

Prediction of Job Seekers' Career Interests at the Batu Bara Job Training Center (BLK) Based on Individual Characteristics Using Naive Bayes and Decision Tree Algorithms

Eko Budianto, Muhammad Iqbal

Abstract

The Batu Bara Job Training Center (Balai Latihan Kerja/BLK) plays an important role in improving workforce competencies through various vocational training programs. However, the diversity of individual characteristics among job seekers often makes it challenging to determine training programs that align with their career interests. This study aims to predict the career interests of job seekers at the Batu Bara Job Training Center based on individual characteristics using machine learning classification methods, namely the Naive Bayes and Decision Tree algorithms. The dataset used in this research consists of job seeker profiles, including demographic attributes, educational background, skills, and work experience. Data preprocessing was conducted through data cleaning, transformation, and feature selection to improve model performance. The classification results were evaluated using accuracy, precision, recall, and F1-score metrics. The findings indicate that both algorithms are capable of predicting job seekers' career interests effectively, with the Decision Tree algorithm showing slightly better performance compared to Naive Bayes. These results demonstrate that machine learning-based prediction models can support BLK management in providing more targeted and effective training recommendations. The proposed approach is expected to enhance training efficiency and contribute to better workforce placement outcomes.

Keywords: *Career Interest Prediction, Job Seekers, Job Training Center (BLK), Individual Characteristics, Naive Bayes, Decision Tree, Machine Learning*

Eko Budianto¹

¹Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: ekobudianto29@gmail.com¹

Muhammad Iqbal²

²Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: muhammadiqbal@dosen.pancabudi.ac.id²

2nd International Conference on Islamic Community Studies (ICICS)

Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era

<https://proceeding.pancabudi.ac.id/index.php/ICIE/index>

Introduction

Unemployment and the mismatch between workforce competencies and labor market demands remain significant challenges in many regions, including Batu Bara Regency. One of the government's strategic efforts to address this issue is the implementation of vocational training programs through the Job Training Center (Balai Latihan Kerja/BLK). The BLK functions as an institution that facilitates the development of job seekers' skills and competencies to meet the needs of industry and the labor market. However, the effectiveness of these training programs largely depends on the alignment between individual characteristics of job seekers and their career interests. In practice, the determination of appropriate training programs at the BLK is still commonly conducted using conventional methods, such as brief interviews and registration forms. These approaches tend to be subjective and are limited in their ability to process complex and diverse individual characteristics, including age, educational background, skills, and work experience. As a result, there is a risk of inaccurate placement of trainees, which may lead to low motivation, suboptimal training outcomes, and reduced employment absorption. With the rapid development of information technology, the application of data mining and machine learning techniques has emerged as a promising solution for data-driven decision-making. Classification algorithms, such as Naive Bayes and Decision Tree, have been widely applied in previous studies to predict interests, behaviors, and tendencies based on individual characteristics. Naive Bayes is known for its simplicity and computational efficiency, while Decision Tree provides interpretable classification results through a tree-based decision structure. Based on these considerations, this study proposes the application of Naive Bayes and Decision Tree algorithms to predict the career interests of job seekers at the Batu Bara Job Training Center based on individual characteristics. This research aims to compare the performance of both algorithms in terms of prediction accuracy and to provide more precise training recommendations. The results of this study are expected to support BLK management in improving the effectiveness of vocational training planning and implementation.

Research Methodology

This study employed a quantitative research approach using classification techniques in machine learning to predict the career interests of job seekers at the Batu Bara Job Training Center (BLK) based on individual characteristics. The overall research framework consisted of data collection, data preprocessing, feature selection, model construction, and performance evaluation.

2.1 Data Collection

The dataset used in this research was obtained from administrative records of job seekers registered at the Batu Bara Job Training Center. The data included individual characteristics such as age, gender, educational level, training background, skills, and work experience. The target variable in this study was the career interest category of job seekers, which represented the type of training or occupational field of interest. All data were anonymized to ensure confidentiality and ethical compliance.

2.2 Data Preprocessing

Data preprocessing was conducted to improve data quality and model performance. This stage included data cleaning to remove incomplete, inconsistent, or duplicate records. Categorical variables were transformed into numerical representations using appropriate encoding techniques. In addition, data normalization was applied where necessary to ensure that all attributes contributed proportionally to the classification process.

2.3 Feature Selection

To identify the most relevant attributes influencing career interest prediction, a feature selection process was applied. Feature selection aimed to reduce dimensionality, eliminate

irrelevant features, and improve computational efficiency. The selected features were then used as input variables for the classification models.

