

Analyzing Public Sentiment Toward Makanan Bergizi Gratis Program Using Machine Learning

Musriatun Napiah^{1*}, Sujiliani Heristian², Mugi Raharjo³, Rachmat Adi Purnama⁴

^{1,4}Informatics Department, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, RT.2/RW.9, Senen, Kota Jakarta
Pusat, Daerah Khusus Ibukota Jakarta, Indonesia

²Informatics Department, Faculty of Engineering and Informatics, Universitas Bina
Sarana Informatika
Jl. Banten No. 1, Karangpawitan, Kec. Karawang Barat, Kabupaten Karawang, Jawa Barat

³Informatics Department, Faculty of Information Technology, Universitas Nusa Mandiri
Jl. Jatiwaringin Raya No. 02 RT 08 RW 13 Kelurahan Cipinang
Melayu, Kecamatan Makasar Jakarta Timur

e-mail: 1musriatun.mph@bsi.ac.id, 2sujiliani.she@bsi.ac.id, 4rachmat.rap@bsi.ac.id
3mugi.mou@nusamandiri.ac.id

(*) Corresponding Author

Article Info: Received: 17-10-2025 | Revised : 10-11-2025 | Accepted : 21-11-2025

Abstrak - Makanan Bergizi Gratis (MBG) program is a strategic initiative of the Indonesian government to improve the nutritional quality of schoolchildren. This research seeks to examine public sentiment regarding the MBG program by leveraging 10,000 tweets obtained from Kaggle. The method used combines Natural Language Processing (NLP) and Machine Learning approaches, several algorithms such as Logistic Regression, Support Vector Machine (SVM), Random Forest, Naive Bayes, XGBoost, and LightGBM were tested to compare classification performance. The dataset contains a collection of public reviews categorized into three sentiment classes: positive, negative, and neutral. The analysis process includes text cleaning, tokenization, stopword removal, and stemming to obtain a cleaner text representation. Text features were then extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The results showed that the Logistic Regression 97% with an F1-score of 0.9552 models showed the most optimal performance. Sentiment analysis revealed 65% positive responses, 25% neutral, and 10% negative, with the dominant keywords being “nutrisi,” “sehat,” “anak sekolah,” and “gratis.” The results visualization, in the form of a Word Cloud and a bar chart, indicate that public opinion tends to be positive towards the implementation of the MBG program, particularly regarding improving the nutrition of schoolchildren. This research is expected to provide input for policymakers in evaluating public perceptions of the implementation of food-based social programs.

Keywords: MBG, Machine Learning, Natural Language Processing

INTRODUCTION

Makanan Bergizi Gratis Program (MBG) is one of the Indonesian government's policies aimed at improving the quality of school nutrition. Along with its implementation, public response to this program has spread widely on social media, especially X (Twitter). This program is expected to help reduce malnutrition rates and support the achievement of sustainable development goals (Sustainable Development Goals) in the health and education sectors. That there are 30% of children in Indonesia under the age of 5 have experienced stunting due to malnutrition (Hidayatullah, 2024) As the stunting rate in Indonesia in 2024 was 14%. Starting from this concern, when Prabowo served as President of the Republic of Indonesia tried to realize the idea of free meals that would be provided to students throughout Indonesia. Furthermore, the free meal idea was called Makan Bergizi Gratis (MBG) (Prasetyo, 2023)

The MBG program launched by the government is a program with a noble goal, it is in line with the spirit of the fifth principle of Pancasila, namely "Social justice for all Indonesian people." The government strives to



meet the nutritional needs of all students equally, from remote areas such as the 3T (underdeveloped, frontier, infected) to big cities without regard to socioeconomic background. Every student has the same right to be able to enjoy nutritious food. The government's steps reflect equitable social justice because all students are treated equally and equally to achieve shared prosperity. (Prasetyo, 2023).

Nutrition plays a crucial role in a child's development. By meeting nutritional needs, such as carbohydrates as a source of energy, protein as a building block, and vitamins and minerals as regulators, a child's health will be maintained. This will help prevent various diseases that can hinder their growth and development, which in turn can affect their intelligence (Andreas Halomoan Tambunan et al., 2025). However, public reception of this program has been mixed. Some consider the MBG program very beneficial, while others criticize its effectiveness and implementation. Therefore, sentiment analysis of these conversations can provide insight into the acceptance and emerging issues within the community.

