J-naker
Jurnal Ketenagakerjaan

Jurnal Ketenagakerjaan
Volume 19 No. 1, 2024
Online ISSN: 2722-8770
Print ISSN: 1907-6096

Sentiment Analysis Siap Kerja Training Services with Naive Models Bayes and SVM

Arif Budi Setiawan

Ministry of Manpower – Republic of Indonesia

Email Correspondences: arifbudi@kemnaker.go.id

Abstract

Training service Siap Kerja has an important function in improving the quality of human resources and developing a competent workforce in various industrial sectors. To ensure trainee success and satisfaction, sentiment analysis in trainee reviews is essential to evaluate the strengths and weaknesses of the service offered. This research intends to compare the performance of algorithms Support Vector Machine (SVM) and Naive Bayes in sentiment analysis of trainee reviews on pages service training Siap Kerja. Review data was collected using web scraping on the Siap Kerja training service page by collecting ratings and reviews of training participants after completing the training they attended on the page training service Siap Kerja. Methods of data pre-processing, data sharing, feature extraction, model training, model evaluation, and results analysis were used in this research. The research results show that the SVM model has higher accuracy (0.93) compared to the Naive model Bayes (0.80). Additionally, positive sentiment was found at 99%, negative at 0.3%, and neutral at 0.7%. ased on the results of this research, it is recommended that the SVM model be used in sentiment analysis in the context of Siap Kerja training services and that the model be optimized to improve performance in classifying reviews with negative or neutral feelings.

Keywords: skillhub, siapkerja, sentiment analysis, naive bayes, web scraping, trainee reviews

DOI: 10.47198/naker.v19i1.303 Received: 14-3-2024 Revised: 25-4-2024 Accepted: 30-4-2024

1. Introduction

Siap Kerja training services are training programs organized by government or private institutions to improve the skills and competencies of the workforce. This training service aims to help trainees improve their readiness to enter the world of work and meet industry needs.

To ensure trainee success and satisfaction, sentiment analysis on trainee reviews is important to identify the strengths and weaknesses of the services provided. Sentiment analysis research needs to be carried out on Siap Kerja training service reviews for several reasons, including: Sentiment analysis research on Siap Kerja training service reviews is very important for understanding customer satisfaction (Naseem et al., 2021; Zhu et al., 2022), improving service quality (Ain et al., 2017), visualizing user opinions (Naseem et al., 2021), monitoring the effectiveness of training services (Pu et al., 2021), and benchmarking against competitor (Liu et al., 2019). By conducting sentiment analysis, providers can make informed decisions to improve the overall quality and effectiveness of their training services. The introductory section explains the disclosure of background, problems, and objectives accompanied by supporting data for conducting research.

This paper offers a study on sentiment analysis of Siap Kerja training service evaluation using two general machine learning methods, Naive Bayes and Support Vector Machine (SVM). This research intends to compare the performance of algorithms Support Vector Machine (SVM) and Naive Bayes in sentiment analysis of Siap Kerja trainee reviews. Based on the literature assessment, the SVM approach is generally superior to Naive Bayes in many applications of sentiment analysis (Dey et al., 2020; Kristiyanti et al., 2018; Rahat et al., 2019; Shivaprasad & Shetty, 2017; Sugitomo et al., 2021; Surya & Subbulakshmi, 2019).

In previous related research, various studies have been conducted to compare the performance of SVM and Naive Bayes in sentiment analysis, such as in Amazon product evaluations, public opinion about West Java Governor candidates for the 2018-2023 period on Twitter, and restaurant review (Dey et al., 2020; Kristiyanti et al., 2018; Rahat et al., 2019)datasets. The results of this research show that the SVM method tends to provide better performance compared to the Naive method Bayes (Dey et al., 2020; Kristiyanti et al., 2018; Rahat et al., 2019; Shivaprasad & Shetty, 2017; Sugitomo et al., 2021; Surya & Subbulakshmi, 2019).

Based on the theorem Bayes and assuming independence between features, Naive Bayes is a probabilistic classifier (Tripathi et al., 2019). It has been successfully implemented in various domains, including spam filtering (Feng et al., 2016)and prediction of student performance (Tripathi et al., 2019). SVM, on the other hand, is a powerful and adaptable classifier that seeks to identify ideal hyperplanes that subdivide multiple classes in feature space (Marianingsih & Utaminingrum, 2018). It has been used for various classification tasks, such as road surface type classification (Marianingsih & Utaminingrum, 2018)and object classification (Jindal et al., 2023).