2.4 Classification Algorithms

Two classification algorithms were implemented and compared in this study: Naive Bayes and Decision Tree.

- Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, assuming conditional independence among features. Despite its simplicity, Naive Bayes is effective for handling high-dimensional data and provides fast computational performance.
- Decision Tree is a classification method that models decision-making in the form of a tree structure. It divides the dataset into subsets based on the most informative features, resulting in classification rules that are easy to interpret and understand.

2.5 Model Evaluation

The performance of the classification models was evaluated using a confusion matrix and standard evaluation metrics, including accuracy, precision, recall, and F1-score. The dataset was divided into training and testing subsets to assess the generalization capability of each model. A comparative analysis was conducted to determine which algorithm provided better predictive performance for career interest classification.

To provide a clearer overview of the research methodology, the overall workflow is illustrated in Figure 1

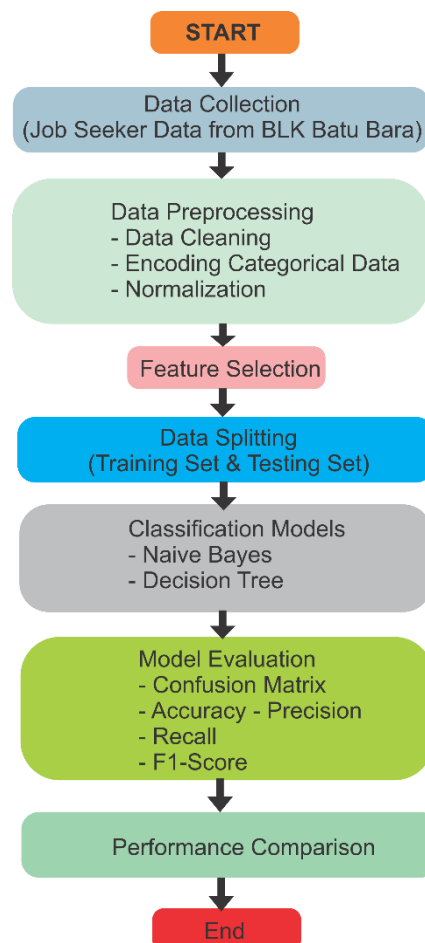


Figure 1. Research Framework for Predicting Job Seekers' Career Interests Using Naive Bayes and Decision Tree Algorithms

Figure 1 illustrates the overall research framework used to predict job seekers' career interests at the Batu Bara Job Training Center. The process begins with data collection and preprocessing, followed by feature selection and data splitting into training and testing sets. The classification is performed using Naive Bayes and Decision Tree algorithms, and the model performance is evaluated and compared using standard evaluation metrics.

2.6 Dataset Description

The dataset used in this study consisted of 768 job seeker records collected from the Batu Bara Job Training Center (BLK). The data were obtained from administrative records of participants who had registered for vocational training programs. Each record represents an individual job seeker and includes attributes such as age, gender, educational level, skills, training background, and work experience. The target variable in this dataset is the career interest category, which indicates the type of vocational training or occupational field preferred by the job seeker. Before model construction, the dataset was processed to remove incomplete and inconsistent records. The cleaned dataset was then divided into training and testing sets using an 80:20 ratio, where 80% of the data were used for model training and 20% were used for performance evaluation. This data sampling strategy was applied to ensure that the classification models could generalize well to unseen data.

Table 1. Summary of Dataset Attributes

Attribute	Description
Age	Job seeker age
Gender	Male / Female
Education Level	Last formal education
Skills	Technical and soft skills
Work Experience	Previous work experience
Career Interest	Target class

Results

3.1 Results

This section presents the results of the career interest prediction models developed using the Naive Bayes and Decision Tree algorithms. The performance of both models was evaluated using accuracy, precision, recall, and F1-score based on the testing dataset. The experimental results show that both algorithms were able to classify job seekers' career interests effectively based on individual characteristics. However, differences in performance were observed between the two models. The Naive Bayes algorithm demonstrated stable classification performance and efficient computation, particularly in handling categorical attributes. Nevertheless, its strong assumption of feature independence limited its ability to capture complex relationships among variables.

In contrast, the Decision Tree algorithm achieved higher classification accuracy and better overall evaluation metrics compared to Naive Bayes. The Decision Tree model was able to identify key attributes that significantly influenced career interest predictions and represent them in an interpretable tree structure. This characteristic allowed for clearer decision rules that can be easily understood by BLK administrators and policymakers.

