

Analysis And Prediction Of Job Training Suitability For Job Seekers' Professions At The Department Of Employment, Industry, And Trade Of Batu Bara Regency Using The Naive Bayes Algorithm And Feature Selection

Eko Budianto^{1,*}, Muhammad Iqbal², Zulham Sitorus³

Master of Information Technology, University Pembangunan Panca Budi, Medan, Indonesia

Email: ^{1,*}ekobudianto29@gmail.com, ²muhammadiqbal@dosen.pancabudi.ac.id, ³zulhamsitorus@dosen.pancabudi.ac.id

Correspondence Author Email: ekobudianto29@gmail.com*

Submitted: 12/06/2026; Accepted: 29/06/2026; Published: 30/06/2026

Abstract- Public vocational training effectiveness depends on alignment between training programs and job seekers' professional profiles. In practice, training placement is often determined manually and subjectively, causing competency mismatches that reduce program effectiveness. This study develops an optimized computational framework to predict the suitability of vocational training programs for job seekers at the Department of Employment, Industry, and Trade of Batu Bara Regency. Using the Knowledge Discovery in Databases (KDD) framework, 1,434 historical records containing demographic data, education, work experience, occupational interests, and competency indicators were analyzed. To address the conditional independence limitation of the Naive Bayes classifier, three filter-based feature selection methods Information Gain, Mutual Information, and Chi-Square were implemented and compared. Results show that feature selection improved model performance, increasing accuracy from 91.26% to 93.01% across all methods. The consistent performance indicates that all methods identified the same dominant predictor, primarily professional interest, while removing redundant attributes. The proposed hybrid model demonstrates strong stability and generalization capability, providing a reliable decision support system for reducing employment mismatches and improving workforce development resource allocation.

Keywords: Classification; Feature Selection; Job Training recommendation; Machine Learning; Naive Bayes Classifier.

1. INTRODUCTION

The acceleration of human resource development remains a critical prerequisite for regional governance to sustain Workforce competitiveness amid rapid economic transformation and shifting skill requirements. Societal and economic progress dictates that vocational training programs managed by regional authorities specifically the Department of Employment, Industry, and Trade (Disnakerperindag) of Batu Bara Regency serve as a strategic instrument to equip the workforce with competencies relevant to market demands. According to Indonesian Law No. 13 of 2003 concerning Employment, occupational training must be systematically designed to develop and enhance competencies, productivity, and work ethics aligned with specific qualification frameworks.

However, empirical evidence highlights a persistent misalignment between the training modalities provided and the professional fields or capability matrices inherent to job seekers. This discrepancy yields a structural competency mismatch, which severely impairs program efficacy, diminishes training utility, and impedes post-training employment absorption. This mismatch induces broader socioeconomic implications, including an inflation in educated unemployment and an inefficient allocation of public expenditure.

A primary institutional challenge resides in the participant assignment framework, which heavily relies on manual, subjective heuristics by officers, lacking data-driven insights from historical profiles. Although Disnakerperindag possesses a comprehensive repository of historical data encompassing educational backgrounds, technical skills, Prior work experience, and occupational interests this asset remains underutilized as an objective foundation for tailored training assignment. Consequently, a definitive gap emerges between data availability and data utilization for producing precise, automated training recommendations. To address this optimization challenge, data mining approaches offer highly robust computational solutions to extract hidden patterns from multidimensional administrative records. In the modern era, leveraging secondary data via structural Knowledge Discovery in Databases (KDD) paradigms allows institutions to systematically clean, transform, and model data to enhance institutional decision support mechanisms [1] [2]. Within the taxonomy of machine learning, classification methods have been widely applied to resolve predictive matching problems across various domains. The Naive Bayes algorithm is recognized as a lightweight and computationally efficient probabilistic model that maintains consistent performance across diverse datasets. Numerous previous studies have validated its predictive effectiveness in the fields of employment, recommendation systems, and education. For example, one study demonstrated excellent capability in mapping individual characteristics to employment opportunities, achieving a predictive accuracy of 97.66%, a precision of 100%, and a recall of 97.59% [3]. Furthermore, the implementation of a web-based Gaussian Naive Bayes algorithm for educational track recommendation achieved a classification accuracy of 87.85%, demonstrating its effectiveness in handling data uncertainty and supporting evidence-based decision making [4].

