DOI: 10.22146/jnteti.v14i3.20516

Interpretable Machine Learning for Job Placement Prediction: A SHAP-Based Feature Analysis

Swono Sibagariang

Batam State Polytechnic, Kota Batam, Kepulauan Riau 2946, Indonesia

[Submitted: 25 April 2025, Revised: 18 June 2025, Accepted: 14 July 2025]
Corresponding Author: Swono Sibagariang (email: swono@polibatam.ac.id)

ABSTRACT — Predictive modeling is important in analyzing graduates' job outcomes, especially in forecasting job placements based on academic performance and courses. This study aims to improve predictive accuracy and interpretability in job placement classification using advanced machine learning models and SHapley Additive exPlanations (SHAP) analysis. Utilizing a dataset containing graduates' academic records, including course grades, grade point average (GPA), and internship duration, this research employed several classification models, including decision tree, random forest, extreme gradient boosting (XGBoost), light gradient-boosting machine (LightGBM), CatBoost, and logistic regression. Evaluation metrics showed that most models achieve 92% precision, 92% recall, and 92% F1 score, with an accuracy of 85%, while logistic regression excelled with 100% recall, 96% F1 score, and 92% accuracy. SHAP analysis identified key features such as Administration, Computer Organization, Information Systems, Entrepreneurship, Professional Ethics, and Web Programming as the most influential in predicting job placement. Other significant contributors include Introduction to Information Technology, Software Engineering II, and Data Mining, although with relatively lower influence. Extracurricular activities and internship experiences were also found to be influential factors, highlighting the importance of academic and nonacademic elements in shaping graduates' career prospects. These findings highlight and emphasize the need to provide students with certain academic courses to better prepare them for the job market. These findings emphasize the importance of interpretable machine learning models in career forecasting, enabling educational institutions to optimize curriculum design and enhance graduates' employability. Future research should explore feature selection techniques, temporal analysis, and personalized recommendation systems to refine predictive accuracy.

KEYWORDS — Machine Learning, Job Suitability Prediction, Shapley Additive Explanations (SHAP), Graduate Job Placement.

I. INTRODUCTION

Competition in the job market is becoming increasingly fierce for university graduates, especially with the rapid industrial transformation and the continuously evolving demand for skills across various sectors. One of the main challenges faced by graduates is finding a job that matches their field of study. According to the World Economic Forum report in 2023, more than 30% of university graduates required more than six months to get their first job [1]. One of the main factors contributing to this issue is the gap between the academic skills acquired during studies and the skills demanded by the industry. The academic curriculum often does not align with technological advancements and labor market needs, leaving graduates less prepared to compete in the professional world.

Currently, most educational institutions still rely on traditional methods for career guidance and predicting graduates' job opportunities. This manual approach has limitations in considering various academic and nonacademic factors that contribute to the career success of graduates. Therefore, a data-driven approach using machine learning techniques has emerged as a potential solution to understanding the patterns linking academic performance, experience, and career success.

Several studies have explored the application of machine learning in predicting job placements for university graduates. Previous research has shown that machine learning techniques can effectively analyze academic factors and classify alumni job statuses with high accuracy. For example, a study using the random forest algorithm achieved an accuracy rate of 98% in predicting job placement based on academic attributes such as cumulative grade point average GPA, internship experience, and involvement in extracurricular activities [2]. Additionally, factors related to internships such as duration and performance during training have proven to play a significant role in predicting job placement, with the ensemble machine learning approach (stacking) achieving an accuracy of up to 91% [3].

Although previous research results are promising, some areas remain unexplored. Most studies only consider general academic factors, such as GPA and internship experience, without investigating the correlation between the performance of specific courses and graduates' career paths. Moreover, the models used often focus solely on accuracy metrics without conducting in-depth interpretability analysis to understand the key factors contributing to the predictions.

This research presents novelty in its approach by emphasizing a detailed analysis of specific courses that affect graduates' job prospects, rather than solely relying on overall GPA. By utilizing feature selection techniques and SHapley Additive exPlanations (SHAP), this research aims to identify the most significant academic and nonacademic factors influencing alumni career success [4], [5]. Previous studies generally focused on aggregate variables such as cumulative GPA, internship status, or extracurricular activities as primary predictors, without considering the specific contributions of certain courses to career outcomes [2], [3]. However, while SHAP has been shown to help understand predictions about academic success and job satisfaction [4], [6], it has not been widely used to directly assess how relevant the curriculum is to getting a job after graduation. Thus, this research not only offers a more focused approach to feature selection but also evaluates the model based on accuracy and emphasizes the importance of interpretability in giving deep insights to policymakers in enhancing graduates' job readiness.

