DIJEMSS: Dinasti International Journal of Education Management and Social Science
E-ISSN: 2686-6331
P-ISSN: 2686-6358
DINASTI PUBLISHER
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/dijemss.v7i1.5119
https://creativecommons.org/licenses/by/4.0/
Received: 07 July 2025 | **Revised:** 21 September 2025 | **Accepted:** 09 October 2025

Comparative Study of the Performance of Naïve Bayes, SVM, and K-NN Algorithms for Sentiment Analysis and Topic Modeling of #KaburAjaDulu Hashtags

Sonia Tikamidia[1], Imam Yuadi[2]
[1]Airlangga University, Surabaya, Indonesia, sonia.tikamidia-2024@pasca.unair.ac.id
[2]Airlangga University, Surabaya, Indonesia, imam.yuadi@fisip.unair.ac.id

Corresponding Author: sonia.tikamidia-2024@pasca.unair.ac.id[1]

Abstract: The #KaburAjaDulu hashtag phenomenon that has been widely discussed on platform X reflects the increasing anxiety of Indonesia's younger generation towards socio-economic conditions and the direction of state policy. This research aims to assess public perception of the hashtag through sentiment analysis and topic modeling approaches. Data was collected from X users' tweets from May to June 2025. The methods used include text preprocessing, sentiment classification using Naïve Bayes, SVM, and K-NN algorithms, and topic modeling with Latent Dirichlet Allocation (LDA). The analysis results show that SVM performs best with 98.93% accuracy and optimal precision-recall balance. The Naïve Bayes model also shows competitive results but tends to favour positive classes. In contrast, K-NN showed the lowest performance due to its inability to overcome the curse of dimensionality in TF-IDF representation. LDA topic modeling identified three main themes: the employment crisis, distrust of institutions due to corruption, and the nationalism vs. migration dilemma. These three topics indicate deep psychological conflicts experienced by youth. The findings support the Self-Determination Theory, which emphasizes the importance of autonomy, competence, and social connection for individual attachment to the environment. Lack of fulfilment of these needs triggers migration intentions as a form of escape or adaptive strategy. This research provides a practical contribution to designing HR policies based on social data. In addition, this approach can be used as the basis for a real-time public perception monitoring system.

Keywords: #KaburAjaDulu, Sentiment Analysis, LDA

INTRODUCTION

Indonesia has the fourth-largest number of social media users in the world, after China, India, and the United States. Most of the population uses social media as one of the leading platforms to voice their aspirations. Starting in February 2025, the hashtag #KaburAjaDulu began to appear on platform X, amid a wave of student and community protests against economic uncertainty, unemployment, and policies that were considered unfair (Marciano, 2025). This reflects public disappointment with domestic socio-economic conditions. This

phenomenon reflects a moral contradiction of love for the homeland but a strong desire to seek opportunities and stability abroad.

Starting from this phenomenon, it is essential to analyze the sentiment of tweets on X to understand the feelings of young people, whether leaving is an escape, a protest, or an adaptation strategy. To answer this question, an approach that can extract information from textual data is needed, namely through text mining. Text mining is a multidisciplinary process that aims to discover patterns, trends, and hidden knowledge from unstructured text data by utilizing techniques from various fields, such as Natural Language Processing (NLP), machine learning, and statistics (Aggarwal & Zhai, 2012). Text mining allows the aspirations and hidden meanings of tweets on platform X to be analyzed systematically and objectively (Xu et al., 2022). Previous studies have demonstrated the effectiveness of NLP and machine learning methods in analyzing public opinion on platform X regarding national issues. For example, Zaqy et al. (2024) used Multinomial Naïve Bayes to analyze the sentiment of netizen comments on Rohingya refugees on Twitter and found a dominant negative signal of 81%. For example, Dey et al. (2016) successfully classified product review sentiment using a combination of Naïve Bayes and kNN. Meanwhile, Munawaroh & Alamsyah (2023) compared the performance of SVM, Naïve Bayes, and kNN in reading public opinion on the COVID-19 vaccine on Twitter and found that SVM excels in accuracy. Zerrouki et al. (2020) compared three algorithms, namely Naïve Bayes, kNN, and Decision Tree, for sentiment analysis on Twitter and found that Naïve Bayes and kNN provide high accuracy of nearly 90%. Furthermore, Huq et al. (2017) used Twitter data to compare the performance of SVM and kNN and found that both effectively recognized public sentiment, with kNN showing superiority in certain situations. These studies prove that text mining is an appropriate method to understand the dynamics of public opinion on social media, including sensitive and socio-politically charged issues.