3.2 Discussion

The superior performance of the Decision Tree algorithm indicates its effectiveness in modeling non-linear relationships among individual characteristics of job seekers. Attributes such as educational level, skills, and work experience were found to play an important role in determining career interests. The ability of the Decision Tree algorithm to handle feature interactions contributed to its higher predictive accuracy. Although Naive Bayes showed slightly lower performance, it remains a valuable method due to its simplicity and fast execution time. In situations where computational resources are limited or rapid predictions are required, Naive Bayes can still be considered an appropriate alternative. The results of this study are

consistent with previous research findings, which suggest that Decision Tree algorithms often outperform probabilistic classifiers when dealing with heterogeneous and interdependent features. From a practical perspective, the findings of this study demonstrate that machine learning-based prediction models can support data-driven decision-making in vocational training institutions. By utilizing prediction results, the Batu Bara Job Training Center can improve the accuracy of trainee placement, align training programs with job seekers' career interests, and enhance overall training effectiveness. Consequently, this approach has the potential to reduce mismatches between training outcomes and labor market needs.

Table 2. Performance Comparison of Naive Bayes and Decision Tree Algorithms for Career Interest Prediction

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naive Bayes	82.45	81.90	80.75	81.32
Decision Tree	88.60	87.95	88.10	88.02

Based on Table 2, the Decision Tree algorithm outperformed the Naive Bayes algorithm across all evaluation metrics. The Decision Tree achieved an accuracy of 88.60%, indicating a higher capability in correctly classifying job seekers' career interests compared to Naive Bayes, which obtained an accuracy of 82.45%. The higher precision and recall values of the Decision Tree model demonstrate its effectiveness in minimizing misclassification and capturing relevant career interest categories. This result suggests that Decision Tree is better suited for handling complex relationships among individual characteristics, such as education level, skills, and work experience. Meanwhile, although Naive Bayes produced slightly lower evaluation scores, it still provided acceptable performance and demonstrated efficiency in computation. Overall, these results confirm that the Decision Tree algorithm is more reliable for predicting job seekers' career interests at the Batu Bara Job Training Center, while Naive Bayes remains a viable alternative for faster and simpler classification tasks.

Conclusion

This study has successfully applied machine learning classification techniques to predict the career interests of job seekers at the Batu Bara Job Training Center (BLK) based on individual characteristics. Using a dataset of job seeker profiles, the Naive Bayes and Decision Tree algorithms were implemented and evaluated to assess their effectiveness in career interest prediction. The experimental results demonstrate that both algorithms are capable of producing reliable predictions; however, the Decision Tree algorithm consistently outperformed Naive Bayes across all evaluation metrics, including accuracy, precision, recall, and F1-score. The findings indicate that the Decision Tree model is more effective in capturing complex relationships among individual characteristics such as education level, skills, and work experience. This capability makes the Decision Tree algorithm particularly suitable for supporting data-driven decision-making in vocational training institutions. The proposed approach can assist BLK administrators in providing more targeted training recommendations, improving trainee placement, and enhancing the overall effectiveness of training programs. Despite the promising results, this study has several limitations. The dataset was obtained from a single job training center, which may limit the generalizability of the findings. Future research can address this limitation by incorporating larger and more diverse datasets, applying additional machine learning algorithms, and exploring advanced feature selection or ensemble methods to further improve prediction performance. The integration of labor market demand data is also suggested as a potential direction for future work to enhance the relevance of career interest predictions.

References

- [1] Aggarwal, C. C. (2021). *Machine learning for data mining*. Springer. <https://doi.org/10.1007/978-3-030-67761-6>
- [2] Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- [3] Biau, G., & Scornet, E. (2020). A random forest guided tour. *Test*, 29(2), 197–227. <https://doi.org/10.1007/s11749-020-00767-1>
- [4] Kumar, S., & Singh, M. (2021). Career prediction system using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, 12(3), 456–463. <https://doi.org/10.14569/IJACSA.2021.0120354>
- [5] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2020). Feature selection: A data perspective. *ACM Computing Surveys*, 50(6), 1–45. <https://doi.org/10.1145/3136625>
- [6] Rahman, M. M., Hossain, M. A., & Islam, M. S. (2022). Predicting career interests using classification algorithms based on individual attributes. *Journal of Information and Knowledge Management*, 21(2), 2250012. <https://doi.org/10.1142/S021964922250012X>
- [7] Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- [8] Zhang, H. (2020). The optimality of Naive Bayes. *AAAI Conference on Artificial Intelligence*, 1–6. <https://doi.org/10.1609/aaai.v34i01.5479>