In the public service sector, the Naive Bayes algorithm has been utilized to optimize the selection process of prospective government employees, significantly improving objectivity and efficiency in applicant evaluation [5]. Furthermore, studies have demonstrated that data-driven classification algorithms are capable of providing highly objective key performance evaluations for digital public services.[6]. Beyond conventional assignments, the probabilistic classifier has shown remarkable versatility in reinforcing workforce integration integrity; for instance successfully deployed Naive Bayes to identify fraudulent job vacancy advertisements with an accuracy of 89% [7]. In trust-aware environments, the algorithm has been extended for application to social network data, achieving an accuracy of 89% while demonstrating its flexibility in addressing various cross-domain classification problems [8]. However, studies have verified the algorithm's capability to accommodate diverse demographic variables for predicting timely student graduation. Furthermore, it has been empirically confirmed that the Naive Bayes algorithm is a highly effective baseline method for processing municipal compliance data [9] [10]. Nevertheless, the traditional Naive Bayes classifier suffers from a foundational limitation: the strict assumption of conditional independence among predictor features. Its predictive performance is highly sensitive to the quality of the input features. When a dataset contains irrelevant or redundant attributes, the accuracy of the model degrades significantly, rendering it susceptible to overfitting. Given that job seeker profiles exhibit complex, multidimensional attributes, including educational levels, competencies, occupational interests, and work experiences, executing Naive Bayes without data dimension preprocessing poses a high risk of suboptimal classification. Hence, integrating an auxiliary feature selection mechanism to isolate the most informative attributes and eliminate data noise is theoretically imperative to preserve the model's predictive stability.

Empirically, combining feature selection techniques with Naive Bayes has been proven to significantly improve accuracy, stability, and prediction metrics across various domains compared to standalone implementations. Subsequent research demonstrated that the integration of filter-based methods with wrapper paradigms within the Naive Bayes framework achieved optimal classification performance, attaining an accuracy of approximately 91.80% in specialized diagnostic classification tasks [11]. In student academic longevity tracking, applied Information Gain and Correlation Attribute selection, proving that models equipped with feature selectors significantly outperform baseline classifiers. Concurrently, utilized Gain Ratio attribute weighting to dramatically enhance the Area Under the Receiver Operating Characteristic (ROC/AUC) and predictive stability of Naive Bayes models [12] [13]. Further substantiating this synergy, reported an optimized accuracy of 97.50% by pairing a Naive Bayes classifier with Forward Selection [1]. Under different paradigms, observed that integrating feature weighting methods can enhance Naive Bayes classification accuracy by up to 20% [14]. In public welfare classification, successfully applied Backward Elimination to prune less informative demographic parameters, boosting classifier performance [15]. Furthermore, the developed computational framework demonstrates that filter-based Information Gain methods can systematically eliminate redundant institutional attributes, thereby improving the predictive performance of classification models [16] [17]. The Particle Swarm Optimization (PSO) algorithm is used to integrate evolutionary optimization in constructing meta-heuristic feature subsets, resulting in significant performance improvements [18]. Beyond demographic data, predictive modeling combined with clustering algorithms is used to analyze institutional performance, while the stability of feature-selected Naive Bayes systems has been validated on localized commercial datasets [19] [20]. The selection of core attributes using Chi-Square techniques is considered capable of producing robust text and sentiment classification profiles [21]. While literature regarding standalone Naive Bayes applications is extensive, its specialized deployment to resolve vocational training-to-profession mismatches within regional government agencies remains profoundly scarce. More importantly, the primary scientific novelty of this study resides in its comprehensive computational evaluation and comparison of three distinct filter-based feature selection paradigms Information Gain (IG), Mutual Information (MI), and Chi-Square (χ^2) to identify the optimal dimensional reduction technique for localized demographic datasets. While major prior works typically deploy a single feature selection technique, this research explicitly addresses the empirical gap by offering a comparative, data-driven optimization framework.

