TECHNIUM

Romanian Journal of Applied Sciences and Technology

ISSN: 2668-778X

vol. 16 / 2023

The 7th International Conference on Science Technology

organized by
Faculty of Social Science and
Law Universitas Negeri Manado and
Consortium of International Conference
on Science and Technology

The Innovation Breakthrough in Digital and Disruptive Era

PLUS COMMUNICATION

International Communication & PR

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 16, 169-175, October, 2023
ISSN: 2668-778X
www.techniumscience.com

Multinomial Optimization of Naïve Bayes Through the Implementation of Particle Swarm Optimization

Made Hanindia Prami Swari[11], *Dio Farrel Putra Rachmawan*[2], and *Chrystia Aji Putra*[3]

[1,2,3]Informatics Department, Faculty of Computer Science, UPN Veteran Jawa Timur, Indonesia

Abstract. Sentiment analysis is widely used in cases of text processing and comments. One of the case studies is about the analysis of a hotel review by the public. The method used in analyzing a sentiment from comments or reviews of a hotel is the Naïve Bayes Classifier. One that can be used is the Multinomial Naïve Bayes method. In improving the results of the accuracy of the method required an optimization method. There are many optimization methods that can be applied to algorithms in sentiment analysis case studies. One well-known method is Particle Swarm Optimization (PSO). This study aims to determine the effect of PSO optimization on the Multinomial Naïve Bayes algorithm in the case of sentiment analysis. From the results of optimization and model testing, the highest accuracy was obtained in the Multinomial Naïve Bayes test with PSO optimization as hyperparameter tunning and feature selection of 97%.

[1] Corresponding author: ajiputra@upnjatim.ac.id

1 Introduction

Sentiment analysis is a scientific discipline that focuses on methods of analyzing opinions, views, and judgments from an individual towards an object (Husada & Paramita, 2021). Sentiment analysis is used to categorize text polarity in documents or sentences, with the aim of determining positive, negative, or neutral sentiments (Samsir et al., 2021). Sentiment analysis is one of the fields in text mining which aims to analyze the sentiment of a problem object automatically by a machine and makes it possible to obtain predictive results for every opportunity related to the results of the analysis.

This research has influenced various aspects, political sentiment analysis was performed to analyze the sentimental relationship between the social media and general pupils (M. Aman Ullah, 2020). More massive usage related to sentiment analysis research is used in the economic field. In research conducted by (Chen et al., 2020) it was explained that online reviews have a major influence on the views and decisions of a consumer in buying a product. Determining algorithms in computing is the most important thing in developing sentiment analysis, this is because the accuracy of sentiment analysis results depends on the computational algorithm used to classify the data to be tested.

Based on previous research, in recent years many studies have been developed related to the development of sentiment analysis using various algorithms to produce an optimal sentiment analysis system. One of the algorithms that is often used to perform sentiment analysis is the naive Bayes classifier algorithm, one example is the use of the naive Bayes multinomial approach.

The Naïve Bayes Multinomial Algorithm is a probability learning technique based on the Bayes theorem, and is often applied in Natural Language Processing (NLP). In this algorithm, the working principle is based on the frequency of occurrence of words in a document known as term frequency (Yuyun et al., 2021). In some cases, it is not uncommon for Multinomial Naïve Bayes to obtain lower accuracy than other algorithms. In research conducted by (Puspita & Widodo, 2020) on a comparison of the KNN, Decision Tree, and Naïve Bayes methods for sentiment analysis cases, it shows that the results for the accuracy of Naïve Bayes get the lowest value compared to the other 2 methods. In this study, KNN obtained an accuracy of 96.01%. The use of the Decision Tree method results in an accuracy of 96.13%. The Naïve Bayes method only gets an accuracy value of 89.14%. If the research accuracy using the Naïve Bayes algorithm gets low accuracy, a solution is needed to increase the accuracy value. The solution used is by using an optimization method.

