A New Sentiment Analysis System Of Tweets Based On Machine Learning Approach

Yousef El Mourabit, Youssef El Habouz, Mustapha Lydiri Hicham Zougagh

Abstract: A very huge amount of data is generated every second for microblogs, content sharing via Social media sites and social networking. Twitter is an important popular microblog where people voice their opinions with regard to daily issues. Recently, analyzing these opinions is the main concern of Sentiment analysis (or opinion mining). Efficiently capturing, gathering and analyzing sentiments has been challenging for researchers. To deal with these challenges, in this paper we propose a highly accurate model for sentiment analysis of tweets. Using the Crowdflower's dataset, we started by data preprocessing (replace missing value, Denoising, tokenization, stemming…). We applied a semantic model with Term Frequency, Inverse Document Frequency weighting for data representation. In the measuring and evaluation step we applied four machine-learning algorithms such as Naive Bayesian, K-Nearest Neighbors, Neural Networks (LSTM), and Support Vector Machine. Afterwards, and based on the results we boiled a highly efficient prediction model with python, we trained and evaluated the classification model according to the most efficient metrics measures in this field, then tested the model on a set of unclassified tweets, to predict the sentiment class of each tweets. Experimental results demonstrate that our model reached a high accuracy compared to the other models.

Index Terms: Neural network, Machine learning, Sentiment analysis system, Twitter

1. INTRODUCTION

With the development of the web and the explosion of data sources such as opinion sites, blogs and microblogs, it has become necessary to analyze millions of posts, tweets or opinions in order to know what Internet users are thinking. Sentiment analysis is a technology that automatically analyses speech, written or spoken, and highlights the different opinions expressed on a specific subject such as a brand, a news item or a product. The importance of sentiment analysis is present in several areas, including politics, marketing, reputation management, etc. Sentiment analysis involves several disciplines; there are mainly three approaches to make this analysis: Approach based on automatic learning, approach based on the automatic processing of natural language (Natural Language Processing), which presented by Chowdhary, [1], and a last one combines the first two approaches. Recently, Artificial Intelligence and machine learning algorithms are applied in various fields: such as for data classification by S Park et al [2], Damage detection in truss bridges by DH Nguyen et al [3], Health monitoring by XW Ye and al [4]…etc. The aims of our work is to create an efficient sentiment analysis model of tweets based on machine learning approach, we explored the steps of classifying the text to discover the secrets of Sentiment analysis by adopting an automatic learning approach, for this we compared several methods including Probabilistic Naïve Bayes (NB), used by many researcher in this field as YU, Thein et al [5] and DEY, Sanjay et al [6], Support Vector Machine (SVM), k-neighboring nearest (KNN) described by MUSTAQIM, T. Et al [7], and artificial neural networks (RNN) presented and desscussed by GIMÉNEZ, Maite [8]. We applied these four approaches on the Twitter data set of customer reviews from many airlines companies. We considered a semantic model with Term Frequency Inverse Document Frequency (TF-IDF) approach explained by Kim, D. et al [9], for data representation. The results obtained in terms of True Positive rate (TP), False Positive rate (FP), Precision, F-Mesure and Accuracy; reveal the best classification method, in order to implement it and build our classification model.

- Yousef EL Mourabit *is currently Assistant Professor, Attached to TIAD Laboratory, on Sciences and Technology Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco, E-mail: y.elmourabit@usms.ma*
- *Youssef El Habouz is à researcher on Institute Igdr Umr 6290 Cnrs-Rennes1 University, Rennes, France. E-mail: youssef.elhabouz@univ-rennes1.fr*
- *Mustapha Lydiri is) PhD Student on TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco, E-mail:mustaphalydiri@gmail.com*
- *Hicham Zougagh is currently Professor on Sciences and Technology Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco, E-mail:*h.zougagh@usms.ma