By establishing this integrated framework, this research aims to offer a robust decision support system for government institutions to generate accurate training recommendations, minimize competency mismatches, enhance regional employment placement rates, and optimize the operational efficiency of public vocational budgets.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study employed the Knowledge Discovery in Databases (KDD) framework to develop a predictive model for determining the suitability of vocational training programs for job seekers registered at the Department of Employment, Industry, and Trade of Batu Bara Regency. The KDD framework was selected because it provides a systematic process for extracting meaningful knowledge from large-scale administrative datasets through data preparation, transformation, classification, and evaluation stages. The research process consisted of five primary stages: data selection, preprocessing, feature selection, classification, and evaluation. Historical job seeker records

were collected from the institutional database and subsequently prepared through cleaning and transformation procedures. The processed dataset was then subjected to feature selection techniques to identify the most informative attributes before classification using the Naive Bayes algorithm. Finally, the predictive performance of each model was evaluated using standard classification metrics. The research workflow is illustrated in Figure 1.

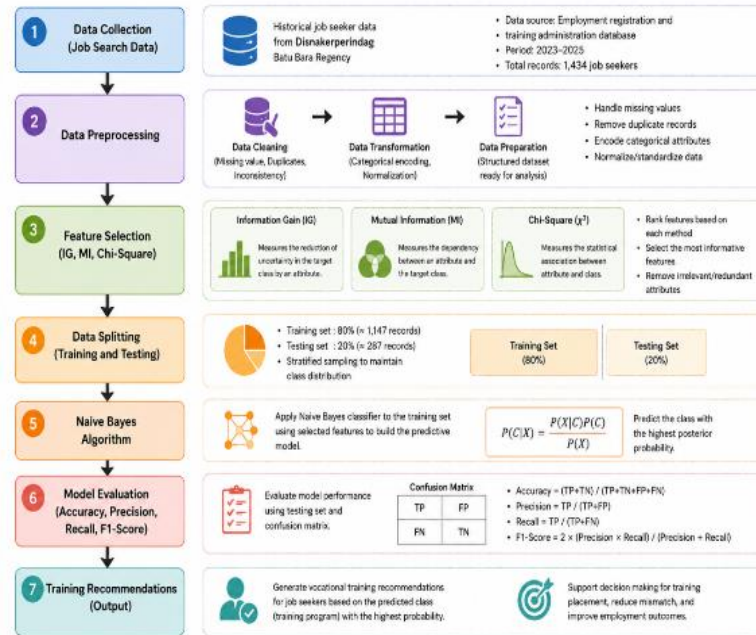


Figure 1. Research Framework Based on KDD Process

2.2 Dataset Description

The dataset utilized in this study was obtained from the Department of Employment, Industry, and Trade (Disnakerperindag) of Batu Bara Regency. The dataset consisted of 1,434 historical records of job seekers collected from employment registration and vocational training administration databases. Each record contained demographic and competency-related attributes representing the characteristics of job seekers. The target variable represented the vocational training recommendation category determined from historical training participation records and competency suitability assessments.

Table 1 presents the attributes used in this study.

Table 1. Variables Used in the Study

No	Variable	Description	Data Type
1	Age	Age of job seekers at the time of registration	Numeric
2	Gender	Gender (Male/Female)	Nominal
3	Education Level	Highest level of formal education attained	Ordinal
4	Work Experience	Duration of work experience (None, <1 year, 1–3 years, >3 years)	Ordinal
5	Professional Interest	Desired job field or profession	Nominal
6	Competency Test Score	Score obtained from competency test (0–100)	Numeric
7	Training Recommendation	Suitable training program (target variable)	Nominal

The selected attributes represent demographic, educational, and competency-related characteristics of job seekers that potentially influence the suitability of vocational training recommendations. These attributes were subsequently used as predictor variables in the feature selection and classification processes.

2.3 Data Preprocessing

Data preprocessing was conducted to improve data quality before model construction. This stage included data cleaning, transformation, and preparation. The data cleaning process involved identifying incomplete records, removing duplicate entries, and correcting inconsistent attribute values. Subsequently, categorical variables such as education level, professional interest, and gender were transformed into structured categorical formats suitable for classification analysis. The dataset was then normalized and prepared for machine learning processing. The final dataset was divided into training and testing subsets using an 80:20 ratio. The training dataset was utilized to construct the classification model, while the testing dataset was employed to evaluate predictive performance.

2.4 Feature Selection Methods

Feature selection was implemented to reduce irrelevant and redundant attributes that may negatively affect classification performance. This study compared three filter-based feature selection techniques, namely Information Gain, Mutual Information, and Chi-Square.

