IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Generative Process Model And Machine Learning For Sentiment Analysis

Ravi Prasad Ravuri
Application Developer, Sriven Technologies, Ashburn, VA, USA

Abstract

Sentiment analysis from text documents has its impact on different real world applications. Latent Diritchlet Allocation (LDA) is widely used generative process model for processing text documents. However, as one size done not fit all, it is important to improve it towards improving performance in sentiment analysis. In this paper we proposed a generative process model with underlying machine learning (ML) for sentiment analysis. Our framework is known as Enhanced LDA based Sentiment Analysis Framework (ELDA-SAF). It makes use of our enhanced LDA model and also a ML classifier known as Support Vector Machine (SVM) for sentiment classification. We proposed an algorithm known Learning based Sentiment Analysis (LbSA) for realizing our framework. Game dataset consisting of comments on different food ball players is used for our empirical study. Experimental results revealed that our algorithm LbSA outperforms existing methods.

Keywords *– Sentiment analysis, machine learning, LDA, enhanced LDA, feature engineering, generative process model*

1. INTRODUCTION

Processing text documents is an important research area. Particularly, the reviews available on products and services including players of games are widely used for sentiment analysis. Even in social media also public opinion is given highest importance by organizations. As explored in [1], [2] and [3], sentiment analysis has plenty of real world applications in different domains. In essence, those applications derive public opinion that provides knowledge required to make well informed decisions. There are many real world applications that are benefited from sentiment classification which is very useful tool for improving quality of products and services.

Many contributions are found in the literature. A generative process model is exploited in [3] for dealing with text documents in order to generate categories. Product reviews are considered for processing in [6] towards discovering sentiments from the data. This could help in ascertaining what is the thinking of customers on specific product. It is aspect based approach. In [7] also GAN models are employed for

analyzing sentiments besides improving data quality with augmentation process. Cross domain data is used in [14] for summarization of abstracts. They used reinforcement learning in order to have the required discriminative power. In [15] there is a novel research effort that resulted in approval network model. This model is used to analyze sentiments over social media data. From the literature review, it is understood that generative process models are widely used for processing text documents. However, as one size does not fit all, we improved LDA model and used it along with ML model towards efficient classification of sentiments. Our contributions in this paper are as follows.

1. We proposed a generative process model with underlying machine learning (ML) for sentiment analysis.
2. Our framework is known as Enhanced LDA based Sentiment Analysis Framework (ELDA-SAF). It makes use of our enhanced LDA model and also a ML classifier known as Support Vector Machine (SVM) for sentiment classification.
3. We proposed an algorithm known Learning based Sentiment Analysis (LbSA) for realizing our framework.
4. We built an application to evaluate LbSA and compared its results with existing methods.

The remainder of the paper is structured as follows. Section 2 reviews literature on different methods that deal with sentiment analysis. Section 3 throws light on our enhanced LDA based method. Section 4 presents sentiment classification results and evaluation. Section 5 concludes the paper and provides scope for the future work.

2. RELATED WORK

This section provides review on state of the art in the field of sentiment analysis. Generative adversarial network based methodology is proposed in [1] for finding public opinions as part of their empirical study. Topic level approach is used in [2] based on social media data for sentiment analysis. Their approach is based on deep learning which exploits the advanced neural networks available. A generative process model is exploited in [3] for dealing with text documents in order to generate categories. Modeling topics in the given documents is explored in [4]. It is designed to improve sentiment prediction performance. It is observed as the research specified in [5] that generative process models are widely used for processing text documents and particularly opinion mining applications. Product reviews are considered for processing in [6] towards discovering sentiments from the data. This could help in ascertaining what is the thinking of customers on specific product. It is aspect based approach. In [7] also GAN models are employed for analyzing sentiments besides improving data quality with augmentation process.

Deep learning approaches are discussed and used for analyzing sentiments in [8]. Their research revealed the utility of deep learning models. Dealing with short texts is the research carried out in [9]. They proposed a topic model to deal with product reviews in this regard. The work found in [10] and [12] is similar to that of [9] with respect to sentiment classification tasks. A hybrid approach that is based on unsupervised classification and topic modeling approach is used in [11]. Ensemble learning approach on social media data

is used along with deep learning models in [13] to achieve efficient opinion mining. Cross domain data is used in [14] for summarization of abstracts. They used reinforcement learning in order to have the required discriminative power. In [15] there is a novel research effort that resulted in approval network model. This model is used to analyze sentiments over social media data. Other important research contributions include deep learning [16], LDA [17], fine-grained approach [18], big data approach [19] and bi-lingual hybrid approach [20]. From the literature review, it is understood that generative process models are widely used for processing text documents. However, as one size does not fit all, we improved LDA model and used it along with ML model towards efficient classification of sentiments.

