ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 6, pp. 184-197
DOI:10.33168/JLISS.2024.0611

Comparative Evaluation of Data Mining Classification Techniques for Assistant Recruitment in a Computer Laboratory

Devia Liany, Tanty Oktavia

Information System Management Department, BINUS Graduate Program – Master of Information System, Bina Nusantara University, Jakarta, Indonesia 11480

devia.liany@binus.ac.id; toktavia@binus.edu

Abstract. This study analyzed different data mining classification techniques for predicting assistant acceptance decisions in a computer lab's recruitment process. Historical data was utilized on 128 accepted and 150 rejected candidates over 4 years. After pre-processing, classifiers like decision trees, Naïve Bayes, and K-Nearest Neighbor were implemented using Python for comparative evaluation. Accuracy metrics revealed KNN as optimal with 92.86% correct prediction over alternatives. However, precision and false positive rates require additional scrutiny from a recruitment perspective. The findings imply that customized algorithm tuning and balanced error analysis allow better leveraging of predictive recruiting to minimize selection inaccuracies despite growing applicant volumes. Though specific to an organizational context, results initiate valuable trajectories for effective data-driven human resources practice.

Keywords: Recruitment, Data Mining, Classification, Decision Tree, Naïve Bayes, K-Nearest Neighbor

1. Introduction

Now, Digital Era 4.0 refers to increasing automation and data exchange in industrial technology and operations. With the support of advances in information technology, business operations become more effective and efficient. In this era, business competition is increasingly fierce. Companies or organizations are competing to use information technology to improve their business. One inseparable part of the business is human resources. They play a key role in achieving business success. Therefore, the human resource recruitment process is a critical point in human resource management, using information technology in analyzing data such as a data mining approach as an effective tool to increase efficiency and accuracy in selecting candidates who best suit the needs of the company or organization.

Recruiting candidates with high potential performance is challenging for organizations or companies. High-potential candidates are organizational or company assets important for increasing business growth, while candidates with bad potential can be destructive to the business. Consequently, the correct recruitment method can support and predict future business success (Amos Pah & Utama, 2020). Recruitment is conducted by assessing candidates, which is an indicator that differentiates between the quality of candidates and how the candidates will perform in the future. This assessment is conducted to select candidates to obtain competent human resources—the more recruitment that is conducted, the more recruitment data must be stored and processed. So, the human resources department needs a system to support decision-making recommendations in recruitment. The automation of the recruitment process can potentially reduce the time and energy spent by the recruitment team (Lewaherilla & Huwae, 2023).

In the selection process, there are usually several stages, and the human resources department has many things to consider in selecting qualified candidates (Karim et al., 2021). In the past, the human resources department extracted recruitment data with a simple pattern manually, but as time passed, the data volume increased, and the pattern became more complex. Managing candidate data manually takes a lot of time to complete. With current technological advances in data management, data mining can be designed as a solution to overcome this problem. Data mining is commonly known as the process of finding useful information from a lot of data (Kumar et al., 2019). With data mining, it is also hoped that it can help in processing enormous amounts of data and finding a pattern or something unique from the processed data (Su & Wu, 2021). Currently, the amount of data stored in recruitment databases is increasing so data mining techniques are needed to find hidden relationships between variables (Algur et al., 2016). Implementation of data mining in the human resources department to support better decisions over time because of the enormous number of datasets available in the database that need to be converted into meaningful information (Shehu & Saeed, 2016). Utilizing data mining helps provide the right information to make predictions based on past trends and current conditions. This research will use classification data mining techniques to analyze historical data from a database and automatically produce a model that can predict future events (Sharda et al., 2021).

The computer laboratory organization is the educational industry. This organization manages and monitors computer practicum activities for college students. The human resources needed in this organization are teachers or what can be called assistants. The selection of the right assistant can predict the success of this business in the future. Every six months the computer laboratory recruits assistants. A series of recruitment processes were carried out over six months. In this case, every year this organization carries out two periods of assistant recruitment activities. Students who want to become computer laboratory assistants must go through several recruitment stages, namely: registration, initial test, pre-training, interview, and core training. This research focuses on pre-training until core training. In the last four periods (two years ago), it was found that there was an increase every year in the number of candidates at the pre-training stage which can be seen in Figure 1. In an even semester, from 112 candidates to 121 candidates. And then, from 295 candidates to 333 candidates. The data collected and analyzed by resource management is also increasing and it takes longer to manage this data. Therefore, human resource management needs better and faster data processing methods.