2.4.1 Information Gain

Information Gain measures the contribution of an attribute in reducing uncertainty regarding the target class. Attributes with higher Information Gain values provide more information and are considered more relevant for classification.

$$IG(S,A) = Entropy(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (1)$$

Where:

- IG(S,A) = Information Gain of attribute A
- Entropy(S) = Entropy of dataset S
- S_v = Subset corresponding to attribute value v

2.4.2 Mutual Information

Mutual Information quantifies the dependency between an attribute and the target class. A higher value indicates a stronger relationship between the predictor and the classification outcome.

$$MI(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (2)$$

Where:

- p(x,y) = Joint probability distribution
- p(x) and p(y) = Marginal probability distributions

2.4.3 Chi-Square

The Chi-Square method evaluates the statistical association between predictor attributes and class labels. Attributes with larger Chi-Square values are considered more significant for classification.

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (3)$$

Where:

- O = Observed frequency
- E = Expected frequency

The selected features obtained from each method were subsequently used as inputs for the Naive Bayes classification model.

2.5 Naive Bayes Classification

The Naive Bayes classifier was employed as the primary predictive model due to its simplicity, computational efficiency, and effectiveness in handling categorical data. The algorithm estimates posterior probabilities based on Bayes' theorem while assuming conditional independence among predictor attributes [1],[5].

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (4)$$

where $P(C | X)$ denotes the posterior probability of class C given predictor attributes X , $P(X | C)$ denotes the likelihood, $P(C)$ denotes the prior probability, and $P(X)$ denotes the evidence.

Example Calculation

Given the following probabilities:

$$\begin{aligned} P(A) &= 0.20 \\ P(B | A) &= 0.85 \\ P(B | \neg A) &= 0.10 \end{aligned}$$

The probability of not being in class A is calculated as:

$$P(\neg A) = 1 - P(A) = 1 - 0.20 = 0.80$$

The evidence $P(B)$ is calculated as:

$$\begin{aligned} P(B) &= P(B | A)P(A) + P(B | \neg A)P(\neg A) \\ P(B) &= (0.85)(0.20) + (0.10)(0.80) = 0.17 + 0.08 = 0.25 \end{aligned}$$

The posterior probability is calculated as:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

$$P(A | B) = \frac{0.85 \times 0.20}{0.25} = \frac{0.17}{0.25} = 0.68$$

Therefore, the posterior probability of class *A* given evidence *B* is 0.68, or 68%.

2.6 Model Evaluation

Model performance was evaluated using a confusion matrix and four classification metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified instances relative to the total number of observations.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Precision evaluates the proportion of correctly predicted positive instances among all positive predictions. Recall measures the model's ability to identify relevant instances from all actual positive observations. The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of classification performance [5], [11], [21]. The performance of the baseline Naive Bayes classifier was compared with Naive Bayes integrated with Information Gain, Mutual Information, and Chi-Square feature selection methods to determine the most effective approach for vocational training recommendation prediction [11], [17], [21].

3. RESULT AND DISCUSSION

3.1 Mapping of Industry Needs to Training Programs

To ensure that vocational training recommendations align with labor market demands, a mapping process was conducted between industrial needs (DUDI) and available training programs administered by the Department of Employment, Industry, and Trade of Batu Bara Regency. The mapping was developed based on employment demand data, regional industrial characteristics, and competency requirements identified from previous workforce development programs.

Table 2. Mapping of Industry Needs (DUDI) to Training Program

No	Company	Required Professions	Training Programs (Output Label)
1	PT Indonesia Asahan Aluminium	Admin, Safety Patrol, Warehouse Admin	Office Administration, K3, Warehouse Management
2	PT Multi Nabati Asahan	Warehouse Admin, Safety Patrol	Warehouse Management, K3
3	PT Bakrie Oleo Chemical	Admin, Operator, Safety Patrol	Office Administration, Heavy Equipment, K3
4	PT Biyu Iyas Malela	Harvester, Operator, Admin	Palm Harvesting, Heavy Equipment, Administration
5	PT Horas Motor	Mechanic, Admin, Sales	Motorcycle Mechanic, Digital Marketing, Admin
6	PT Socfin Indonesia	Operator, Admin, Warehouse	Heavy Equipment, Administration, Warehouse
7	PT Moeis	Harvester, Operator, Safety Patrol	Palm Harvesting, Heavy Equipment, K3

Feature selection techniques were applied to identify the most relevant attributes affecting classification performance. Three methods were used and compared, namely Information Gain, Mutual Information, and Chi-Square. These methods evaluate the importance of each feature based on its relationship with the target variable, allowing the model to focus on relevant attributes and reduce noise from irrelevant data.