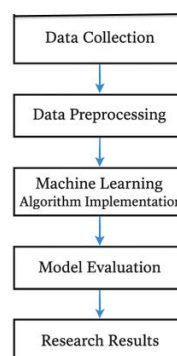
Proper nutrition for school-age children is essential to ensure optimal growth and development. This stage of life is a critical period marked by rapid physical and cognitive progress, making balanced nutrient intake highly important. Sufficient nutrition helps maintain overall health, strengthens the immune system, and supports cognitive abilities.. (Andreas Halomoan Tambunan et al., 2025). Therefore, this policy will be an indicator in differentiating the quality between cities and rural areas in adapting to the industrial revolution 4.0.(Albaburrahim et al., 2025).

Machine Learning (ML) holds transformative potential for addressing the energy challenges in the mining industry. Supervised learning algorithms are commonly applied in predictive maintenance to enable early detection of equipment failures and minimize operational downtime (Sakiru Folarin Bello et al., 2024). Advances in Natural Language Processing (NLP) and machine learning technology enable automated sentiment analysis of text extracted from social media, news, and public comments. Through this approach, public opinion can be categorized as positive, negative, or neutral.

This study aims to identify public opinion trends related to the Free Nutritious Food (MBG) program and measure the accuracy of various machine learning algorithms in classifying sentiments by offering a comprehensive comparison between machine learning algorithms and demonstrating the effectiveness of a transformer-based model (IndoBERT) in understanding the Indonesian language sentiment context, thereby enriching local NLP studies and its application in AI-based policy evaluation, thus contributing to policy and science by presenting a data-driven analysis of public opinion towards the MBG program, which can provide input for the government in evaluating public acceptance.

RESEARCH METHOD

In this study, the authors employed a quantitative approach based on machine learning. The analysis was conducted experimentally by comparing several classification models on a dataset of public opinion on the MBG program.



Source: Research Result (2025)

Figure 1. Research Method

1. Data Collection

The data obtained from this research comes from the kaggle.com website which can be accessed publicly [Sentimen Publik Terhadap Makan Bergizi Gratis](#). A total of 10,000 tweets containing keywords related to MBG. The data includes columns such as full_text (tweet text), favorite_count (retweet_count), and engagement metrics (username and location).

The data is stored in CSV format with two main attributes: text (the content of the comment or review), and sentiment (positive, negative, and neutral sentiment category labels).

2. Pre-Processing

The tasks carried out in this section are text cleaning (URL, mention, hashtag), tokenization and normalization, stopword removal and stemming, and finally feature extraction using TF – IDF and

CountVectorizer to represent text in numeric form that can be processed by the classification algorithm.

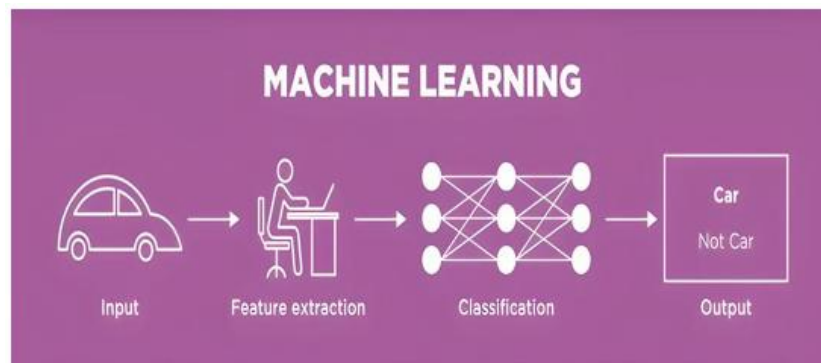
3. Algorithm Implementation

a. Machine Learning

Machine learning (ML) has become a leading area of study within computer science and related disciplines, fostering significant progress across various other fields (Ezugwu et al., 2025). Machine Learning is a sub-unit of Artificial Intelligence that enables machines to learn independently using data without having to be programmed repeatedly by humans. (Alfarizi M. Riziq Sirfatullah et al., 2023).

Machine learning It is a system that can learn to make decisions on its own without having to be repeatedly programmed by humans, allowing the computer to become increasingly intelligent by learning from the data it has. Based on the learning technique, supervised learning can be distinguished as using a labeled dataset (training data), while unsupervised learning draws conclusions based on the dataset. (Retnoningsih & Pramudita, 2020).

