ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse171942020.pdf
https://doi.org/10.30534/ijatcse/2020/171942020

Application of Multilayer Perceptron for Digital Society Sentiment Analysis towards Indonesia Biggest E-commerce Platforms

Kevin Kelly Isyanta[1], Julio Christian Young[2], Nunik Afriliana[3]
[1]Universitas Multimedia Nusantara, Indonesia, kevin.isyanta@student.umn.ac.id
[2]Universitas Multimedia Nusantara, Indonesia,julio.christian@umn.ac.id
[3]Universitas Multimedia Nusantara, Indonesia, nunik@umn.ac.id

ABSTRACT

Indonesia is one from many countries in the world which has the highest growth of e-commerce platforms. By the mean of dynamic website technology, through their social media accounts, digital societies often express their opinion on e-commerce platforms. Oftentimes, on a large scale, capital leading e-commerce companies try to evaluate sentiment polarity of their new promotion or sale strategy through such data. However, as a copious amount of opinions from e-commerce's various social media accounts, the manual labelling process might devour a large number of resources and prone to human errors. For tackling the aforementioned problem, previous researches show a promising result by encoding such problem into a machine learning classification task. Henceforth, in this research, we try to utilize multilayer perceptron (MLP), as one of machine learning classification algorithm, for tackling such problem and analyze the performance of it. In this research, we also present the effect of a comparatively new word embedding model, fastText and aTF–IDF as a well-known feature extraction method for MLP method. Our experiment shows that the developed MLP model able to achieves an accuracy score of 89.24% and F1-score of 89%.

Keywords: e-commerce, fastText, multilayer perceptron, sentiment analysis.

1. INTRODUCTION

By definition, e-commerce is a process related to all business transactions (the selling and purchasing process of products) with the help of computer technology. In contrast to the traditional commerce process, e-commerce has several advantages such as reducing overall business transaction costs and time [5,24]. In 2018, Indonesia has over than 100 million e-commerce transactions which have an estimated value of USD 20 billion. Aforementioned growth is mainly generated by some of the largest e-commerce platforms in Indonesia, for instance, Tokopedia, Shopee, Bukalapak, etc.

Along with e-commerce transactions growth, each individual in the digital society starts to develop their preference in an e-commerce platform over the rest. Oftentimes, users of an e-commerce platform express their online transactions' experience through their social media posts or comments. From a business perspective, mining some useful information such as users' sentiment toward your e-commerce platform from their social media accounts might be beneficial for the company's growth [7, 25]. However, as the copious amount of data from e-commerce's various social media accounts, the manual sentiment analysis process by a human might devour a huge amount of resources and prone to human errors.

For tackling the aforementioned problem, previous researches show promising results by encoding such problem into a natural language processing (NLP) task, viz. sentiment analysis. A sentiment analysis process can be represented as a binary classification problem and solved by the help of some machine learning techniques [23]. Machine learning is needed to help classify opinion [10]. In this research, we will be focusing on evaluating one of machine learning techniques, known as multi-layer perceptron (MLP) for solving the aforementioned problems. Moreover, as different feature extraction techniques might affect classification methods' performance, in this research we will be also focusing on using fastText as feature extraction technique and comparing its use to other well-known technique, term frequency – inverse document frequency (TF–IDF).

2. RELATED STUDIES

In the field of NLP, sentiment analysis referring to all automated processes that extracting the subjective information in the unstructured text dataset such as social media data streams, emails, etc. [2]. In its simplest form, the sentiment analysis process can be carried out as a binary classification task, where each given text will be classified into a positive or negative class exclusively [10] as in [21, 22].

For analyzing sentiment towards a brand or products, nowadays, corporate marketers often use sentiment analysis tools over various social media platforms (Facebook, Instagram, Twitter, etc) [8]. As a social media platform, through its microblogging features, Twitter often used by a digital society in their daily lives for expressing their opinions or sentiments towards something [12]. For e-commerce platforms, Twitter is an engagement goldmine, where they can analyze users' experiences and sentiments toward their services. Therefore, in our research, we choose to use user-generated texts (tweets) which its sentiment has been manually labelled in [15].