The classification process was performed using the Naive Bayes algorithm, which calculates posterior probabilities to determine the most probable class. Model performance was evaluated using a confusion matrix and performance metrics, including accuracy, precision, recall, and F1-score. A sample of the dataset used in this study is presented in Table 3.

3.2 Sample of Job Seeker Data

The dataset consists of historical records of job seekers registered at the Department of Employment, Industry, and Trade of Batu Bara Regency. Each record contains demographic information, educational background, work experience, professional interests, competency test scores, and corresponding training recommendations.

Table 3. Sample of Job Seeker Data

No	Gender	Age	Education	Experience	Interest	Test Score	Training Recommendation
1	M	20	High School	< 1 year	Digital Marketing	67	Office Administration
2	M	17	MA	None	Electrical	83	Electrical Installation
3	M	31	High School	> 3 years	Automotive	75	Motorcycle Mechanic
4	M	36	Vocational	1–3 years	Digital Marketing	84	Digital Marketing
5	F	20	MA	< 1 year	Administration	94	Office Administration

The sample data presented in Table 3 illustrates the characteristics of job seekers used in this study. Each record contains demographic and competency-related attributes, including gender, age, educational background, work experience, professional interest, competency test score, and the corresponding training recommendation. The variation of attributes demonstrates the diversity of job seeker profiles, which is essential for building a robust classification model. The dataset also reflects the relationship between professional interests and the recommended training programs. For example, applicants with interests in administration and digital marketing are generally directed toward Office Administration or Digital Marketing training programs, while applicants with automotive interests are recommended for Motorcycle Mechanic training. These patterns indicate that the dataset contains meaningful relationships that can be utilized by the Naive Bayes algorithm to generate accurate predictions. Furthermore, the inclusion of competency test scores and work experience provides additional predictive value for the classification process. By combining demographic, professional, and competency-based variables, the model is expected to identify the most suitable training program for each job seeker more effectively.

3.3 Classification Performance Results

The classification process was performed using the Naive Bayes algorithm as the baseline model. To improve predictive performance, three feature selection techniques were integrated into the classification framework, namely Information Gain, Mutual Information, and Chi-Square. The performance of each model was evaluated using classification accuracy

Table 4. Comparison of Naive Bayes Model Performance

No	Method	Accuracy (%)	Precision	Recall	F1-Score
1	Naive Bayes without Feature Selection	91.26	91.7462	91.2587	91.1919
2	Naive Bayes + Information Gain	93.01	93.5805	93.0070	92.9368
3	Naive Bayes + Mutual Information	93.01	93.5805	93.0070	92.9368
4	Naive Bayes + Chi-Square	93.01	93.5805	93.0070	92.9368

Based on Table 4, the application of feature selection improves the performance of the Naive Bayes classifier. The baseline model achieves an accuracy of 91.26%, while all feature selection methods increase accuracy to 93.01%. The results show that Information Gain, Mutual Information, and Chi-Square produce identical performance values. This indicates that all three methods select a similar subset of dominant features, suggesting that the dataset contains strong and consistent attributes that significantly influence classification outcomes.

3.4 Discussion

The results demonstrate that the integration of feature selection techniques improves the effectiveness of the Naive Bayes classifier in predicting vocational training suitability for job seekers. The baseline model already achieved a relatively high classification accuracy, indicating that demographic characteristics, educational background, professional interests, work experience, and competency test scores provide meaningful information for training recommendation prediction. Among the evaluated attributes, Professional Interest was identified as the most influential factor affecting training suitability recommendations. This finding is reasonable because vocational training programs are designed to support specific occupational pathways. Consequently, job seekers whose professional interests align with industry requirements are more likely to receive appropriate training recommendations. The implementation of feature selection reduced data redundancy while preserving relevant information. As a result, classification accuracy improved from 91.26% to 93.01%. Although all three feature selection methods produced the same accuracy value, this consistency indicates that the selected features have strong predictive power and that the dataset contains a stable attribute structure. These findings support previous studies that reported improvements in Naive Bayes performance after incorporating feature selection techniques. Therefore, the proposed framework can serve as a practical decision support system for government institutions in recommending vocational training programs more objectively, efficiently, and consistently.