Machine Learning works when data is available which is used for analysis of large data sets (BigData) to find certain patterns. Machine Learning has three types, namely Supervised Learning, Unsupervised Learning, and Reinforcement Learning (R.H. Zer et al., 2022). In this study, the system was given a training dataset in the form of desired input and output information, so that the system would learn based on existing data. (R.H. Zer et al., 2022). The system will look for patterns from the dataset, where these patterns will be used as a reference for subsequent data collection.

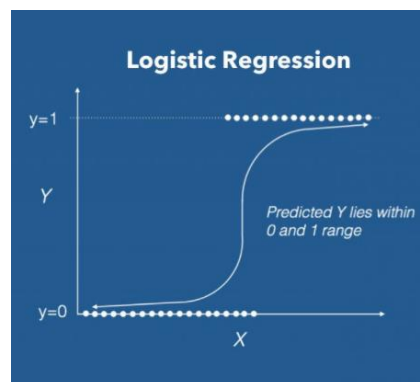


Source : (Napiah et al., 2022)

Figure 2. Machine Learning

b. Logistic Regression

Logistic Regression is commonly applied because it is easy to implement and requires minimal computational resources (Chase et al., 2022) A machine learning classification algorithm used to predict the probability of a categorical dependent variable. In Logistic Regression, the dependent variable is a binary variable containing data coded as 1 (Yes) or 0 (No) (Ramadhy & Sibaroni, 2022). Logistic Regression is the appropriate regression analysis to perform when the dependent variable is binary (two possibilities)(Junifer Pangaribuan et al., 2021)



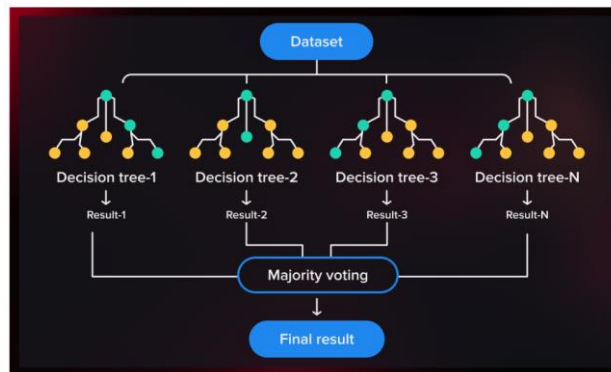
Source : (Learning+, 2024)

Figure 3. Logistic Regression

c. Random Forest

The Random Forest algorithm can be used to classify big data. Variable pruning, like decision trees, is not available in the random forest algorithm. However, the advantage of random forests is that they can

combine multiple trees and use a single tree to perform classification and class prediction (Prasojo & Haryatmi, 2021).



Source : (Serokell, 2022)

Figure 4. Random Forest

d. SVM

Support Vector Machine (SVM) is a method used to find the best hyperplane that can separate two classes in an input space. In the classification process, SVM uses training data to build a model, which is then used to predict the class of new, previously unseen data, also known as testing data (Valentino et al., 2020)

SVM is a powerful and effective classification algorithm, especially in cases of linear and non-linear separation. Sentiment analysis is a technique used to identify, understand, and classify sentiments or opinions in text (Munandar et al., 2023).

e. XGBOOST

XGBoost is a development of the ensemble-based gradient tree boosting algorithm, capable of efficiently handling large-scale machine learning problems. This method was chosen because it has various additional features that can speed up the computational process while reducing the risk of overfitting. XGBoost can be applied to various types of problems such as classification, regression, and ranking. This algorithm works by combining several decision trees (CARTs) to form a more robust model. The main advantage of XGBoost lies in its ability to adapt to various conditions, which comes from a series of improvements to its predecessor algorithm. (Yulianti et al., 2022)

f. LightGBM

LightGBM is a popular method and has been consistently proven to be able to solve various classification problems (Febriantoro et al., 2023). Thus, this method is able to produce more effective models in various classification tasks and regression studies, and achieve excellent detection rates.

g. Naïve Bayes

The Naive Bayes (NB) classifier is a widely utilized supervised learning algorithm in Machine Learning (ML). Its performance, however, depends on the strong assumption of conditional independence, an assumption that is frequently not met in real-world applications (Gohari et al., 2023), Naïve Bayes is an algorithm in machine learning that works based on probability calculations (Susana & Suarna, 2022).