In the previous research [6], a performance comparison of Naïve Bayes (NB), Support Vector Machines (SVM) and Multilayer Perceptron (MLP) for multi-domain sentiment analysis dataset [16] shows that MLP has outperformed the other two algorithms with slightly better accuracy. For two different groups (movies and products sentiments) in the dataset, MLP has achieved maximum accuracy of 0.8160 and 0.7947, respectively. On the other hand, for the same groups, NB has only achieved a maximum accuracy of 0.7935 and 0.6917, while SVM has only a maximum accuracy of 0.811 and 0.7940.

As researchers in [6] used n-gram based feature extractor, another research state that a representation learning technique, such as fastText, can produce a more representative features compared to the n-gram method, thus lead to a better classification performance when used in a classifier [17]. Therefore, in this research, we will use a representation learning technique as a feature extractor. Moreover, in the research which comparing various representation learning techniques (Word2Vec, fastText and GloVe) for twitter sentiment classification using various classification algorithms (Gaussian NB, Linear SVM, etc.), the results show that for the most classification algorithms, fastText yields a better classification performance [4].

3. FUNDAMENTALS

3.1 Word Embedding

As machine learning models take vectors (array of numbers) as input,word embedding is one of preprocessing techniques for representing a word into a vector [14]. Beside of word embedding, there are other two conventional language model used for representing text as a vector, namely, one-hot encodings and represent each token (word) in the corpus (collection of text inputs) as a number. Unlike the conventional methods which map each unique token in the corpus manually, word embeddings work by trying to capture words representation by their relationship with one another. Hence, word embedding techniques offermore efficient and dense representation compared to the other conventional methods.

3.2fastText

fastText is one of popular library for word embeddings learning process which released by Facebook [13]. The learning of word embeddings technique used in fastText is an improvement from an earlier technique known as word2vec. Unlike word2vec and GloVe, when fastText does word embeddings learning process, fastText break down each input token into a bag of n-grams of character. Using the aforementioned principle, fastText has the ability to map a word outside of the model's training corpus. As fastText is able to produce dense vector representation for each given word, some text processing techniques (stopwords removal, stemming and lemmatization) might be no longer needed for producing a word (or text) vectors. As a consequence, when a classifier uses fastText as a feature extractor, it might gain a better performance score and reduce the preprocessing time needed (particularly for text dataset with morphological rich language characteristics). As the stemming process reducing the input word into its root form, the stemming process might remove the real meaning or mislead the context of the input word [9].

3.3Term Frequency – Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical technique to signifies the importance of each token within a sentence based on a corpus (collection of sentences). TF-IDF for a word/ token (t) in a sentence (d) with respect to a collection of sentences/ corpus (D) is given in the equation below,

$$tf(t, s, Doc) = tf(t, s) \cdot idf(t, Doc)$$

where $tf(t, s)$is number of token t appears in a sentence s, and $idf(t, Doc)$is given by,

$$idf(t, Doc) = \log \frac{1 + n}{tf(t, Doc)} + 1$$

3.3Multilayer Perceptron

Multilayer Perceptron (MLP) is avariety of feedforward artificial neural network (ANN). An MLP consists of an input layer, a single or stack of hidden layer(s) and an output layer [3] where each layer consist of a collection of neurons. Beside of the input layer, each neuron within an MLP layer uses a nonlinear activation function to perform a non-linearity mapping from an input to its corresponding output [6]. Figure 1 shows an example of a multilayer perceptron architecture.

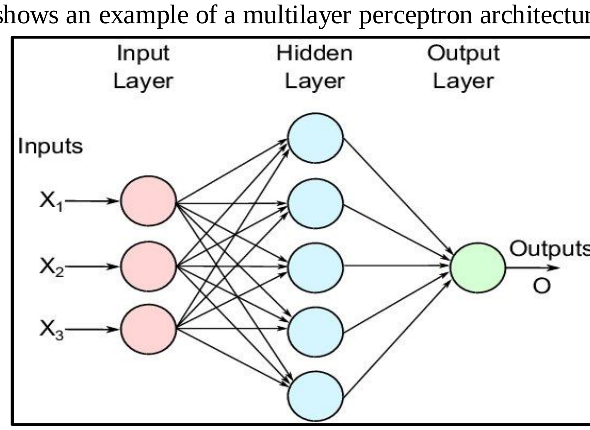

Figure 1: An example of MLP model with 3 input neurons, one hidden layer with 5 neurons and 1 output neurons.