3. PROPOSED FRAMEWORK

We proposed a novel model based on an improved form of topic model based on LDA which is widely used for processing text corpora. we proposed a generative process model with underlying machine learning (ML) for sentiment analysis. Our framework is known as Enhanced LDA based Sentiment Analysis Framework (ELDA-SAF). It makes use of our enhanced LDA model and also a ML classifier known as Support Vector Machine (SVM) for sentiment classification. We proposed an algorithm known Learning based Sentiment Analysis(LbSA) for realizing our framework. Game dataset consisting of comments on different food ball players is used for our empirical study. Figure 1 shows our framework.

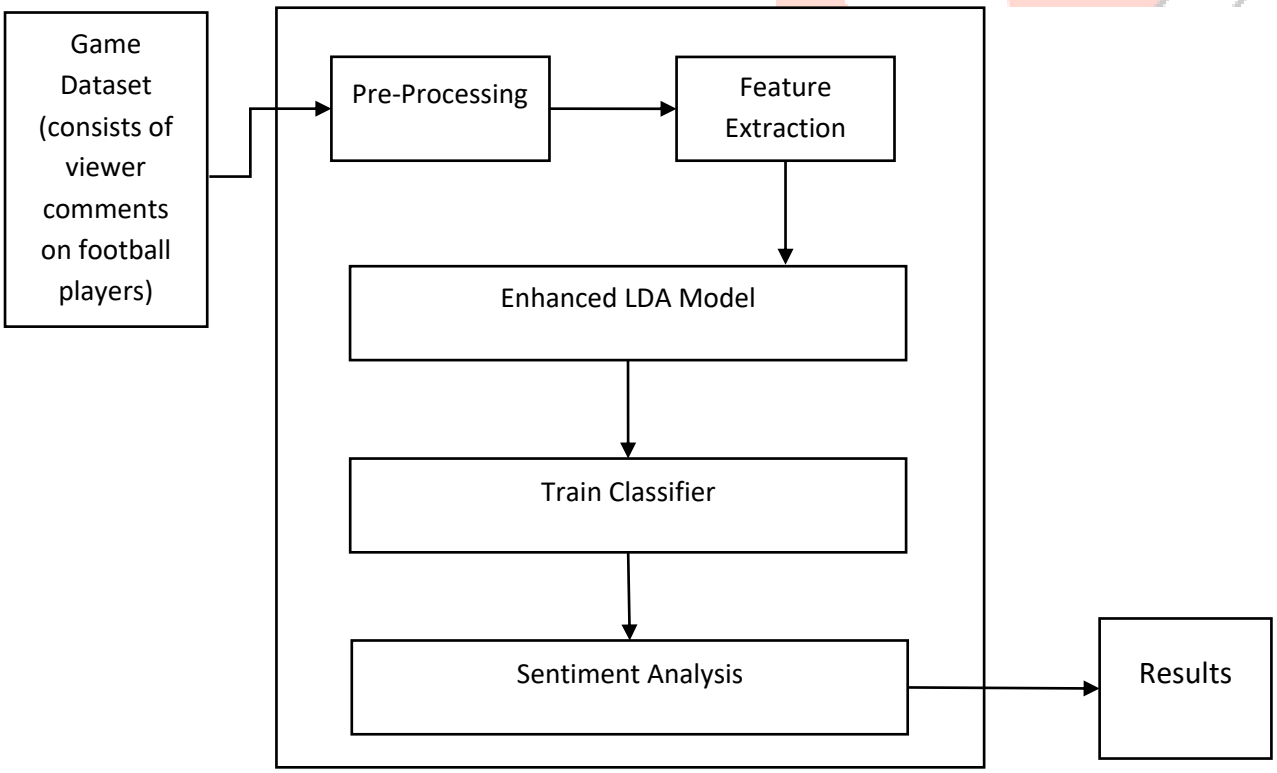

Figure 1: Proposed framework named Enhanced LDA based Sentiment Analysis Framework (ELDA-SAF)

As presented in Figure 1, Game

Dataset (consists of viewer comments on football players) is used as input to the framework. It has not only user comments but also ID for the document and also the ground truth. The data is in textual format and needs preprocess such as removal of stop words and tokenization. Then feature is extracted from the data and the enhanced LDA model shown in Figure 2 is employed to have final content to be used to train a classifier. In other words, our approach is a hybrid in nature which has both enhanced LDA and also ML. with enhanced LDA, training quality of data is enhanced leading to better performance in sentiment analysis.