Few studies have specifically investigated the sentiment of Indonesians towards the hashtag #KaburAjaDulu on platform X. Previous studies focus on significant issues such as politics or disasters, and few have examined grassroots phenomena such as this hashtag. For example, Agustina & Hendry (2021) research is effective in sentiment classification but does not address the topic modeling or LDA aspects to explore the deeper context of tweets. On the other hand, (Wati & Yusuf, 2021)used LDA to identify topics from tweets. However, the focus was on different issues, such as public policy and did not involve integrated sentiment analysis. Research on #KaburAjaDulu in X has been conducted by Khristianto et al. (2025); this study used a qualitative Naïve Bayes approach for sentiment analysis. However, the study did not include topic modeling to explore the important issues behind the sentiment on the X platform.

Based on the literature review of previous studies, there is a significant research gap, namely that no study has comprehensively integrated sentiment analysis and topic modeling to understand the discourse of migration aspirations among Indonesians, especially those expressed through the hashtag #KaburAjaDulu on the X platform. Departing from the literature gap, this research aims to answer the questions of how the perceptions of Indonesian youth towards the #KaburAjaDulu phenomenon can be classified into positive and negative sentiment categories using machine learning algorithms and what are the main topics that trigger the emergence of positive and negative sentiments behind the #KaburAjaDulu hashtag based on topic modeling analysis (LDA).

This research will use a comprehensive text-mining methodology to answer these questions. The structure of this paper is divided into several parts: Data Collection, where tweets from platform X are obtained using crawling techniques; Data Pre-processing by cleaning and normalizing text; modeling by applying Naïve Bayes, SVM and K-NN machine learning classification algorithms for sentiment analysis, and using LDA for topic modeling; Model evaluation by comparing algorithm performance and analyzing sentiment results and topics found; and Visualization by presenting findings in visual form. The contribution of this research is expected to be in the form of mapping sentiments and topics underlying youth aspirations,

insight into the urgency of policies to maintain the talents of the nation's children, and analysis models that can be used for similar issues in the future.

METHOD

This research uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework to build a structured and organized sentiment analysis procedure (IBM, 2021).

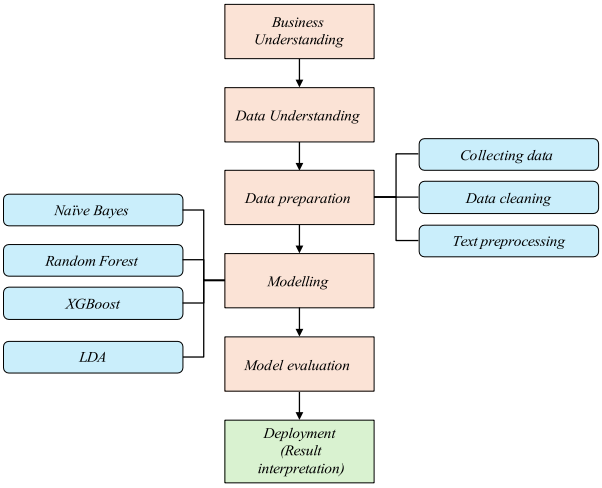

Figure 1. Process Flow using CRISP-DM

Business Understanding

The #KaburAjaDulu hashtag phenomenon on social media reflects the passion and disillusionment among Indonesia's younger generation towards domestic conditions. Understanding the sentiment and context behind this hashtag is crucial for policymakers and stakeholders to identify root causes and formulate relevant solutions. Manual analysis of large volumes of data is inefficient and prone to bias. Therefore, this research aims to automate sentiment analysis and map the key topics behind the hashtags to provide objective insights into the perceptions of youth and the important topics they are voicing.