2 RELATED WORK

Many researches have been done in Sentiment analysis field. They analyze the behaviours of users live data to extract the feelings of ordinary people towards any subject, trend, product, etc. several studies focus mainly on extracting useful information from the users' natural language and processing it to get the real feelings. It has generated interest with the ever-increasing use of the Internet by people to share their opinions. Hatzivassiloglou and McKeown [10] working at the document level and using "World Street Journal" as a data source, their work is based on conjunctions and adjectives and creates a Log Linear Regression model. In the same level document Pang et al. [11] conducted an analysis with learning models Naive Bayes (NB), Support Vector Machine (SVM), Maximum Entropy (ME), they used Unigram, bigram, contextual effect of negation, and frequencies, and they applied several models on film reviews, we can also cite other work on the analysis of feelings at the document level: Das and Chen [12], Turney , Morinaga et al [13], Turney and Littman [14] and Pang et Lee [15]. At the word level, Melville et al [16], performed a Bayesian classification with lexicons and learning documents using posts from blogs, opinion sites, political blogs and film reviews. A survey was presented by Kharde and Sonawane[17], it covered the techniques of the sentiment analysis on Twitter data, and compared the existing approaches. Another survey provided by Ravi and Ravi [18], they presented a detailed survey on the tasks, the applications and the approaches of the opinion mining that included a separate section for sentiment analysis. Agrawal and Mittal [19] explored various selection techniques to extract the prominent features in a machine learning based sentiment analysis. Fouad and al [20], presented an example of twitter sentiment analysis system using Information Gain (IG) feature selection technique. Recently, Akchi Kumar and Arunima Jaiswal [21], presented a systematic literature review of sentiment analysis on twitter using soft computing techniques. Guo, Xinyi, and al [22], proposed a novel social networks sentiment analysis model based on Twitter sentiment score (TSS).

3 METHODOLOGY

This section provides a general overview of the modelling of our tweets classification system by showing all the processes and treatments used to build our classification model. The methodological approach can be summarized in four main steps:

3.1 Data pre-processing

Tweets must be cleaned during the pre-processing process, in this phase we applied a number of cleanings and filters on these tweets such as removing links, identifiers, deleting words that contain less than 3 characters, filtering empty words…

3.2 Data vectorisation

Transformation of texts to digital vectors, we used a transformation of text to digital vectors based on the bag-of-words technique with the Term Frequency times Inverse Document Frequency (TF-IDF) method, for calculating the score of each word. It's a data transformation and a scoring scheme typically used in text analyses for measuring whether or not and how concentrated into relatively few tweets the occurrences of an input word are [23].

3.3 Classification model building

When we finished data preparing phases, we choose four machine learning algorithms among the most used and efficient ML algorithms for sentiment analysis of Tweets, based on recent researches on this field. Support vector machine algorithm (SVM): SVM is a non-probabilistic binary linear classifier (two class), originally proposed by and Cortes &Vapnik [24] , and Vapnik [25]. SVM separates data across a decision boundary (the hyperplane) f(x) = 0 , by solving a constrained quadratic optimization problem based on the structural risk minimization.
(1)

$$y = f(x) = W^T x + b = \sum_{i=1}^{N} W_i x_i + b$$

Naive Bayesian algorithm (NB)
Naïve Bayes is a machine learning algorithm that uses probability calculations, based on the concept of a Bayesian approach. The use of Bayes theorem in the Naïve Bayes algorithm is by combining conditional probability and prior probability in the following formula, which can be used to calculate the probability of each possible classification [26].
(2)

$$P(c|x) . \frac{P(x|c)\hat{P}(c)}{P(x)}$$

K-Nearest Neighbours algorithm (KNN)
K-Nearest Neighbour (K-NN) algorithm is a method for classifying samples based on learning data that are closest to the same groups of samples [27] . To perform predictions with K-NN, we need to determinate a metric to measure the distance between the query point and the case from the example sample using the following formula (3) :

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Recurrent neural network algorithm RNN (LSTM)

LSTM neural network, as a specific type of recurrent neural networks RNN, was first proposed by Hochreiter and Schmidhuber [28].