The learning process of an MLP refers to the adjustment process of each neurons' connection weights, based on the cumulative errors of expected outputs and yielded outputs [6]. Moreover, the learning process of an MLP can be seen in the following section.

1. In an MLP model, the inputs for a layer is pushed forward to the next layer by taking the dot product of each input with the weights that exist between it and its preceding layer.
2. Moreover, as each neuron in the preceding layer takes the calculated output as an input, this input will be pushed through an activation function to decide whether a neuron should be "fired" (used for the preceding calculation process) or not.
3. The first and second step are repeated until the last layer (the output layer) is reached.
4. At the output layer, the yielded output will be used to calculate the error value by using the expected output. The resulted error then will be used to adjust the weights by the use of a backpropagation algorithm. Depending on the value of batch size in the MLP parameterization, the produced error might be accumulated before the weight updating process occurs.

In addition to the batch size configuration in an MLP model, there are other parameter which will affecting an MLP model learning steps, namely, the learning rate and learning iterations. The learning rate is a parameter which control how much the corresponding weight moving toward minimum error value for a given batch size. On the other hand, learning iterations control how much the learning process occurs for all data instance within the dataset [6].

4. EXPERIMENT AND ANALYSIS

As previously mentioned, we will be focusing on experimenting with MLP for solving an e-commerce sentiment analysis task which represented as binary classification problem. In the experiment, we use the dataset from [15] which contains 12212 feedbacks. From the previous research, every feedback has been systematically labelled into positive and negative class. The systematic labelling process have resulted imbalance class proportion with 10010 positive and 2202 negative class.

Our experiment using MLP is consist of 4 main part where detail explanation that differ and similes each scenario are described in the following section.

1. For every scenario, we randomly distribute the dataset into 5 different groups. Later, by using each single group from the 5 resulted groups as a testing set and the rest as a training set, an fastText embedding model and MLP model are trained and evaluated.
2. Considering the imbalance class proportion in the dataset used, to prevent accuracy paradox, each model will be evaluated by its average F1 score.
3. In the first scenario, the MLP model with one hidden layer with 250 neurons, initial learning rate of 0.01, batch size of 40, epoch of 400, Rectified Linear Unit (ReLU) as hidden neurons' activation function and Adam as MLP's optimizer is trained and evaluated by using the provided training and testing set as it is.
4. In the second scenario, the MLP with same configuration in the first scenario is trained by the dataset where a collection of data with negative class have been upsampledso the training set will consist of the equal amount of negative and positive class proportion.
5. In the third scenario, the MLP with same configuration in the first scenario is trained by the dataset where a collection of data with positive class have been downsampled (randomly selected with respect to the amount of negative class) into half of its amount and the data with negative class have been upsampled to equalize its amount to the downsampled positive class amount.
6. By the best resulting score in the first to third scenario, the hyperparameter tuning is conducted to show the best performance of MLP model for sentiment analysis task. As an addition, by using same scenario, the hyperparameter tuning for finding another MLP model which utilize TF-IDF as word embedding is also conducted. From these two experiments, we might conclude whether an MLP model with a fastText word embedding model will outperform an MLP model with TF-IDF.

The experimentation result regarding to the first scenario is given in Figure 2 where Prec. Pos and Rec. Pos are referring to the average precision and recall for the positive class from each data group, respectively; while Prec. Neg and Rec. Neg are representing the result for the negative class. On the other hand, F1 score in Fig. 2 is referring to the average of F1 score macro-average for positive and negative class from each data group.

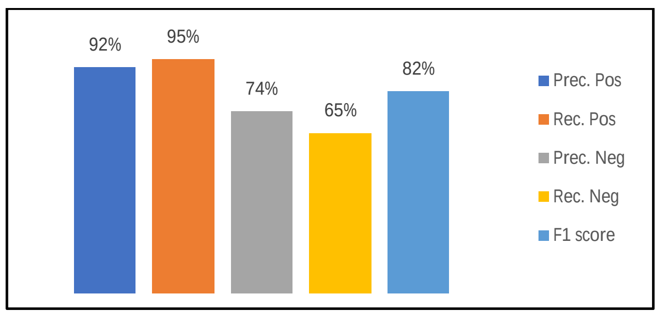

Figure 2: First scenario result

From Fig. 2, an MLP-based model shows a promising result for an e-commerce sentiment analysis task as it achieves the average F1 score of 0.82 from 1.00. As positive sentiment data is dominating the dataset used for each training phase, an MLP-based model is able to predict sentences with positive sentiment accurately (by the given Prec. Pos and Rec. Pos in Fig. 2). However, the resulted model might not perform to well for predicting sentences with negative sentiment as Fig. 2. show Prec. Neg and Rec. Neg score of 0.74 and 0.65.

