Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

Prediction of Student Work Readiness using Artificial Neural Network and Decision Tree

Bilqiis Shahieza Andini, Kemas Muslim Lhaksmana*

School of Computing and Informatics, Telkom University, Bandung, Indonesia
Email: [1]bilqiisshahieza@students.telkomuniversity.ac.id, [2,*]kemasmuslim@telkomuniversity.ac.id
Correspondence Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **08/05/2025**; Accepted: **31/05/2025**; Published**: 01/06/2025**

Abstract−The readiness of students for the workforce is a critical metric in evaluating the quality of higher education. Forecasting students' work readiness before their entrance into the industry calls for a data-driven approach since they often lack the necessary skills and experience until graduation. Using two machine learning techniques—Artificial Neural Network (ANN) and Decision Tree (DT)—this research aims to create a classification model to predict students' employment readiness. Among the several aspects the data covers are academic knowledge, professional attitude, soft skills, hard skills, and socio-economic background. Data preparation, data cleansing, feature selection, model training, and performance evaluation make up the study approach. The ANN model comprised four hidden layers, while the DT was refined with RandomizedSearchCV. The test results showed that DT had an accuracy of 90.80%, and ANN had 90.69%, indicating that both performed very well and can be selected based on what the user needs most. This research contributes a predictive model for educational institutions to assess students' employment preparedness in a more objective and systematic way.

Keywords: Classification; Work Readiness; Artificial Neural Network; Decision Tree; Machine Learning

1. INTRODUCTION

Advances in science and information technology in this era of globalization have led to fierce competition in the provision of human resources [1]. The labor market has become increasingly competitive, especially in fields that require specialized knowledge. Basically, people analytics is the application of business analysis and data mining to human resource data with the aim of improving the effectiveness of human resource management through the use of data [2]. Evaluating the readiness of human resources to enter the workforce, especially students, is a challenging task. Various factors need to be considered, such as academic knowledge, practical skills, professional attitudes, soft skills, hard skills, and others. Therefore, it is crucial to develop a classification model that can help predict how prepared students are for entering the workforce. Such knowledge is beneficial for graduates who are about to join the professional workforce and also for businesses actively seeking new talent. Previous research [3] indicates that project-based and case study learning methodologies effectively enhance students' work preparation, particularly in acquiring 21st-century competencies such as cooperation, communication, and problem-solving. This technique requires a greater examination in particular contexts, such as online learning and specialised study programs, to assess its impact on learning outcomes and overall career preparedness. By assessing the effectiveness of Artificial Neural Network (ANN) and Decision Tree (DT) approaches in raising on-time graduation rates and predicting graduates' work readiness, this study seeks to clarify the factors of student preparedness for employment and to address the challenges college graduates face in the job market. Many earlier studies on people analytics point to the need to create a predictive classification model for variables for machine learning-driven human resource management. This study uses Artificial Neural Network to examine student performance during their active study period, aiming to construct a model that predicts work readiness. This study uses the Decision Tree algorithm as a comparative measure to enhance the model's accuracy values. The length of study of students was predicted with very high accuracy at 99% in previous research [4], which used ANN backpropagation. In addition, there is research [5] that uses ANN for heart disease classification, using 304 patient data from Kaggle, achieving an accuracy of 73.77%. A study [6] predicted student work readiness using the DT C4.5 algorithm, achieving a good accuracy of 100%. However, study [7] showed that the Naive Bayes algorithm had the highest accuracy (53.7%) compared to DT (50.9%) and K-Nearest Neighbor (51.1%) in determining student work readiness.

Based on the background described above, the research problem in this study is to develop a classification model that can predict how ready students are to enter the workforce based on factors such as academic knowledge, professional attitudes, soft skills, hard skills, and socio-economic context, as well as to evaluate the results between the DT and ANN models in predicting students' work readiness. This study is limited to the 2022 Tracer Study dataset, which contains 8,542 rows and 48 columns. The dataset was obtained from an alumni survey at Telkom University. Overall, the data contains academic and non-academic assessments from alumni who participated in the survey. The objective of this study is to develop an ANN and DT classification model that can predict the work readiness of students. The factors used in developing this model include aspects of academic knowledge, professional attitudes, soft skills, hard skills, and the socio-economic context of the alumni. Additionally, this study aims to evaluate the performance differences between the two models. As a result, this study can provide further insights into the best and most practical approach for predicting students' job readiness.