Data Understanding

The data used comes from social media X (Twitter), which collected tweets containing the hashtag #kaburajadulu. The data was collected using the social media API during the May - June 2025 time period. The total number of tweets obtained was 1,005, but adjustments were made to get tweets that were purely related to the phenomenon to be studied, namely 552 tweets. The language used is a mixture of Indonesian and English.

Table 1. Sample Data Tweets on X

User ID	Tweets on X
1,8602E+18	@azeotrop_ Ayo #kaburajadulu gaji luar negeri lebih menggoda
1,17385E+18	Makin menjadi aja ini a little hope for #kaburajadulu mana orang tua juga mau support ya allah
224904950	Ga pengen ngomporin orang buat #KaburAjaDulu jalur WHV Australia tapi negara ini sudahlah dekat dari Indonesia gajinya melebihi US dan kalo di luar jam kerja kamu ga bakalan direcokin urusan kerjaan di WA. https://t.co/pRRwiNiNMQ

Data Preparation

This stage involves collecting data obtained from data crawling on X and then cleaning and transforming the raw data into a format ready for modeling. The preprocessing steps that will be carried out include:

a. Collecting tweet data in Excel form
b. Removing irrelevant data by removing duplicate tweets, retweets without comments, links, irrelevant hashtags, and special characters such as emojis, symbols, or unnecessary punctuation.
c. Tokenization involves breaking down each tweet into words (tokens). For example, "gaji luar negeri lebih menggoda" becomes ["gaji", "luar", “negeri”, “lebih”, “menggoda”].
d. Transform Cases involves converting all letters into lowercase letters to homogenize the data. For example, "Gaji," "GAJI," becomes "gaji".
e. Filter Stopwords, which is removing words that do not have significant meaning for sentiment analysis, such as "dan", "di", "yang", etc. In this research, the stopwords used sastrawi stopwords.
f. Feature Extraction, i.e. converting text into numerical representation using TF-IDF (Term Frequency-Inverse Document Frequency) or Word2Vec method to give weight to each word based on its relevance in the text.
g. SMOTE is an oversampling technique used to rebalance the class distribution in the dataset by equalizing the number of positive and negative samples.

Modeling

This research will use three modeling approaches, namely classification to label sentiment, text classification to predict sentiment from testing data, and topic modeling using Python applications, as shown in Figure 2. Before modeling, the data is divided into 2, namely training data and testing data, with a division of 80. Labelling is done manually by researchers to create training data. The labels used are positive labels that contain positive comments towards Indonesia or show optimism about building the country and negative labels that show disappointment, frustration, or desire to leave Indonesia, as shown in Table 2. After that, in the prediction model stage, the labelled training data will be used to train the model. This research will compare the performance of three machine learning classification algorithms: Naïve Bayes, SVM and K-NN. The model will be trained to predict the sentiment label of the testing data. The topic modeling stage will be divided into three main topics often discussed in X tweets.

Table 2. Results of Labeling on Tweets

Tweets on X	Entiment Label
Kalo saat ini di Nusantara ada two state Indonesia dan Hindia Belanda gw jamin mayoritas warganya lebih pilih ikut pemerintahan Hindia Belanda daripada pemerintahan Indonesia Sekarang aja di bawah pemerintahan Indonesia pada pingin KaburAjaDulu	Positive
emang lu udah ngasih apa kenegara gw rasa lu cuma beban negara mening lu yang kaum lu sering koar koar eh lupa yang diluar juga pada balik	Negative

```
nb_model = MultinomialNB()
nb_model.fit(X_train_bal, y_train_bal)

svm_model = LinearSVC()
svm_model.fit(X_train, y_train)

knn_model = KNeighborsClassifier(n_neighbors=5)
knn_model.fit(X_train_bal, y_train_bal)
```

```
KNeighborsClassifier
KNeighborsClassifier()
```

```
def evaluate_model(y_true, y_pred, model_name):
    print(f"\n=== Evaluasi Model: {model_name} ===")
    print("Accuracy :", accuracy_score(y_true, y_pred))
    print("Precision:", precision_score(y_true, y_pred, average='weighted'))
    print("Recall   :", recall_score(y_true, y_pred, average='weighted'))
    print("F1 Score :", f1_score(y_true, y_pred, average='weighted'))
    print("\nClassification Report:\n", classification_report(y_true, y_pred, target_names=le.classes_

    cm = confusion_matrix(y_true, y_pred)
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=le.classes_, yticklabels=le.classes
    plt.title(f'Confusion Matrix - {model_name}')
    plt.xlabel('Predicted')
    plt.ylabel('Actual')
    plt.show()

# Prediksi
y_pred_nb = nb_model.predict(X_train_bal)
y_pred_svm = svm_model.predict(X_train_bal)
y_pred_knn = knn_model.predict(X_train_bal)


# Inisialisasi dictionary untuk menyimpan metrik
evaluation_scores = {}

# Evaluasi
evaluate_model(y_train_bal, y_pred_nb, "Naive Bayes")
evaluate_model(y_train_bal, y_pred_svm, "Support Vector Machine")
evaluate_model(y_train_bal, y_pred_knn, "K-Nearest Neighbors")
```

