JOURNAL LA MULTIAPP

VOL. 06, ISSUE 02 (341-352), 2025
DOI: 10.37899/journallamultiapp.v6i2.2050

Implementation of Machine Learning-Based Classification Model in Employee Recruitment Decision Prediction

Muhammad Fauzan Nur Adillah[1], Sinung Suakanto[1], Nur Ichsan Utama[1]

[1]*Master of Information Systems, Faculty of Industrial Engineering, Telkom University, Bandung, Indonesia*

**Corresponding Author: Muhammad Fauzan Nur Adillah*
Email: mfauzanna@student.telkomuniversity.ac.id

Article Info

Article history:
Received 3 March 2025
Received in revised form 21 March 2025
Accepted 8 April 2025

Keywords:
Recruitment Prediction
Machine Learning
Recruitment Decision
Classification
Civil Servant Selection

Abstract

Employees are vital assets for any organization, and accurate recruitment decision-making is crucial for the organization's long-term success. Incorrect decisions can lead to high costs due to re-hiring processes, onboarding, and decreased productivity. This study aims to develop a recruitment decision prediction model using data obtained from the Final Results of the 2024 CPNS Recruitment in the Ministry of Finance. The data includes attributes such as educational background, age, GPA, SKD Score, and SKB Score. To understand the relationships between variables, correlation analysis was conducted using a correlation matrix and heatmap visualization. Additionally, data exploration was performed using histograms to show the influence of attributes on recruitment decisions. This study employs five machine learning algorithms for prediction: Linear Support Vector Machine, Decision Tree (C5.0), Random Forest, k-Nearest Neighbor (k-NN), and Naïve Bayes Classifier. The results indicate that some attributes significantly influence recruitment decisions, and machine learning models can identify candidates who are more suitable for the available positions. Among the five models tested, Naïve Bayes proved to be the most effective, achieving an accuracy of 88% and an AUC of 0.97, demonstrating its strong performance in distinguishing positive and negative classes. The key factors contributing to the model's success include relevant feature selection, data quality, as well as appropriate preprocessing and validation techniques. This model is expected to enhance objectivity, efficiency, and accuracy in employee recruitment processes, thereby assisting organizations in making more precise and fair decisions.

Introduction

Employee Recruitment is a crucial process for any organization as it directly impacts workforce quality and efficiency (Pampouktsi et al., 2021). Recruitment decisions involve evaluating candidates based on various criteria, such as educational qualifications, professional experience, skills, and personality traits (Goretzko & Israel, 2022). While, in theory, candidate selection predictions can be made using simple calculations and logic, the complexity and large volume of data in modern recruitment make manual approaches less effective and prone to human bias.

In this context, applying machine learning-based classification models offers a more objective, efficient, and accurate solution (Smelyakov et al., 2023). Simple algorithms may handle basic cases, but they often fail to address non-linear relationships and complex interactions between various factors influencing recruitment decisions. For instance, work

ISSN: 2716-3865 (Print), 2721-1290 (Online)

experience and technical skills may have different weights depending on the required position, and factors such as personality or organizational cultural fit are challenging to measure manually. Machine learning algorithms excel at identifying patterns and relationships in large datasets, making them highly effective for this research. By utilizing historical recruitment data, these models can predict a candidate’s likelihood of success, thereby accelerating and enhancing the recruitment process efficiency (Reddy et al., 2020).

Moreover, errors in recruitment decision-making can result in significant costs for organizations, including re-hiring, training new employees, and decreased productivity (Pekdas et al., 2024). By leveraging machine learning, organizations can minimize these risks by identifying candidates most likely to succeed in a given role based on historical data (Gupta et al., 2024). Machine learning algorithms can also adapt to changing trends and organizational needs, providing flexibility that traditional methods lack.

This study aims to explore the use of machine learning-based classification models in predicting employee recruitment decisions. Although simple predictions can be made using basic logic, this approach is often insufficient to handle the complexity and dynamics of modern recruitment processes. By utilizing historical data and more advanced algorithms, this research seeks to provide a more accurate and efficient solution, improving recruitment decision quality and supporting better workforce planning.

Methods

K-Nearest Neighbour

K-Nearest Neighbour (KNN) is an instance-based machine learning algorithm used for classification and regression tasks. This algorithm works by identifying the k nearest neighbors of a new data point based on a certain distance metric, such as Euclidean Distance, Manhattan Distance, or Minkowski Distance. After calculating the distance, the class or value of the new data is determined based on the majority class of the nearest neighbors (for classification tasks) or the average value of the neighbors (for regression tasks) (Halder et al., 2024).