Various related studies have employed the concept of people analytics to analyze performance data using appropriate classification techniques. Since machine learning is faster and cheaper than human experience, this

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

technique has been used for prediction and classification tasks in various industries. Research related to classification prediction in recent years has seen developments in both methods and related topics, such as employee performance prediction, population growth prediction, and others. Several studies related to classification prediction apply several machine learning methods of the supervised learning type, such as linear regression, k-NN, Naive Bayes, Decision Tree, Support Vector Machine, and neural networks. As in study [8], which used multiple linear regression to investigate that soft skills and hard skills contributed 37.3% to the work readiness of students in the Information Systems Study Program at ABC University Class of 2017. The results showed that both types of skills contributed 37.3% to work readiness. For addition using another algorithm for predicting student work readiness, research [9] merged supervised and unsupervised learning methodologies, used K-Means for initial labeling, and determined that Random Forest achieved the highest accuracy of 98%, indicating the significant potential of machine learning in enhancing employability predictions. However, in study [7], it was concluded that Naive Bayes had the highest accuracy (53.7%) compared to DT (50.9%) and K-Nearest Neighbor (51.1%) in determining student work readiness. The study used data taken from the Business Placement Center database of Amikom University Yogyakarta with a sample size of 1,664 data points. Research [10] employs algorithms including Support Vector Machine, Decision Tree, Logistic Regression, and Artificial Neural Network to forecast graduate work readiness, using features similar to those in this study. This study used attributes such as training, soft skills, hard skills, and in-demand abilities derived from graduate and employer questionnaires. The decision tree method attained the best accuracy of 100% in comparison to other algorithms. Moreover, there is research [11] that employs algorithms such as Random Forest, Naive Bayes, CART, and Neural Networks to forecast student adaptability in online education, although they have not optimised model fusion. The research introduced an enhanced ensemble model that incorporates factors such as financial status and class time, achieving an accuracy of 93.96%. In a study [12] addressing the same subject, specifically predicting the work readiness of Telkom University students, the optimal Random Forest model achieved an accuracy of 48.5%, while the most effective Multinomial Logistic Regression model attained an accuracy of 53.9% after implementing data balancing with SMOTE-ENN.

People analytics is a specific use of big data analytics that transforms human resource management tactics [13]. Advanced analyses, including perspectives and predictions, are essential for classifying students based on their readiness to enter the workforce. Such information provides crucial support to businesses and students in making judgments regarding human resource recruiting. The implementation of people analytics facilitates enhanced transparency concerning an individual's performance, competencies, weaknesses, risks, and potential. These insights can be advantageous throughout an employee's career, from recruitment to retirement [14]. Understanding the implications of an individual's organizational persistence and outcomes can also serve to characterize team dynamics and communication networks [15].

Work readiness is defined as the condition in which new graduates are prepared to thrive in their professional careers [12]. Although this idea is still relatively new, it has been discussed in scientific publications as a factor for selecting candidates based on their graduation prospects [16]. Work readiness evaluation is the process of assessing an individual's preparedness to meet the requirements of a specific project or job function. A person is deemed work-ready if they have the technical and interpersonal abilities required to carry out duties and cooperate efficiently with others [17]. Essentially, work readiness is a condition in which graduates are sufficiently qualified to reach maximum readiness in their professional pursuits.

Data mining is the process of gathering and analyzing data to extract significant data. Software can facilitate the collection and extraction of data through statistical analysis, mathematical formulas, or artificial intelligence (AI) technologies [18]. This project will employ many data mining techniques, including association, clustering, regression, and classification. Artificial Neural Network (ANN) is a computational model capable of processing information to handle tasks such as classification and regression. It is a component of artificial intelligence inspired by the human brain and nervous system, similar to the way neurons are arranged in the human brain [19]. Millions of interconnected nerve cells form neurons in human brain, which act as impulse conductors to enable data classification in humans [20]. Prediction model called Decision Tree (DT) uses a tree structure or hierarchy. Converting data into a decision tree and decision rules is the idea behind DT [21]. This model is useful for breaking down complex decision-making processes into simpler ones. By simplifying difficult decision-making processes, decision-makers can evaluate solutions to problems more effectively. Additionally, DT can be used to explore data and uncover hidden correlations between target variables and potential input factors. Even when used as the final model of other techniques, DT serves as an excellent starting point for modeling because it integrates data exploration and modeling [22]. One of the advantages of ANN is its ability to model both linear and non-linear relationships with a wide range. In addition, ANN also can recognize interactions that may occur between predictor variables. Even in non-linear relationships between independent and dependent variables, ANN is able to detect them thoroughly without hesitation [23].