Furthermore, as we are balancing the positive and negative data proportion in our dataset by the second scenario,MLP-based model able to achieve a quite significantly better average F1 score (0.88 from 1.00) as given in the Figure 3.

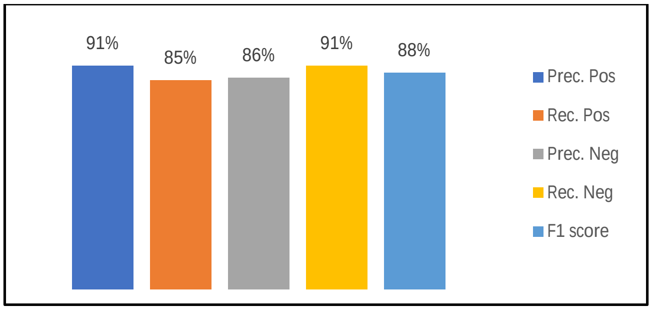

Figure 3: Second scenario result

Moreover, by downsampling the data with positive class, the resulted model is able to achieve better performance score for predicting data with negative sentiment (as precision and recall score for the negative class significantly increased from 0.74 to 0.86 and 0.65 to 0.91, respectively). However, such performance gains are sacrificing a little precision score for the positive class (from 0.92 to 0.91) and quite significant recall score (from 0.82 to 0.88).

As the second scenario showing a fruitful result, we try to balance the positive and negative classes with a slightly different approach as previously described (the third scenario). Unfortunately, in the third scenario, the resulted model is producing inferior overall performance scores compared to the resulted model in the second scenario as can be seen in Figure 4.

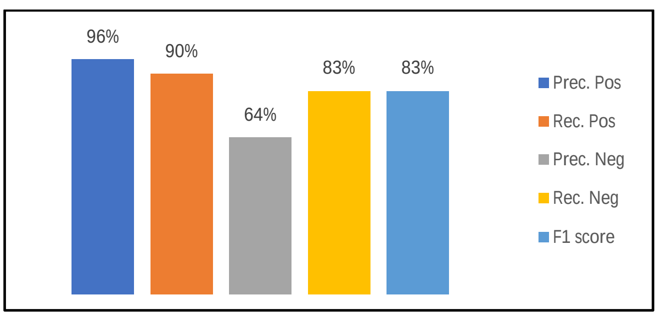

Figure 4: Third scenario result

By previous experiments, then we decide to do a grid search hyperparameter tuning using the second scenario to find the best F1 score performance of the MLP based model. Each parameter and its possible value is given in the Table 1.

Table 1: Grid search Hyperparameter Space

Parameter's name	Possible values
Neurons Amount	250, 350, 450
Batch size	20, 40, 60
Epoch	200, 300, 500
Learning rate initialization	0.01, 0.001

By grid search hyperparameter tuning on the second scenario, the resulted model (with 350 neurons in single hidden layer, batch size of 60, epoch of 300, and learning rate of 0.001) able to achieve F1 score of 0.89 with a slightly stable precision and recall for both of positive and negative class as given in Figure 5.

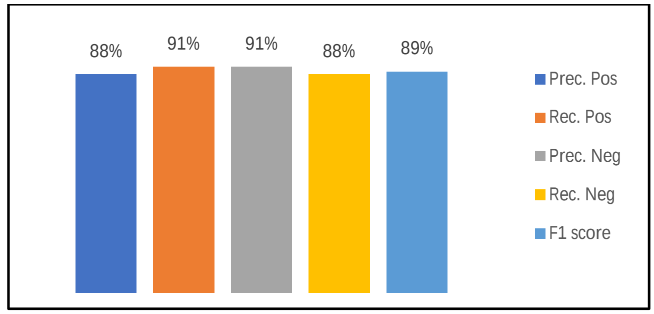

Figure 5: Hyperparameter Tuning Result

Furthermore, to ensure FastText is better as a feature extractor compared to TF-IDF, we conduct another experiment, using TF-IDF as feature extractor and set the other settings equal to the previous experiment. For the TF-IDF part, we use a various word level n-gram configuration with n equals to one (unigram), two (bigram) and three (trigram). From the last experiment part, the best result is obtained when n is equal to one (uni-gram). However, through Figure 6, it can be seen that an MLP-based model with FastText is able to achieve a slightly better performance score compared to an MLP based model with TF-IDF.