4. CONCLUSION

This study aims to analyze and predict the suitability of job training programs with the professional fields of job seekers using the Naive Bayes algorithm combined with feature selection methods. Based on the experimental results, it can be concluded that the application of feature selection significantly improves the performance of the Naive Bayes classifier. The baseline model without feature selection achieved an accuracy of 91.26%, while the models with feature selection methods Information Gain, Mutual Information, and Chi-Square improved the accuracy to 93.01%. The results also indicate that all three feature selection methods produce identical performance values, suggesting that the selected features have strong and consistent relevance to the target variable. This demonstrates that the dataset used in this study contains dominant attributes that significantly influence classification outcomes. Furthermore, the integration of industry needs (DUDI) into the classification process ensures that the recommended training programs are aligned with Job market demands. This enhances the practical applicability of the model as a decision support system for government institutions in designing more targeted and effective job training programs. In conclusion, the combination of Naive Bayes and feature selection provides an accurate and reliable approach for predicting job training suitability. Future research is recommended to explore other classification algorithms or hybrid methods, as well as to utilize larger and more diverse datasets to further improve model performance.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Employment, Industry, and Trade of Batu Bara Regency for providing the data and information required for this research. The authors also wish to thank the Master of Information Technology Program, Universitas Pembangunan Panca Budi, for the academic support, guidance, and facilities provided throughout the completion of this study. Special appreciation is extended to the supervisors, lecturers, and administrative staff whose valuable assistance contributed significantly to the successful completion of this research. The authors are also grateful to all parties involved in the data collection, validation, and evaluation processes. Their support and cooperation were invaluable in conducting this study.

REFERENCES

- [1] Noliza Safitri, "Implementasi Algoritma Naive Bayes Classifier Menggunakan Feature Forward Selection Dalam Memprediksi Ketepatan Masa Studi Mahasiswa," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 11, no. 2, pp. 62–68, Oct. 2023, doi: 10.21063/jtif.2023.v11.2.62-68.
- [2] A. F. Sallaby and E. Suryana, "Penerapan Data Mining untuk Menentukan Jumlah Pencari Kerja Terdaftar Berdasarkan Umur dan Pendidikan Menggunakan K-Means Clustering (Studi Kasus di Dinas Tenaga Kerja Dan Transmigrasi Provinsi Bengkulu)," *J. Technopreneursh. Inf. Syst.*, vol. 1, no. 1, pp. 35–38, 2018, doi: 10.36085/jtis.v1i2.28.
- [3] A. Purnomo and A. Sururi, "Prediksi Kemampuan Siswa Dalam Bersaing di Dunia Kerja Menggunakan Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor," *ICIT J.*, vol. 8, no. 1, pp. 34–45, Feb. 2022, doi: 10.33050/icit.v8i1.2171.
- [4] K. B. P. Y. Perkasa and F. Eka Purwiantono, "Sistem Rekomendasi Jurusan Menggunakan Algoritma Naive Bayes Gaussian Berbasis Web," *J-INTECH*, vol. 11, no. 2, pp. 361–370, Dec. 2023, doi: 10.32664/j-intech.v11i2.1090.
- [5] Y. Bustomi, A. Nugraha, C. Juliane, and S. Rahayu, "Data Mining Selection of Prospective Government Employees with Employment Agreements using Naive Bayes Classifier," *Sinkron*, vol. 8, no. 1, pp. 1–8, 2023, doi:

