sampleaverage (asa): A new distant supervision approach for twitter sentiment analysis," in 22nd European Conference on Artificial Intelligence (ECAI), vol. 285. IOS Press, 2016, pp. 498–506.

[6] B. Pang, L. Lee et al., "Opinion mining and sentiment analysis," Foundations and Trends R in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[7] K.-L. Liu, W.-J. Li, and M. Guo, "Emoticon smoothed language models for twitter sentiment analysis," in Proceedings of the Twenty- Sixth AAAI Conference on Artificial Intelligence, 2012.

[8] J. Zhao, L. Dong, J.Wu, and K. Xu, "Moodlens: an emoticon-based sentiment analysis system for chinese

tweets," in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012, pp. 1528–1531.

[9] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, vol. 1, no. 12, 2009.

[10] D.-T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features." in IJCAI, 2015, pp. 1347– 1353.

[11] E. Cambria, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.

[12] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 1555–1565.

[13] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813–830, 2016.

[14] J. Zhao and X. Gui, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.

[15] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," Expert Systems with Applications, 2018 .

[16] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in twitter events," Journal of the Association for Information Science and Technology,vol. 62, no. 2, pp. 406–418, 2011.

[17] N. Du, Y. Liang, M. Balcan, and L. Song, "Influence function learning in information diffusion networks," in International Conference on Machine Learning, 2014, pp. 2016–2024.

[18] M. Tsytsarau, T. Palpanas, and M. Castellanos, "Dynamics of news events and social media reaction," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014, pp. 901–910.

[19] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social mediasentiment of microblogs and sharing behavior," Journal of Management Information Systems, vol. 29, no. 4, pp. 217–248, 2013.

[20] Y. Fu, Y. Ge, Y. Zheng, Z. Yao, Y. Liu, H. Xiong, and J. Yuan, "Sparse real estate ranking with online user reviews and offline moving behaviors," in 2014 IEEE International Conference on Data Mining (ICDM). IEEE, 2014, pp. 120–129.

[21] Navatha, S., et al. "Multitask Learning Architecture For Vehicle Over Speed As Traffic Violations Detection And Automated Safety Violation Fine Ticketing Using Convolution Neural Network And Yolo V4 Techniques." Chinese Journal of Computational Mechanics 5 (2023): 431-435.

[22] Prakash, G., P. Logapriya, and A. Sowmiya. "Smart Parking System Using Arduino and Sensors." NATURALISTA CAMPANO 28.1 (2024): 2903-2911.

[23] Kumar, E. Boopathi, et al. "Predicting the Fake Review of Products Using Graph Recurrent Neural Network." International Journal of Interdisciplinary Organizational Studies 18.1 (2023).

[24] Reddy, C. S., Yookesh, T. L., & Kumar, E. B. (2022). A Study On Convergence Analysis Of Runge-Kutta Fehlberg Method To Solve Fuzzy Delay Differential Equations. Journal of Algebraic Statistics, 13(2), 2832-2838.