KNN does not require a training process because it is a non-parametric algorithm, meaning it operates directly on the training data when making predictions. The prediction process begins by determining the distance between the new data and all training data, then sorting the data based on the calculated distance to select the k nearest neighbors. The value of k significantly influences the algorithm's results, as it affects the classification or regression outcomes.

Support Vector Machine

Support Vector Machine (SVM) is a supervised learning machine learning algorithm used for classification and regression tasks. The main principle of SVM is to find the optimal hyperplane that separates the data into two or more classes with the maximum margin. The hyperplane is a line (in two-dimensional data) or a plane (in higher-dimensional data) that divides the data space in such a way that the distance between the hyperplane and the nearest data points from each class (called support vectors) is maximized. By maximizing the margin, SVM aims to produce good generalization on unseen data.

In SVM, data that cannot be linearly separated can be handled using kernel functions, such as polynomial kernel, radial basis function (RBF), or sigmoid kernel. These kernel functions allow the data to be mapped into a higher-dimensional space where it can be linearly separated. Due to its flexibility in handling both linear and non-linear data, SVM is frequently used in various fields, such as face recognition, text classification, and medical diagnosis (Tian et al., 2022).

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Naïve Bayes Classifier

The Naïve Bayes Classifier is a probabilistic machine learning algorithm used for classification tasks. It is based on Bayes' Theorem, which calculates the probability of a class based on the evidence provided by the features in the data. It is called "naïve" because the algorithm assumes that all features are independent of each other, even though, in reality, these features are often correlated. Despite this simple assumption, Naïve Bayes often provides accurate and efficient results, particularly in large datasets.

The algorithm works by calculating the posterior probability for each class using prior probability and conditional probabilities for the features in the data. Once the probabilities are computed, the new data is classified into the class with the highest probability. Naïve Bayes is widely used in tasks such as text classification, sentiment analysis, and spam filtering due to its ability to handle large data sets quickly and efficiently (Blanquero et al., 2021).

Random Forest

Random Forest is an ensemble learning machine learning algorithm used for classification and regression tasks. This algorithm works by building a large number of decision trees during training, then combining the prediction results from each tree to make a final decision through a voting mechanism (for classification) or averaging (for regression). Each tree in the Random Forest is trained using a random subset of data obtained through the bagging method (Bootstrap Aggregating), which increases variability among the trees and reduces the risk of overfitting.

At each data split in the decision tree, only a small subset of features is randomly selected to be considered. This technique, known as feature randomness, helps improve model accuracy by reducing correlations among the trees. With a diverse set of decision trees, Random Forest is able to produce a robust, stable, and efficient model for handling complex data and datasets with many features or classes (Amaliah et al., 2022).

Research Systematics

This research adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as a systematic approach in the development of a machine learning-based classification model for predicting recruitment decisions (Saputra et al., 2024). The method consists of six interrelated stages within an iterative cycle, allowing the research process to be structured and optimized based on the evaluations conducted (Rianti et al., 2023).

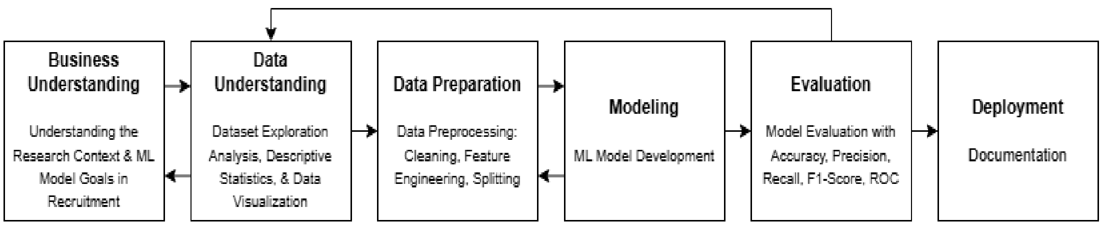

Figure 1. Research Systematics with CRISP-DM

The stages of the research systematics using the CRISP-DM method are as follows:

Business Understanding, which aims to understand the research context and define the objectives of the machine learning model in the employee selection process. In this stage, the research identifies the problems faced in recruitment, such as subjectivity in decision-making and inefficiency in the manual selection process. Additionally, success criteria for the model are established, including prediction accuracy, result reliability, and effectiveness in reducing recruitment bias.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Data Understanding, where the dataset used is analyzed to determine its structure, quality, and distribution. Exploratory data analysis is conducted to understand patterns in the dataset, identify missing values, and determine relationships between available variables and employee recruitment decisions. Data visualization is also used to identify potential issues in the dataset and ensure optimal data representation before moving to the modeling stage.