There are many optimization methods that can be applied to algorithms for sentiment analysis cases, one of which is the Particle Swarm Optimization (PSO) method. Particle Swarm Optimization (PSO) is a method used to optimize a result. PSO optimization can control the determination of a subset to produce the best accuracy (Hayuningtias & Sari, 2019). In the research conducted (Hayuningtias & Sari, 2019) on sentiment analysis using Naïve Bayes and Particle Swarm Optimization for TMII tourist attractions resulted in higher accuracy than the Naïve Bayes method without optimization, which is 94%. In the above research, modeling only uses 1 data sharing scenario, the role of PSO is used as feature selection only. Research conducted by (Maulana, 2022) on sentiment analysis of users of protective applications on the Twitter platform using the CNN and PSO algorithms shows that PSO has an effect on increasing accuracy in sentiment analysis cases where PSO is used to search for the best parameters/hyperparameter tuning for the CNN algorithm. Test results using the PSO optimization method get an accuracy of 81%. While the test results without PSO optimization get an accuracy value of 77%.

In addition to using the optimization method, the acquisition of high accuracy can also be determined by training data and test data. This is shown in research conducted by (Putri, 2022) regarding sentiment analysis on BCA mobile application reviews using the Naïve Bayes Classifier. There are 12 test scenarios with different distribution of training data and test data with different accuracy results. The results of the highest accuracy in this study reached 96% in the efficiency aspect with 90% training data and 10% test data. The lowest accuracy was obtained for the Learnability and Satisfication aspects with 80% training data and 20% test data. Based on the things explained in previous studies, this research was carried out to optimize the multinomial naive Bayes method using particle swarm optimization in the hypertunning and feature selection processes. Testing will be carried out by dividing the training data and test data to produce analysis sentiment with the highest accuracy. The case study used in this research is an analysis of 4.355 comments from hotel guests at Favehotel Kusumanegara Yogyakarta.

2 Methods

The steps to be taken in this research are shown in Figure 1.

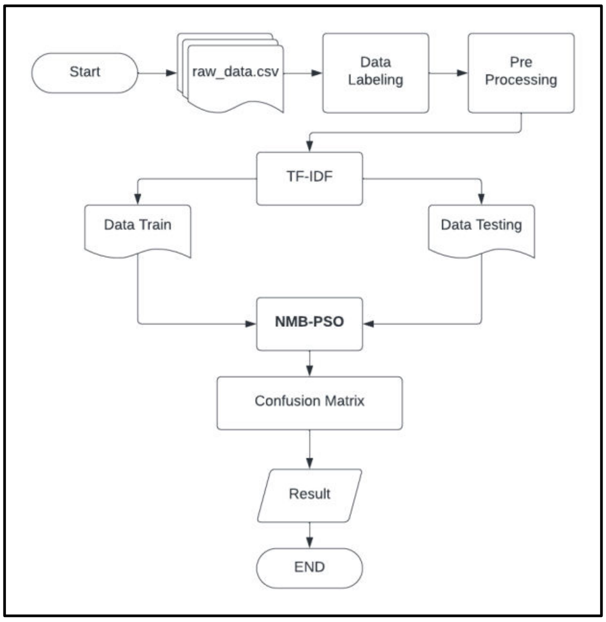

Fig. 1. Application Flow Chart

The data in this research were taken directly from the hotel management as much as 4.355 data. In the data labeling process, labeling is done automatically using the Vader Lexicon. Then in the preprocessing stage, the dataset is cleaned to remove noisy data or unnecessary data. The next step is the word weighting process using the TF-IDF method. Making models for 18 types of testing using Multinomial Naïve Bayes and Multinomial Naïve Bayes based on Particle Swarm Optimization. In the final stage, an analysis of the results of positive and negative sentiments will be carried out using the Confusion Matrix.

1.1 Labelling Data

The dataset from the hotel manager is raw data that does not have a label. Therefore, a data labeling process is needed to provide label information for each data. The dataset used is Indonesian data, so it is first converted into English using the Translator library before calculating the compound value using the Vader Lexicon. The compound value is used to determine whether the data is a positive, negative or neutral sentiment. If the compound value of the 19 sentence data set is above 0 then the sentence is considered a positive sentiment, otherwise if the compound value of a sentence is below 0 then the sentence is considered a negative sentiment. Meanwhile, if the compound value of the sentence is equal to 0 then the sentence is considered a neutral sentiment.