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\
c_t &= f_t * c_{t-1} + i_t * \tanh(W_{xc}x_t + b_c) \\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\
h_t &= o_t * \tanh(c_t)
\end{aligned}$$

LSTM is ideal for text processing, because it considers the order and dependencies of tokens. We use LSTM as the example to explore the potential of including contextual characteristics. The output of each cell in LSTM is decided by a set of gates that are represented by the functions above [29]

3.4 Model evaluation and testing

On the Training and evaluation phase, we based on four metrics (True Positive Rate (TP), False Positive Rate (FP), Accuracy, Confusion Matrix), then we tested the model on a set of test data that represents a set of unclassified tweets, to predict the sentiment class (Positive, negative, neutral) of each tweet. Figure 1 shows the architecture of our system.

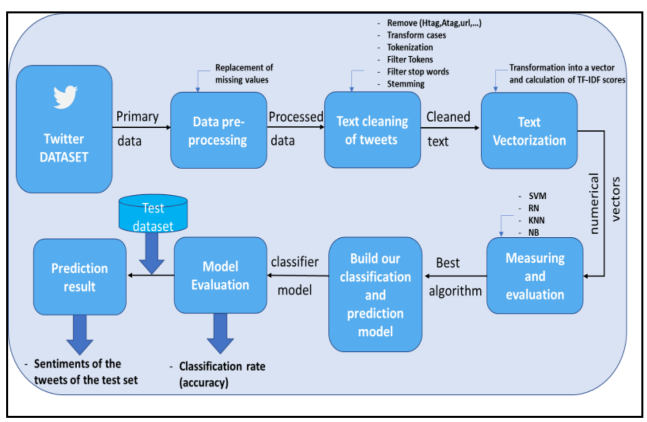


Figure 1. *Classification System Architecture*

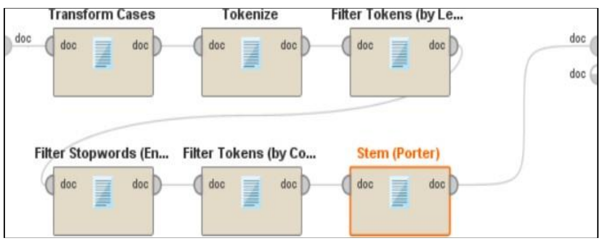
3.5 Twitter Dataset

The US airlines' Twitter data set represents the predictive modelling and analysis competition platform where data processing companies download various data sets to compete and develop the best models. This data set is the reformatted version of the original source (Crowdflower's Data for Everyone library) retrieved in February 2015 from Twitter. The dataset contains tweets with a feeling defined as "positive", "negative" or "neutral", for six major US airlines, it includes 3485 records with 11 independent attributes.

4 EXPERIMENT AND RESULTS

4.1 Experiment on Rapid Miner environment

For the data pre-processing we used Rapid Miner, which is an efficient tool, presented in many researches in this field , such as Pavarsi, H. J. Et al [30], [Anandarajan, M., et al [31]. We builded our System according the following steps: Replace missing values, Denoising or Selection, Pre-processing tweets text (Transform cases, Tokenization, Filtering Tokens, Filtering Tokens, Filter stop words, Stemming). These operations are illustrated in the following figure: Figure 2. Text preprocessing phase