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study builds a classification system with stages starting from dataset exploration, followed by data preprocessing (consisting of data cleaning and feature selection), division of data into training data and test data, and then the training

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

data is processed with RandomizedSearchCV for DT and SMOTE and normalization for ANN before the model is trained. Finally, the results of both models will be evaluated based on the metrics of accuracy, precision, recall, F1 score, and confusion matrix. Figure 1 displays the flowchart for the system.

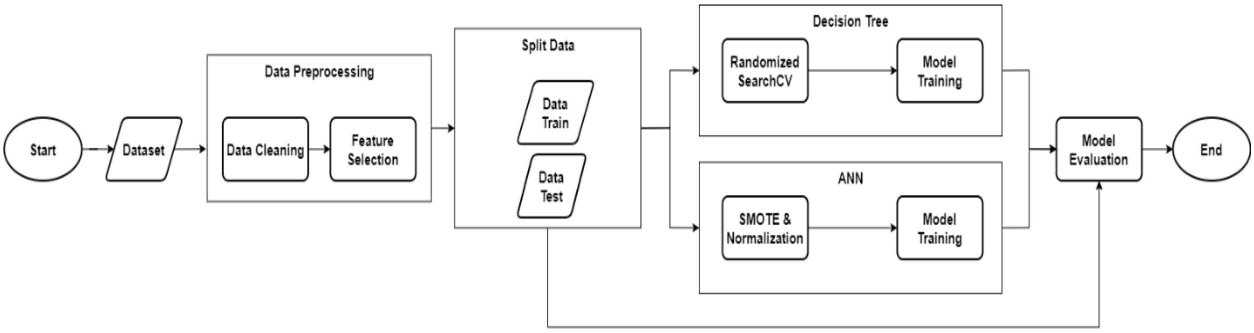

Figure 1. Flowchart of stage research

2.2 Dataset

This study uses the 2022 Tracer Study dataset, which contains 48 columns and 8,542 rows. Overall, the data contains assessments based on academic, non-academic, and external factors from the respondents.

2.3 Data Preprocessing

Preprocessing is the initial stage in dataset preparation that is done before the classification process. The raw, often unstructured data is first transformed at this point into a more appropriate format for analysis and assessment. This stage is defined by data cleaning and the selection of important qualities from academic records, while extraneous characteristics are ignored. The following describes the data preprocessing steps:

a. Data Cleaning
   Data cleansing removes unnecessary data in the classification process. For example, remove variables like student ID numbers that lose relevance after duplicate filtering. The next steps are to handle absent values and standardize the data format.
b. Feature Selection
   Building a prediction model requires considering several factors: the correlation between features and job readiness, the relevant information provided by each feature, and the feature's relevance to the target variable. Feature selection is thus done to choose the most pertinent variables such that the model training process is more efficient and its accuracy rises. In this study, feature selection was done manually by considering the context of the problem and its relationship with student work readiness. The features selected for use in the model can be seen in Table 1.

Table 1. Selected Features

Feature Name	Description
Ethical competence	Level of ethical competence during the study period
Competence based on field of study	Level of competencies based on respondent’s field of study
Time management	Individual time management skills
Use of information technology	Level of information technology usage by students during their studies
Communication	Level of communication skills of students during their studies
Type of company	Type of company where respondents work (e.g., local, national, multinational)
Teamwork	Level of teamwork ability during study period
Self-development competence	Level of self-development during studies
Emphasis on apprenticeship learning methods	Level of emphasis on apprenticeship learning methods during the study period
Emphasis on fieldwork learning methods	Level of emphasis on field learning methods during the study period
Income range	The income range earned by respondents
Industry sector of employment	Categories of companies according to their industry sector
Source of educational funding	Where the respondent’s educational expenses were funded from (e.g., parents, scholarships, self-funded)

Following the data cleaning and feature selection phases, transform all chosen attributes into a numerical format. This procedure entails converting category data, like type of company, industry sector of employment, source of educational funding, and income range, into a standardized numerical format. This transformation ensures that the

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

employed machine learning algorithm can effectively utilize all data. All categorical features are initially transformed using an encoding technique because both models—Decision Tree and Artificial Neural Network—call for numerical inputs in their execution. Though conceptually able to manage categorical data, Decision Tree requires numerical input from the libraries in use, including scikit-learn. The same preprocessed data was used for both models to preserve consistency and fairness in the model's performance assessment.

2.4 Split Data

A key step in the analysis and development of machine learning models is splitting data. By dividing the dataset into two parts—train data and test data—this method seeks to objectively evaluate the model's performance. This approach guarantees that the model can recognize patterns in new data and does not just memorize the training data. This work used 20% test data and 80% training data in a random distribution. From the total dataset, 6,827 data were set aside for training, and 1,707 data were selected for testing. Achieving reliable findings that are relevant in many settings depends on the choice of appropriate splitting ratios and techniques.