This research combines sentiment analysis using SVM and Naive approaches Bayes, as well as using Word2Vec as a feature extension. The originality of this research involves the application of Word2Vec to improve sentiment classification performance in the context of Siap Kerja training services. Word2Vec has proven useful in improving sentiment classification performance for political viewpoints on Twitter, therefore it is expected to provide performance improvements in the context of Siap Kerja training services (Dimas Lutfiyanto & Setiawan, n.d.).

In this study, we will examine the performance of the algorithm Naive Bayes and SVM in the context of sentiment analysis in the evaluation of Siap Kerja training services. The comparison will be based on several evaluation measures, such as accuracy, precision, recall, and F1 score. Additionally, we will analyze the advantages and limitations of each strategy and provide insight into their relevance in real-world circumstances.

In this regard, this research will make an important contribution in finding the most effective method for sentiment analysis in Siap Kerja trainee reviews, as well as provide recommendations for improving the quality of training services.

2. Research Methods

In the context of sentiment analysis in SiapKerja trainee reviews, the research method used involves several stages, from data collection to model evaluation. Method study shown in Figure 1.

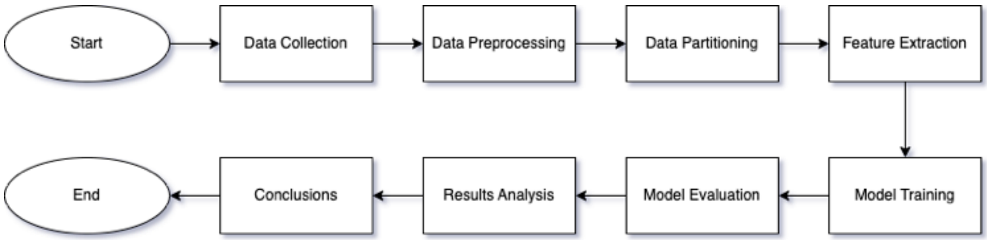

Figure 1. Flowchart Method Study

Data used in the study This originates from reviewing participants who have been finishing their training . Follow that total of 2000 reviews. Data obtained with web scraping on the page service training Siap Kerja, with take participant ratings and reviews training on the page service training Siap Kerja. This data will used as input For training and testing classification models' sentiment.

Data preprocessing is carried out To remove noise, such as deleting characters specifically, changing text to make letters small, and removing the stop word. Next, tokenization and stemming / lemmatization are done To reduce variations of the same word. Data is imported using the Panda's library and read from a CSV file. Review and rating columns were chosen as variable independent and dependent. Stemmer and eraser objects stop words made use of Literati library. Function preprocess defined For do pre-processing text, which includes:

a) Change text to letter small.
b) Wipe non- alphanumeric characters.
c) Separate text into word tokens.
d) Wipe stop words of tokens.
e) Perform stemming on the token.
f) Merge tokens back to text.

A review column is provided as preprocess function. Then it separated into two sets, called training data and testing data. The Classification model trained using training data, and model performance is evaluated using test data. The data separated into training data and test data with ratio 80:20.

Text converted becomes a feature vector that can used by the algorithm classification using Word2Vec. Representation vector from the word that describes it connection semantics between words in text generated by Word2Vec. Text changed becomes a word list for Word2Vec. The Word2Vec model is trained on pre-trained data. The word2vec_mean function is defined For convert a list of words to mean vector. Word list changed become average vector for training data and testing data.

Classification model sentiment trained using training data and existing features extracted with SVM and Naive Bayes algorithms. Algorithm parameters customized For reach performance best. Object GaussianNB made For do classification with naïve Bayes method. The Naïve Bayes model is trained with training data. Predicted test data labels using a naive Bayes model. The SVC object is created For do classification with SVM method. The SVM model is trained with training data. Test data labels predicted using the SVM model

Test data is used For evaluate performance of the trained model. Indicator evaluation like accuracy, precision, recall, and F1-score are calculated For evaluate model performance.

The performance of SVM and Naive Bayes models is compared For determine method best in context service training Siap Kerja. Besides the, influence of Word2Vec is deep increase performance classification sentiment analyzed.

Based on analysis result, created conclusion about method best For analysis sentiment in context service training Siap Kerja. Recommendation For study more carry on or development application practical from selected method given.