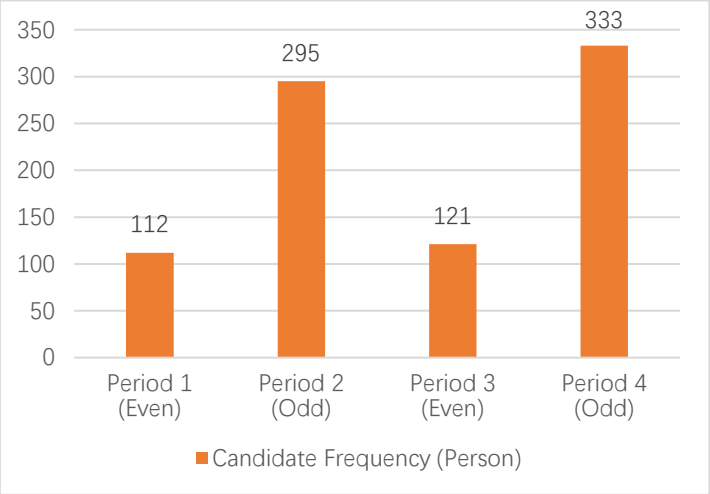

Fig. 1: Graph of the Number of Candidates at the Pre-training Stage.

Every period, computer laboratory resource management carries out data processing using Excel to combine recruitment data such as candidate data, test scores, teaching presentation scores, attendance, voting between candidates, and interview scores. As the number of candidates increases, it takes longer to prepare the data. Sometimes human resource management finds it difficult to determine the right candidate to recruit. Additionally, there is a risk of bias in selecting assistant candidates based on subjective factors such as personal impressions or recruiter preferences. This research case was conducted by comparing three algorithm models for assistant recruitment in a particular computer laboratory organization using data mining classification: Decision Tree, Naive Bayes, and K-Nearest Neighbor with Python. Data mining classification algorithms, such as Decision Tree, Naive Bayes, and K-Nearest Neighbor, are very relevant choices in dealing with this research problem. Decision tree algorithms can break down candidate decisions into a series of steps that are easy to interpret, while Naive Bayes algorithms use probability models to classify candidates based on training data. Then, K-Nearest Neighbor utilizes the similarities between candidates to make predictions. Using these algorithms can increase efficiency in the recruitment process, reduce subjective bias, and optimize the discovery of candidates who best suit the organization's needs. The considerations in choosing these three algorithms are due to the type of data used where the data source is labeled and these three algorithms are included in supervised learning in machine learning Therefore, this research compares the level of accuracy in predicting assistant acceptance decisions based on several attributes that have been determined by a computer laboratory.

The rest of the paper is organized as follows. Section 2 represents some related works that existed before the proposed work. Then, section 3 provides the proposed methodology. Section 4 shows results using the classification models built with three classification algorithms. This section also provides performance accuracy metrics and result analysis. The final section represents the conclusion of this research.

2. Literature Review

This research aims to analyze three algorithms of data mining classification techniques for predicting assistant acceptance decisions in a computer lab's recruitment. The important theories and instruments from previous research that have been applied to this research will be summarized in this section.

2.1. Recruitment

Recruitment is a phase or action carried out by an organization or company to obtain human resources through several stages including identification and evaluation of sources, determining their need, selection process, placement, and orientation. Recruitment is one of the cores of human resource

management (Abdul et al., 2020). Recruitment aims to provide sufficient human resources so management can select employees who meet the qualifications they need. The assistant recruitment process is an important stage in workforce management in this computer laboratory organization. Currently, the recruitment process takes place in several quite lengthy stages. The number of assistant candidates from year to year also continues to increase. The data collected and processed from several stages is increasingly large and complex. Therefore, this recruitment process takes a long time and can reduce the quality of the assistants recruited. Thus, applying technology in the assistant recruitment process is becoming increasingly important to optimize the efficiency and quality of candidate selection. Technology can be used in various stages of the recruitment process, from collecting and screening applications to evaluation and final decision-making. This research will implement data mining technology. This supports references in building systems to process candidate data and find quality candidates more efficiently, especially in data mining.

2.2. Knowledge Discovery in Database

Knowledge discovery in databases is a machine-learning process that induces rules or procedures regarding establishing knowledge in large databases (Sharda et al., 2021). Knowledge discovery and data mining have become increasingly important fields due to the recent increased demand for knowledge discovery techniques in databases, including those used in knowledge acquisition, machine learning, databases, statistics, data visualization, and high-performance computing. The use of knowledge discovery and data mining can be used in the field of artificial intelligence in many fields, such as education, industry, commerce, government, and so on (Kumar et al., 2019). Figure 2 shows the knowledge discovery process starting from selection, pre-processing, transformation, data mining, and interpretation/evaluation. Finally, the result of this process is knowledge or knowledge information that is found.