This research contributes significantly to supporting higher education institutions in creating data-driven policies relevant to the job market demands. With more transparent prediction results and model interpretation, institutions can design more targeted curricula and training.

In term of programs and career guidance services, this research has the potential to help graduates secure jobs that match their educational background more quickly and efficiently, thereby increasing their competitiveness in the global job market and reducing the rate of educated unemployment.

The main objective of this research is to develop a machine-learning model capable of predicting the employability of graduates based on their academic records and nonacademic experiences. To achieve this, the study examined the effectiveness of various machine learning algorithms, including decision tree, random forest, extreme gradient boosting (XGBoost), light gradient-boosting machine (LightGBM), CatBoost, gradient boosting machines (GBM), and logistic regression, in classifying the employment status of graduates. The developed model was evaluated using metrics such as accuracy, precision, recall, and F1 score to ensure the reliability of predictions.

The research methodology included collecting academic and nonacademic data from alumni, such as job information, course grades, GPA, internship experience, and organizational involvement. The collected data underwent preprocessing techniques, including value normalization, categorical feature encoding, and feature selection, to identify the most influential variables on job placement. Additionally, the predictive model was tested using cross-validation techniques and analyzed with SHAP to interpret the key factors contributing to job placement predictions.

II. RELEATED WORKS

Predicting job placement outcomes for university graduates has become a prominent area of research, particularly in light of the growing adoption of data-driven decision-making within higher education and human resource management. Traditional prediction methods have primarily relied on subjective assessments, historical employment trends, and manual evaluations of student qualifications. However, these approaches are often limited in accuracy, scalability, and objectivity. In contrast, machine learning techniques have emerged as a powerful alternative, capable of processing large datasets and uncovering complex patterns that influence job placement outcomes. Numerous studies have examined the application of various machine learning algorithms in this domain and reported encouraging results regarding predictive accuracy and reliability.

Several studies have explored using machine learning models to predict job placement outcomes for graduates. A study examined the prediction of student placements through diverse machine-learning methodologies. It demonstrated that neural networks attained an accuracy rate of 85% in forecasting employment within six months post-graduation [7]. A separate study utilized logistic regression, integrating academic performance with demographic data, achieving a prediction accuracy of 87% [8]. A study used the decision tree model to forecast job placement outcomes based on academic achievement and extracurricular activities, achieving an accuracy of 89% [9]. These studies underscore the efficacy of machine learning methodologies in assessing graduate employability and enhancing job placement forecasts.

Supplementary research has corroborated these findings. A study demonstrated that decision tree classifiers can accurately estimate students' employment capacities based on academic achievement data [10]. A separate study indicated that utilizing the XGBoost model on graduates' academic records accurately forecasts students' career trajectories [11]. A job recommendation system using a recurrent neural network (RNN) methodology has been presented, wherein the input data is converted into vector representations through doc2Vec to maintain semantic significance [12]. A hybrid convolutional neural network (CNN)-RNN model augmented with a self-attention mechanism was created to forecast employee performance. This model effectively encapsulates semantic and structural information, yielding enhanced predictive accuracy [13].

Although many studies have validated the effectiveness of machine learning techniques in predicting job placement outcomes, variations in accuracy and performance depend on several factors, including feature selection, dataset quality, and model selection. This study aims to build on this research by developing a better predictive model that integrates various machine learning approaches, including decision tree, random forest, XGBoost, LightGBM, CatBoost, GBM, and logistic regression. This study aims to provide a robust and interpretable framework for predicting graduate employment outcomes by leveraging various features encompassing academic, extracurricular, and personal attributes.

Overall, the existing literature emphasizes the potential of machine learning in optimizing job placement predictions. However, further research is necessary to refine these models and improve their generalization across various academic disciplines and geographical contexts. This study contributes to the field by exploring the comparative performance of multiple machine learning algorithms and proposing an optimized model that can assist universities and employers in streamlining the job placement process.

III. METHODOLOGY

This research was conducted using a structured methodology, encompassing stages of data collection, preprocessing, model development, evaluation, and interpretability analysis. The software used in this research included various tools and libraries within the Python ecosystem that support data analysis and the implementation of machine learning algorithms. The collected data were processed using the Pandas library, allowing for the efficient import and manipulation of data in spreadsheet format. For numerical analysis, this research utilized NumPy, which provided various mathematical functions to process data optimally.