2.5 Modeling

The modeling phase used two approaches, Artificial Neural Network (ANN) and Decision Tree (DT), to evaluate the accuracy and performance of the models. DT was chosen because it could handle both categorical and numerical data and was understandable. ANN is typically used for big classification tasks and is superior at detecting complicated patterns.

The DT algorithm converts the target features into binary labels: "Ready" and "Not Ready." Using the input class_weight='balanced', DecisionTreeClassifier initializes the model to correct class imbalance. Then, RandomizedSearchCV with cross-validation using StratifiedKFold runs optimization by means of several hyperparameter combinations, including max_depth, min_samples_split, min_samples_leaf, criterion, and ccp_alpha. The best model obtained is used to run evaluations and predictions on the test data.

Then, for the ANN algorithm started with target feature conversion similar to DT. Afterwards, the features with StandardScaler and SMOTE were used to balance the data. The ANN architecture used is made up of numerous hidden layers of 256, 128, 64, and 32 neurons with batch normalization and dropout. Using an Adam optimizer with a learning rate of 0.0003, the model was constructed with a ReLU activation function in the hidden layer and a sigmoid activation function in the output layer. In addition, model training was also equipped with EarlyStopping and ReduceLROnPlateau to improve stability and prevent overfitting.

2.6 Evaluation

Model performance evaluation is the final and crucial step in understanding the extent to which a model can predict accurately. This assessment will help ensure that the selected model can accurately and reliably predict students' readiness to enter the workforce in various scenarios. The evaluation is conducted using a confusion matrix, which displays important overall values such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These four values form the basis for calculating the following evaluation metrics:

a. Accuracy
   Accuracy is used to measure how many predictions are correct compared to the total amount of data.

$$Accuracy = \frac{True\ positive + True\ negative}{Positive + Negative} \tag{1}$$

b. Precision
   Precision is used to measure how many positive predictions are correct (true positive) compared to the total positive predictions.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \tag{2}$$

c. Recall
   Recall is used to measure how much positive data was successfully identified by the model (true positive) compared to the total actual positive data.

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{3}$$

d. F1-Score
   F1-score is the harmonic mean between precision and recall.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

3. RESULT AND DISCUSSION

In predicting student readiness for work, benchmarks based on student competencies during their studies are required. The features used in this study have been adapted to the prediction requirements and incorporated into a machine learning model for calculation. The data was then divided into two categories: features and targets. The features section contains all attributes used as input, except for the target variable, while the target section refers to the status feature, which is the variable to be predicted. The status section provides information regarding students' post-graduation conditions, such as employment, entrepreneurship, further education, and unemployment. For binary classification purposes, the statuses “employed” and “self-employed” are categorized as “Ready,” while “continuing studies” and “unemployed” are categorized as “Not Ready.”

3.1 Test Result

The testing process was conducted using two algorithms, DT and ANN as a comparison. The purpose of this test was to evaluate the performance of both algorithms based on selected features in predicting student work readiness. Both models were tested using the same data to ensure that the evaluation was equivalent and fair.

In the Decision Tree model based on the tuning results, the best parameters obtained were min_samples_split=15, min_samples_leaf=2, max_depth=20, criterion='entropy', and ccp_alpha=0.0. The evaluation results on the test data showed an accuracy of 90.80%. The evaluation results in the form of Precision, Recall, and F1-score for the Decision Tree model on the test data are presented in Table 2.

Table 2. Precision, Recall, and F1-score of Decision Tree

	Precision	Recall	F1-Score
Ready	100%	87%	93%
Not Ready	70%	100%	86%

Meanwhile, the training results from the ANN model show that the model achieved an accuracy of 90.69% on the test data. The evaluation results in the form of Precision, Recall, and F1-score for the ANN model on the test data are presented in Table 3.