Figure 2. Prediction Model using Naïve Bayes, SVM, and K-NN Algorithms and Topic modeling using LDA

Evaluation

Evaluation is done to measure the performance of the built model and validate its reliability. The review is done by looking at the confusion matrix to calculate each algorithm's accuracy, precision, and recall. After that, the metrics from Naïve Bayes, SVM and K-NN will be compared to determine which algorithm is most accurate in classifying netizen sentiment. Meanwhile, the evaluation of LDA is seen from the topics generated and then interpreted.

Deployment

In the deployment stage, sentiment analysis and topic modeling results will be visualized to provide actionable insights into the perceptions and topics behind the #KaburAjaDulu desire, which will be discussed in depth in Results and Discussion.

RESULTS AND DISCUSSION

After training the model using 448 tweets consisting of 376 positively labelled tweets and 72 negatively labelled tweets and then performing the SMOTE technique to develop the distribution of positive and negative sentiments, the performance evaluation results of each algorithm are obtained based on accuracy, precision and recall metrics.

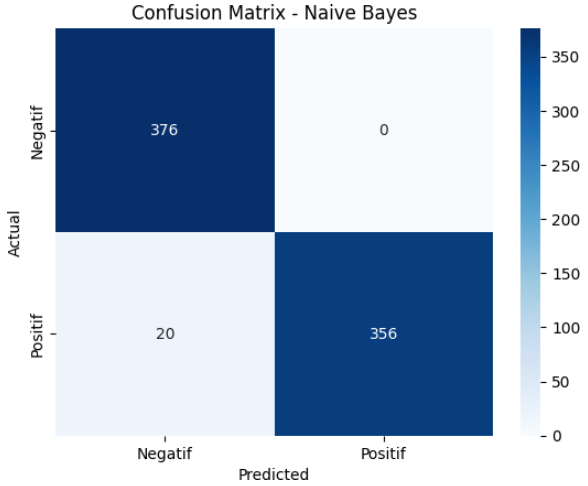

Figure 3. Confusion Matrix of Naïve Bayes Algorithm

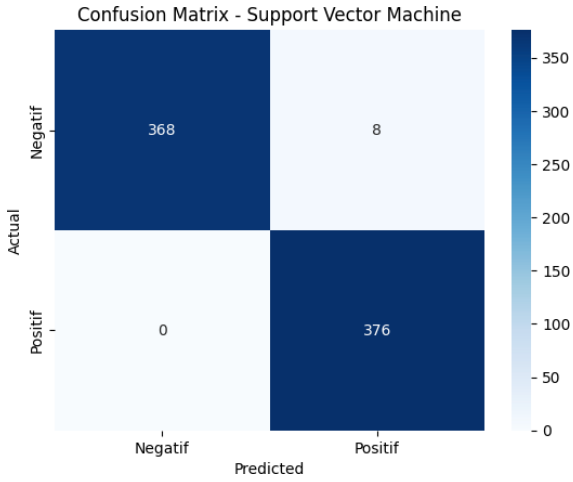

Figure 4. Confusion Matrix of SVM Algorithm

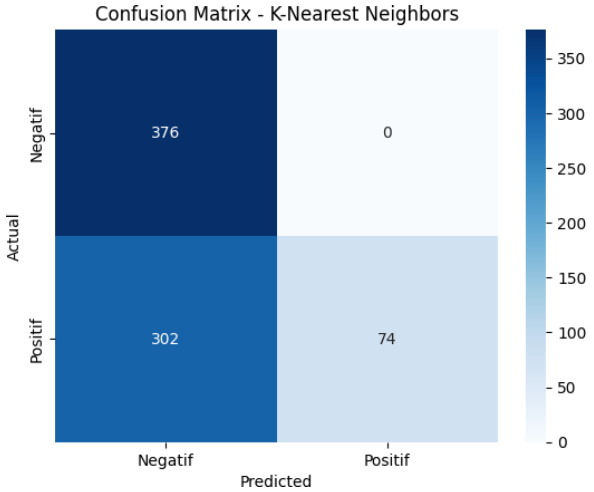

Figure 5. Confusion Matrix of K-NN Algorithm

Table 3. Visualization of Algorithm Comparison

Algorithm	Accuracy	Precision (%)	Recall
Naïve Bayes	97,34	97,47	97,34
SVM	98,93	98,95	98,93
KNN	59,84	77,72	59,84

The evaluation results show that the three algorithms have very different performance: Naïve Bayes recorded 97.34% accuracy, 97.47% precision, and 97.34% recall, while SVM excelled the highest with 98.94% accuracy, 98.96% precision, and 98.94% recall. In contrast, KNN has a low performance with an accuracy of only 59.84%, precision of 77.73%, and recall of 59.84%. From the confusion matrix, it can be seen that Naïve Bayes and SVM capture both positive and negative classes equally. At the same time, KNN successfully detects negative classes where negative recall is 100% but fails to recognize many positive labels where positive recall is 20%. This finding is relevant to the TF-IDF-based tweet analysis, where linear models such as Naïve Bayes and SVM are more stable. In contrast, distance-based methods such as KNN tend to be biased towards the dominant class due to the texture distribution of the features (Le & Nguyen, 2015).