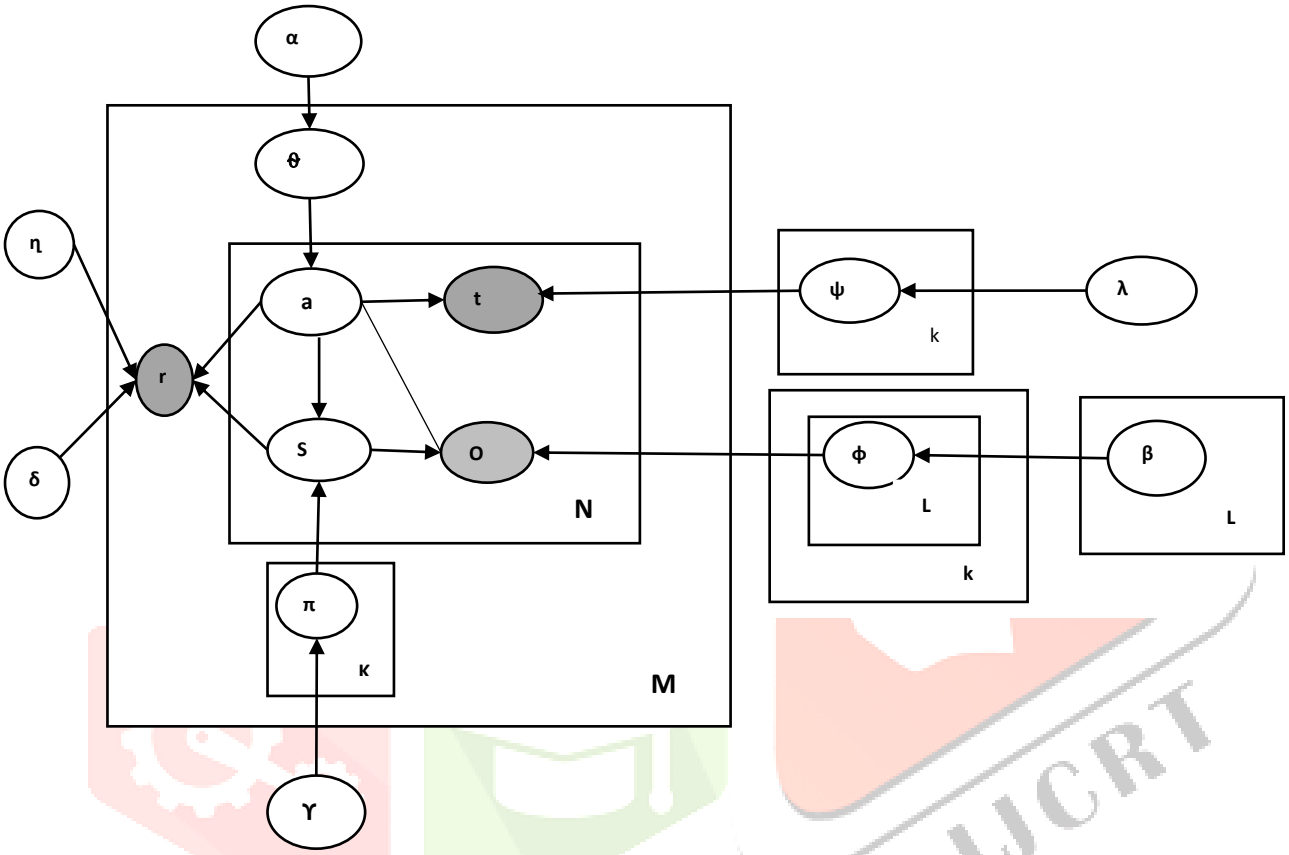

Figure 2: Enhanced LDA model

As presented in Figure 2, in the proposed enhanced LDA model, each place is represented as a box. The outer one indicates processing of text documents (each instance in the dataset is a text document) and the inner one indicates the underlying process. Total documents in the given dataset is denoted as M. N denotes the opinion pairs associated with the documents. Sentiment aspects are denoted by K while L refers to actual sentiments. The hidden sentiment aspects in the textual documents is computed as in Eq. 1.