A. DATA COLLECION

This study used data from graduates of the D3 Informatics Engineering program at Politeknik Negeri Batam in 2019 through surveys, institutional records, and datasets from alumni tracer studies. This dataset included various academic and nonacademic factors that could affect graduate employment placement.

The academic factor in this study showed students' academic achievements during their studies. The data consisted of the cumulative GPA and grades from various courses relevant to the D3 Informatics Engineering curriculum. The courses analyzed include Final Project I and II, English I and II, Professional Ethics, Work Reporting, Administration, Basic Programming, Data Mining, Computer Networks, Occupational Safety & Health, Entrepreneurship, Mathematics, Database Programming, Hardware Programming, Object-Oriented Programming, Web Programming, Introduction to Databases, Introduction to Information Technology, Information Systems, Statistics, Pancasila, Artificial Intelligence, Multimedia, Mobile Device Programming, Computer Organization, Operating Systems, Advanced Computer Networks, and Software Engineering I and II. The grades from these courses were converted into a numerical scale (0–4).

The job information category included various data about the employment status of alumni after graduation. These data included information whether alumni were employed, unemployed, self-employed, interning, or freelancing. In addition, information regarding the workplace, type of company (such as industry sector, educational institution, or non-profit organization), as well as the scale of the company (local, national, or multinational) was also analyzed. Other factors also considered include income in rupiah, the waiting time to get a job after graduation, and the degree of job relevance to the field of study pursued.

This research also considered nonacademic data that could influence graduates' success in obtaining employment. These data included organizational experiences that alumni had participated in, including whether they had ever been members, administrators, chairpersons, or held other roles within the organization. Internship experience was also an important aspect in this category, including the duration of the internship and the industry field in which it was conducted, such as information technology, manufacturing, or finance.

B. DATA PREPROCESSING

Data preprocessing is a crucial step in the development of machine learning models. It ensures that the data used are of high quality and ready for further analysis. The main stages in data preprocessing include data cleaning, feature encoding, and normalization.

Data cleaning is the process of identifying and handling anomalies in a dataset, such as missing values, duplicates, or inconsistencies. This step is important to improve model accuracy and prevent bias in the analysis. Commonly used techniques include imputing missing values, removing duplicates, and correcting typographical errors. According to [14], effective data cleaning significantly improved the performance of machine learning models.

The dataset often contains categorical variables that needed to be converted into a numerical format so machine learning algorithms can process them. The encoding method commonly used includes label encoding in machine learning and data preprocessing. Label encoding was a method that converts categorical data represented as text labels into a numerical format. In other words, this involves assigning a unique numerical value to each category or label in the categorical variable [15]. One-hot encoding in data processing and machine learning refers to a technique used to represent categorical variables as binary vectors. Each category or label is transformed into a binary vector with a length equal to the total number of distinct categories in the variable. All elements of the vector have a value of zero, except for the index corresponding to the category, which is represented by the number 1 [16]. The selection of the appropriate encoding method can affect the model's overall performance.

Normalization is the process of scaling numerical features into a certain range, usually between 0 and 1, to ensure that each feature contributes proportionally to the model. One common normalization method is the min-max scaling, which is formulated as in (1).

$$x' \frac{x-x_{min}}{x_{max}-x_{min}} \quad (1)$$

In the normalization process, x' represented the normalized value, x was the original value, and x_{min} and x_{max} denoted the minimum and maximum values of the feature, respectively [17]. Proper normalization prevented features with a larger numerical range from dominating the analysis [18].

C. MACHINE LEARNING

Developing machine learning models was important in building a reliable prediction system. Various machine learning algorithms were applied and compared to identify the model with the best performance in predicting graduate job placements. The models applied included decision tree, random forest, XGBoost, LightGBM, CatBoost, GBM, and logistic regression.

Decision tree is an algorithm characterized by a rule-based structure resembling a tree that hierarchically expressed features [19]. This algorithm divides data into subsets based on specific attributes, forming a tree-like structure that aids decision-making. The main advantages of decision tree are its simplicity and its ability to be easily interpreted.

Random forest is an ensemble learning technique that builds several decision trees during training and combines their results to make the final decision. This method effectively handles large volumes of data and various variables, and it can reduce overfitting by combining predictions from multiple decision trees, resulting in a stronger and more generalized model [20].

XGBoost is a boosting algorithm known for its high accuracy. XGBoost works by building models iteratively, where each new model attempts to correct the errors of the previous model. This technique has proven superior in various competitions and machine learning applications, especially in handling complex and diverse data [21].