4. Training Process

Training is the process by which a machine learning or deep learning model learns from labeled data to recognize patterns and relationships between input (text) and output (sentiment) features. The data used constitutes 80% of the dataset.

5. Testing Process

The testing process is the stage of evaluating model performance using data never seen during training to measure how well the model can generalize. The data used is 20% of the dataset.

6. Model Evaluation

Model evaluation is the practice of testing a trained machine learning model on new data to measure its effectiveness and identify its strong and weak points (Rainio et al., 2024) In the evaluation phase, an assessment is conducted to determine whether the data mining results have met the stated objectives. Next, each cluster formed is profiled to determine its characteristics. Furthermore, to assess its suitability to the interest pathway, further

analysis is conducted by linking it to the interest attributes. This is expected to yield useful information or patterns that can serve as a basis for future data updates. The three main metrics at this stage are accuracy, F1-score, and confusion matrix.

RESULTS AND DISCUSSION

Testing using Google Colab on a data set taken from kaggle, 10,000 Twitter User Sentiments Towards the Free Nutritious Meal Program taken.

1. Data Pre-processing

The data obtained from online and social media contained public opinion regarding the Free Nutritious Meals (MBG) program. Text preprocessing included cleaning irrelevant elements (punctuation, URLs, numbers), tokenization, stopwords removal, and stemming using NLTK. The stemming quality was verified through manual inspection and word distribution analysis to ensure semantic consistency. TF-IDF comparison and model performance evaluation confirmed that sentiment-bearing features were retained, indicating that preprocessing effectively standardized the text without losing essential information.



Source : Research Result (2025)
Figure 5. Text Preprocessing

2. Feature Extraction

After the data is cleaned, the text is converted into a numerical representation using the TF-IDF (Term Frequency–Inverse Document Frequency) method. This technique assesses the importance of a word in a document based on its frequency across the entire corpus. Words that appear frequently in one document but rarely in another are weighted higher.

The results of this extraction produce a feature matrix that is used as input to the classification model.

3. Model Training Results

In model testing, the dataset was divided into two parts, a training set of 80% and a testing set of 20%, and cross-validation was used during model training to tune hyperparameters and evaluate generalization performance. To ensure robustness and reproducibility, the experiments were repeated with multiple random seeds (e.g., 42, 100, and 123), and the results showed stable accuracy and F1-score values across runs. Several machine learning algorithms were applied to compare sentiment classification performance, including Logistic Regression, Support Vector Machine (SVM), Random Forest, Naive Bayes, XGBoost, and LightGBM.

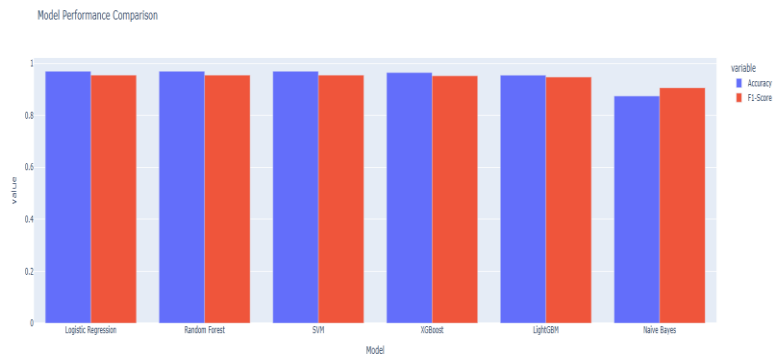
Table 1. Machine Learning Model Evaluation

| No | Model | Accuracy | F1Score |
|----|------------------------------|----------|---------|
| 1 | Logistic Regression | 0.9700 | 0.9552 |
| 2 | Random Forest | 0.9700 | 0.9552 |
| 3 | Support Vector Machine (SVM) | 0.9700 | 0.9552 |

| No | Model | Accuracy | F1Score |
|----|-------------|----------|---------|
| 4 | XGBoost | 0.9650 | 0.9527 |
| 5 | LightGBM | 0.9550 | 0.9477 |
| 6 | Naive Bayes | 0.8750 | 0.9063 |

Source: Research Result (2025)

The three best-performing models based on the test results were Logistic Regression, Random Forest, and SVM, each achieving a maximum accuracy of 97% with an F1-score of 0.9552. Meanwhile, XGBoost also performed very well with an accuracy of 96.5%, followed by LightGBM with a slightly lower accuracy of 95.5%. Naive Bayes achieved the lowest result with an accuracy of 87.5%, although it still managed to maintain an F1-score of 0.9063.