In data vectorization step, most approaches are based on the vector representation of documents; here we used TF-IDF coding which gives a view to documents (or tweets in our case) in the form of rows and terms in the form of columns. To build our classification model, we processed the data using the text mining operators available in Rapid Miner before applying the classification algorithms. We displayed the FP rate, TP Rate and accuracy measurements to find the most efficient model with the highest measurement values. The cross-validation approach which is a standard evaluation technique, is a systematic way to perform repeated percentage splits, which divides the data set into 10 pieces (folds) and then takes each piece in turn to test it and takes the remaining 9 pieces as training data, 9 subsets of data is used as a learning data set to form a model and the remaining subset is used to validate the model. This gives 10 evaluation results, and takes the average of these results, then we applied on our data set four algorithms, usually using for this area of research. The nearest Neighbours K algorithm presented by Luque, A., et al [32], and Maraziotis, I. A., et al [33]. Support Vector Machine algorithm used by Mehta, R. P. Et al [34], and Pham, B. T., et al [35], Naive Bayesian algorithm presented by El Mourabit, Y., and al [36] to build a classifier for intrusion detection, also used by Wang, Q., et al [37], for an emotional analysis of public opinions. The results are presented in graphics form to clearly visualize this performance evaluation between the models built with the different algorithms on the Twitter dataset.