$$\bar{z}_m = \frac{1}{C}\sum_{n=1}^{N}(a_{mn} \times (\omega^T \times s_{mn}))(1)$$

Associated with the sentiment analysis, there is need for generation of response variable from given distribution is computed as in Eq. 2.

$$p(r|\rho,\delta) = h(r,\delta)\exp\{\frac{\rho r - A(\rho)}{\delta}\}(2)$$

The aspect distribution in each document is computed as in Eq. 3.

$$\theta_{m,k} = \frac{N_{m,k}+\alpha_k}{N+\sum_{k`=1}^{K}\alpha_{k`}}. \quad (3)$$

The distribution of aspect word in the given document is computed as in Eq. 4.

$$\psi_{k,u} = \frac{N_{k,u}+\lambda}{N_k+|U|\lambda}. \quad (4)$$

Afterwards, the opinion word distribution is expressed as in Eq. 5.

$$\phi_{klv} = \frac{N_{k,l,v}+\beta_{l,v}}{N_{k,l}+\sum_{v`=1}^{|V|}\beta_{l,v`}}. \quad (5)$$

Dirichlet priors are appropriately used in the enhanced LDA model in order to have an iterative approach in processing text documents that will help in training data quality improvement. Then such data is fed to SVM to learn from data.

Algorithm1: Learning based Sentiment Analysis (LbSA)

Input: Dataset D

Output: Sentiment analysis results R

1. Begin
2. D'←PreProcess(D)
3. F←FeatureExtraction(D')
4. topicModel←EnhnacedLDA(D', F)
5. Train SVM with topicModel
6. For each test instance in T
7. r←Predict(SVM model)
8. add r to R
9. End For
10. Return R
11. End

As presented in Algorithm 1, it takes dataset as input and produces sentiment prediction results. It has provision for pre-processing data to improve its quality. Then it makes use of feature extraction method prior to the application of enhanced LDA model proposed in this paper. Afterwards, SVM is trained to gain knowledge with improved quality in training. Thus the knowledge model is well equipped with sentiment classification knowledge. Finally, the algorithm results in sentiment analysis that are further used for knowing performance statistics.