Source: Research Result (2025)

Figure 6. Model Performance Comparison

4. Visualization of Results

Hasil This section uses a Word Cloud (to display the dominant words per sentiment category) and a bar chart to compare model performance. Then, a sentiment distribution analysis shows a predominance of positive opinions.

The most frequently occurring words in the sentiment category are: **Positive:** “gizi”, “anak”, “sehat”, “gratis”, “program”, “pemerintah”. **Negative:** “tidak merata”, “kurang”, “masalah distribusi”. **Neutral:** “pelaksanaan”, “kegiatan”, “menu”.

=== TOPIC ANALYSIS ===

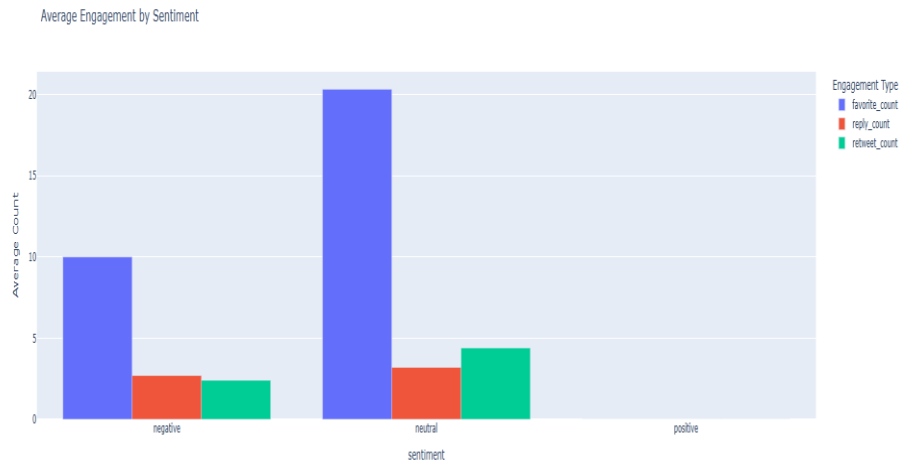
Topic 1: grati, makan, program, bergizi, bergizi grati, makan bergizi, meningkatkan, kualitas, meningkatkan kualita, maksin
 Topic 2: program, bergizi, grati, bergizi grati, makan, prabowo, makan bergizi, presiden, mbg, grati mbg
 Topic 3: makan, grati, siang, makan siang, bergizi, siang grati, bergizi grati, makan bergizi, pemerintah, program
 Topic 4: makan, siang, grati, makan siang, siang grati, pendidikan, pendidikan kesehatan, kesehatan, pemberdayaan masyarakat, pemberdayaan
 Topic 5: polri, makan, grati, bergizi, makan bergizi, bergizi grati, program, masyarakat, mendukung, program makan



Source: Research Result (2025)

Figure 7. WordCloud

From the image above, In addition to the WordCloud visualization, a quantitative analysis was performed using **word frequency and TF-IDF weighting**. The results showed that the words “*bergizi*” (TF-IDF = 0.082), “*gratis*” (0.079), “*makan*” (0.071), “*program*” (0.067), and “*pemerintah*” (0.059) obtained the highest TF-IDF scores across sentiment classes. These values indicate that the highlighted words are the most dominant and informative features distinguishing sentiment categories, thereby providing quantitative support for the visual WordCloud findings.



Source: Research Result (2025)

Gambar.7 Average Engagment by Sentiment

Visualization using a bar chart shows that 65% of the public supports the program, highlighting the benefits of nutrition and education. 25% are neutral regarding factual news and program information. Finally, 10% are negative, with concerns about implementation, delays, and quality. This indicates that the majority of the public has responded positively to the implementation of the MBG program, particularly in terms of improving the nutrition of schoolchildren.