1.2 Preprocessing Data

The initial stage of data preprocessing is to carry out the cleaning process. This process reduces non-alphabetical data to reduce noise in the data. The removed characters are punctuation marks, symbols, emoticons, and website address links. Followed by the case folding stage, this is the process of changing the alphabetic characters that have gone through the cleansing stage into lower case letters. The tokenizing step aims to break sentences into small parts called terms or tokens based on each word that makes them up. The tokenizing process is done by breaking sentences based on the spaces between the words. The stopword removal process is a stage that aims to remove words that often appear, but have less informative value or are not significant in text analysis, such as prepositions, conjunctions, and other common words. This is intended so as not to affect the final result of the analysis. The final process is stemming, this process aims to change the words that have affixes to become the base word for each word that has been selected.

1.3 Data Weighting

This research discusses about sentiment analysis, and word weighting step is a process that cannot be missed. Word weighting is a method for giving weight to each word in a text. This process improves the ability of sentiment analysis in the text mining process. In this study, the authors used the Term Frequency Inverse Document Frequency (TF-IDF) method for weighting words in datasets. Data that has passed the preprocessing stage must be in numerical form so that it can be processed. TF-IDF is used to convert data into numeric. In the process of calculating the weight of the first word, it is necessary to know the value of tft,d, namely the value of the term frequency which is the weight of term t in document d. Then, determine the value of idf which is the number of documents that contain the term you are looking for. If both values have been found then multiply the two values to get the weight (ω). The less the frequency of occurrence of a term in the document, the lower the weight value (Yulita et al., 2021). Equation 1 is the computation used to calculate the TF-IDF value.

$$TF - IDF = tf_{t,d} x \log \log \frac{N}{dft} \tag{1}$$

Where,

$TF - IDFt,f$	=	The weight of each word of term t contained in the document d
tft,d	=	The value of the term frequency which is the weight of the term t in the document d
idf	=	The number of documents containing the search term
t	=	Term or Word

1.4 Optimation and Model Testing

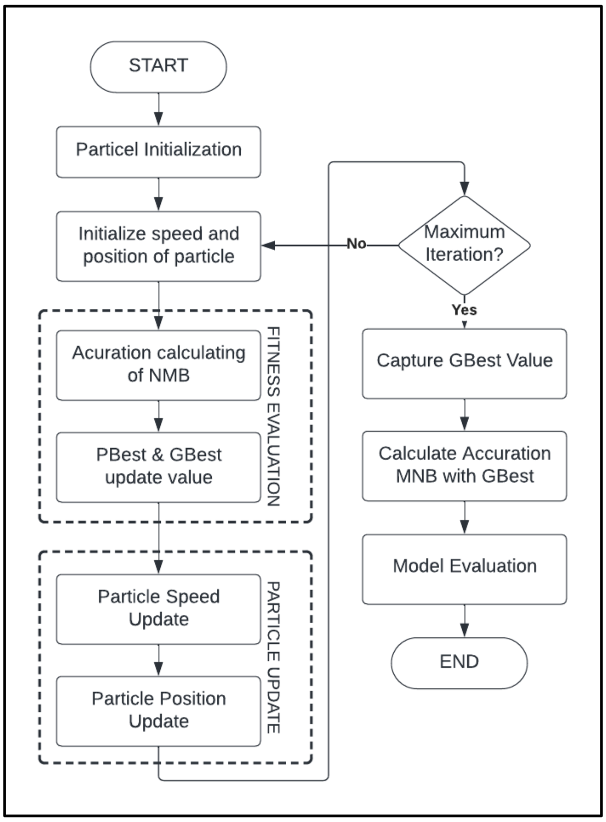

Fig. 2. Optimation Diagram and Model Testing

The first stage in PSO optimization is to initialize random particles based on the dataset. Particle initialization includes the process of initializing lower bound, upper bound, swarm size, and random particle positions. At each iteration, the fitness value is checked. In this trial the fitness value is calculated based on the accuracy value of the Multinomial Naïve Bayes algorithm. After that, it is followed by the initialization or updating of the Pbest and GBest values based on the particles. The Pbest value will be updated based on the particle results while the Gbest value will be reviewed based on the overall particle value. The process for updating the values is carried out in the particle, the latest velocity and the latest position with the acceleration equation formula (2).