Naïve Bayes shows high and balanced performance, which is especially suitable for short texts such as tweets. This aligns with research from Transiska et al. (2024), who recorded 96.55% accuracy when analyzing opinions about ChatGPT using Naïve Bayes on Twitter. Research on the sentiment of the COVID-19 vaccine also shows that the accuracy of Naïve Bayes reaches 73%-91 (Ressan & Hassan, 2022; Villavicencio et al., 2021). A similar implementation was also used for sentiment analysis of presidential elections with Naïve Bayes, resulting in 96.63% accuracy (Tamara et al., 2023). This consistency is supported by global literature that states Naïve Bayes is effective for text classification with sparsity data (Liu et al., 2013).

SVM appears to be the most dominant, with almost 99% accuracy. Global studies such as Mujahid et al. (2021) and Pavitha et al. (2022) show SVM with TF-IDF getting an accuracy of 90 - 98%. The SVM approach also successfully classified COVID-19 tweets and movies recommendation with accuracy in the 91-98% range (Nurkholis et al., 2022; Rahardi et al., 2022). The superiority of SVM is explained by its ability to maximize margins in high-feature spaces, so it is often the first choice for short and high-dimensional text (AlBadani et al., 2022; Bansal et al., 2022).

KNN shows low performance, with an accuracy of only 59.84% and a very low positive recall of 20%. These results are consistent with the study Bansal et al. (2022), where KNN has an accuracy below Naïve Bayes tweet classification. Aprilianti et al. (2025) also noted that KNN precision was only 66.9%, while Naïve Bayes was 73.2%. The vulnerability of KNN to the curse of dimensionality in TF-IDF space causes it to fail to recognize minority classes effectively (Pestov, 2011; Raeisi Shahraki et al., 2017).

When compared, SVM is the best algorithm based on 98.94% accuracy and class balance. Naïve Bayes excels as a highly efficient option with fast computation and performance close to SVM, making it ideal for real-time monitoring. Global research supports this, with studies Mujahid et al. (2021) and Transiska et al. (2024) illustrating the reliability of Naïve Bayes implementation. In contrast, KNN is not recommended due to its bias towards majority classes and sensitivity to feature sparsity. Therefore, SVM becomes the primary model with Naïve Bayes as a practical alternative. At the same time, KNN is not feasible for sentiment analysis of X tweets based on existing empirical evidence. This finding aligns with recent studies that consistently rank SVM and Naïve Bayes above KNN in both accuracy and robustness on short-text datasets. Considering the dynamic nature of social media content, selecting a model with strong generalization and minimal tuning like SVM is critical for sustainable deployment (Defit et al., 2024; Israt Jahan et al., 2024).