Table 3. Precision, Recall, and F1-score of Artificial Neural Network

	Precision	Recall	F1-Score
Ready	100%	87%	93%
Not Ready	75%	100%	86%

Based on the evaluation results in Tables 2 and 3, both models exhibited good classification performance; the decision tree reached an accuracy of 90.80%, and the artificial neural network was slightly lower at 90.69%. Essential for guaranteeing that at-risk students are not ignored, both models correctly identified all students labeled as "Not Ready," hence attaining a recall of 100% in that group. The decision tree, though, was more accurate than the ANN in recognizing the “Ready" category (100%) students; hence, it made fewer errors by mislabeling students as "Ready." By comparison, the ANN showed a slightly better balance between the two categories, with a higher recall for "Ready" (87% vs. 87%) and a better accuracy for "Not Ready" (75% vs. 70%), indicating a more cautious and generalized approach. The results show that while the decision tree is good for quick and easy decisions, the ANN model might be more reliable in different or complicated data situations, especially when it's important to reduce mistakes in identifying unprepared students. The results remain inferior to those of study [4], which utilized an ANN system to predict student preparation with an accuracy of 99%. Likewise, studies [6] and [10] demonstrated that the decision tree algorithm achieved a maximum accuracy of 100% in predicting student preparation and graduation. Conversely, study [7] indicated an accuracy of merely 50.9% for the decision tree algorithm. The discrepancies in accuracy may be ascribed to changes in the utilized characteristics, data quantity and distribution, along with the divergent problem contexts among investigations. Consequently, while the accuracy in this investigation has not attained its peak value, it yet exhibits robust and consistent performance on the utilized data, preserving a balance between accuracy and interpretability.

3.2 Analysis of Test Results

Based on the testing results, both models produced excellent accuracy and were almost equivalent, 90.80% for the Decision Tree and 90.69% for the ANN. Both models showed very similar capabilities. For precision, recall, and F1-score for “Ready,” the results were the same, Precision 100%, Recall 87%, and F1-Score 93%. Similarly, the recall and F1-score for “Not Ready” remain the same, with Recall at 100% and F1-Score at 86%. However, the Decision Tree model produces slightly more false positives (FP) compared to the ANN model, as evidenced by the FP value of 2 in the Decision Tree, while the ANN model does not produce any FP at all, resulting in a slight difference in the Precision for “Not Ready”.

Besides evaluating the models using standard performance measures like accuracy, precision, and recall, it is crucial to examine several practical and technical factors that affect the application of each model. Table 4 defines a comparison between decision trees and artificial neural networks regarding interpretability, pattern recognition

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

capability, training time, model complexity, and resource requirements. These aspects are critical when determining a model's applicability in various operational scenarios.

Table 4. Additional aspects

Additional aspects	Decision Tree	Artificial Neural Network
Interpretability	High	Low
Pattern recognition ability	Limited to linear patterns	Good at capturing non-linear patterns
Training time	Very fast (< 1 second)	Moderate (~10 seconds, depending on specifications)
Model Complexity	Low (simple structure)	High (multi-layer architecture requires tuning)
Resource requirements	Low	High (requires strong GPU/CPU)

Each model has its own advantages, Decision Tree surpasses in training time and model interpretability, facilitating understanding for non-technical stakeholders. Conversely, the ANN demonstrated superiority in classifying the data, as indicated by the lack of false positives. Nevertheless, ANN models necessitate increased training duration and processing resources and exhibit greater complexity in tuning and result interpretation. Although both produce almost equivalent performance, the final model selection can be adjusted to user needs, whether prioritizing interpretation (Decision Tree) or model flexibility (ANN).

3.3 Visualization

This section aims to provide a more profound understanding of the behavior and interpretation of the two models. To support the analysis of the test results, several visualizations are presented to show the distribution of predictions and feature importance of the two models.

A feature importance analysis was performed to enhance understanding of the decision-making processes of each model. Figures 2 and 3 illustrate the significance of each input feature in the decision tree and artificial neural network models, respectively. This research identifies the elements that most significantly influenced the forecast of students' work preparedness.

Figure 2 illustrates that the Decision Tree model predominantly depended on the "Industry sector of employment," which accounted for approximately 75% of the total prediction. Additional attributes, including "Type of company," "Emphasis on fieldwork learning methods," and "Income range," exhibited markedly lower yet still discernible contributions. Concurrently, the remaining features exerted low influence, indicating that the model's decisions were predominantly controlled by a limited array of salient traits. This exemplifies the characteristic behavior of Decision Trees, wherein splits occur on the most informative traits early in the tree structure, potentially disregarding others that do not yield substantial information gain.