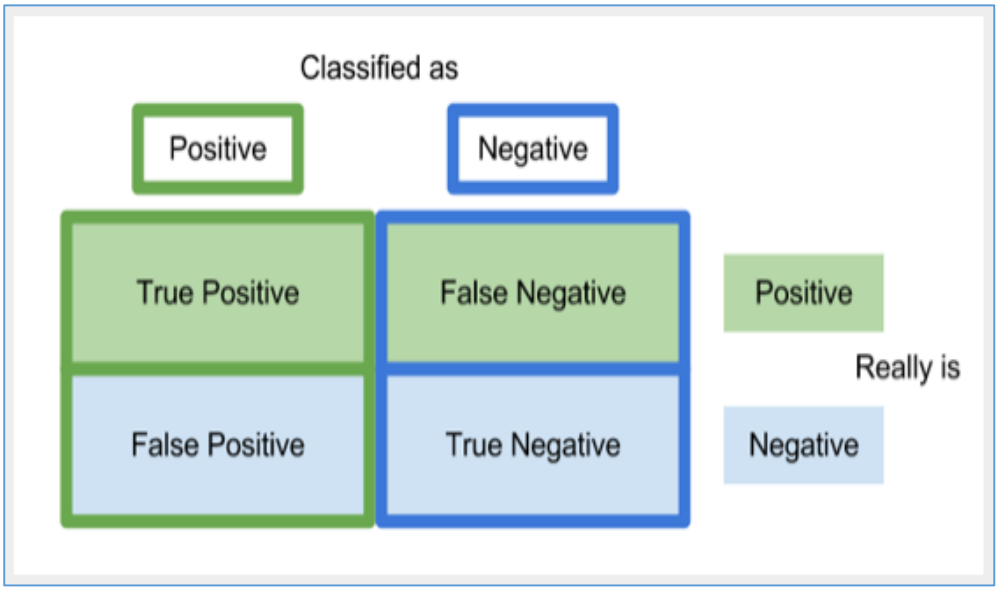

Figure 3: Illustrates confusion matrix

Afterwards, the resultant model is used to classify sentiments. Instead of using SVM directly, the proposed model exploits SVM after the enhanced LDA provides its quality inputs to SVM. Confusion matrix shown in Figure 3 is used to evaluate performance of our method.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Accuracy is one of the metrics obtained from experimental results in terms of confusion matrix. It reveals how accurate the proposed algorithm is dealing with sentiment analysis.

5. EXPERIMENTAL RESULTS

This section presents results of our experiments in terms of sentiment analysis and also accuracy of the proposed method compared with many existing methods. In the process of enhanced LDA different aspect numbers are associated with the sentiment analysis.

26306	positive	May I ask what\u2019s so cool about A$ap Rocky?
26408	positive	Oooh yha and the DMC and A$AP ROCKY one is 2-4th
10570	negative	Merril Hoge is stupid. I'm sick of hearing what Tebow can't do. He may not be a #1 right now but it took Aaron Rodgers 3 yrs to start
10255	negative	RT @MNVikingsGuy: Is it possible for Aaron Rodgers to suffer a career ending injury tonight? Just curious.
10004	neutral	"Aaron Rodgers jersey battery room, under the windmill. Wagnoli - 4 August 2011 - Blog - designer wedding dresses sale,"
100000	neutral	"And on the very first play of the night, Aaron Rodgers is INT'd by UDFA CB Brandian Ross, who returns it for a pick-six touchdown."

Table 1: An excerpt from training data from football game dataset containing user comments on players

As presented in Table 1, an excerpt from training data from football game dataset containing user comments on players. It has an ID, ground truth sentiment and actual comment on the player.

S.no	Id	Actual Sentiment	Predicted Sentiment
1	26418	positive	positive
2	26424	negative	negative
3	26410	positive	positive
4	26416	neutral	neutral
5	26319	neutral	neutral
6	26365	positive	positive
7	26412	negative	negative
8	26235	negative	negative
9	26386	neutral	neutral
10	26417	positive	positive

Table 2: An excerpt from the results of the proposed algorithm towards sentiment analysis on the given test data

As presented in Table 2, the experimental results revealed that the ground truth and prediction result are matching. It shows the ID of the instance in the test dataset, the actual sentiment of the user comment on player based on the ground truth and predicted sentiment.

Sentiment Analysis Method	Accuracy (%)							
	Number 5	Number 10	Number 15	Number 20	Number 25	Number 30	Number 35	Number 40
ASUM	0.7638	0.7839	0.77385	0.7437	0.7437	0.73365	0.73365	0.73365
JST	0.7839	0.7839	0.804	0.7839	0.77385	0.79395	0.77385	0.77385
Lexicon	0.6633	0.6633	0.6633	0.6633	0.6633	0.6633	0.6633	0.6633
Pooling	0.8241	0.8241	0.8241	0.8241	0.8241	0.8241	0.8241	0.8241
SJASM	0.8844	0.86731	0.86530	0.86631	0.8643	0.8643	0.85425	0.85425
SLDA	0.8241	0.8442	0.83415	0.8241	0.8241	0.81405	0.79395	0.79395
SVM	0.804	0.804	0.804	0.804	0.804	0.804	0.804	0.804
Proposed	0.9447	0.9246	0.91455	0.89445	0.89445	0.89445	0.87435	0.87435

Table 3: Sentiment analysis results

As presented in Table 3, the experimental results are provided for different existing methods and also the proposed method in terms of overall accuracy in sentiment prediction.