LightGBM is a gradient-boosting framework optimized for efficiency and scalability in machine learning applications. It uses histogram-based learning to accelerate training, allowing the management of large datasets with less memory usage. Unlike conventional boosting techniques, LightGBM builds tree leaf-wise rather than a level-wise approach, which improves accuracy and reduces overfitting [22].

CatBoost is a gradient-boosting algorithm specifically designed to efficiently handle categorical features. Unlike traditional boosting algorithms that require extensive preprocessing for categorical variables, CatBoost combines advanced encoding mechanisms that allows the direct processing of categorical data while maintaining its predictive power. The main feature of CatBoost is ordered boosting, which reduces target leakage and overfitting by ensuring that each training instance only use past information for prediction [23].

GBM is an ensemble learning methodology that builds predictive models by gradually combining several weak learners, usually in decision trees. Each subsequent model is trained to correct the residual errors of the previous ensemble, thereby improving the overall predictive accuracy [24].

Logistic regression was a commonly used statistical technique for modelling binary outcomes, where the dependent variable can take one of two possible values, usually denoted as 0 or 1. It uses the logistic (sigmoid) function on a linear combination of the input data to determine the likelihood that a specific input is associated with a particular class. Unlike linear regression, which predicts continuous values, logistic regression produces probability scores ranging from 0 to 1. Each input characteristic is assigned a weight (coefficient), and the weighted sum is transformed using the sigmoid function to produce a probability. A threshold, often set at 0.5, is then used to classify observations into one of two categories.

This approach effectively predicts the probability of an event occurring based on one or more predictor variables. One of the main advantages of logistic regression is its interpretability; the model coefficients reveal the strength and direction of the relationship between each variable and the predicted outcome. Logistic regression is based on the assumption that the independent variables are linearly related to the log odds of the dependent variable. However, this assumption does not hold in complex or nonlinear situations. Additionally, it requires the observations to be independent and assumed minimal multicollinearity among the predictors [25].

D. SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

The SHAP approach is a method used to interpret black-box or hard-to-understand machine learning prediction models [26]. The goal of SHAP is to explain the prediction of a feature xxx by calculating the contribution of each feature to the prediction [27]. The SHAP method calculates Shapley values in a manner similar to coalition game theory [28]. In this context, the feature values of a data instance acted as players in a coalition. The Shapley value fairly distributes the prediction contributions among the features. Each player represents the value of an individual feature, enabling a deeper understanding of the factors that influences the prediction outcome.

The SHAP approach is presented as follows:

$$\phi_i = \sum_{S\subseteq\{1,\dots,p\}\setminus\{i\}} \frac{|S|!\,(p-|S|-1)!}{p!} \times [Val(S \cup \{i\}) - Val(S)] \quad (2)$$

ϕ_i is the Shapley value of a feature member with respect to the prediction outcome. $Val(S)$ is the output of the machine learning model that will be explained using a set of features S, and p is the total number of all features.

The final contribution or Shapley value of a feature $i(\phi i)$ is also defined as the average of its marginal contributions across all possible permutations of the feature set. This process uses Shapley value calculations by incorporating all possible combinations of features and measuring how the contribution of each feature changes as other features change.

IV. RESULTS AND DISCUSSION

This section analyzes in detail specific courses that affect graduates' job opportunities and compares machine learning algorithms applied to identify the best-performing models in predicting graduates' job placements.

A. DATASET DESCRIPTION

Data for these models comes from the alumni dataset. The variables used in this study encompass all parameters that affect the job placement of graduates. In the dataset, there are 43 attributes that include specific academic information about graduates such as job information, academic performance, involvement in organizations, and internship experience. The dataset information can be seen in Table I.

Encoding was carried out on several categorical attributes to prepare the data for modelling. Label encoding was applied to binary or ordinal features such as employment status, job relevance to study, industry internship, and organizational activity. This method converted categorical values into simple numeric representations, such as "Yes" to 1 and "No" to 0. Meanwhile, one-hot encoding was used for features with multiple nominal categories with no inherent order, such as workplace, type of company, company scale, position in organization, internship duration, and internship field. One-hot encoding transformed each unique category within these features into separate binary columns, allowing the machine learning algorithms to process the information without implying any ordinal relationship. This preprocessing step ensured that all input data were in a numerical format suitable for the machine learning models used in the study.