Data Preparation, which involves preprocessing the data so it can be used in machine learning models. This process includes data cleaning to handle missing values and outliers, feature engineering to create more informative features, and data transformation such as normalizing numerical variables. After the data is ready, the dataset is split into a training set and a testing set to ensure that the model can be tested with previously unseen data.

Modeling, which is the core of this research, where various machine learning algorithms are applied to build the classification model. The algorithms used include Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbour (KNN), and Naive Bayes Classifier. Each model is tested and compared based on predefined evaluation metrics. Cross-validation is used to improve model performance and avoid overfitting.

Evaluation is conducted to assess the performance of the models based on various metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. The best model is selected by considering the balance between precision and recall to generate more accurate predictions and minimize classification errors. Additionally, an analysis of the model's interpretability is performed to ensure that the decisions generated can be explained rationally.

Deployment, which includes documentation of the research results. This research documents the main findings and provides recommendations for further development, such as integrating the model into AI-based systems or testing with a broader dataset.

Data Sources

In this study, the data collection process was carried out to obtain relevant information for building a machine learning-based classification model to predict recruitment decisions for the 2024 Ministry of Finance Civil Servant Candidate (CPNS) recruitment (Anand & Dubey, 2022). The dataset used in this study was obtained from the announcement of the recruitment results for CPNS in the Ministry of Finance in 2024. This dataset consists of 613 samples and includes several features relevant to determining the CPNS recruitment outcomes. The selection of this dataset is based on its high relevance to the recruitment process in the public sector, particularly within the Ministry of Finance (Kemenkeu), which has a stringent and systematic recruitment process (Zuhri & Harani, 2023). Additionally, CPNS recruitment is one of the most competitive, making the prediction of candidate success based on historical data crucial to optimize the selection process (Barokah & Gunawan, 2023).

The use of the CPNS Kemenkeu 2024 recruitment dataset is strongly justified, as it contains a wide variety of variables. Variables such as age, GPA, Basic Competency Test (SKD) score, Field Competency Test (SKB) score, and final recruitment decision (Status) provide comprehensive and detailed information about the candidate profiles needed in CPNS selection (Titisari & Ikhwan, 2021). These variables significantly impact the selection results and are commonly used in many government agencies for CPNS recruitment or similar processes.

By using this dataset, the model developed can not only make predictions for Kemenkeu but also be applied to other government agencies with similar selection processes (Yuliani & Aliyyah, 2024). The selection process in many government agencies, whether in terms of academic performance (GPA), basic competencies, or field competencies, often involves criteria similar to those used in Kemenkeu (Kharisma & Wening, 2023). Therefore, the classification model built in this study can be widely applied in the public sector and even in

ISSN: 2716-3865 (Print), 2721-1290 (Online)

the private sector, as long as the selection criteria are similar. This flexibility allows the model to benefit various organizations that conduct recruitment based on similar or identical data (Vivek, 2023).

The variables collected in this study include: age, GPA, Basic Competency Test (SKD) score, Field Competency Test (SKB) score, and the target variable, which is the Status (final recruitment decision). The selection of age as an input variable is based on the fact that age can be associated with experience and maturity, which may affect a candidate's future performance. GPA was chosen because it is an important indicator often used to assess a candidate's academic quality (Syafaruddin, 2024). Meanwhile, the SKD and SKB scores are highly relevant as they are key tests designed to measure basic competencies and technical skills, which directly relate to the candidate's ability to perform tasks at the Ministry of Finance. The target variable, Status, reflects the final decision regarding the candidate's pass or fail status, which is the ultimate result of the entire selection process (Hein et al., 2024).

By incorporating these variables into a machine learning-based classification model, it is expected that more accurate predictions can be made regarding the candidates most likely to succeed in the 2024 Ministry of Finance CPNS selection. These predictions not only help speed up the recruitment process but also increase objectivity and reduce the risk of human error in assessments (Yildizhan et al., 2023).