=== MODEL EVALUATION ===

Best Model: Logistic Regression

Best F1-Score: 0.9552

Detailed Classification Report for Logistic Regression:

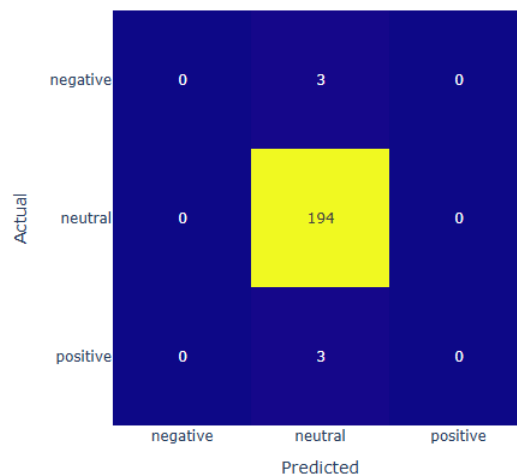
| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| negative | 0.00 | 0.00 | 0.00 | 3 |
| neutral | 0.97 | 1.00 | 0.98 | 194 |
| positive | 0.00 | 0.00 | 0.00 | 3 |
| accuracy | | | 0.97 | 200 |
| macro avg | 0.32 | 0.33 | 0.33 | 200 |
| weighted avg | 0.94 | 0.97 | 0.96 | 200 |

Source: Research Result (2025)

Figure 8. Logistic Regression

The best model obtained from the experiments was logistic regression in sentiment classification. The model performed very well in recognizing neutral sentiment, but was unable to effectively

generate negative and positive sentiment due to the imbalance in the amount of data between classes. For further studies, we can try combining text augmentation and balancing techniques (oversampling/undersampling), applying class weighting or focal loss, and evaluating with imbalance-sensitive metrics (macro-F1, PR-AUC). Experiments should be repeated with multiple random seeds and stratified cross-validation to ensure robustness and reproducibility of the results. This can also be seen in Figure 9 for model evaluation using a confusion matrix..



Source: Research Result (2025)

Figure 9. Confusion Matrix

CONCLUSION

The experimental results demonstrate that the application of Natural Language Processing (NLP) with the TF-IDF technique can transform public opinion text into effective numerical features for sentiment classification. Logistic Regression and LightGBM models demonstrated optimal performance. Logistic Regression excels in its ability to linearly separate classes in text data processed using TF-IDF, resulting in accurate classification. LightGBM, on the other hand, stands out for its efficient training time and better memory utilization, especially when processing data with a large number of features.

Other algorithms, such as Random Forest and Naive Bayes, still produce quite good results, but their performance is less than optimal when faced with text data with a high level of word variation. Overall, the analysis results indicate that public sentiment toward the Free Nutritional Meals program tends to be positive, with the majority of opinions supporting the program as an appropriate step to improve child health and nutritional equity in schools.

This research successfully demonstrates that machine learning can be used effectively to analyze public perceptions of national-scale social policies.

REFERENCES

- Albaburrahim, A., Putikadyanto, A. P. A., Efendi, A. N., Alatas, M. A., Romadhon, S., & Wachidah, L. R. (2025). Program Makan Bergizi Gratis: Analisis Kritis Transformasi Pendidikan Indonesia Menuju Generasi Emas 2045. *Entita: Jurnal Pendidikan Ilmu Pengetahuan Sosial Dan Ilmu-Ilmu Sosial*, 767–780. <https://doi.org/10.19105/ejpis.v1i.19191>
- Alfarizi M. Riziq Sirfatullah, Al-farish Muhamad Zidan, Taufiqurrahman Muhamad, Ardiansah Ginan, & Elgar Muhamad. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- Andreas Halomoan Tambunan, K., Nababan, R., Anisa Siagian, R., Naiborhu, R., Harianti, S., & Jamaludin, J. (2025). Tinjauan Kritis Tentang Program Makan Bergizi Gratis Terhadap Produktivitas Belajar Siswa. *Katalis Pendidikan: Jurnal Ilmu Pendidikan Dan Matematika*, 2(2), 21–31. <https://doi.org/10.62383/katalis.v2i2.1428>
- Chase, R. J., Harrison, D. R., Burke, A., Lackmann, G. M., & McGovern, A. (2022). A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional Machine Learning. *Weather and Forecasting*, 37(8), 1509–1529. <https://doi.org/10.1175/WAF-D-22-0070.1>
- Ezugwu, A. E., Greeff, J., & Ho, Y. S. (2025). A comprehensive study of groundbreaking machine learning research: Analyzing highly cited and impactful publications across six decades. *Journal of Engineering Research (Kuwait)*, 13(1), 371–383. <https://doi.org/10.1016/j.jer.2023.10.027>