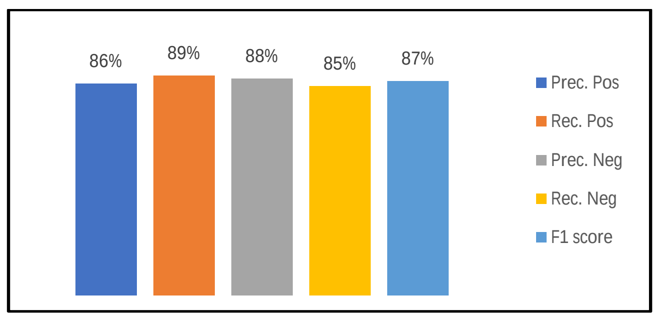

Figure 6: Scenario result using TF-IDF feature extraction

5. CONCLUSION AND FUTURE WORKS

Our experimentations show that, for a classification task with imbalanced dataset, balancing the training data by oversampling the minority class and right parameterization could lead to a more robust classifier .Furthermore, with the MLP grid search hyper parameter tuning in our experiment, the resulted model is able to achieve the best performance accuracy of 89.24% andF1-score macro average of 89% for e-commerce sentiment analysis task.

For a further works, as a sentiment analysis task oftentimes has a highly imbalance class proportions, instead of just simple down sampling and up sampling techniques, a more sophisticated sampling techniques such as Random Over-Sampling Examples (ROSE) Synthetic Minority Over-Sampling Technique (SMOTE), or Majority Weighted Minority Over-Sampling Technique (MWOTE) might be used to improve the MLP model performances. Other than that, by a more detail look each false prediction resulted by an MLP model, the misclassification cases often happen for a group of sentences that contains negative words. Therefore, the synonym extraction as a preprocessing technique to remove the negative words will presumably improving model classification performances.

On the other hand, instead of using standard MLP, it is worth a shot to use "similarity-based learner" such as siamese MLP. Many researches in various problem domains [18, 19, 20] show a promising result of using siamase MLP for imbalanced dataset. Moreover, for a small dataset, siamase MLP able to achieve better performance scores compared to standard MLP.

ACKNOWLEDGMENT

We would like to express our special thanks to Universitas Multimedia Nusantara Research and Community Outreach Center especially to Mrs. Dina Rosiana who helped us a lot for research administration related stuffs.