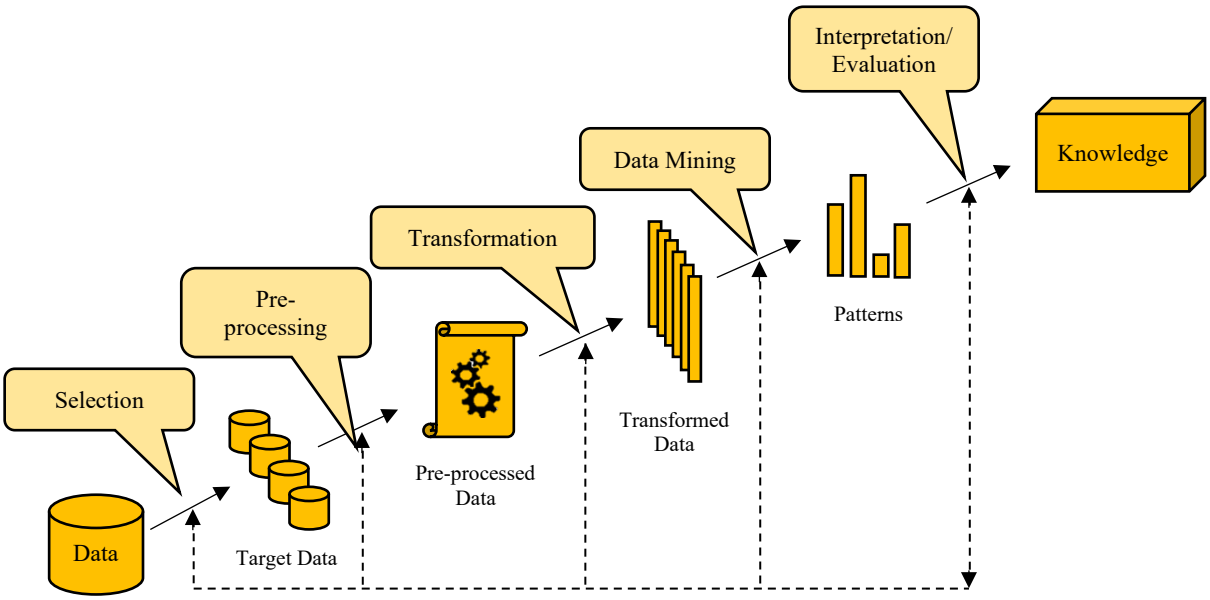

Fig. 2: Knowledge Discovery Process.

In this research, the knowledge discovery concept will be implemented. The process of discovering knowledge from databases has become increasingly important in recent years (Bal & Bal, 2022). The process of discovering useful knowledge from raw data sets, through the application of exploratory data mining algorithms. The knowledge discovery concept provides a structured framework and systematic approach to exploring, evaluating, and deriving knowledge from data.

2.3. Data Mining

Data mining is the process of finding useful information from a lot of data (Kumar et al., 2019). With data mining, it is also hoped that it can help in processing large amounts of data and finding a pattern or something unique from the processed data (Su & Wu, 2021). Data mining involves processes, methodologies, tools, and techniques that can be used to discover and extract patterns, knowledge, insights, and valuable information from data sets. Implementation of data mining in the human resources department to support better decisions over time due to the large number of datasets available in the database that need to be converted into meaningful information (Shehu & Saeed, 2016). There are several data mining techniques, namely classification, regression, time series, market basket, link analysis, clustering, and outlier analysis (Sharda et al., 2021). In this research, a data mining classification technique will be used. Classification is the phase of finding a model that explains and differentiates data classes and concepts. The process of converting raw datasets into new features and then selecting the best features to improve the performance of the classification technique (Abedin et al., 2023). Classification techniques make it possible to classify data into different categories or classes based on certain attributes or features. In the computer laboratory for predictions, assistant candidates are classified into appropriate categories, such as "Accepted" or "Not Accepted". Classification results can provide recommendations or decisions based on historical data analysis, thereby helping these organizations to make faster and more accurate decisions. The algorithms to be used are Decision Tree, Naïve Bayes, and K-Nearest Neighbor. The algorithm selection is also based on previous research that has been carried out. The first previous research was research conducted by Amos Pah and Utama in 2020. This journal discusses the use of data mining classification methods for employee recruitment using historical data with a comparison of four data mining classification algorithms, namely Decision Tree, Naive Bayes, Support Vector Machine, and Random Forest. The second previous research was research conducted in 2016 by Algur, Bhat, and Kulkarni. This journal discusses a model that was built to predict whether a student will be recruited or not using Decision Tree and J48. They also provide suggestions for further research to use other data mining classification algorithms such as k-NN, and Naïve Bayes. In this research, we want to test other classification algorithms to see whether their accuracy level is better or worse than the Decision Tree algorithm. So, these three algorithms were chosen because these algorithms are most widely used in predictions, especially in the field of recruitment.