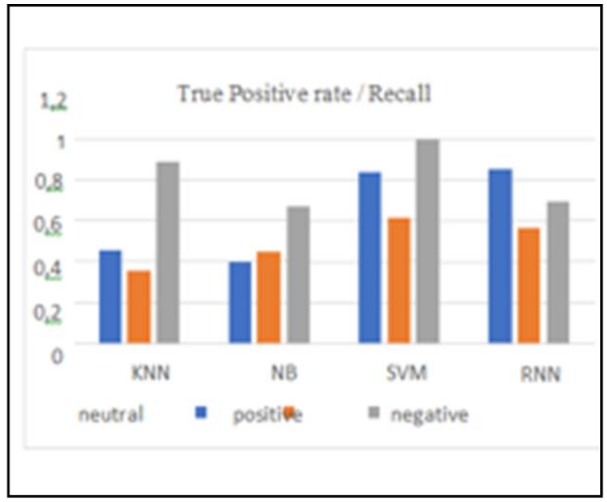


***Figure 3**. Cross validation operator*

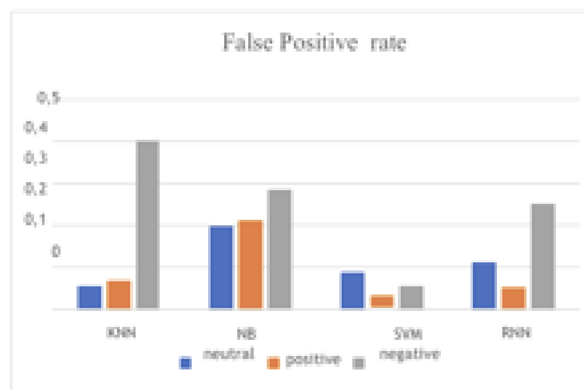


***Figure 4.** True Positive rate of KNN, RNN, NB and SVM*

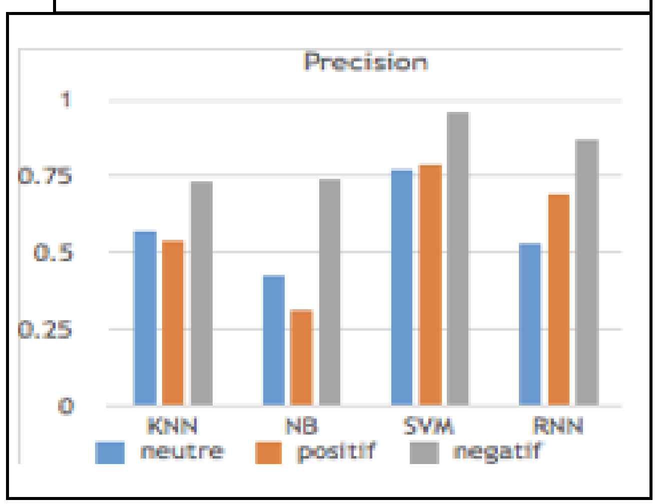


***Figure .5** False Positive rate of KNN, RNN, NB and SVM*

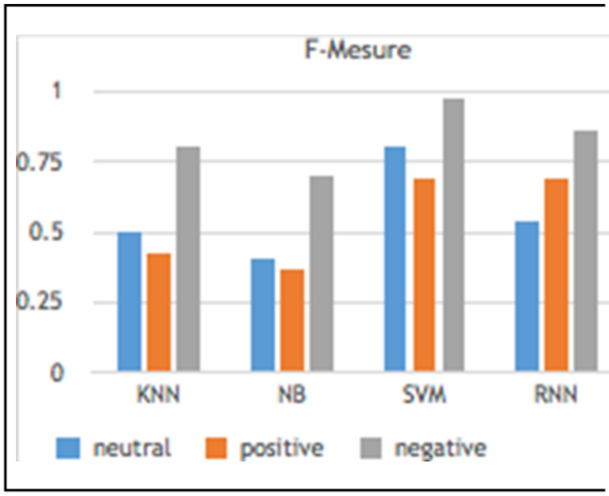


Figure 6. *Precision rate of KNN, RNN, NB and SVM*

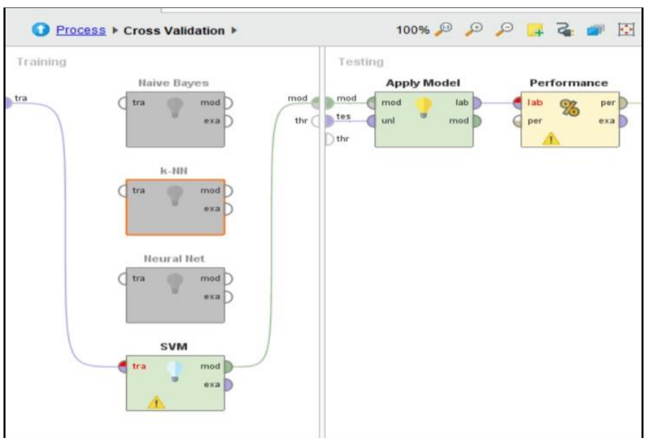


Figure 7. *F-Mesure rate of KNN, RNN, NB and SVM*

The Performance operator is used to evaluate the performance of a classifier. It provides a list of performance criteria values, this operator can be used for all types of learning tasks. The confusion matrix: shows the relationship between the actual and predicted values. Figures below, present the Experiment results:

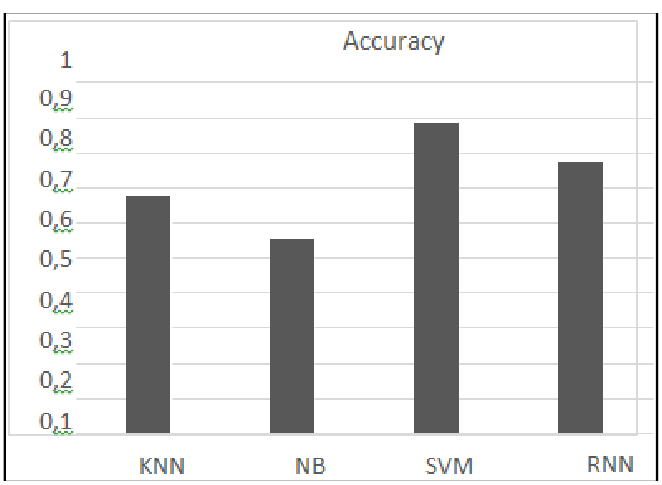


Figure 8. *Accuracy of KNN, RNN, NB and SVM*

We used four different classifiers: KNN, RN, NB and SVM, and from the comparison of the different measures we find that SVM performs better than other learning methods, SVM remains the most efficient tool in this case, based on accuracy rate, it gives a very high accuracy of 87%. Consequently, we will implement SVM algorithm on the python environment to build an efficient predictive sentiment analysis model of tweets.