- Febriantoro, E., Setyati, E., & Santoso, J. (2023). Pemodelan prediksi kuantitas penjualan mainan menggunakan Light Gradient Boosting Machine. *SMARTICS Journal*, 9(1), 7–13. <https://doi.org/10.21067/smartics.v9i1.8279>
- Gohari, K., Kazemnejad, A., Mohammadi, M., Eskandari, F., Saberi, S., Esmaili, M., & Sheidaei, A. (2023). A Bayesian latent class extension of naive Bayesian classifier and its application to the classification of gastric cancer patients. *BMC Medical Research Methodology*, 23(1), 1–15. <https://doi.org/10.1186/s12874-023-02013-4>
- Hidayatullah, M. . (2024). "Adik Prabowo Ungkap Asal-usul Ide Makan Bergizi Gratis, Tercetus Sejak 2006". (D. R. Meilanova, Ed.) *Bisnis.com*. Retrieved from. <https://ekonomi.bisnis.com/read/20241007/12/1805463/adik-prabowo-ungkap-asal-usul-ide-makan-bergizi-gratis-tercetus-sejak-2006>
- Junifer Pangaribuan, J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression. *Journal Information System Development (ISD)*, 06(02), 1–10.
- Learning+, M. (2024). *Logistic Regression – A Complete Tutorial With Examples in R*. <https://www.machinelearningplus.com/machine-learning/logistic-regression-tutorial-examples-r/>
- Munandar, A. A., Farikhin, F., & Widodo, C. E. (2023). Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM. *JOINTECS (Journal of Information Technology and Computer Science)*, 8(2), 77. <https://doi.org/10.31328/jointecs.v8i2.4747>
- Napiah, M., Purnama, R. A., Raharjo, M., & Bismi, W. (2022). Komparasi Algoritma Untuk Klasifikasi Penyakit Ispa (Infeksi Saluran Pernapasan Akut). *Jurnal Infortech*, 4(2), 105–110. <https://ejournal.bsi.ac.id/ejurnal/index.php/infortech/article/view/13644/5720>
- Prasetyo, D. E. (2023). Pancasila : Jurnal Keindonesiaan. *Jurnal Keindonesiaan*, 3(2), 1–10.
- Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 79–89. <https://doi.org/10.25077/teknosi.v7i2.2021.79-89>
- R.H. Zer, P. P. P. A. N. . F. I., Hayadi, B. H., & Damanik, A. R. (2022). Pendekatan Machine Learning Menggunakan Algoritma C4.5 Berbasis Pso Dalam Analisa Pemahaman Pemrograman Website. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3). <https://doi.org/10.23960/jitet.v10i3.2700>
- Rainio, O., Teuvo, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 1–14. <https://doi.org/10.1038/s41598-024-56706-x>
- Ramadhya, I. F., & Sibaroni, Y. (2022). Analisis Trending Topik Twitter dengan Fitur Ekspansi FastText Menggunakan Metode Logistic Regression. *JURIKOM (Jurnal Riset Komputer)*, 9(1), 1. <https://doi.org/10.30865/jurikom.v9i1.3791>
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- Sakiru Folarin Bello, Ifeoluwa Uchechukwu Wada, Olukayode B Ige, Ernest C Chianumba, & Samod Adetunji Adebayo. (2024). AI-driven predictive maintenance and optimization of renewable energy systems for enhanced operational efficiency and longevity. *International Journal of Science and Research Archive*, 13(1), 2823–2837. <https://doi.org/10.30574/ijrsra.2024.13.1.1992>
- Serokell. (2022). *Random Forest Classifier: Basic Principles and Applications*. <https://serokell.io/blog/random-forest-classification>
- Susana, H., & Suarna, N. (2022). PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET Program Studi Teknik Informatika STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 3) Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon Jl Perjuangan No 1. *Jurnal Sistem Informasi Dan Teknologi Informasi*, 4(1), 1–8.
- Valentino, K. S. Q., Ade, I., & Hindriyanto, D. P. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis 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. www.tripadvisor.com,
- Yulianti, S. E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics: Theory and Applications*, 4(1), 21–26.