2.3.1. Decision Tree (DT)

Decision Tree is one of the most popular and understandable algorithms in data mining classification (Jayadi et al., 2019). Decision Tree is an algorithm for classification methods in data mining. Decision Tree has the concept of changing data and then forming rules for decisions based on the processed data. In the Decision Tree algorithm, predictions are made using a tree structure. Decision Trees can make decisions from complex processing processes simpler. So that later the results of predictions using Decision Trees can better interpret solutions to existing problems (Song & Lu, 2015).

In the Scikit-learn library with Python, the Decision Tree algorithm uses the DecisionTreeClassifier class which has several parameters including:

- criterion: used to determine the quality of separation at each node in the model formation process. There are three types, namely “gini”, “entropy” and “log_loss”. In general, “gini” and “entropy” are used. Gini measures the level of uniformity of a sample set, while entropy measures the level of uncertainty in a sample set.
- max_depth: used to determine the maximum depth of the tree.
- min_sample_leaf: used to determine the minimum number of samples required as a leaf node.
- min_sample_split: used to determine the minimum number of samples required to divide a node.

2.3.2. Naïve Bayes (NB)

Naïve Bayes is an algorithm in the data mining classification method. This algorithm is based on Bayes'

theorem put forward by British scientist Thomas Bayes. This algorithm is one of the purest of the Bayesian type (Jayadi et al., 2019). The concept applied in the Naïve Bayes algorithm is that all the variables used are independent, and also these variables are independent variables that are not related to each other. Naïve Bayes predicts future opportunities based on what has happened in the past. The main characteristic of this algorithm is that there is a very strong assumption of the independence of each event.

In the Scikit-learn library with Python, the Naïve Bayes algorithm has three types of classification including Gaussian Naive Bayes, Multinomial Naive Bayes, and Bernoulli Naive Bayes, for this research uses GaussianNB because the dataset used is continuous. Therefore, the class used is GaussianNB and there are no parameter settings.

2.3.3. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor commonly abbreviated as k-NN is a classification technique that is commonly used to classify input data into pre-defined classes (k) (Zhang et al., 2017). It was proposed by Cover and Hart in 1968. The direct mechanism of the k-NN algorithm is to calculate the Euclidean distance function between a predefined class and each varying sample. After that, the k-NN algorithm selects the minimum nearest neighbors according to each category. Samples are assigned to their categories based on k nearest neighbors (Ali et al., 2020).

In the Scikit-learn library with Python, the K-Nearest Neighbor algorithm uses the KNeighborsClassifier class which has several parameters including:

- n_neighbors: used to determine the number of nearest neighbors that will be used in the prediction process. The general value chosen is odd to avoid difficulties in determining the majority.
- Weights: used to determine the weight given to each neighbor in the prediction process. There are two weights, namely: "uniform" meaning that each neighbor has the same weight, and "distance" meaning that the neighbor's weight is closer and gives greater results.

3. Research Methodology

This section explains the flow of conducting this research with the knowledge discovery concept. Figure 3 shows the flow for predicting assistant recruitment using data mining classification. The flow consists of data collection, data selection, data pre-processing, data transformation, data mining modeling, and evaluation to choose the highest accuracy of the data mining algorithm.

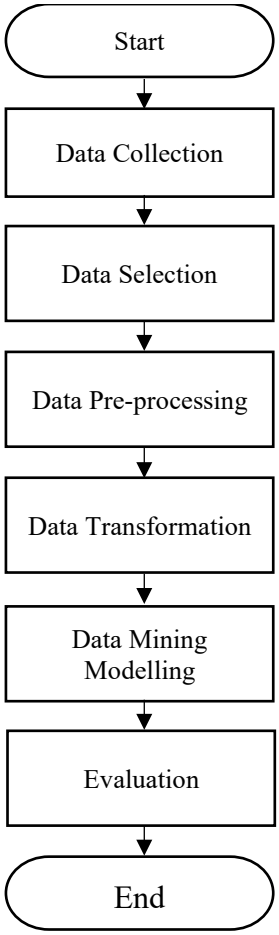

Fig. 3: Flow of This Research.

3.1. Data Collection

The assistant candidate data used to carry out data mining classification techniques is obtained directly from the computer laboratory's management resources. The data obtained is in Excel form. This data was selected for the last two years, from 2020 to 2022 (the last four periods). This historical data is very supportive of using data mining techniques to analyze data patterns because this data is real data from the recruitment process in this organization. Hopefully, the resulting prediction results will achieve the maximum level of accuracy.