3. Results and Discussion

Naive model results The Bayes shown in Figure 2 show lower performance compared to the SVM model. The naive Bayes model has an accuracy of 0.7985714285714286. Although this accuracy is quite high, the model performs poorly in classifying reviews with negative or neutral sentiment. The Fusion Matrix and Classification report shows relatively low performance for classes 1-4, with low f1 scores and low recall. This shows that the Naive Bayes model was less effective at classifying reviews with negative or neutral sentiment.

```
Naive Bayes
Accuracy: 0.8
Confusion Matrix:
 [[  0   0   0   0   0]
 [  0   2   0   0   1]
 [  0   1   1   0   3]
 [  3   9   3   0  29]
 [ 18  55  13   5 557]]
Classification Report:
              precision    recall  f1-score   support

           1       0.00      0.00      0.00         0
           2       0.03      0.67      0.06         3
           3       0.06      0.20      0.09         5
           4       0.00      0.00      0.00        44
           5       0.94      0.86      0.90       648

    accuracy                           0.80       700
   macro avg       0.21      0.35      0.21       700
weighted avg       0.87      0.80      0.83       700
```

Figure 2. *Naïve Bayes results*

In the context of sentiment analysis in Siap Kerja trainee reviews, the Naive model Bayes may not be powerful enough to identify existing patterns in text data. This may be due to several factors, such as the assumption of independence between the features used by Naive Bayes, which may not always be true in the context of text.

Although the overall accuracy of the Naive model Bayes is quite high (around 80%), poor performance in classifying reviews with negative or neutral sentiment suggests that this model may not be suitable for sentiment analysis in this context.

From the results of this research, it can be concluded that the Naive model Bayes was less effective in classifying Siap Kerja trainee reviews with negative or neutral sentiment. Therefore, researchers use another, more effective classification model, namely SVM.

Support Models Vector Machine (SVM) shows better performance compared to Naive model Bayes. The SVM model has an accuracy of 0.9257142857142857, which is higher than the accuracy of the Naive model Bayes. As shown in Figure 3.

```
SVM
Accuracy: 0.9257142857142857
Confusion Matrix:
 [[  0   0   0   3]
 [  0   0   0   5]
 [  0   0   0  44]
 [  0   0   0 648]]
Classification Report:
              precision    recall  f1-score   support

           2       0.00      0.00      0.00         3
           3       0.00      0.00      0.00         5
           4       0.00      0.00      0.00        44
           5       0.93      1.00      0.96       648

    accuracy                           0.93       700
   macro avg       0.23      0.25      0.24       700
weighted avg       0.86      0.93      0.89       700
```

Figure 3. *SVM results*

Fusion's Matrix and Classification reports show better performance than Naive Bayes, especially for class 5, with high f1 scores and high recall. This shows that the SVM model is more effective in classifying reviews with positive sentiment.

In the context of sentiment analysis in Siap Kerja trainee reviews, the SVM model is able to identify existing patterns in text data better than Naive Bayes. This may be due to the ability of SVM to find the best hyperplane that separates classes in a higher feature space, resulting in better classification performance.

As Figure 4 suggests, this study obtained 99% positive, 0.3% negative, and 0.7% neutral sentiment from Siap Kerja training participant reviews. The following is an explanation of the impact of this research on Siap Kerja training services:

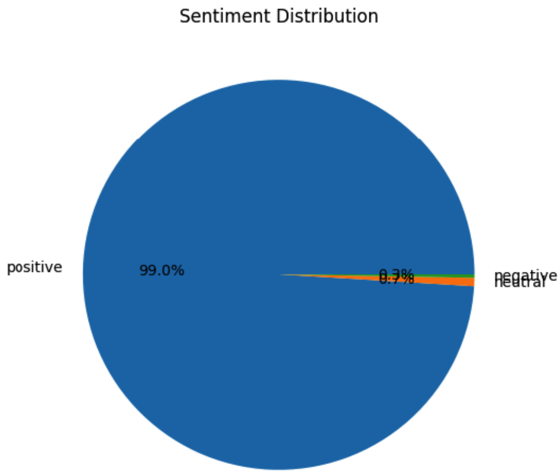

Figure 4. Distribution Sentiment

Positive Sentiment (99%), with a very high percentage of positive sentiment, shows that the majority of trainees are satisfied with Siap Kerja training services. This indicates that the training services provided have succeeded in meeting the needs and expectations of the training participants. In this case, training organizers can continue and maintain the good practices that have been implemented.