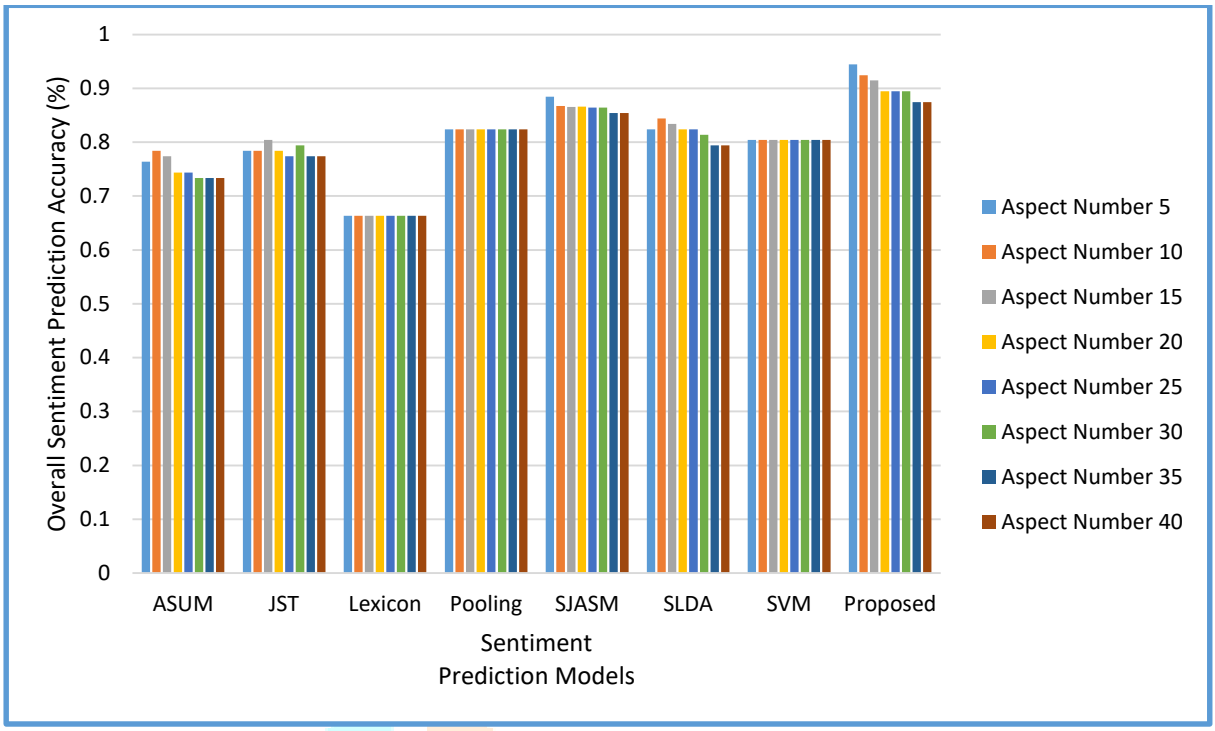

Figure 3: Performance comparison

As presented in Figure 3, there are many existing methods used for sentiment analysis along with the proposed method. Since the proposed method is based on enhanced LDA model along with ML model like SVM, it outperforms other methods due to its generative process model and also strong pre-processing approach. When aspect number 5 is considered, the ASUM model exhibited 76.38% accuracy, JST 78.39%, Lexicon approach 66.33%, Pooling method 82.41%, SJASM 88.44%, SLDA 82.41% and SVM 80.40% accuracy. The proposed method outperformed all existing methods with 94.47% accuracy.

6. CONCLUSION AND FUTURE WORK

In this paper we proposed a generative process model with underlying machine learning (ML) for sentiment analysis. Our framework is known as Enhanced LDA based Sentiment Analysis Framework (ELDA-SAF). It makes use of our enhanced LDA model and also a ML classifier known as Support Vector Machine (SVM) for sentiment classification. We proposed an algorithm known Learning based Sentiment Analysis (LbSA) for realizing our framework. Game dataset consisting of comments on different food ball players is used for our empirical study. Experimental results revealed that our algorithm LbSA outperforms existing methods. Our method has up to 94.47% accuracy. In future, we intend to improve our method using deep learning approaches along without enhanced LDA model.

References

[1] Haihong, E.; Yingxi, Hu; Haipeng, Peng; Wen, Zhao; Siqi, Xiao and Peiqing, Niu (2019). Theme and sentiment analysis model of public opinion dissemination based on generative adversarial network. Chaos, Solitons & Fractals, 121, 160–167. http://doi:10.1016/j.chaos.2018.11.036.

[2] Ajeet Ram Pathak; Manjusha Pandey and Siddharth Rautaray; (2021). Topic-level sentiment analysis of social media data using deep learning . Applied Soft Computing. http://doi:10.1016/j.asoc.2021.107440.