3.2. Data Selection

Research data consists of selected attributes and values that have been determined based on laboratory computer resource management criteria. Table 1 shows a selection of attributes that will be used in this research.

Table 1: Attributes of This Research

Attributes	Description	Data Type - Value
Status	Marks whether or not who is accepted as an assistant	Non-numerical Value: Accepted, Not Accepted
Period	Semester level of the candidate	Numerical Value: 1-6
Major	Faculty of the candidate	Non-numerical Value:

		• Business Information Technology • Computer Engineering • Computer Science • Computer Science & Mathematics • Computer Science & Statistics • Cyber Security • Data Science • Game Application and Technology • Informatics • Information Systems • Information Systems & Management • Mobile Application and Technology • Software Engineering
Average Material Score	Average scores obtained when working on the questions given during core training	Numerical Value: 1-100
Average Presentation Score	Average values obtained when presenting teaching simulations during core training	Numerical Value: 1-100
Permission Attendance Frequency	Number of permitted attendances during core training	Numerical Value: 1-30
Absent Attendance Frequency	Number of absentees during core training	Numerical Value: 1-30
Late Attendance Frequency	Number of late attendances during core training	Numerical Value: 1-30
Neglect Attendance Frequency	Number of attendees who forgot to record their attendance during core training	Numerical Value: 1-30
Vote Top Frequency	number of other candidates felt that the selected candidate had a good performance	Numerical Value: 1-100
Vote Bottom Frequency	number of other candidates felt that the selected candidate had a poor performance	Numerical Value: 1-100
Interview Score	The final value of the interview assessment results given by the interviewer during the interview session	Numerical Value: 1-100

3.3. Data Pre-processing

Several parts of the research data that have been selected will be deleted, such as the title in Excel which

will interfere with the data reading process, as well as tidying up the header and the data to be processed. After completing this stage, we will continue by paying attention to the details of the data, whether there is repeated or incomplete data. If there is, the rows of data will be immediately removed or deleted. Lastly, make changes to the Excel file format from (.xls) format to (.csv) format so that reading the file when modeling will be easier and the analysis process will be optimal.

3.4. Data Transformation

In data transformation, change the attribute value data type, and define target attributes and predictor attributes. Based on Table 1, there are 2 attributes whose data type is non-numerical, namely "Status" and "Major". This can cause data anomalies. On, eliminating data anomalies, we change the type of data values by grouping them using numbers. After that, we determine there are 2 types of attributes, namely target attributes and predictor attributes. For this research, the target attribute is status accepted with an “Accepted” label and a “Not Accepted” label. Meanwhile, the predictor attributes are status accepted, period, major, test scores, teaching presentation scores, attendance, number of votes, and interview scores. In this step, the data will be divided into 2 parts, commonly called training data and testing data. The commonly used ratio is 80:20, which means 80% of the data is for training data and 20% for testing data (Joseph, 2022).

3.5. Data Mining Modelling

The data that has been divided will then be processed using a data mining classification algorithm using the sci-kit-learn library with Python programming. The use of Decision Tree, Naïve Bayes, and K-Nearest Neighbor algorithms in data mining classification in predicting label results. The considerations in choosing these three algorithms were seen from the type of data used in this study where the data sources were labeled. The decision tree algorithm uses several parameter configurations to get the best accuracy according to the dataset. For "criterion" use the "entropy" value because it can capture more complex relationships and adapt to finer structures in the data. Then the “max_depth” value is 5, the “min_sample_leaf” value is 3. And the “min_sample_split” value is 3. The Naive Bayes algorithm used with this research dataset is "GaussianNB" so there is no need to configure parameters to get the best accuracy according to the dataset. This type of algorithm was chosen because the data is continuous. The K-Nearest Neighbor algorithm also uses several parameter configurations to get the best accuracy according to the dataset. The n value can be filled with a minimum of 1 and a maximum according to the total amount of data. For this case, the value n = 1.

3.6. Evaluation

Last in evaluation, compared the results of the best accuracy among the three data mining classification algorithms. The value of the comparison results can be obtained through the accuracy score. The requirement for selecting the algorithm to be chosen is to produce the highest accuracy value. This research uses an accuracy score to measure the performance of the model. The accuracy score is the total of how often the model correctly classified the data by adding up the TP (True Positive) and TN (True Negative) and dividing by the total all data (Thakar et al., 2015). The formula for accuracy is shown in Equation (1).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

TP (True Positive) is a condition when the model correctly predicts the positive class, whereas TN (True Negative) is a condition when the model correctly predicts the negative class. Moreover, FP (False Positive) is a condition when the model incorrectly predicts the positive class, whereas FN (False Negative) is a condition when the model incorrectly predicts but as a negative class (Ren & Liu, 2019).