4.2 Experiment on Python environment

We started to build our Support-Vector Machine (SVM) classifier. The concept of our classifier is to separate data points using a hyperplane with the largest margin. This is why an SVM classifier is also called a discriminative classifier; SVM builds the hyperplane in a multidimensional space to separate the different classes, and generates the optimal hyperplane iteratively, which helps to minimize the error. The general idea of SVM is to find a maximum marginal hyperplane (MMH) that better divides the data set into classes. Our automatic learning model can only process numerical values as vectors or matrices. To prepare our tweets for the automatic learning model, we create a reverse document frequency vectorization term (TF-IDF). The result of this vectorization is a matrix that contains a representation of each sentence as a vector; the vector has the same length as our vocabulary, i.e. the list of all the words observed in our learning data, each word representing an entry in the vector. For the evaluation of the model's performance we will perform a test on a separate test set, in order to estimate the performance of the generalized model, this is done with the scikit-learn train_test_split function integrated in python. We obtain the following results:

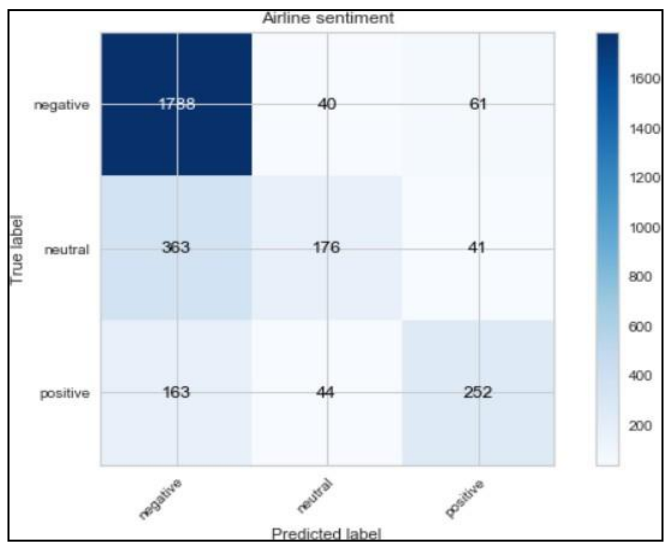


Figure 9*. Confusion matrix of our system*

Finally we shows the metrics of our classifier used on the cleaned data set, and we notice that the classification rate is high (75%) and TP Rate (76%) Accuracy (84%) F-Measure (78%) are also high, while FP Rate (3%) is low, which expresses the efficiency of our model. According to the confusion Matrix that measures the quality of a classification system, we notice that 2216 tweets are well classified out of 2928. Experimental results demonstrate that our model reached a high accuracy (84%), compared to the other models.

	Precision	TP Rate	FP Rate	F-Mesure	Accuracy
Class (negative)	0.95	0.77	0.10	0.85	
Class (neutral)	0.30	0.68	0.11	0.42	
Class (Positive)	0.55	0.71	0.19	0.62	
Model	0.84	0.76	0.13	0.78	0.75

Table 1. *Results table*

Prediction model
To validate our model we used a set of unclassified tweets, in order to predict their classes, vectors results contain the probability that a tweets belongs to a specific class (negative, neutral, positive), for example for the tweet "what a great airline, the trip was a pleasure !" we notice that the classifier has given the probability 0.73 that he belongs to the positive class and 0.20 and 0.05 for the negative and neutral classes. The following Figure shows the probability of the other tweets:

```
sentences = count_vectorizer.transform([
    "What a great airline, the trip was a pleasure!",
    "My issue was quickly resolved after calling customer support. Thanks!",
    "What ! My flight was cancelled again. This sucks !",
    "Service was awful. I'll never fly with you again.",
    "I have mixed feelings about airlines. I don't know what I think.",
])
clf.predict_proba(sentences)

array([[0.20756126, 0.05764636, 0.73479238],
       [0.13931069, 0.06994379, 0.79074552],
       [0.91909175, 0.046192  , 0.03471624],
       [0.8908144 , 0.07061903, 0.03856657],
       [0.46518174, 0.50258776, 0.03223049]])
```