Table 4. Topic Modeling Visualization Result using LDA

Topic	Word
0	Job, nasionalisme, bobrok
1	Hati, koruptor, bilang
2	Negara, Indonesia, kaburajadulu

Latent Dirichlet Allocation (LDA) results on #KaburAjaDulu tweets revealed three main topics: job, nationalism, and dilapidated; heart, corruptor, and said; country, Indonesia, and kaburajadulu. The first topic shows the young generation's frustration with the inability of the national employment system to meet the expectations of a decent career while still maintaining the values of love for the country. This finding is similar to a study Yunitasari et al. (2021) which found that disappointment with domestic job opportunities triggered migration thoughts among millennials. The second topic, showing an emotional reaction to the issue of corruption, is in line with the findings of Zulqarnain et al. (2022), who highlighted that moral damage by corruptors could trigger collective alienation for the younger generation. Meanwhile, the third topic, which links the state, Indonesia, and escapes, illustrates the dilemma of nationalism versus the desire to release attachment to the state. This aligns with the rising phenomenon of human capital flight in Indonesia, where many young professionals, especially those aged 25-35, choose to change citizenship or work abroad.

The interpretation of these topics has strong relevance to HR development theory, particularly Self-Determination Theory (SDT) by Ryan and Deci (1985), which asserts that individuals need autonomy, competence, and emotional connectedness to remain performant and loyal. Identity and motivational conflicts arise When these needs are unmet due to unfair career competition, corruption that undermines trust, and unsupportive public policies, as seen through the #KaburAjaDulu hashtag. Research by (Indrawati & Kuncoro, 2021) supports this view, finding that efficient human capital development requires psychological justice and empowerment, not just technical training. Therefore, the #KaburAjaDulu phenomenon can be interpreted as a signal for governments and organizations to focus on improving technical skills and creating a fair and empowering social and work environment. In addition, brain drain theory warns that systemic failures in meeting HR needs will encourage talent leakage, which is now increasingly evident in the #KaburAjaDulu trend. These three topics not only map the public conversation but also present the state of the art of Indonesia's problematic HR development, namely the need for policy design and organizational culture that facilitates three basic psychological needs so that young people feel engaged, motivated, and have a confident future in the country.

CONCLUSION

Based on the results of sentiment analysis and topic modeling of tweet data on platform X using the hashtag #KaburAjaDulu, this study found that the perceptions of Indonesian youth towards the phenomenon can be effectively classified into two main sentiment categories: positive and negative through the application of machine learning algorithms. Among the three algorithms tested, SVM proved to be the most balanced and consistent model, with 98.93% accuracy, 98.95% precision, and 98.93% recall, making it the most appropriate algorithm for real-time classification of short texts typical of social media. SVM also performed well but had weaknesses in correctly classifying positive sentiment or maintaining balanced precision and recall. In contrast, the K-Nearest Neighbors (KNN) algorithm showed the lowest performance with an accuracy of only 59.84% due to its sensitivity to data distribution and its inability to overcome the curse of dimensionality in TF-IDF representation.

Through topic modeling analysis using LDA, three main topics were identified behind the #KaburAjaDulu hashtag: (1) disappointment with the national employment system (‘job’, ‘nationalism’, ‘dilapidated’); (2) emotional anger toward corruption and social inequality (‘heart’, ‘corruptor’, ‘said’); and (3) identity conflict between love for the country and the desire to leave (‘country’, ‘Indonesia’, ‘kaburajadulu’). n and social inequality with the keywords heart, corruptor, said; and identity conflict between love for the country and the desire to leave with the keywords country, Indonesia, kaburajadulu. These three topics show that negative sentiments are triggered mainly by a crisis of trust in state institutions and systemic failures to fulfil young people's basic needs for a career, justice and a future. On the other hand, positive

sentiments emerge from Indonesian youth who see migration as an adaptive strategy and a form of seeking a better quality of life, not merely an escape.