$$\vartheta_{n+1} = \omega\vartheta_n + c_1 r_1(\rho_{best} - x_n) + c_2 r_2(g_{best} - x_n) \quad (2)$$

Where $c1$ and $c2$ are constant numbers, $r1$ and $r2$ are random numbers, ω are inertia, and n are the iterations. The displacement formula uses equation (3).

$$x_{n+1} = x_n + v_{n+1}$$

Recalculation of the fitness value is carried out to obtain the latest accuracy of the Multinomial Naïve Bayes algorithm. This will continue to be repeated in an iteration until it enters the Stop Criteria condition. After the stop criterion condition occurs, the Gbest value will be captured and a final test will be carried out using Multinomial Naïve Bayes.

The model evaluation process is carried out using the Confusion Matrix. This is useful for knowing the value of accuracy, precision, recall, and f1-score from the model testing process. Stop criteria in Particle Swarm Optimization (PSO) are conditions for stopping the iteration process in the PSO algorithm. This condition is determined to ensure that the PSO algorithm has produced a good enough solution. In other words, this condition is close to the optimal solution. Examples of stop criteria conditions in PSO include the maximum number of iterations that have been determined, the fitness value that has reached a certain value, or the difference in fitness value between the last iteration and the previous iteration that is considered quite small. With the stop criterion conditions, the PSO algorithm can stop automatically when an adequate solution has been found. This saves time and resources used in the optimization process.

2 Experiments and Result

To get the results of the Multinomial Naïve Bayes algorithm optimization test using Particle Swarm Optimization, a number of test scenarios are performed which are shown in Table 1.

Table 2. Testing Scenario

Scenario Number	Classification Model	Data Training	Data Testing
1	Multinomial Naive Bayes	90%	10%
2		80%	20%
3		70%	30%
4	MNB & PSO (Hyperparameter Tuning and Feature Selection)	90%	10%
5		80%	20%
6		70%	30%
7	NMB & PSO (Hyperparameter Tuning)	90%	10%
8		80%	20%
9		70%	30%
10	NMB & PSO (Feature Selection)	90%	10%
11		80%	20%
12		70%	30%

The analysis stage in this research is the calculation of the test results with the confusion matrix. It aims to compare the results of several test scenarios that have been carried out. The values compared from the test scenarios are accuracy, precision, recall, and F1 score.

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 16, 169-175, October, 2023
ISSN: 2668-778X
www.techniumscience.com

2.1 Model Evaluation Senario 1

In the first scenario, sentiment analysis is tested using Multinomial Naïve Bayes. The dataset is a customer assessment of the Favehotel Kusumanegara Yogyakarta hotel. In this test scenario, data separation is carried out, 90% as training data and 10% as test data. The results of the confusion matrix table in the first scenario can be seen in Figure 3.

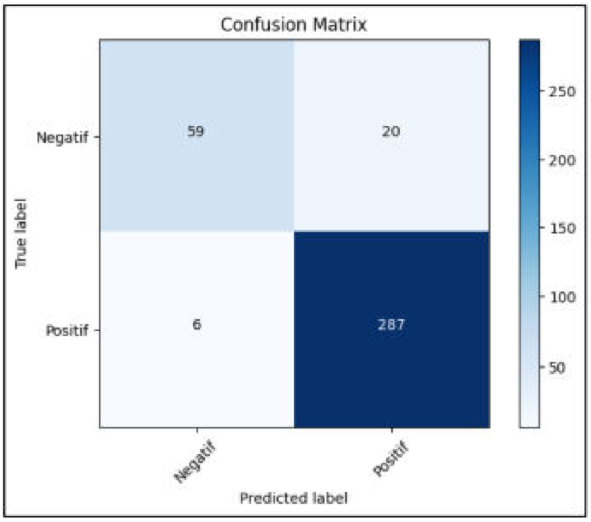

Fig. 3. Confusion Matrix of Scenario 1

Based on data testing in the first scenario, this model correctly predicts 346 data out of 372 data. There are 287 data with positive labels and 59 data with negative labels. Meanwhile, the number of data that failed to predict was 26 data. Figure 4 is the result of the classification report for scenario 1.

```
Classification Report :
              precision    recall  f1-score   support

           0       0.91      0.75      0.82        79
           1       0.93      0.98      0.96       293

    accuracy                           0.93       372
   macro avg       0.92      0.86      0.89       372
weighted avg       0.93      0.93      0.93       372
```