REFERENCES

1. Cahyono, Y. (2017). Analisis Sentiment Pada Sosial Media Twitter MenggunakanNaïve Bayes Classifier Dengan Feature Selection Particle Swarm Optimization Dan Term Frequency. JurnalInformatikaUniversitasPamulang, 2(1), pp.14–19. https://doi.org/10.32493/informatika.v2i1.1500
2. Fiarni, C., Maharani, H. and Pratama, R. (2016). Sentiment Analysis System for Indonesia Online Retail Shop Review Using Hierarchy Naive Bayes Technique. International Conference on Information and Communication Technology (ICoICT).IEEE.
3. Jotheeswaran, J. and Koteeswaran, S. (2015). Decision Tree Based Feature Selection And Multilayer Perceptron For Sentiment Analysis. Asian Research Publishing Network Journal, 10(14), pp.5883–5894.
4. Kaibi, I., Nfaoui, E.H. and Satori, H. (2019). A Comparative Evaluation of Word Embeddings Techniques for Twitter Sentiment Analysis. 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS).IEEE. https://doi.org/10.1109/WITS.2019.8723864
5. Laudon, K. and Laudon, J. (2009). Management Information Systems: International Edition. 11th ed. Pearson Higher Education.
6. Bakliwal, A., Arora, P., Patil, A., & Varma, V. (2011, November). Towards Enhanced Opinion Classification using NLP Techniques. In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011) (pp. 101-107).
7. Machine, M. (2019). Saturated Sectors: Finding Gaps In The Ecommerce Market In 2020. [online] Merchant Machine. Avaiable at: https://merchantmachine.co.uk/saturated-sectors/ [Accessed 31 Jan. 2020].
8. Mathapati, S., Manjuka, S.H. and K R, V. (2016). Sentiment Analysis and Opinion Mining from Social Media : A Review. Global Journal of Computer Science and Technology: C Software & Data Engineering, 16(5), pp.77–90.
9. Mladenović, M., Mitrović, J., Krstev, C. and Vitas, D. (2015). Hybrid sentiment analysis framework for a morphologically rich language. Journal of Intelligent Information Systems, 46(3), pp.599–620.
10. Nurhuda, F., Sihwi, S.W. and Doewes, A. (2013). AnalisisSentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkanOpinidari Twitter MenggunakanMetode Naive Bayes Classifier. JURNAL ITSMART, 2(2), pp.35–42. https://doi.org/10.20961/its.v2i2.630
11. Nurjannah, Musfiroh, dkk. "PenerapanAlgoritma Term Frequency-Inverse Document Frequency (Tf-Idf) Untuk Text Mining." JurnalInformatikaMulawarman, vol. 8, no. 3, 2013, pp. 110-113.
12. Pak, A. and Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. DBLP Computer Science Bibliography, 7(17), pp.1320–1326.
13. Rehurek, R. (2019). Gensim: Topic Modeling for Humans. [online] radimrehurek.com. Avaiable at: https://radimrehurek.com/gensim/auto_examples/tutorials/run_fasttext .html#sphx-glr-auto-examples-tutorials-run-fasttext-py [Accessed 2 Feb. 2020].
14. Waykole, R.N. and Thakare, A.D. (2018). A Review Of Feature Extraction Methods For Text Classification. International Journal of Advance Engineering and Research Development, 5(4), pp.351–354.
15. Wijayanti, R. and Arisal, A. (2017). Ensemble approach for sentiment polarity analysis in user-generated Indonesian text. 2017 International Conference on Computer, Control, Informatics and its Applications (IC3INA).
16. Blitzer, J., Dredze, M., & Pereira, F. (2007, June). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th annual meeting of the association of computational linguistics (pp. 440-447).
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
18. Prayogo, K. A., Suryadibraya, A., & Young, J. C. (2020). Classification of pneumonia from X-ray images using siamese convolutional network. Telkomnika, 18(3).
19. Bedi, P., Gupta, N., & Jindal, V. (2020). Siam-IDS: Handling class imbalance problem in Intrusion Detection Systems using Siamese Neural Network. Procedia Computer Science, 171, 780-789. https://doi.org/10.1016/j.procs.2020.04.085
20. De Baets, L., Develder, C., Dhaene, T., &Deschrijver, D. (2019). Detection of unidentified appliances in non-intrusive load monitoring using siamese neural networks. International Journal of Electrical Power & Energy Systems, 104, 645-653.
21. Wiratama, G. P. &Rusli, A. "Sentiment Analysis of Application User Feedback in Bahasa Indonesia Using Multinomial Naive Bayes," *2019 5th International*

Conference on New Media Studies (CONMEDIA), Bali, Indonesia, 2019, pp. 223-227, doi: 10.1109/CONMEDIA46929.2019.8981850.

22. Elvina, N., Rusli, A. &Hansun, Seng. Line Bot Chat Filtering using Naïve Bayes Algorithm. (2019). International Journal of Innovative Technology and Exploring Engineering Regular Issue, 8(12), 4877–4882. doi: 10.35940/ijitee.l3726.1081219.
23. Syamala, M. & N. J. Nalini. (2019). A Deep Analysis on Aspect based Sentiment Text Classification Approaches. International Journal of Advanced Trends in Computer Science and Engineering, 8(5), 1795-1801. https://doi.org/10.30534/ijatcse/2019/01852019
24. Bahafid, A., El Guemmat, K., Ben Lahmar, E. H.&Talea, M. (2019). Towards a Semantic Search Engine for E-commerce. International Journal of Advanced Trends in Computer Science and Engineering, 8(6), 3412-3422. https://doi.org/10.30534/ijatcse/2019/116862019
25. Matthews, M. D., Abraham, S. (2019). Human Annotation and Interpretation of Public Sentiments about Jio Coin marked in Social Networks using Machine Learning Algorithms. International Journal of Advanced Trends in Computer Science and Engineering, 8(4), 1162-1167. https://doi.org/10.30534/ijatcse/2019/25842019