Negative Sentiment (0.3%), The low percentage of negative sentiment indicates that only a small percentage of trainees are dissatisfied with Siap Kerja training services. Even though the number is small, training organizers must still pay attention to input from training participants who leave negative reviews. In this case, training providers need to identify and address existing problems or deficiencies in training services to improve overall service quality.

Neutral Sentiment (0.7%), A percentage of neutral sentiment indicates that a small percentage of trainees have a neutral view of Siap Kerja training services. Training providers should consider feedback from trainees who provide neutral reviews and look for ways to improve services to better meet trainees' needs and expectations.

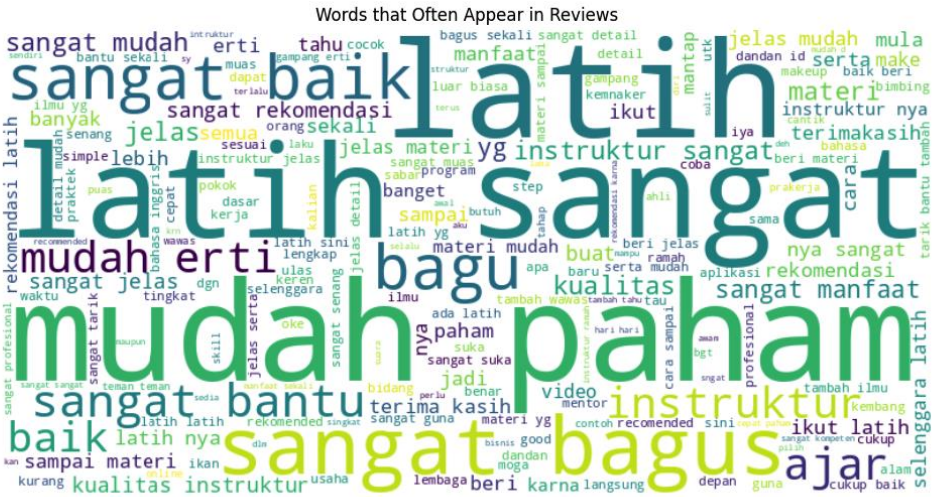

Figure 5. Frequently Appearing Words in Reviews

With a word cloud as shown in Figure 5, words that represent positive sentiment will be displayed with a larger font size, such as very good, easy, understandable and so on. This indicates that the majority of training participants are satisfied with the services provided. Meanwhile, words representing negative and neutral sentiments will be displayed in a smaller font size, indicating that only a small percentage of trainees are dissatisfied or have a neutral view of Siap Kerja training services.

Using word clouds, Siap Kerja training providers can more easily identify trainee satisfaction and understand which service aspects have successfully met trainee needs. Additionally, the cloud can help training providers make more informed decisions to develop and improve the services they offer.

The results of this research show that the majority of training participants are satisfied with the services provided in the context of Siap Kerja training services. However, training providers should remain alert to both negative and neutral reviews and strive to continuously improve the quality of training services. In addition, the results of this research also show that the SVM model is more effective in classifying the sentiment of trainee reviews, so it is recommended that this model be used in sentiment analysis in the context of Siap Kerja training services.

Sentiment analysis of Siap Kerja trainee reviews has important implications in the employment context. By understanding trainee sentiment, training service providers can improve the quality of their programs to better prepare a skilled and Siap Kerja workforce.

Effective training is critical to developing a competent and adaptive workforce. By analyzing participant feedback, training providers can identify areas that need improvement, such as curriculum, instructors, or facilities. This can ultimately increase participants' success in entering the job market and contribute to overall workforce productivity.

Overall, this research highlights the importance of considering participant feedback in designing and improving workforce training programs. By applying sentiment analysis techniques, training providers can take a data- driven approach to improving employment outcomes and supporting the development of a skilled and competitive workforce.

4. Conclusion and Recommendations

Analysis sentiment on skillhub Siap Kerja reviews using the Support method Vector Machine (SVM) and Naive Bayes shows different performance. The SVM model has higher accuracy (0.9257142857142857) compared to the Naive model Bayes (0.7985714285714286). Additionally, the SVM model is more effective in classifying reviews with positive sentiment, whereas Naive's performance Bayes in classifying reviews with negative or neutral sentiment is still low.

Sentiment analysis of Siap Kerja trainee reviews has important implications in the employment context. By understanding trainee sentiment, training service providers can improve the quality of their programs to better prepare a skilled and Siap Kerja workforce.