Results and Discussion

Evaluation Metrics

The comparison of model performance is summarized in the following table to make the analysis easier and more comprehensive. The table presents key metrics including accuracy, precision, recall, F1-score, and AUC-ROC for each machine learning model evaluated in the study.

Table 1. SVM Classification Report

Components	***Accuracy***	***Precision***	***Recall***	***F1-Score***
P/L	0.86	1.00	1.00	1.00
TH		1.00	1.00	1.00
TL		0.71	0.86	0.77
TMS-1		0.80	0.62	0.70
Macro avg.		0.88	0.87	0.87
Weighted avg.		0.86	0.86	0.85

Table 2. Decision Tree Classification Report

Components	***Accuracy***	***Precision***	***Recall***	***F1-Score***
P/L	0.83	1.00	0.97	0.98
TH		1.00	0.92	0.96
TL		0.69	0.79	0.73
TMS-1		0.71	0.65	0.68
Macro avg.		0.85	0.83	0.84
Weighted avg.		0.83	0.83	0.83

Table 3. Random Forest Classification Report

Components	***Accuracy***	***Precision***	***Recall***	***F1-Score***
P/L	0.86	1.00	0.97	0.98
TH		1.00	1.00	1.00
TL		0.73	0.86	0.79
TMS-1		0.77	0.65	0.71

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Macro avg.		0.88	0.87	0.87
Weighted avg.		0.86	0.86	0.86

Table 4. K-NN Classification Report

Components	***Accuracy***	***Precision***	***Recall***	***F1-Score***
P/L	0.74	0.94	0.91	0.92
TH		0.92	1.00	0.96
TL		0.61	0.61	0.61
TMS-1		0.58	0.58	0.58
Macro avg.		0.76	0.77	0.77
Weighted avg.		0.75	0.74	0.74

Table 5. Naïve Bayes Classification Report

Components	***Accuracy***	***Precision***	***Recall***	***F1-Score***
P/L	0.88	1.00	1.00	1.00
TH		1.00	1.00	1.00
TL		0.72	0.93	0.81
TMS-1		0.89	0.62	0.73
Macro avg.		0.90	0.89	0.88
Weighted avg.		0.89	0.88	0.87

The differences in results between each classification model can be explained by several key factors, such as the characteristics of the dataset (e.g., class imbalance, relationships between features, and data distribution), as well as the intrinsic strengths and weaknesses of each model. Naïve Bayes performs well due to its ability to handle class imbalance, despite the often inaccurate assumption of independence between features. SVM and Random Forest work well with data that has complex relationships between features, but still face challenges with minority classes. Decision Trees tend to be vulnerable to overfitting, especially with poorly structured data. On the other hand, K-NN shows poor results due to its reliance on proximity between data points, which becomes problematic in high-dimensional and imbalanced data.

Confusion Matrix

The results obtained from the evaluation of the classification models can be explained by considering several factors that influence the performance of each model in predicting recruitment decisions for employees.

Data Quality and Structure: The data used in this study, particularly in the TMS-1 category, shows an imbalanced distribution. This class imbalance plays a significant role in affecting model performance. Probabilistic models, such as Naïve Bayes, perform better in handling data imbalance due to their robust approach to uneven class distributions. In contrast, decision tree-based models like Decision Tree and Random Forest are more prone to overfitting when dealing with imbalanced data. This shows that although Random Forest is more stable than Decision Tree, both struggle with handling minority classes such as TMS-1.

Advantages of Probabilistic Models: Naïve Bayes relies on a probabilistic approach that is more effective at handling class imbalance and unstructured data distributions. This model calculates the probability of each class based on available data, making it able to deliver more stable results even when the data distribution is uneven. This makes Naïve Bayes highly effective in predicting recruitment decisions, where the distribution between accepted and rejected candidates is often imbalanced.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Figure 2. Confusion Matrix

Limitations of Distance-Based Models: K-NN, which works by measuring the proximity between data points, faces difficulties when the data has many dimensions or when the data distribution is imbalanced. Its reliance on proximity makes it more susceptible to outliers and irregular distributions. The TMS-1 class, which has fewer samples and is more difficult to predict, presents an additional challenge for K-NN in correctly separating the classes, resulting in lower performance.