- 10.33395/sinkron.v8i1.11968.
- [6] N. Parhusip, M. Iqbal, and Z. Sitorus, "Analysis Of Licensing Data Using Naive Bayes And Decision Tree Algorithms To Evaluate Analysis Of Licensing Data Using Naive Bayes And Decision Tree Algorithms To Evaluate The Performance Of Digital Public Services (Case Study: Investment and One-Stop Inte," vol. 07, no. 02, pp. 24–35, 2025, [Online]. Available: <https://journal.cattleyadf.org/index.php/jatilima/index>
- [7] M. Hairu Dzikri, I. Rizal Setiawan, and D. Indrayana, "penerapan algoritma naive bayes untuk mendekteksi penipuan lowongan pekerjaan," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 9, no. 2, pp. 102–109, Oct. 2024, doi: 10.51876/simtek.v9i2.392.
- [8] K. Rrmoku, B. Selimi, and L. Ahmedi, "Application of Trust in Recommender Systems—Utilizing Naive Bayes Classifier," *Computation*, vol. 10, no. 1, p. 6, Jan. 2022, doi: 10.3390/computation10010006.
- [9] A. D. Rahman, A. Ichsan Pradana, and D. Hartanti, "Prediksi Kelulusan Mahasiswa Pendidikan Kimia Universitas Sebelas Maret Dengan Metode Naive Bayes," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 11, pp. 419–429, Dec. 2024, doi: 10.52436/1.jpti.501.
- [10] Z. Sitorus, S. Pranoto, and S. Sutiono, "Comparison of Accuracy between Naïve Bayes and Decision Tree Methods for Property Tax (PBB-P2) Compliance in Tebing Tinggi City," *J. Inf. Technol. Comput. Sci. Electr. Eng.*, vol. 1, no. 2, pp. 121–128, 2024, doi: 10.61306/jitcse.v1i2.
- [11] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, pp. 1361–1368, Dec. 2023, doi: 10.25126/jtiik.2023107467.
- [12] muhammad nur akbar, hariani, and asep indra syahyadi, "analisis prediksi ketepatan masa studi mahasiswa menggunakan algoritma naive bayes classifier dan feature selection," *J. INSTEK (Informatika Sains dan Teknol.*, vol. 7, no. 2, pp. 306–315, Oct. 2022, doi: 10.24252/instek.v7i2.32576.
- [13] R. H. Sukarna and Y. Ansori, "implementasi data mining menggunakan metode naive bayes dengan feature selection untuk prediksi kelulusan mahasiswa tepat waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, Feb. 2022, doi: 10.47080/saintek.v6i1.1467.
- [14] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovensional dalam Prediksi Komplikasi Hipertensi," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 6, no. 1, pp. 37–49, Apr. 2024, doi: 10.35746/jtim.v6i1.488.
- [15] Y. Aprillia, Z. Alawi, and I. Sa'ida, "Implementasi Algoritma Naive Bayes Dengan Feature Selection Backward Elimination Dalam Pengklasifikasian Status Penderita Stunting Pada Balita," *J. Multidiscip. Appl. Quantum Inf. Sci.*, vol. 4, no. 2, pp. 1–6, 2023, doi: 10.32665/almantiq.v4i2.3238.
- [16] S. Hartati, N. A. Ramdhan, and H. A. SAN, "penerapan algoritma naive bayes untuk mendekteksi penipuan lowongan pekerjaan," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 223–234, Dec. 2022, doi: 10.46772/intech.v4i02.889.
- [17] A. Budianita, "Information Gain Berbasis Algoritma Naive Bayes Classifier Pada Pemodelan Prediksi Kelulusan," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 5, no. 1, pp. 1–10, May 2023, doi: 10.46772/intech.v5i1.1116.
- [18] S. Siswanto, I. Kurniawan, and S. A. Thamrin, "Naive Bayes Algorithm with Feature Selection Using Particle Swarm Optimization," *J. Varian*, vol. 7, no. 2, pp. 127–136, Jun. 2024, doi: 10.30812/varian.v7i2.2409.
- [19] Eko Budianto and Muhammad Iqbal, "Model Predictive Analysis of Performance in Training and Course Institutions Using Naive Bayes and K-Means Clustering," *J. Comput. Sci. Res.*, vol. 3, no. 1, pp. 10–16, Jan. 2025, doi: 10.65126/jocosir.v3i1.68.
- [20] F. M. Anto, L. S. Abimanyu, and T. Herdi, "Penerapan Algoritma Naïve Bayes Dengan Feature Selection Pada Data Penjualan Konstruksi," *J. Ilm. FIFO*, vol. 15, no. 2, p. 102, Apr. 2024, doi: 10.22441/fifo.2023.v15i2.002.
- [21] A. P. Permana, T. Chamidy, and C. Crysdian, "Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes dengan Seleksi Fitur Chi-Square," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 2, pp. 112–124, May 2023, doi: 10.14421/jiska.2023.8.2.112-124.