4. Result and Discussion

The candidate data has been collected by the computer laboratory organization. This data consists of 278 rows of data. In this study, 12 attributes were used in the candidate data, namely “Status”, “Major”, “Period”, “Average Material Score”, “Average Presentation Score”, “Permission Attendance Frequency”, “Absent Attendance Frequency”, “Late Attendance Frequency”, “Neglect Attendance Frequency”, “Vote Top Frequency”, “Vote Bottom Frequency”, and “Interview Score”. These attributes are used under the instructions from the resource management computer laboratory organization. There were 150 candidates with the status “Not Accepted” and 128 candidates with “Accepted”. The status attribute value will be used as the target attribute. The target attribute is also called the label class. The total amount of data for each class label is declared balanced. This dataset is sufficient because it was obtained over the last two years, so the results are more relevant. Even though it is a small dataset, it is ideal for this research because it allows us to more quickly test and understand various data mining techniques.

This research used Python programming with the library sci-kit-learn to process data mining classification algorithms. In pre-processing, several types of attributes used need to be in numerical form to optimize the results of data mining algorithm calculations. Therefore, attributes that do not have a numeric value will be grouped and the value will be replaced with a numeric type. The attribute "status" and "Major" attributes will change their values to numeric. After that, the data is divided into training and testing. 80% training data and 20% testing data. By using a ratio of 80:20, the size of the training data is larger so that the data results obtained are closer to the results that match the existing data. After that, perform data modeling with the three algorithms with the sci-kit-learn library in Python programming. With the use of this library, the processing of classification data mining can be completed more effectively and efficiently.

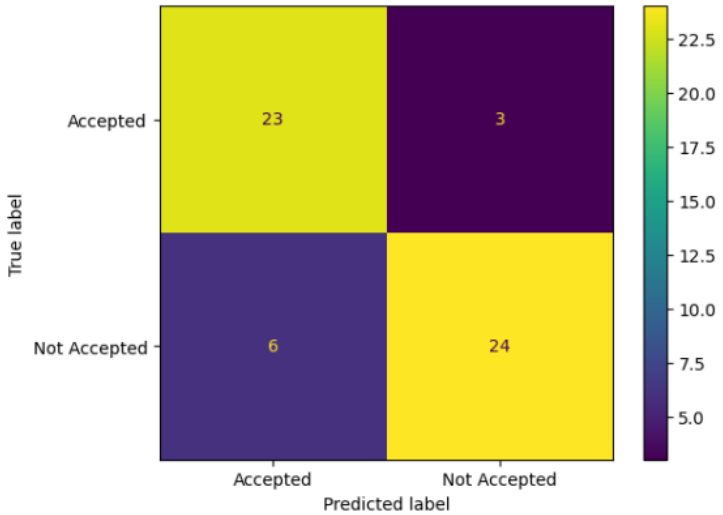

Fig. 4: Confusion Matrix of Decision Tree Algorithm Model.

Figure 4 shows clearly that the model incorrectly predicted the accepted class correctly for 23 data (TN). Then the condition when the model predicts the not accepted class correctly is 24 data (TP). Then the condition when the model incorrectly predicts the accepted class is 6 data (FN) and the condition when the model incorrectly predicts but the class is not accepted is 3 data (FP). This also shows that the model correctly classifies 47 data, and the model classifies 9 data incorrectly.

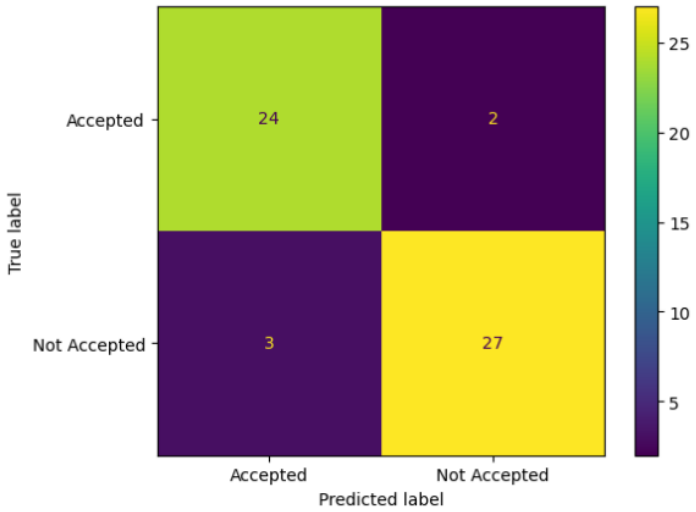

Fig. 5: Confusion Matrix of Naïve Bayes Algorithm Model.