Fig. 4. Confusion Matrix of Scenario 1

The accuracy obtained in testing scenario 1 is 93%. The Precision value obtained is 91% on negative sentiment and 93% on positive sentiment. In other words, based on 65 test data that get negative sentiment, there are 59 data that are truly negative and 6 data that fail to be predicted. Then, based on 307 test data that received positive sentiment, there were 287 data that were truly positive, and 20 data that failed to be predicted. Recall value in scenario 1 results in 75% negative sentiment and 98% positive sentiment. In other words, based on 79 data that should have negative sentiments, there are 59 data that the model was able to detect and 20 data that the model did not know. Based on 293 data that should have positive sentiments, there were 287 data that the model managed to find out and 6 data that the model didn't know. After obtaining the Precision and Recall values, the F1-Score is obtained which is the harmonic average value of precision and recall. The results of the F1-Score for scenario 1 are 82% negative sentiment and 96% positive sentiment.

2.2 Result

The next step is carried out after classifying and evaluating the model. Comparative analysis of the results of model testing is carried out based on the scenarios that have been made. The purpose of this section is to find out the differences and also the effect of the Particle Swarm Optimization optimization method on sentiment analysis. The data used is the customer assessment of the Favehotel Kusumanegara Yogyakarta hotel.

Table 2. Testing Scenario

No.	Classification Model	Data Training	Data Testing	Accuracy	Precission	Recall	F-1 Score
1	Multinomial Naive Bayes	90%	10%	93%	92%	86%	89%
2		80%	20%	91%	90%	82%	86%
3		70%	30%	89%	90%	79%	83%
4	MNB & PSO (Hyperparameter Tuning and Feature Selection)	90%	10%	97%	96%	96%	96%
5		80%	20%	93%	91%	90%	90%
6		70%	30%	93%	92%	89%	91%
7	NMB & PSO (Hyperparameter Tuning)	90%	10%	96%	94%	94%	94%
8		80%	20%	93%	91%	89%	90%
9		70%	30%	92%	80%	88%	89%
10	NMB & PSO (Feature Selection)	90%	10%	95%	96%	89%	92%
11		80%	20%	91%	91%	83%	86%
12		70%	30%	90%	92%	85%	85%

In Table 2, in scenarios 1, 2 and 3, model testing was carried out with Multinomial Naïve Bayes (MNB) without using optimization methods. Testing 75 methods using MNB with a comparison of training data and test data of 90% and 10%. The results of this test produce an accuracy value of 93% and an F1-Score value of 89%. Tests using the MNB method with a comparison of training data and test data of 80% and 20% produce an accuracy value of 91% and an F1-Score value of 86%. Tests using the MNB method with a comparison of training data and test data of 70% and 30% produce an accuracy value of 89% and an F1-Score value of 83%. In scenarios 4, 5, and 6, model testing was carried out with Multinomial Naïve Bayes and using the Particle Swarm Optimization optimization method for hyperparameter tunning and feature selection. Testing the MNB and PSO methods for hyperparameter tunning and feature selection with a comparison of training data and test data of 90% and 10% resulted in an accuracy value of 97% and an F1-Score of 96%.