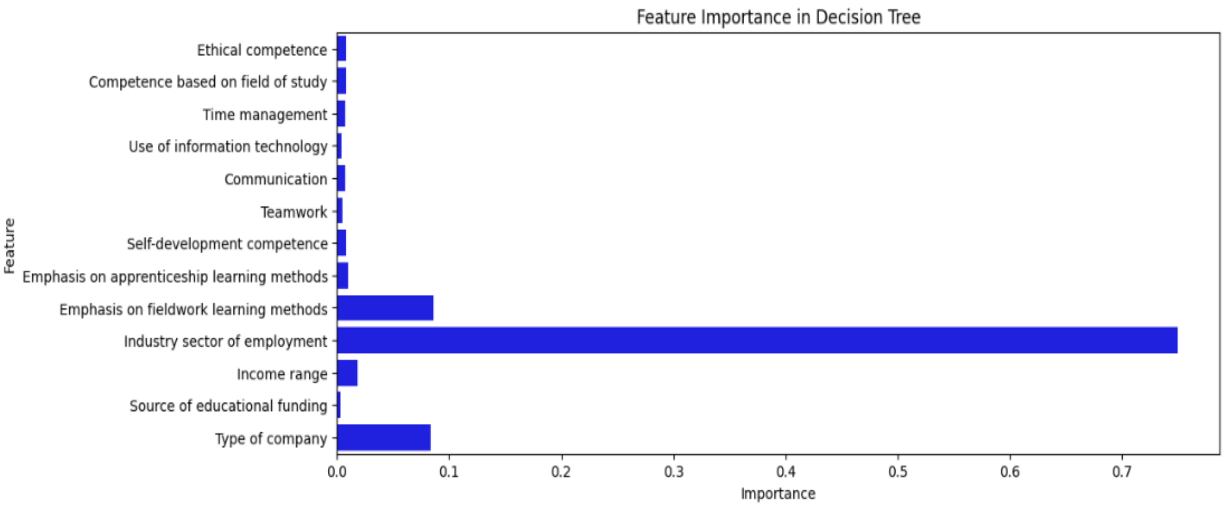

Figure 2. Feature Importance in Decision Tree Model

In contrast, the distribution of feature importance in the ANN model in Figure 3 showed greater balance. Although “Industry sector of employment,” “Emphasis on apprenticeship learning methods,” and “Emphasis on fieldwork learning methods” were the most significant, other factors such as “Self-development competence,” “Communication,” “Type of company,” and “Income range” also carried substantial importance. This indicates that the ANN successfully identified a more intricate pattern across a wider array of features. The uniform distribution of significance demonstrates ANN's proficiency in modeling non-linear interactions and combining various signals, perhaps resulting in enhanced generalization across varied datasets.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

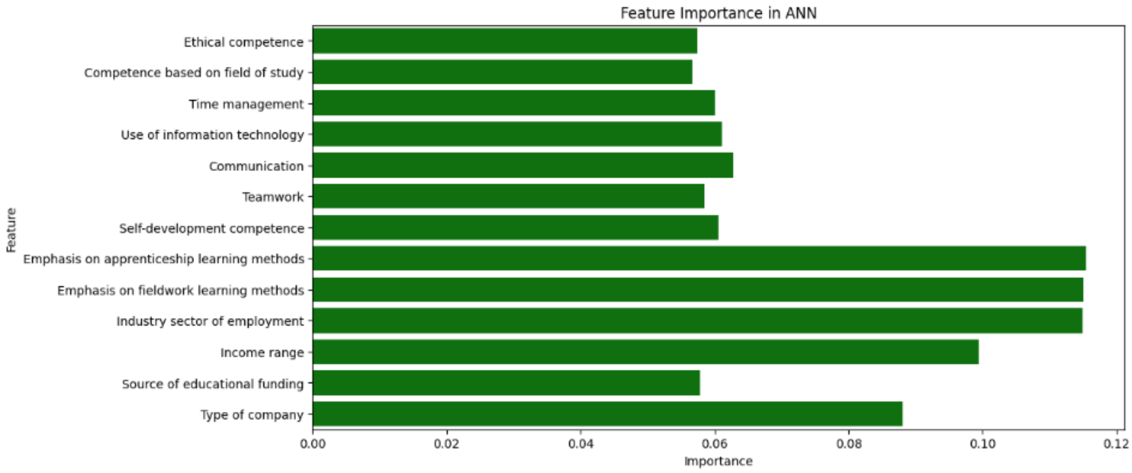

Figure 3. Feature Importance in ANN Model

The difference between the two models highlights the compromise between interpretability and adaptability of patterns. Emphasizing a small number of important variables, the decision tree model delivers clearer insights. The ANN model, on the other hand, offers a more complicated understanding of numerous interacting elements but compromises some interpretability.

Next, Figure 4 provides a thorough view of how the artificial neural network model learns and performs consistently over 60 epochs, showing the training and validation loss on the left and the training and validation accuracy on the right. These graphs facilitate the visualization of the generalization of the model to unseen data across the training phase. For models like the artificial neural network, which goes through iterative optimization across several epochs, this visual assessment is vital. On the other hand, the decision tree model is trained in a single run without iterative updates; thus, it does not produce similar loss or accuracy curves.