The research method used involves several stages, starting from data collection to model evaluation. The data used comes from 2000 reviews of trainees who have completed their training. Data preprocessing is carried out to remove noise and prepare data for analysis.

By conducting sentiment analysis, training service providers can make informed decisions to improve the overall quality and effectiveness of their training services. This can help reduce skills gaps and improve the match between job seekers and available job vacancies.

This research makes an important contribution in finding the most effective method for sentiment analysis in Siap Kerja trainee reviews, as well as providing recommendations for improving the quality of training services in the employment context.

References

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment Analysis Using Deep Learning Techniques: A Review. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 8, Issue 6). www.ijacsa.thesai.org

Dessler, G. (2009). Personnel planning and recruiting. In *A framework for human resource management*.

Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020). A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 217–220. https://doi.org/10.1109/IC3A48958.2020.233300

Dimas Lutfiyanto, M., & Setiawan, E. B. (n.d.). *Expansion Feature dengan Word2Vec untuk Analisis Sentimen pada Opini Politik di Twitter dengan Klasifikasi Support Vector Machine, Naïve Bayes, dan Random Forest*.

Feng, W., Sun, J., Zhang, L., Cao, C., & Yang, Q. (2016). A support vector machine based naive Bayes algorithm for spam filtering. *2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC)*, 1–8. https://doi.org/10.1109/PCCC.2016.7820655

Hubbard, G. (2009). Measuring organizational performance: Beyond the triple bottom line. *Business Strategy and the Environment*, *18*(3), 177–191. https://doi.org/10.1002/bse.564

Jindal, P., Parikh, S., Sikka, R., Alatba, S. R., Babu, S., & Sriramakrishnan, G. V. (2023). Analyzing the differences between SVM and Naive Bayes for Feature Extraction. *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 775–778. https://doi.org/10.1109/ICACITE57410.2023.10183068

Kristiyanti, D. A., Umam, A. H., Wahyudi, M., Amin, R., & Marlinda, L. (2018). Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–6. https://doi.org/10.1109/CITSM.2018.8674352

Liu, R., Shi, Y., Ji, C., & Jia, M. (2019). A Survey of Sentiment Analysis Based on Transfer Learning. *IEEE Access*, *7*, 85401–85412. https://doi.org/10.1109/ACCESS.2019.2925059

Marianingsih, S., & Utaminingrum, F. (2018). Comparison of Support Vector Machine Classifier and Naïve Bayes Classifier on Road Surface Type Classification. *2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, 48–53. https://doi.org/10.1109/SIET.2018.8693113

Naseem, S., Mahmood, T., Asif, M., Rashid, J., Umair, M., & Shah, M. (2021, September). *Survey on Sentiment Analysis of User Reviews*. https://doi.org/10.1109/ICIC53490.2021.9693029

Pu, X., Yan, G., Yu, C., Mi, X., & Yu, C. (2021). Sentiment Analysis of Online Course Evaluation Based on a New Ensemble Deep Learning Mode: Evidence from Chinese. *Applied Sciences*, *11*(23). https://doi.org/10.3390/app112311313

Rahat, A. M., Kahir, A., & Masum, A. K. M. (2019). Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset. *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 266–270. https://doi.org/10.1109/SMART46866.2019.9117512

Shivaprasad, T. K., & Shetty, J. (2017). Sentiment analysis of product reviews: A review. *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 298–301. https://doi.org/10.1109/ICICCT.2017.7975207

Sugitomo, J. C., Kevin, N., Jannatri, N., & Suhartono, D. (2021). Sentiment Analysis using SVM and Naïve Bayes Classifiers on Restaurant Review Dataset. *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, *1*, 100–108.

https://doi.org/10.1109/ICCSAI53272.2021.9609776

Surya, P. P. M., & Subbulakshmi, B. (2019). Sentimental Analysis using Naive Bayes Classifier. *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 1–5. https://doi.org/10.1109/ViTECoN.2019.8899618

Tripathi, A., Yadav, S., & Rajan, R. (2019). Naive Bayes Classification Model for the Student Performance Prediction. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, *1*, 1548–1553. https://doi.org/10.1109/ICICICT46008.2019.8993237

Zhu, L., Xu, M., Bao, Y., Xu, Y., & Kong, X. (2022). Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science*, *8*. https://doi.org/10.7717/PEERJ-CS.1044