This finding confirms the relevance of the Self-Determination Theory, which states that when autonomy, competence, and social connectedness are not fulfilled in the domestic system, there will be a strong urge to seek actualization space abroad. Thus, the #KaburAjaDulu phenomenon is not just a viral hashtag but an essential indicator for governments and organizations to evaluate the success of HR development strategies, especially in responding to the aspirations and dissatisfaction of Indonesia's younger generation..

REFERENCES

Aggarwal, C., & Zhai, C. (2012). *Mining Text Data*. https://doi.org/10.1007/978-1-4614-3223-4_6

Agustina, M., & Hendry, H. (2021). Sentimen Masyarakat Terkait Perpindahan Ibukota Via Model Random Forest dan Logistic Regression. *AITI*, *18*, 111–124. https://doi.org/10.24246/aiti.v18i2.111-124

AlBadani, B., Shi, R., & Dong, J. (2022). A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating the Universal Language Model Fine-Tuning and SVM. *Applied System Innovation*, *5*(1). https://doi.org/10.3390/asi5010013

Aprilianti, H., Mustofa, H., Umam, K., & Handayani, M. R. (2025). Comparative Study of SVM, KNN, and Naïve Bayes for Sentiment Analysis of Religious Application Reviews. *Journal of Applied Informatics and Computing (JAIC)*, *9*(3), 2548–6861. http://jurnal.polibatam.ac.id/index.php/JAIC

Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071. https://doi.org/https://doi.org/10.1016/j.dajour.2022.100071

Defit, S., Windarto, A. P., & Alkhairi, P. (2024). Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization. *Telematika*, *17*(1), 52–67. https://doi.org/10.35671/telematika.v17i1.2824

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-NN Classifier. *International Journal of Information Engineering and Electronic Business*, *8*(4), 54–62. https://doi.org/10.5815/ijieeb.2016.04.07

Huq, M. R., Ali, A., & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM. *International Journal of Advanced Computer Science and Applications*, *8*(6), 19–25. https://doi.org/10.14569/ijacsa.2017.080603

IBM. (2021). *A practical guide to the Cross-Industry Standard Process for Data Mining (CRISP-DM)*.

Indrawati, S., & Kuncoro, A. (2021). Improving Competitiveness Through Vocational and Higher Education: Indonesia’s Vision For Human Capital Development In 2019–2024. *Bulletin of Indonesian Economic Studies*, *57*, 29–59. https://doi.org/10.1080/00074918.2021.1909692

Israt Jahan, Md Nakibul Islam, Md Mahadi Hasan, & Md Rafiuddin Siddiky. (2024). Comparative analysis of machine learning algorithms for sentiment classification in social media text. *World Journal of Advanced Research and Reviews*, *23*(3), 2842–2852. https://doi.org/10.30574/wjarr.2024.23.3.2983

Khristianto, R., Chalvani, R., Alauddin Ramadhan, H., & Rakhmawati, N. (2025). *Analisis Sentimen terhadap #KaburAjaDulu sebagai Ekspresi Frustrasi Ekonomi Anak Muda dalam Sosial Media X*. https://doi.org/10.5281/zenodo.15639676

Le, B., & Nguyen, H. (2015). Twitter Sentiment Analysis Using Machine Learning Techniques. *Advances in Intelligent Systems and Computing*, *358*, 1–415. https://doi.org/10.1007/978-3-319-17996-4

Liu, B., Blasch, E., Chen, Y., Shen, D., & Chen, G. (2013). Scalable sentiment classification for Big Data analysis using Naïve Bayes Classifier. *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 99–104. https://doi.org/10.1109/BigData.2013.6691740

Marciano, H. (2025). *Kabur Aja Dulu, bentuk frustasi generasi muda terhadap kondisi bangsa*.