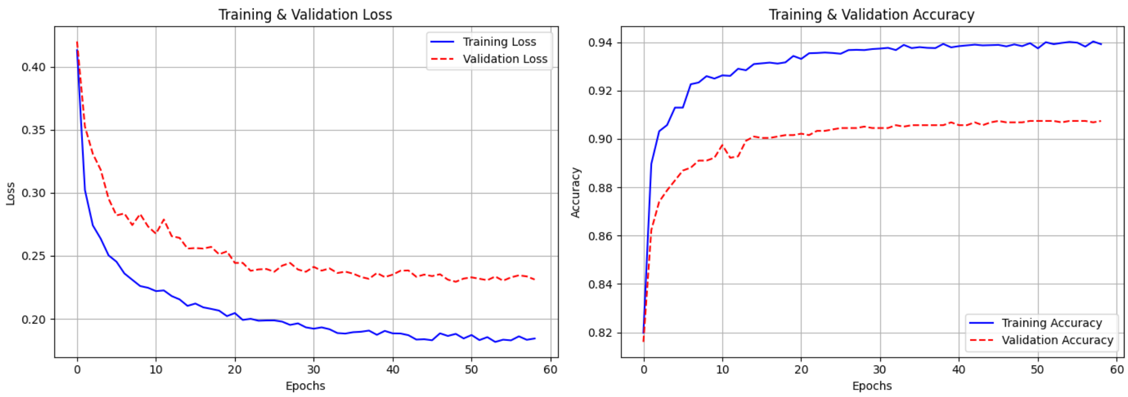

Figure 4. Training and validation loss and accuracy of the ANN model

The training and validation loss graph indicates that the training loss consistently decreases with an increase in epochs, whereas the validation loss exhibits a downward trend with minor fluctuations. The tight and consistent gap between the two curves shows that the model is learning quickly without much overfitting. The accuracy graph indicates consistent enhancement in both training and validation accuracy. The training accuracy attains approximately 94%, whereas the validation accuracy levels off at approximately 90%, suggesting effective generalization performance. The lack of significant divergences between the two curves indicates that the ANN model achieved effective convergence and showed consistent performance during training. The decision tree model is not applicable for epoch-based visualization since it is non-iterative. The performance is assessed using final measures, including accuracy, precision, recall, and the confusion matrix, once the model is fully developed.

4. CONCLUSION

This study successfully developed a model to predict students' readiness for work by achieving an accuracy of 90.80% with a decision tree and 90.69% with an artificial neural network (ANN), demonstrating that both models exhibit nearly equivalent performance. The decision tree model is superior in interpretability, whereas the artificial neural network demonstrates greater flexibility in managing intricate non-linear patterns, albeit necessitating a more stringent training regimen and extended computation duration. The findings indicate that both models are appropriate for further advancement as systems for predicting student job readiness, each possessing distinct strengths and shortcomings that

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 222−229
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7316

may be customized to meet end-user requirements. Future research should involve testing the models on bigger and more diverse datasets to evaluate generalization capabilities, as well as incorporating ensemble learning approaches, automatic feature selection, and AUC-ROC and fairness analyses to improve model validity and performance. In practical terms, higher education institutions can employ these models as an early detection mechanism to identify students who are not yet prepared for the workforce, facilitating intervention through individualized training or soft skills enhancement and providing more tailored career guidance services via career centers. Further development could incorporate this model into a web-based dashboard for real-time input and output, improving accessibility and usability for non-technical staff. This study enhances educational data science by evaluating student work preparedness through the amalgamation of academic, behavioral, and socioeconomic factors; hence, it facilitates the incorporation of machine learning in institutional decision-making within higher education.

REFERENCES

[1] U. N. Fajriah and K. Sudarma, "Pengaruh Praktik Kerja Industri, Motivasi Memasuki Dunia Kerja, dan Bimbingan Karir pada Kesiapan Kerja Siswa," *Econ. Educ. Anal. J.*, vol. 6, no. 2, pp. 421–432, 2017, [Online]. Available: http://journal.unnes.ac.id/sju/index.php/eeaj

[2] M. J. Kavanagh and Richard David Johnson, *Human resource information systems : Basics, applications, and future directions*. Los Angeles: Sage, 2018.

[3] B. Gabrikova, L. Svabova, and K. Kramarova, "Machine Learning Ensemble Modelling for Predicting Unemployment Duration," *Appl. Sci.*, vol. 13, no. 18, 2023, doi: 10.3390/app131810146.