B. EVALUATION METHOD

This study assessed the model's performance through a confusion matrix, a crucial instrument for quantifying the model's accuracy in data classification. This matrix allows for ascertaining the number of correct and incorrect predictions, facilitating the calculation of critical metrics such as accuracy, precision, recall, and F1 score [29]. The computation of these measures adheres to widely accepted formulas by contrasting the total accurate predictions (true positives and true negatives) with the overall number of forecasts, as delineated in (3).

$$Accuracy = \frac{TP+TN+FP+FN}{TP+TN} \quad (3)$$

True positive (TP) denotes the quantity of positive instances accurately identified by the model as positive. Conversely, a false positive (FP) denotes the quantity of negative instances erroneously classified as positive. A false negative (FN) refers to the amount of positive data that the model erroneously categorizes as negative. True negative (TN) refers to the quantity of negative instances accurately identified by the model as negative [30].

C. MODEL PARAMETERS

Several machine learning models were implemented in Python to perform classification and prediction tasks. The models were built and evaluated using scikit-learn, XGBoost, LightGBM, and CatBoost libraries. In addition, the classification-metrics framework was used to assess model performance, while Pandas and NumPy were utilized for data manipulation and comparison. Each model was configured with specific parameters to optimize performance and ensure consistency during the training and testing. The parameter configurations used in this study are presented in Table II.

D. MODEL ANALYSIS

A correlation analysis was carried out to better understand the relationships between variables in the dataset. This step

Jurnal Nasional Teknik Elektro dan Teknologi Informasi
Volume 14 Number 3 August 2025

TABLE I
DATASET INFORMATION

No	Attribute Name	Variable Description	Type of Attribute
1	NIM	Student Identification Number	String
2	Name	Student's Full Name	String
3	GPA	Grade Point Average	Numeric
4	Final Project I	Score for Final Project I (0-4)	Numeric
5	Final Project II	Score for Final Project II (0-4)	Numeric
6	English I	Score for English I (0-4)	Numeric
7	English II	Score for English II (0-4)	Numeric
8	Professional Ethics	Score for Professional Ethics (0-4)	Numeric
9	Work Reporting	Score for Work Reporting (0-4)	Numeric
10	Administration	Score for Administration (0-4)	Numeric
11	Programming Basics	Score for Programming Basics (0-4)	Numeric
12	Data Mining	Score for Data Mining (0-4)	Numeric
13	Computer Networks	Score for Computer Networks (0-4)	Numeric
14	Occupational Health & Safety	Score for Occupational Health & Safety (0-4)	Numeric
15	Entrepreneurship	Score for Entrepreneurship (0-4)	Numeric
16	Mathematics	Score for Mathematics (0-4)	Numeric
17	Database Programming	Score for Database Programming (0-4)	Numeric
18	Hardware Programming	Score for Hardware Programming (0-4)	Numeric
19	Object-Oriented Programming	Score for Object-Oriented Programming (0-4)	Numeric
20	Web Programming	Score for Web Programming (0-4)	Numeric
21	Introduction to Databases	Score for Introduction to Databases (0-4)	Numeric
22	Introduction to IT	Score for Introduction to IT (0-4)	Numeric
23	Information Systems	Score for Information Systems (0-4)	Numeric
24	Statistics	Score for Statistics (0-4)	Numeric
25	Pancasila	Score for Pancasila (0-4)	Numeric
26	Artificial Intelligence	Score for Artificial Intelligence (0-4)	Numeric
27	Multimedia	Score for Multimedia (0-4)	Numeric
28	Mobile Programming	Score for Mobile Programming (0-4)	Numeric
29	Computer Organization	Score for Computer Organization (0-4)	Numeric
30	Operating Systems	Score for Operating Systems (0-4)	Numeric
31	Advanced Computer Networks	Score for Advanced Computer Networks (0-4)	Numeric
32	Software Engineering I	Score for Software Engineering I (0-4)	Numeric
33	Software Engineering II	Score for Software Engineering II (0-4)	Numeric
34	Have you joined an organization?	Whether the student has joined an organization	Boolean (Yes/No)
35	Position in Organization	Role in the organization (Member, Committee, Leader, Other)	String
36	Industry Internship	Whether the student has done an internship	Boolean (Yes/No)
37	Internship Duration	Duration of the internship (in months)	Numeric
38	Internship Field	Field of internship (IT, Manufacturing, Finance, etc.)	String
39	Employment Status	Current job status (Employed, Unemployed)	String
40	Workplace	Name of the workplace	String
41	Type of Company	Industry sector (Industry, Education Institution, Nonprofit, etc.)	String
42	Company Scale	Size of the company (Local, National, Multinational/International)	String
43	Job Relevance to Study	Whether the job is relevant to the field of study (Relevant, Not Relevant)	String