Figure 5 shows clearly that the model correctly predicts the accepted class for 24 data (TN). Then the condition when the model correctly predicts the not accepted class is 27 data (TP). Then the condition when the model incorrectly predicts the accepted class as much as 2 data (FN) and the condition when the model incorrectly predicts it as a not accepted class as much as 2 data (FP). This also shows that the model correctly classifies 51 data and the model classifies 5 data incorrectly.

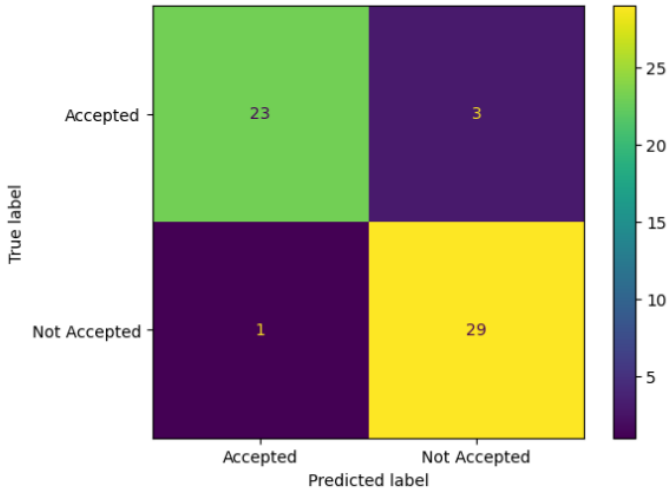

Fig. 6: Confusion Matrix of K-Nearest Neighbor Algorithm Model.

Figure 6 shows clearly that the model correctly predicts the accepted class for 23 data (TN). Then the condition when the model correctly predicts the not accepted class is 29 data (TP). Then the condition when the model incorrectly predicts the accepted class by 1 data (FN) and the condition when the model incorrectly predicts but the class is not accepted by 3 data (FP). This also shows that the model correctly classifies 52 data and the model classifies 4 data incorrectly.

The results of the accuracy values of the three algorithms are compared to conclude which algorithm has the highest accuracy.

Table 2: Result Accuracy Score Comparison

Algorithm	Accuracy Value (%)
Decision Tree	83.93
Naïve Bayes	91.07
K-Nearest Neighbor	*92.86*

Table 2 compares the accuracy values of the data processing models of the three data mining classification algorithms. From the table, it can be concluded that the highest accuracy value of the three algorithms in this study is the K-Nearest Neighbor algorithm with a value of 92.86%. The lowest accuracy value is the Decision Tree algorithm with a value of 83.93%. The second highest order is the Naïve Bayes algorithm with a value of 91.07%. Calculation of good accuracy values for classification algorithms aims to find the best performance of the algorithm that can be used (Ramaswami et al., 2019). This high accuracy shows that the model used in the data mining process has good abilities in predicting or finding relevant patterns in the data. Therefore, the results of this research were obtained from a comparison of algorithms that K-Nearest Neighbor can predict the dataset on the feasibility of accepting assistants better than the other two algorithms.

5. Implications

This research has explained the examination of the data mining classification algorithms to predict whether a candidate assistant will be accepted or not based on historical data from the computer laboratory. This research is based on attributes that have been determined by this organization's resource management in determining the feasibility of accepting computer laboratory assistants. Decision Tree is a learning algorithm that separates data into smaller groups based on a series of hierarchical decisions. It is suitable for situations where the decision-making rules are easily understood by humans. In recruiting assistants, Decision Trees can help identify important criteria in the recruitment process. The dataset produces three attributes that influence the results of this algorithm model, the most influential of which are Average Material Score, Vote Top Frequency, and Interview Score. Naive Bayes is a probability algorithm based on Bayes' Theorem. It assumes that the features in the data are independent of each other, although this may not always be the case in practice. In this research, Naive Bayes can help classify candidates based on specified features. The dataset produces three attributes that influence the results of this algorithm model, the most influential of which are the Interview Score, Average Presentation Score, and Average Material Score. K-Nearest Neighbor is an algorithm that uses an approach that looks for training data that is most similar to new test data points and predicts labels based on the majority of nearest neighbors in the classification. K-Nearest Neighbor is suitable for cases where the data does not have a complex structure. Concerning this assistant recruitment research, K-Nearest Neighbor can be used to find candidates who are most similar to previously successful assistants based on criteria determined by the laboratory computer. The dataset produces three attributes that influence the results of this algorithm model, the most influential of which are Neglect Attendance Frequency, Vote Bottom Frequency, and Interview Score. In this research, it was found that the results of the comparison accuracy values of the three algorithms showed that the K-Nearest Neighbor algorithm was the highest. These attributes and classifications are used as a tool to help make decisions to determine feasibility more accurately and save time for computer laboratory resource management.