[4] D. Kartini, J. A. Yani Km, and K. Selatan, "Penerapan Data Mining dengan Algoritma Neural Network (Backpropagation) Untuk Prediksi Lama Studi Mahasiswa," *Pros. Semin. Nas. Sisfotek*, vol. 3584, pp. 235–241, 2017.

[5] D. Galih Pradana, M. L. Alghifari, M. Farhan Juna, and S. Dwisiwi Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 55–60, 2022.

[6] H. Noviyanto and B. Mukti, "Prediksi Kesiapan Kerja Mahasiswa menggunakan Algoritme K-Means dan C4.5," *JOSTECH J. Sci. Technol.*, vol. 2, no. 2, pp. 179–188, 2022, doi: 10.15548/jostech.v2i2.4422.

[7] F. R. Yudana, M. Suyanto, and A. Nasiri, "Model Klasifikasi Untuk Menentukan Kesiapan Kerja Mahasiswa Dan Kelulusan Tepat Waktu Dengan Metode Machine Learning," *IJITECH Indones. J. Inf. Technol.*, vol. 1, no. 1, pp. 1–12, 2023, doi: 10.37680/ijitech.v1i1.xx.

[8] I. Supriadi, A. Hariyanti, M. Z. Abidin, Rinrin, and D. Gustian, "Penerapan regresi linier berganda dalam kesiapan kerja mahasiswa," *Semin. Nas. Inform. 2020*, vol. 1, no. 1, pp. 204–211, 2020.

[9] Muhammad Hadiza Baffa, Muhammad Abubakar Miyim, and Abdullahi Sani Dauda, "Machine Learning for Predicting Students' Employability," *UMYU Sci.*, vol. 2, no. 1, pp. 001–009, 2023, doi: 10.56919/usci.2123_001.

[10] G. ElSharkawy, Y. Helmy, and E. Yehia, "Employability Prediction of Information Technology Graduates using Machine Learning Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, pp. 359–367, 2022, doi: 10.14569/IJACSA.2022.0131043.

[11] A. Malik *et al.*, "Forecasting students' adaptability in online entrepreneurship education using modified ensemble machine learning model," *Array*, vol. 19, no. July, p. 100303, 2023, doi: 10.1016/j.array.2023.100303.

[12] V. Prikshat, A. Nankervis, J. Burgess, and S. Dhakal, "Conceptualising graduate work-readiness: theories, concepts and implications for practice and research," *Transit. from Grad. to Work challenges Strateg. Twenty-first Century Asia Pacific beyond*, pp. 15–29, 2019.

[13] F. Nasril, D. Indiyati, and G. Ramantoko, "Talent performance analysis using people analytics approach," *Budapest Int. Res. Critics Inst. Humanit. Soc. Sci.*, vol. 4, no. 1, pp. 216–230, 2021.

[14] M. Thite, *e-HRM: Digital approaches, directions & applications*. Routledge, 2018.

[15] R. L. Cross, J. Singer, S. Colella, R. J. Thomas, and Y. Silverstone, *The organizational network fieldbook: Best practices, techniques and exercises to drive organizational innovation and performance*. Hoboken, NJ: John Wiley & Sons, 2010.

[16] F. K. Wardani *et al.*, "The Construction of Work Readiness Scale," *14th Univ. Res. Colloqium*, pp. 78–87, 2021, [Online]. Available: http://repository.urecol.org/index.php/proceeding/article/view/1774

[17] F. F. Kirani, P. R. S. Prakasa, F. F. Vivany, G. A. Arsi, and H. D. Humairo, "Description and Efforts to Increase Work Readiness in Early Adult Individuals Who Have Not Get A Job," *J. Abdi Insa. Univ. Mataram*, vol. 9, no. 1, pp. 160–168, 2022.

[18] I. Ahmad, S. Samsugi, and Y. Irawan, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/107

[19] O. V. William Chan, Navdeep Jaitly, Quoc Le, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," *IEEE Int. Conf. Acoust. Speech, Signal Process. 2016*, pp. 4960–4964, 2016.

[20] Z. ``Leo'' Liu, "Artificial Neural Networks," in *Artificial Intelligence for Engineers: Basics and Implementations*, Cham: Springer Nature Switzerland, 2025, pp. 175–190. doi: 10.1007/978-3-031-75953-6_7.

[21] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.

[22] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch. Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015, doi: 10.11919/j.issn.1002-0829.215044.

[23] H. D. Bhakti, "Aplikasi Artificial Neural Network (ANN) untuk Memprediksi Masa Studi Mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik," *Eksplora Inform.*, vol. 9, no. 1, pp. 88–95, 2019, doi: 10.30864/eksplora.v9i1.234.