Ensemble Model Characteristics: Random Forest, as an ensemble model that combines several Decision Trees, is able to reduce the overfitting issue found in a single Decision Tree. However, this model still struggles with handling more extreme class imbalances, such as those present in TMS-1. This suggests that while Random Forest is more robust in handling data complexity, it still needs further optimization to manage datasets with highly imbalanced class distributions.

Overall, the differences in results indicate that each model has its own strengths, but Naïve Bayes stands out due to its simplicity and effectiveness in handling class imbalance, as well as its ability to provide stable and efficient results in predicting recruitment decisions.

ROC/AUC Curve

Based on the results of the ROC curve and AUC for the five classification models tested in this study, it can be concluded that SVM and Naïve Bayes are the most effective models for predicting employee recruitment decisions. Both models showed very high AUC values, 0.98 and 0.97 respectively, indicating their outstanding ability to distinguish between positive and negative classes. SVM and Naïve Bayes exhibited nearly identical ROC curves, demonstrating stable and excellent performance across all categories.

Although Random Forest was slightly lower with an AUC of 0.96, it still demonstrated very good and reliable classification results, especially due to its ability to handle overfitting issues commonly found in decision tree-based models.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

K-NN showed a lower AUC (0.90) compared to the other models. This indicates that while it still provides reasonably good results, K-NN is more prone to prediction errors, particularly when the data has an imbalanced class distribution or contains outliers (Cholil et al., 2021).

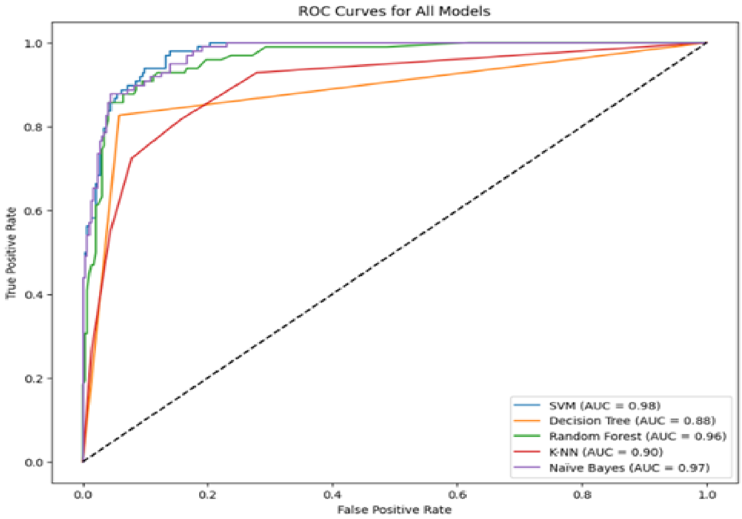

Figure 3. ROC/AUC Curve

Finally, Decision Tree showed the lowest AUC among the tested models (0.88). This model struggled to differentiate between the existing classes, indicating that Decision Tree is less effective in handling imbalanced or more complex class distributions, resulting in lower performance compared to the other models tested.

Overall, SVM and Naïve Bayes proved to be the best choices for predicting recruitment decisions, as both models demonstrated excellent performance in distinguishing classes and provided more consistent and efficient results compared to the other models.

Selection of the Best Model

Based on the evaluation of the evaluation metrics, confusion matrix, and ROC curve, Naïve Bayes emerges as the best model for predicting employee recruitment decisions. Naïve Bayes demonstrates the highest accuracy (0.88) and exhibits a very good balance between precision and recall, providing stable results across all classes, including the minority class. Its ability to handle class imbalance without sacrificing accuracy makes it an extremely effective choice in this context.

Additionally, Naïve Bayes also achieves a very high AUC (0.97), further confirming its ability to effectively separate positive and negative classes. This high AUC indicates that the model is not only accurate in classification but also maintains a high True Positive Rate with a low False Positive Rate, making it excellent at identifying suitable candidates for recruitment.

Although SVM has a slightly higher AUC (0.98) than Naïve Bayes and performs exceptionally well in separating classes, SVM is more complex to implement and requires more intricate parameter tuning. The complexity of SVM in terms of parameter optimization and handling non-linear data poses a challenge in its application in this scenario, especially if the model is to be used by individuals who need simplicity and efficiency. Furthermore, SVM is more prone to errors in the minority class (TMS-1), despite having a high AUC, which causes its stability to be slightly lower than that of Naïve Bayes in terms of overall performance.