Testing the MNB and PSO methods for hyperparameter tunning and feature selection with a comparison of training data and test data of 80% and 20% produces an accuracy value of 93% and an F1-Score value of 90%. While testing the MNB and PSO methods for hyperparameter tunning and feature selection with a comparison of training data and test data of 70% and 30% produces an accuracy value of 93% and an F1-Score value of 91%. In scenarios 7, 8, and 9, model testing was carried out with Multinomial Naïve Bayes and using the Particle Swarm Optimization optimization method for hyperparameter tunning. Testing the MNB and PSO methods for hyperparameter tuning with a comparison of training data and test data of 90% and 10% resulted in an accuracy value of 96% and an F1-Score value of 94%. Then testing the MNB and PSO methods for hyperparameter tuning with a comparison of training data and test data of 80% and 20% produces an accuracy value of 93% and an F1-Score value of 90%. Testing the MNB and PSO methods for hyperparameter tuning with a comparison of training data and test data of 70% and 30% resulted in an accuracy value of 92% and an F1-Score value of 89%. In scenarios 10, 11, and 12, model testing was carried out with Multinomial Naïve Bayes and using the Particle Swarm 76 Optimization optimization method for feature selection.

Testing the MNB and PSO methods for feature selection is processed with a comparison of training data and test data of 90% and 10%. This test produces an accuracy value of 95% and an F1-Score value of 92%. Testing the MNB and PSO methods for feature selection with a comparison of training data and test data of 80% and 20% produces an accuracy value of 91% and an F1-Score value of 86%. Meanwhile, testing the MNB and PSO methods for feature selection with a comparison of training data and test data of 70% and 30% produces an accuracy value of 90% and an F1-Score value of 85%. Based on the results of the tests that have been carried out, all accuracy values have increased after optimization. This shows that the Particle Swarm Optimization optimization method has an influence in increasing the accuracy value of the Multinomial Naïve Bayes method for sentiment analysis cases of customer appraisal of the Favehotel Kusumanegara Yogyakarta hotel. The results of the highest accuracy of all tests were obtained in scenario 4. Optimization of hyperparameter tunning and feature selection with an accuracy value of 97% with a total comparison of the distribution of training data and test data of 90% and 10%. The highest accuracy results were obtained in the optimization of hyperparameter tunning in scenario 7, namely 96% with a comparison of the distribution of training data and test data of 90% and 10%. In optimizing feature selection, the highest accuracy is obtained in scenario 10, which is 95% with a comparison of the distribution of training data and test data of 90% and 10%.

3 Conclusion

The distribution of test data and training data has an influence on the accuracy value. The highest accuracy results in this study were obtained in the distribution of data with a ratio of 90% for training data and 10% for test data. The Particle Swarm Optimization optimization method has an effect on increasing the accuracy of the Multinomial Naïve Bayes algorithm in the case of sentiment analysis of hotel customer ratings at Favehotel Kusumanegara Yogyakarta. All accuracy values have increased after optimization. The highest accuracy value of all scenarios is obtained in the hyperparameter tunning and feature selection optimization scenario for Multinomial Naïve Bayes with an accuracy value of 97%. The highest accuracy result in the Naïve Bayes Multinomial test without optimization is 93%. The highest accuracy results in the hyperparameter tuning optimization scenario are 96%. Meanwhile, the highest accuracy results in the feature selection optimization scenario are 95%.

References

1. Chen, M. Y., Teng, C. I., & Chiou, K. W. (2020). The helpfulness of online reviews: Images in review content and the facial expressions of reviewers' avatars. Online Information Review, 44(1), 90-113. https://doi.org/10.1108/OIR-08-2018-0251
2. Dainamang, S. A., Hayatin, N., & Chandranegara, D. R. (2022). Analisis Sentimen Media Sosial Twitter Terhadap RUU Omnibus Law dengan Metode Naïve Bayes dan Particle Swarm Optimization. Komputika: Jurnal Sistem Komputer, 11(2), 211-218. https://doi.org/10.34010/komputika.v11i2.6037
3. Darmawan, R., Indra, & Surahmat, A. (2022). Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru Di Twitter. Jurnal Kajian Ilmiah, 22(2), 143-152. https://doi.org/10.31599/jki.v22i2.1130
4. Hayuningtias, R. Y., & Sari, R. (2019). ANALISIS SENTIMEN OPINI PUBLIK BAHASA INDONESIA TERHADAP WISATA TMII MENGGUNAKAN NAÏVE BAYES DAN PSO. Jurnal TECHNO Nusa Mandiri, 16(1), 37-42. https://doi.org/10.33480/techno.v16i1.115
5. Hendriyanto, M. D., Ridha, A. A., & Enri, U. (2022). ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY

STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. INTECOMS: Journal of Information Technology and Computer Science, 5(1), 1-7. https://doi.org/10.31539/intecoms.v5i1.3708

6. Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). TEKNIKA, 10(1), 18-26. https://doi.org/10.34148/teknika.v10i1.311
7. Indrayuni, E. (2016). Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization. Jurnal Evolusi, 4(2), 20-27. https://doi.org/10.31294/evolusi.v4i2.697
8. Indrayuni, E. (2019). Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes. JURNAL KHATULISTIWA INFORMATIKA, 7(1), 29-36. https://doi.org/10.31294/jki.v7i1.5740
9. Jung, H., & Lee, B. G. (2020). Research Trends in Text Mining: Semantic Network and Main Path Analysis of Selected Journals. Expert Systems with Applications, 162, 1-16. https://doi.org/10.1016/j.eswa.2020.113851
10. Maulana, Y. (2022). ANALISIS SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI PADA TWITTER MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DAN PARTICLE SWARM OPTIMIZATION (Skripsi). UPN "Veteran" Jawa Timur.
11. Miner, G., Hill, T., Nisbet, R., Delen, D., Fast, A., & Elder, J. (2012). Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications (G. Miner, Ed.). Elsevier Science.
12. Ou, G., He, Y., Fournier-Viger, P., & Huang, J. Z. (2022). A Novel Mixed-Attribute Fusion-Based Naive Bayesian Classifier. Applied Sciences, 12(20), 1-16. https://doi.org/10.3390/app122010443
13. Puspita, R., & Widodo, A. (2020). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. Jurnal Informatika Universitas Pamulang, 5(4), 646-654. http://dx.doi.org/10.32493/informatika.v5i4.7622
14. Putri, R. A. (2023). ASPECT-BASED SENTIMENT ANALYSIS PADA ULASAN APLIKASI BCA MOBILE DENGAN MENGGUNAKAN METODE TF-IDF DAN ALGORITMA NAÏVE BAYES CLASSIFIER (Skripsi). UPN "Veteran" Jawa Timur.
15. Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. JURNAL MEDIA INFORMATIKA BUDIDARMA, 5(1), 157-163. http://dx.doi.org/10.30865/mib.v5i1.2580
16. Saraswati, M., & Riminarsih, D. (2020). ANALISIS SENTIMEN TERHADAP PELAYANAN KRL COMMUTERLINE BERDASARKAN DATA TWITTER MENGGUNAKAN ALGORTIMA BERNOULLI NAIVE BAYES. Jurnal Ilmiah Informatika Komputer, 25(3), 225-238. http://dx.doi.org/10.35760/ik.2020.v25i3.3256
17. Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. 2019 International Conference on Automation, Computational and Technology Management (ICACTM), 593-596. https://doi.org/10.1109/ICACTM.2019.8776800
18. Sipayung, E. M., Maharani, H., & Zefanya, I. (2016). PERANCANGAN SISTEM ANALISIS SENTIMEN KOMENTAR PELANGGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER. Jurnal Sistem Informasi (JSI), 8(1), 958-965. https://doi.org/10.36706/jsi.v8i1.3250
19. Sitanayah Que, V. K., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis
20. Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). Jurnal Nasional Teknik Elektro dan Teknologi Informasi, 9(2), 162-170. https://doi.org/10.22146/jnteti.v9i2.102
21. Titania, B. (2020). Penerapan Metode Text Mining dan Social Network Analysis pada Jejaring Sosial Twitter (Studi terhadap: Dugaan Korupsi Asuransi Jiwasraya dan Dugaan Korupsi Asuransi Sosial Angkatan Bersenjata Republik Indonesia). Universitas Islam Indonesia.
22. Trisna Lestari, A. R., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 1(12), 1718-1724. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/627
23. Yulita, W., Nugroho, E. D., & Alghifari, M. H. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. JDMSI: Jurnal Data Mining dan Sistem Informasi, 2(2), 1-9. https://doi.org/10.33365/jdmsi.v2i2.1344