Figure 10*. Tweets prediction*

4. CONCLUSION

We worked on a dataset of tweets of reviews from some airlines companies, to get consistent and well cleaned data, for this we applied a number of cleanups and filters on these tweets using the RapidMiner tool such as removing links, identifiers, and deleting words that contain less than 3 characters, filtering empty words..., then we transformed these tweets to digital vectors based on the bag-of-words technique with the TF-IDF method for calculating the score of each word. The classification rate of the classes predicted by the KNN, Naïve Bayesien, SVM and Neuron Networks algorithms are respectively 67% ,55% ,88% ,77%, and after a discussion of the results obtained we chose the SVM model as the best model that gives us the best performance compared to the other models, and we implemented it with the python environment. Finally we trained and evaluated the classification model using the metrics; TP rate, FP rate, Accuracy, Confusion matrix. Then we tested the model on a set of test data that represents a set of unclassified tweets to predict the sentiment class of each tweets among this set. In the future work , we will use a large and complex dataset, the number of labels (classes) can also be increased. We can include other languages also and use special characters and non-letter characters as well. It would be valuable to include the Emoticons as it widely used in social media to represent the expressions. Also, we will try to use the Twitter Streaming API to retrieve tweets in real time in order to do a real time sentiment analysis and exploring other social networks.

REFERENCES

[1] Chowdhary, K. R. "Natural language processing." Fundamentals of Artificial Intelligence. Springer, New Delhi, 2020. 603-649.
[2] Park, Seungtae, et al. "Wavelet-like convolutional neural network structure for time-series data classification." Smart Structures and Systems 22.2 (2018): 175-183.
[3] Nguyen, Duong Huong, et al. "Damage detection in truss bridges using transmissibility and machine learning algorithm: Application to Nam O bridge." Smart Structures and Systems 26.1 (2020): 35-47.
[4] Ye, X. W., T. Jin, and C. B. Yun. "A review on deep learning-based structural health monitoring of civil infrastructures." Smart Structures and Systems 24.5 (2019): 567-585.
[5] YU, Thein et NWET, Khin Thandar. Sentiment Analysis System for Myanmar News Using Support Vector Machine and Naïve Bayes. In : International Conference on Genetic and Evolutionary Computing. Springer, Singapore, 2019. p. 551-557.
[6] DEY, Sanjay, WASIF, Sarhan, TONMOY, Dhiman Sikder, et al. A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. In : 2020 International Conference on Contemporary Computing and Applications (IC3A). IEEE, 2020. p. 217-220.
[7] [MUSTAQIM, T., UMAM, K., et MUSLIM, M. A. Twitter text mining for sentiment analysis on government’s response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm. In : Journal of Physics: Conference Series. IOP Publishing, 2020. p. 032024.
[8] Giménez, M., Palanca, J., & Botti, V. (2020). Semantic-based padding in convolutional neural networks for improving the performance in natural language processing. A case of study in sentiment analysis. Neurocomputing, 378, 315-323.
[9] Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations:TF– IDF,LDA,and Doc2Vec. Information Sciences,477,15-29.
[10] Hatzivassiloglou, V., & McKeown, K. R. (1997, July). Predicting the semantic orientation of adjectives. In Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics (pp. 174-181). Association for Computational Linguistics.
[11] Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. InProceedings of the 42nd

annual meeting on Association for Computational Linguistics (p. 271). Association for Computational Linguistics.

[12] Das, S., & Chen, M. (2001, July). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific finance association annual conference (APFA) (Vol. 35, p. 43).

[13] Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002, July). Mining product reputations on the web. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 341-349).

[14] Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), 21(4), 315-346.

[15] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.