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during covid-19. *Applied Sciences (Switzerland)*, *11*(18). https://doi.org/10.3390/app11188438

Munawaroh, K., & Alamsyah, A. (2023). Performance Comparison of SVM, Naïve Bayes, and KNN Algorithms for Analysis of Public Opinion Sentiment Against COVID-19 Vaccination on Twitter. *Journal of Advances in Information Systems and Technology*, *4*(2), 113–125. https://doi.org/10.15294/jaist.v4i2.59493

Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI*, *6*(2), 227–233. https://doi.org/10.29207/resti.v6i2.3906

Pavitha, N., Pungliya, V., Raut, A., Bhonsle, R., Purohit, A., Patel, A., & Shashidhar, R. (2022). Movie recommendation and sentiment analysis using machine learning. *Global Transitions Proceedings*, *3*(1), 279–284. https://doi.org/10.1016/j.gltp.2022.03.012

Pestov, V. (2011). Is the -NN classifier in high dimensions affected by the curse of dimensionality? *Computers & Mathematics with Applications*, *65*. https://doi.org/10.1016/j.camwa.2012.09.011

Raeisi Shahraki, H., Pourahmad, S., & Zare, N. (2017). K Important Neighbors: A Novel Approach to Binary Classification in High Dimensional Data. *BioMed Research International*, *2017*, 7560807. https://doi.org/10.1155/2017/7560807

Rahardi, M., Aminuddin, A., Abdulloh, F. F., & Nugroho, R. A. (2022). Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia. *International Journal of Advanced Computer Science and Applications*, *13*(6), 534–539. https://doi.org/10.14569/IJACSA.2022.0130665

Ressan, M. B., & Hassan, R. F. (2022). Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets. *Indonesian Journal of Electrical Engineering and Computer Science*, *28*(1), 375–383. https://doi.org/10.11591/ijeecs.v28.i1.pp375-383

Ryan, R. M., & Deci, E. L. (1985). *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being Self-Determination Theory*. Ryan.

Tamara, R., Putro, H. P., & Herry Wahyono. (2023). Analisis Sentimen Terhadap Pilpres 2024 Berdasarkan Opini Dari Twitter Menggunakan Naïve Bayes Dan Svm. *Teknokris*, *26*(1), 23–32. https://doi.org/10.61488/teknokris.v26i1.245

Transiska, D., Febriawan, D., & Hasan, F. N. (2024). Analisis Sentimen Terhadap Penggunaan Chatgpt Berdasarka Twitter Menggunakan Algoritma Naïve Bayes. *Jurnal Media Informatika Budidarma*, *8*(2), 1077. https://doi.org/10.30865/mib.v8i2.7540

Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G. (2021). Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes. *Information (Switzerland)*, *12*(5). https://doi.org/10.3390/info12050204

Wati, P., & Yusuf, M. (2021). Analisis Topik Modelling Terhadap Penggunaan Sosial Media Twitter oleh Pejabat Negara. In *Building of Informatics, Technology and Science (BITS)* (Vol. 3). https://doi.org/10.47065/bits.v3i3.1012

Xu, Q., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, *3*, 100073. https://doi.org/10.1016/j.dajour.2022.100073

Yunitasari, D., Khotimah, K., & Fathorrazi, M. (2021). The Implication of Brain Gain on Brain Drain Phenomenon in Overcoming The Problem of Educated Unemployment in Indonesia. *Sosiohumaniora*, *23*(1), 133. https://doi.org/10.24198/sosiohumaniora.v23i1.26749

Zaqy, M., Marlina, L., & Farta Wijaya, R. (2024). Analysis of Indonesian Netizen Sentiment on Platform X Regarding the Arrival of Refugees in Indonesia Using the Multinominal Naive Bayes Method. *Sinkron*, *8*, 1945–1952. https://doi.org/10.33395/sinkron.v8i3.13940

Zerrouki, K., Hamou, R. M., & Rahmoun, A. (2020). Sentiment Analysis of Tweets Using Naïve Bayes, KNN, and Decision Tree. *International Journal of Organizational and Collective Intelligence*, *10*(4), 35–49. https://doi.org/10.4018/ijoci.2020100103

Zulqarnain, Z., Ikhlas, M., & Ilhami, R. (2022). Perception of college students on civic and anti-corruption education: Importance and relevance. *Integritas : Jurnal Antikorupsi*, *8*(1), 123–134. https://doi.org/10.32697/integritas.v8i1.854