[3] Li, Yang; Pan, Quan; Wang, Suhang; Yang, Tao and Cambria, Erik (2018). A Generative Model for Category Text Generation. Information Sciences, S0020025518302366–. http://doi:10.1016/j.ins.2018.03.050.

[4] Vamshi Krishna. B, Dr. Ajeet Kumar Pandey and Dr. Siva Kumar A. P. (2018). Topic Model Based Opinion Mining and Sentiment Analysis. IEEE, pp.1-4.

[5] Cambria, Erik; Das, Dipankar; Bandyopadhyay, Sivaji; Feraco, Antonio (2017). A Practical Guide to Sentiment Analysis Volume 5 || Generative Models for Sentiment Analysis and Opinion Mining. , 10.1007/978-3-319-55394-8(Chapter 6), 107–134. http://doi:10.1007/978-3-319-55394-8_6.

[6] Amplayo, Reinald Kim; Lee, Seanie and Song, Min (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. Information Sciences, 454-455, 200–215. http://doi:10.1016/j.ins.2018.04.079.

[7] Gupta and Rahul (2019). ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - Data Augmentation for Low Resource Sentiment Analysis Using Generative Adversarial Networks, 7380–7384. http://doi:10.1109/ICASSP.2019.8682544.

[8] Habimana, Olivier; Li, Yuhua; Li, Ruixuan; Gu, Xiwu and Yu, Ge (2020). Sentiment analysis using deep learning approaches: an overview. Science China Information Sciences, 63(1), 111102. http://doi:10.1007/s11432-018-9941-6.

[9] Xiong, Shufeng; Wang, Kuiyi; Ji, Donghong and Wang, Bingkun (2018). A Short Text Sentiment-Topic Model for Product Reviews. Neurocomputing, S0925231218301693. http://doi:10.1016/j.neucom.2018.02.034.

[10] Ashima Yadav and Dinesh Kumar Vishwakarma. (2019). Sentiment analysis using deep learning architectures: a review. Springer, pp.1-51. https://doi.org/10.1007/s10462-019-09794-5

[11] Blair, Stuart J.; Bi, Yaxin and Mulvenna, Maurice D. (2017). IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI) - Unsupervised Sentiment Classification: A Hybrid Sentiment-Topic Model Approach, 453–460. http://doi:10.1109/ICTAI.2017.00076.

[12] Haoyue Liu; Ishani Chatterjee; MengChu Zhou; Xiaoyu Sean Lu and Abdullah Abusorrah; (2020). Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods . IEEE Transactions on Computational Social Systems. http://doi:10.1109/TCSS.2020.3033302.

[13] Araque, Oscar; Corcuera-Platas, Ignacio; Sánchez-Rada, J. Fernando and Iglesias, Carlos A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77, 236–246. http://doi:10.1016/j.eswa.2017.02.002.

[14] Yang, Min; Qu, Qiang; Shen, Ying; Lei, Kai and Zhu, Jia (2018). Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning. Neural Computing and Applications. http://doi:10.1007/s00521-018-3825-2.

[15] Fersini, E.; Pozzi, F. A. and Messina, E. (2017). Approval network: a novel approach for sentiment analysis in social networks. World Wide Web, 20(4), 831–854. http://doi:10.1007/s11280-016-0419-8.

[16] Iqbal H. Sarker; (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions . SN Computer Science. http://doi:10.1007/s42979-021-00815-1.

[17] (2020). Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter . 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS). http://doi:10.1109/icimcis51567.2020.9354320.

[18] Kumar, Akshi; Srinivasan, Kathiravan; Cheng, Wen-Huang and Zomaya, Albert 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–. http://doi:10.1016/j.ipm.2019.102141.

[19] Sohangir, Sahar; Wang, Dingding; Pomeranets, Anna and Khoshgoftaar, Taghi M. (2018). Big Data: Deep Learning for financial sentiment analysis. Journal of Big Data, 5(1), 3–. http://doi:10.1186/s40537-017-0111-6.

[20] Liu, Guolong; Xu, Xiaofei; Deng, Bailong; Chen, Siding and Li, Li (2016). 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) - A hybrid method for bilingual text sentiment classification based on deep learning. 93–98. http://doi:10.1109/SNPD.2016.7515884.