6. Conclusion, Limitation, and Future Works

This examination of the data mining classification algorithms for assistant recruitment assists in showcasing techniques that enhance predictive precision. Tailoring tools to account for dataset specifics and particular organizational needs appear vital. By supplementing decision-maker judgments with evidence-based insights, the screening processes can become more robust to intensifying volumes. However, techniques should combine automation with human oversight regarding metrics

interpretation and verification of predictions where needed. As digitalization accelerates, responsible development mandates mitigating potential biases through transparency safeguards too. Beyond efficiency gains, optimizing recruit function productivity warrants cultivating supportive learning cultures centered on ethics.

From this research, there are several limitations and suggestions for future work. First, future research can focus on different organizations. This is because every organization has different criteria for recruitment. Some automatic techniques are also required at the pre-processing phase to identify the best attributes for the best classifier performance. Selecting the right features regarding the features that are important in predicting assistant acceptance can produce an optimal model.

References

Abedin, M. Z., Hajek, P., Sharif, T., Satu, M. S., & Khan, M. I. (2023). Modelling bank customer behaviour using feature engineering and classification techniques. *Research in International Business and Finance*, *65*. https://doi.org/10.1016/j.ribaf.2023.101913

Algur, S., Bhat, P., & Kulkarni, N. (2016). Educational Data Mining: Classification Techniques for Recruitment Analysis. *International Journal of Modern Education and Computer Science*, *2*, 59–65. https://doi.org/10.5815/ijmecs.2016.02.08

Ali, A., Alrubei, M., Hassan, L. F. M., Al-Ja'afari, M., & Abdulwahed, S. (2020). Diabetes classification based on KNN. *IIUM Engineering Journal*, *21*(1). https://doi.org/10.31436/iiumej.v21i1.1206

Amos Pah, C. E., & Utama, D. N. (2020). Decision Support Model for Employee Recruitment Using Data Mining Classification. *International Journal of Emerging Trends in Engineering Research*, *8*(5). https://doi.org/10.30534/ijeter/2020/06852020

Bal, M., & Bal, Y. (2022). A Novel Approach for the Recruitment Process in Human Resources Management: Decision Support System Based on Formal Concept Analysis. *International Journal of Decision Support System Technology*, *14*(1). https://doi.org/10.4018/IJDSST.292447

Jayadi, R., Firmantyo, H. M., Dzaka, M. T. J., Suaidy, M. F., & Putra, A. M. (2019). Employee performance prediction using naïve bayes. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6). https://doi.org/10.30534/ijatcse/2019/59862019

Karim, M. M., Bhuiyan, A., Kumer, S., Nath, D., & Latif, W. Bin. (2021). Conceptual Framework of Recruitment and Selection Process. *International Journal of Business and Social Research*, *September*.

Kumar, N., Jain, S., & Chauhan, K. (2019). Knowledge Discovery from Data Mining Techniques. In *IJERT Journal International Journal of Engineering Research and Technology*.

Lewaherilla, N. C., & Huwae, V. E. (2023). The Use of Information Technology in Improving the Recruitment and Selection Process of Human Resources. *Jurnal Minfo Polgan*, *12*(1). https://doi.org/10.33395/jmp.v12i1.12537

Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. *Information and Learning Science*, *120*(7–8). https://doi.org/10.1108/ILS-03-2019-0017

Ren, J. H., & Liu, F. (2019). Predicting software defects using self-organizing data mining. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2927489

Sharda, R., Delen, D., & Turban, E. (2021). *Business Intelligence, Analytics, and Data Science: A Managerial Perspective* (4th edition). Pearson.

Shehu, M. A., & Saeed, F. (2016). An adaptive personnel selection model for recruitment using domain-driven data mining. *Journal of Theoretical and Applied Information Technology*, *91*(1).

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2). https://doi.org/10.11919/j.issn.1002-0829.215044

Su, Y. S., & Wu, S. Y. (2021). Applying data mining techniques to explore user behaviors and watching video patterns in converged IT environments. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02712-6

Thakar, P., Dr., Prof., & Manisha, Dr. (2015). Role of Secondary Attributes to Boost the Prediction Accuracy of Students' Employability Via Data Mining. *International Journal of Advanced Computer Science and Applications*, *6*(11). https://doi.org/10.14569/ijacsa.2015.061112

Zhang, K., Li, G., & Zhang, M. (2017). Assembly data mining platform based on python. *Proceedings of Science*, *2017-December*. https://doi.org/10.22323/1.300.0056