Furthermore, while Random Forest also achieves a high AUC (0.96) and performs well, it is more complex and requires more resources for implementation and parameter tuning compared to Naïve Bayes, which offers simplicity with nearly equivalent effectiveness.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

K-NN and Decision Tree show lower performance, with K-NN being less effective in handling high-dimensional data and Decision Tree being prone to overfitting.

Overall, Naïve Bayes is the most stable, effective, and efficient model for the task of predicting employee recruitment decisions. Its combination of high accuracy, excellent AUC, and balance between precision and recall makes it the best choice for application in this scenario. Naïve Bayes offers better simplicity and stability, making it more suitable for situations that require quick and reliable modeling compared to SVM, even though SVM has a slightly higher AUC.

Most Influential Factors in Prediction

In this study, an analysis was conducted to identify the most influential factors in predicting the graduation of CPNS (Civil Servant Candidates) participants using a feature importance approach. This approach aims to determine how much each feature contributes to the decisions made by the machine learning model. To perform this analysis, three models were selected, each with an inherent mechanism for determining feature importance: Support Vector Machine (SVM), Decision Tree, and Random Forest.

The analysis results show that the Selection of Competency Field (SKB) score is the most dominant factor in determining the CPNS graduation, followed by the Basic Competency Selection (SKD) score, Cumulative Grade Point Average (IPK), and age. The significant role of SKB in this prediction indicates that the specific competencies tested in the field are more critical than academic factors or the age of the participants.

However, not all models used in this study were employed for feature importance analysis. The k-Nearest Neighbor (k-NN) and Naïve Bayes models were excluded from this analysis due to limitations in their methodology for providing explicit weights to individual features.

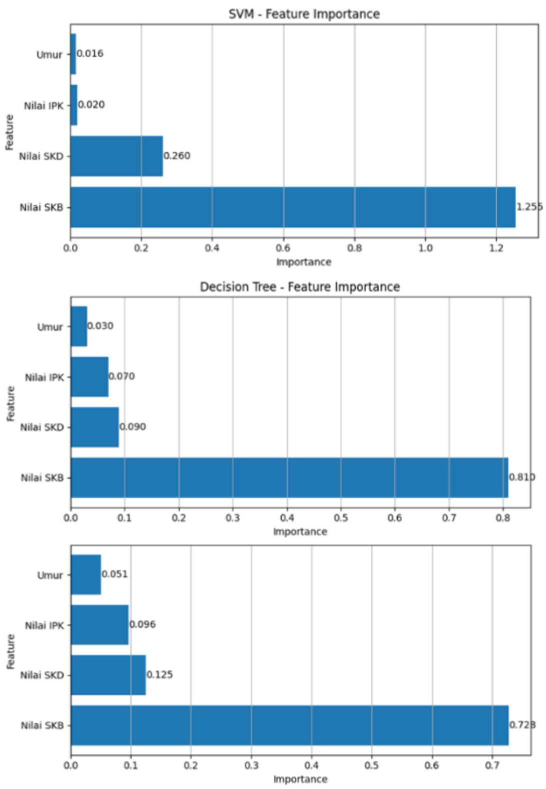

Figure 4. Feature Importance of Each Classification Model

In this study, although both k-NN and Naïve Bayes models were utilized, they were not used for the feature importance analysis because of limitations in their methodologies to measure the explicit contribution of each feature in the classification process.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

The k-NN model classifies samples based on their proximity to other samples in the feature space (e.g., Euclidean distance). In this approach, there is no explicit weighting for individual features, as classification decisions are solely based on the similarity of distances between samples (Cholil et al., 2021). This means that k-NN cannot provide information on how important or influential each feature is in affecting the model's decisions. Previous research indicates that while k-NN relies more on the proximity of features in a multidimensional space, it lacks the ability to assess or measure the relative contribution of each feature in the classification process (Muchtar et al., 2024). Therefore, k-NN cannot be used for feature importance analysis, which requires an explicit evaluation of the contribution of each feature.

On the other hand, Naïve Bayes is a probabilistic model that assumes independence between features. This assumption simplifies the calculation of class probabilities but eliminates the relationships between features when predicting the class (Hartono et al., 2021). Consequently, Naïve Bayes does not have a mechanism to measure the relative influence between one feature and another in affecting the classification decision (Tangkelobo et al., 2023). In Naïve Bayes, each feature is calculated separately under the assumption of independence, which leads to a limitation in determining the collective contribution of all features to the prediction outcome. Further studies suggest that Naïve Bayes is not designed to provide feature importance but rather to handle data with simple probabilistic distributions (Kurniadi et al., 2022). Therefore, Naïve Bayes also cannot be used for feature importance analysis.