TABLE II
PARAMETER CONFIGURATION

Classification Model	Variable Configuration
RandomForestClassifier	n_estimators=100, random_state=42
XGBClassifier	use_label_encoder=False, eval_metric='logloss', random_state=42
DecisionTreeClassifier	random_state=42
GradientBoostingClassifier	random_state=42
LGBMClassifier	random_state=42, verbose=-1
CatBoostClassifier	verbose=0, random_state=42
LogisticRegression	max_iter=1000,random_state=42

aimed to identify which attributes were strongly associated and relatively independent, providing insight into potential feature importance and redundancy. The results of this analysis were visualized using a correlation heatmap, as presented in Figure 1. The correlation heatmap showed how each dataset attribute was related to all other attributes. Lighter colors indicate a lack

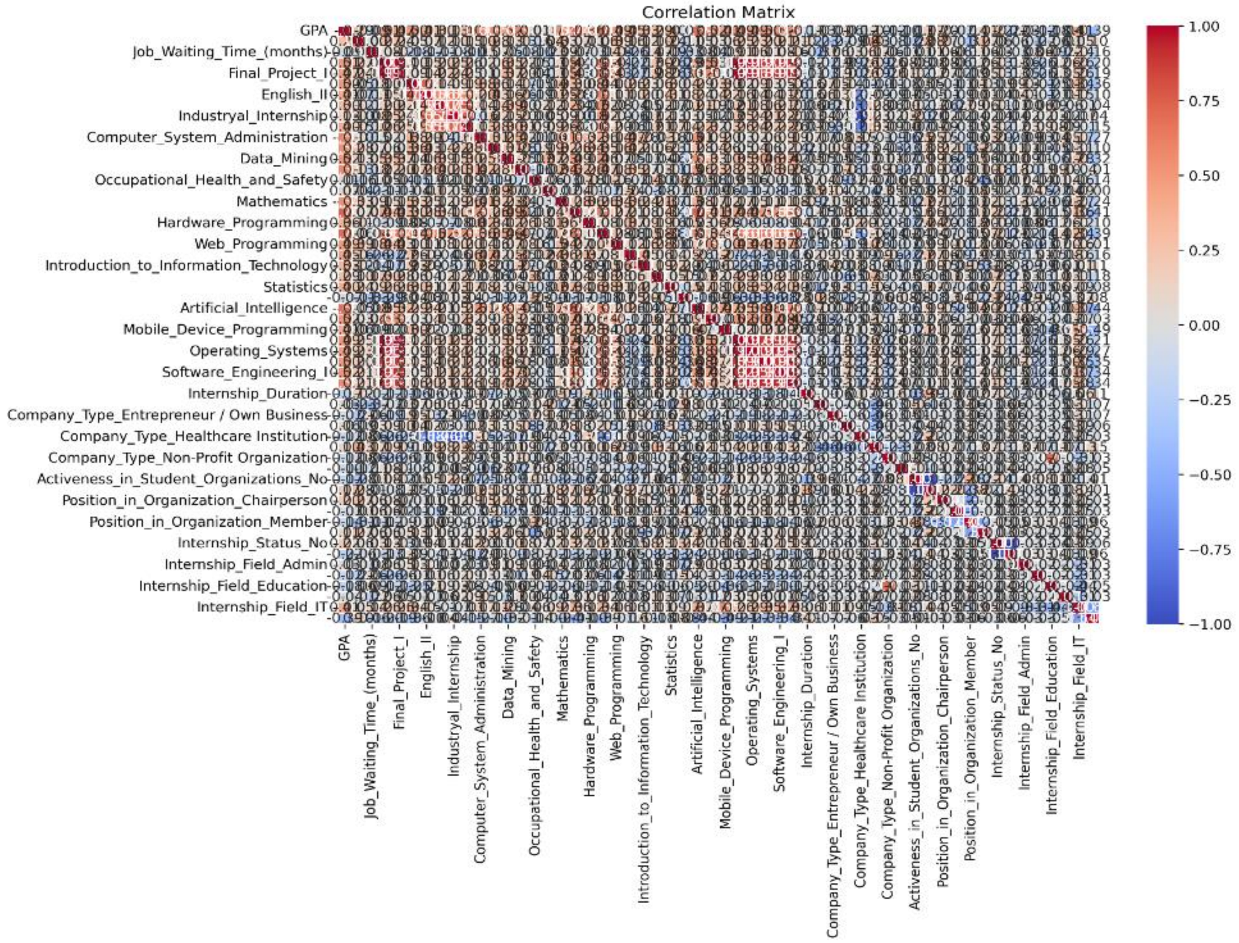

Figure 1. Correlation matrix.

of correlation between two variables, while darker blue shades indicate a stronger correlation within the range of -1.0 to 1.0. The increasing sample correlation over time reflected the quality of the synthetic dataset, with most attributes exhibiting high correlation with others across the dataset. The performance of the machine learning model was evaluated based on accuracy, recall, precision, and F1 score. The detailed evaluation results for each model can be seen in Table III.