[16] Turney, P. D. (2002, July). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. InProceedings of the 40th annual meeting on association for computational linguistics (pp. 417-424). Association for Computational Linguistics.

[17] Kharde, V., Sonawane, S.: Sentiment analysis of twitter data: a survey of techniques. Int. J. Comput. Appl. 139(11), 5–15 (2016)

[18] Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl. Based Syst. 89, 14–46 (2015)

[19] Agarwal, B., Mittal, N.: Prominent Feature Extraction for Sentiment Analysis. Socio- Affective Computing Series. Springer International Publishing (2016).

[20] Fouad, M. M., Gharib, T. F., & Mashat, A. S. (2018, February). Efficient twitter sentiment analysis system with feature selection and classifier ensemble. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 516-527). Springer, Cham.

[21] Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience, 32(1), e5107.

[22] Guo, Xinyi, and Jinfeng Li. "A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency." 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2019.

[23] Artama, M., Sukajaya, I. N., & Indrawan, G. (2020, April). Classification of official letters using TF-IDF method. In Journal of Physics: Conference Series (Vol. 1516, No. 1, p. 012001). IOP Publishing.

[24] Cortes, C.; Vapnik, V.N. Support-Vector networks. Mach. Learn. 1995, 20, 273–297.

[25] Vapnik, V.N. Statistical Learning Theory; John Wiley & Sons Inc.: New York, NY, USA, 1998.

[26] P. Tripathi, S. K. Vishwakarma, and A. Lala, “Sentiment Analysis of English Tweets Using Rapid Miner,” in 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015, pp. 668–672

[27] S. Wahyuningsih, D. R. Utari, U. B. Luhur, D. Tree, and K. Validation, “Perbandingan Metode K-Nearest Neighbor , Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit,” Konf. Nas. Sist. Inf. 2018, pp. 8–9, 2018

[28] Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural. Comput. 1997, 9, 1735–1780.

[29] Liu, Y.; Guan, L.; Hou, C.; Han, H.; Liu, Z.J.; Sun, Y.; Zheng, M.H. Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform. Appl. Sci. 2019, 9, 1108.

[30] Pavarsi, H. J., Hariri, N., Alipour-Hafezi, M., Al-Hawaeji, F. B., & Khademi, M. (2020). Machine Indexing of Documents in the Field of Information Retrieval Using Text Mining in the RapidMiner Software.

[31] Anandarajan, M., Hill, C., & Nolan, T. (2019). Learning-Based Sentiment Analysis Using RapidMiner. In Practical Text Analytics (pp. 243-261). Springer, Cham.

[32] Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition, 91, 216- 231.

[33] Maraziotis, I. A., Perantonis, S., Dragomir, A., & Thanos, D. (2019). K- Nets: Clustering through nearest neighbors networks. Pattern Recognition, 88, 470-481.

[34] Mehta, R. P., Sanghvi, M. A., Shah, D. K., & Singh, A. (2020). Sentiment Analysis of Tweets Using Supervised Learning Algorithms. In First International Conference on Sustainable Technologies for Computational Intelligence (pp. 323-338). Springer, Singapore.

[35] Pham, B. T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P. T., Ngo, T. Q., ... & Bui, D. T. (2019). A comparison of Support Vector Machines and Bayesian algorithms for landslide susceptibility modelling. Geocarto International, 34(13), 1385-1407.

[36] El Mourabit, Y., Bouirden, A., Toumanari, A., & Moussaid, N. E. (2015). Intrusion detection techniques in wireless sensor network using data mining algorithms: comparative evaluation based on attacks detection. International Journal of Advanced Computer Science and Applications, 6(9), 164- 172.

[37] Wang, Q., Liu, K., & Ma, K. (2019, April). Emotional Analysis of Public Opinions in Colleges and Universities: Based on Naive Bayesian Classification Method. In Journal of Physics: Conference Series (Vol. 1187, No. 5, p. 052042). IOP Publishing.