Conclusion

The application of machine learning-based classification models in predicting recruitment decisions can significantly enhance efficiency and objectivity in the selection process. In this study, five classification models—Naïve Bayes, SVM, Random Forest, Decision Tree, and k-NN—were tested to predict recruitment decisions based on a dataset from the CPNS (Civil Servant Candidate) Announcement of the Ministry of Finance in 2024. The dataset consists of 613 samples with four feature attributes and one target attribute.

After data preprocessing steps, including cleaning, normalization, and encoding, as well as model training using training data and testing with test data, the evaluation results indicated that Naïve Bayes is the most effective model. This model achieved the highest accuracy (88%) and demonstrated excellent balance between precision, recall, and F1-Score, making it the best choice for predicting recruitment decisions. Additionally, Naïve Bayes also achieved a very high AUC (0.97), demonstrating its ability to effectively differentiate between positive and negative classes.

Key factors influencing the success of the classification model include the selection of relevant features, data quality, appropriate preprocessing techniques, and model validation methods such as cross-validation to avoid overfitting. Based on the comprehensive evaluation results, Naïve Bayes has proven to be the most effective and efficient model for enhancing data-driven decision-making and ensuring fairness in the recruitment process.

References

Amaliah, S., Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, *4*(3), 121–127. https://doi.org/10.35580/variansiunm31

Anand, A., & Dubey, Mr. S. (2022). CV Analysis Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *10*(5), 1316–1322. https://doi.org/10.22214/ijraset.2022.42295

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Barokah, F. U., & Gunawan, A. (2023). Strategi Rekrutmen dan Seleksi yang Efektif untuk Meningkatkan Kualitas Tenaga Kerja. *GLOBAL: Jurnal Lentera BITEP*, *1*(02), 60–65. https://doi.org/10.59422/global.v1i02.145

Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable Selection for Naïve Bayes Classification. *Computers & Operations Research*, *135*, 105456. https://doi.org/10.1016/j.cor.2021.105456

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT: (Indonesian Journal on Computer and Information Technology*, *6*(2), 118–127. https://doi.org/10.62951/modem.v2i3.130

Goretzko, D., & Israel, L. S. F. (2022). Pitfalls of Machine Learning-Based Personnel Selection. *Journal of Personnel Psychology*, *21*(1), 37–47. https://doi.org/10.1027/1866-5888/a000287

Gupta, A., Gupta, S., Mall, P. K., & Srivastava, S. (2024). ML-CPC: A Pathway for Machine Learning Based Campus Placement Classification. *Journal of Electrical Systems*, *20*(3s), 1453–1464. https://doi.org/10.52783/jes.1521

Halder, R. K., Uddin, M. N., Uddin, Md. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest Neighbor Algorithm: a Comprehensive Review and Performance Analysis of Modifications. *Journal of Big Data*, *11*(1), 113. https://doi.org/10.1186/s40537-024-00973-y

Hartono, H., Hajjah, A., & Marlim, Y. N. (2023). Penerapan Metode Naïve Bayes Classifier untuk Klasifikasi Judul Berita. *Jurnal SimanteC*, *12*(1), 37–46.

Hein, A. Z., Elving, W. J. L., Koster, S., & Edzes, A. (2024). Is Your Employer Banding Strategy Effective? The Role of Employee Predisposition in Achieving Employer Attractiveness. *Corporate Communications: An International Journal*, *29*(7), 1–20. https://doi.org/10.1108/CCIJ-07-2022-0070

Kharisma, I. M., & Wening, N. (2023). Peran Rekrutmen dan Seleksi Terhadap Kinerja Karyawan Perusahaan: Sebuah Tinjauan Literatur Sistematik. *Jurnal E-Bis*, *7*(1), 61–80. https://doi.org/10.37339/e-bis.v7i1.1111

Kurniadi, D., Nuraeni, F., & Lestari, S. M. (2022). Implementasi Algoritma Naïve Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana. *Jurnal Sistem Cerdas*, *5*(2), 63–82. https://doi.org/10.37396/jsc.v5i2.215