Evaluation of various machine learning models for classification showed consistent and high performance across multiple metrics. Results obtained all achieved identical scores, with precision 92%, recall 92%, and F1 score 92%, while their accuracy remained at 85%. This consistency indicated that these models effectively balanced precision and recall but had room for improvement in overall accuracy. Among these models, Logistic regression stood out with a perfect recall of 100% while maintaining a precision of 92%. This resulted in an F1 score of 96% and an overall accuracy improvement of 92%, indicating that this model accurately identified all positive cases without missing any relevant instances. Given these results, logistic regression appeared to be the most reliable model for capturing all relevant instances while maintaining substantial precision. However, the trade-off between recall and accuracy needed to be carefully considered based on the specific needs of the classification task.

TABLE III
RESULTS OF EACH MODEL EVALUATION

Model	Precision	Recall	F1 Score	Accuracy
Decision tree	92%	92%	92%	85%
Random forest	92%	92%	92%	85%
Extreme gradient boosting	92%	92%	92%	85%
LightGBM	92%	92%	92%	85%
CatBoost	92%	92%	92%	85%
Gradient boosting machines	92%	92%	92%	85%
Logistic Regression.	92%	100%	96%	92%

E. MODEL INTERPRETABILITY ANALYSIS

In the development of predictive systems based on machine learning, model interpretability has become an important aspect of understanding how input features influence prediction outcomes. Interpretability allows users, decision-makers, and system developers to understand the reasons behind each decision made by the model, thereby increasing trust and accountability. To support interpretability analysis, this research uses the SHAP approach, which is an explainable AI (XAI) method designed to illustrate the output of machine learning models both individually and aggregately. SHAP is based on the Shapley value theory from cooperative game theory, which provides a "fair" contribution of each feature to the predictive output. In the context of this research, SHAP is used to measure the contribution of each feature (such as specific courses, GPA, and internship duration) to the prediction of graduates' employment status. The mean absolute SHAP value is used to evaluate the overall importance of each feature without considering the direction of its influence

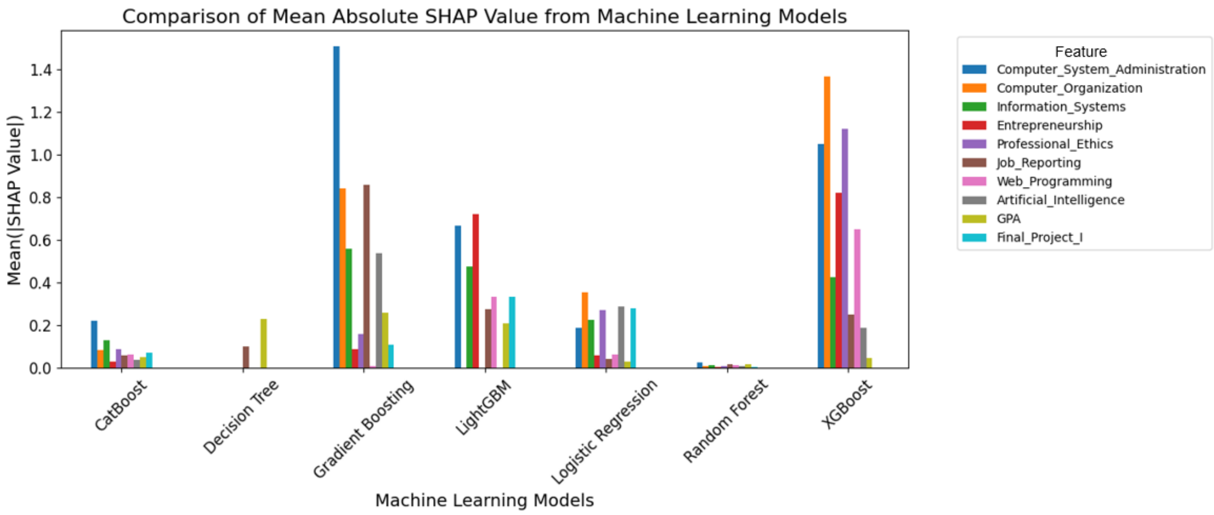

Figure 2. The mean absolute SHAP results for each model.