Muchtar, M., & Muchtar, R. A. (2024). Perbandingan Metode KKN dan SVM dalam Klasifikasi Kematangan Buah Mangga Berdasarkan Citra Hsv dan Fitur Statistik. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4010

Pampouktsi, P., Avdimiotis, S., Maragoudakis, M., & Avlonitis, M. (2021). Applied Machine Learning Techniques on Selection and Positioning of Human Resources in the Public Sector. *Open Journal of Business and Management*, *09*(02), 536–556. https://doi.org/10.4236/ojbm.2021.92030

Pekdas, I. G., Uflaz, E., Tornacı, F., Arslan, O., & Turan, O. (2024). Developing a Machine Learning-Based Evaluation System for the Recruitment of Maritime Professionals. *Ocean Engineering*, *313*, 119406. https://doi.org/10.1016/j.oceaneng.2024.119406

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Reddy, D. J. M., Regella, S., & Seelam, S. R. (2020). Recruitment Prediction using Machine Learning. *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 1–4. https://doi.org/10.1109/ICCCS49678.2020.9276955

Rianti, A., Majid, N. W. A., & Fauzi, A. (2023). CRISP-DM: Metodologi Proyek Data Science. *SENATIB: Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 107–114.

Saputra, D. B., Atina, V., & Nastiti, F. E. (2024). Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit Menggunakan Algoritma Random Forest. *IDEALIS : InDonEsiA JournaL Information System*, *7*(2), 240–247. https://doi.org/10.36080/idealis.v7i2.3244

Smelyakov, K., Hurova, Y., & Osiievskyi, S. (2023). Analysis of the Effectiveness of Using Machine Learning Algorithms to Make Hiring Decisions. *7th International Conference on Computational Linguistics and Intelligent Systems*, *1*(3387).

Syafaruddin, S. (2024). Recruitment, Job Placement and Employee Performance. *Advances in Human Resource Management Research*, *2*(3), 179–190. https://doi.org/10.60079/ahrmr.v2i3.163

Tangkelobo, E., Mayaut, W., Listanto, H., Binanto, I., & Sianipar, N. F. (2023). Perbandingan Algoritma Klasifikasi Random Forest, Gaussian Naive Bayes, dan K-Nearest untuk Data Tidak Seimbang dan Data yang diseimbangkan dengan metode Random Undersampling pada dataset LCMS Tanaman Keladi Tikus. *8th Seminar Nasional Teknik Elektro, Informatika & Sistem Informasi (SINTaKS)*.

Tian, X., Pavur, R., Han, H., & Zhang, L. (2023). A Machine Learning-Based Human Resources Recruitment System for Business Process Management: Using LSA, BERT and SVM. *Business Process Management Journal*, *29*(1), 202–222. https://doi.org/10.1108/BPMJ-08-2022-0389

Titisari, M., & Ikhwan, K. (2021). Proses Rekrutmen dan Seleksi: Potensi Ketidakefektifan dan Faktornya. *JMK (Jurnal Manajemen Dan Kewirausahaan)*, *6*(3), 11. https://doi.org/10.32503/jmk.v6i3.1848

Vivek, R. (2023). Enhancing Diversity and Reducing Bias in Recruitment through AI: a Review of Strategies and Challenges. *Информатика. Экономика. Управление - Informatics. Economics. Management*, *2*(4), 0101–0118. https://doi.org/10.47813/2782-5280-2023-2-4-0101-0118

Yildizhan, H., Hosouli, S., Yılmaz, S. E., Gomes, J., Pandey, C., & Alkharusi, T. (2023). Alternative Work Arrangements: Individual, Organizational and Environmental Outcomes. *Heliyon*, *9*(11), 1–16. https://doi.org/10.1016/j.heliyon.2023.e21899

Yuliani, S., & Aliyyah, R. R. (2024). Seleksi Tenaga Pendidik: Penerapan Rekrutmen pada Sekolah Dasar. *Karimah Tauhid*, *3*(3), 2685–2702. https://doi.org/10.30997/karimahtauhid.v3i3.12203

Zuhri, B., & Harani, N. H. (2023). Aplikasi Rekrutmen Karyawan Menggunakan Artificial Neural Network dan Flask. *Jurnal Sisfotenika*, *13*(2), 125–138. http://dx.doi.org/10.30700/jst.v13i2.1367

ISSN: 2716-3865 (Print), 2721-1290 (Online)