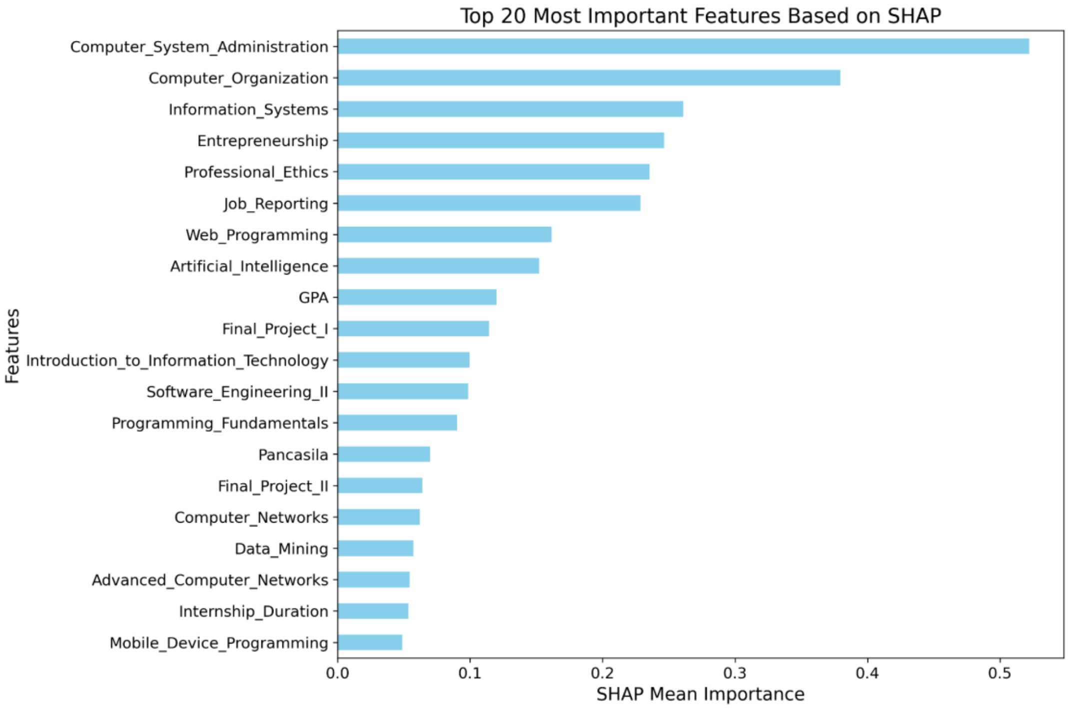

Figure 3. Most influential features.

(positive or negative). The higher the SHAP value of a feature, the greater its influence on the model's prediction outcome.

F. SHAP ANALYSIS

Model interpretability analysis was conducted by calculating the mean absolute SHAP value for each feature. This value represents the average absolute contribution of each feature to the model's prediction outcome, so features with the greatest influence are indicated by higher SHAP values. This approach allows not only the identification of the most influential features but also provides a quantitative understanding of the extent of each feature's contribution to the overall model output.

The visualization of SHAP values was created using the SHAP Python library, which was integrated with the Matplotlib and Seaborn visualization libraries. The entire analysis and visualization process was carried out in a Python 3-based Jupyter Notebook development environment. Figure 2 display the mean absolute SHAP values for each feature used in the machine learning model.

Visualization of the mean absolute SHAP values comparison among machine learning models provides a quantitative picture of the contribution of each feature to the prediction results. This value reflects the average absolute contribution of each feature regardless of the direction of its influence (positive or negative), so features with high SHAP values can be interpreted as the most influential variables in the model's decision. Figure 2 shows that Computer System Administration, Computer Organization, and Entrepreneurship features showed high mean SHAP values in several models, particularly in gradient boosting and XGBoost. This indicates that these three features consistently contribute significantly to the prediction process and can be considered the most important in determining graduates' employment status. On the other hand, other features such as Professional_Ethics or Job_Reporting appeared to have relatively low SHAP values in most models, indicating their lesser influence. With this approach, not only can important features be identified, but their magnitude of impact can also be measured quantitatively, providing a strong foundation for transparent and explainable model interpretation.

Among the analyzed features, 20 important features were obtained using mean importance, representing each feature's average absolute SHAP value and indicating its contribution to the model's predictions. Features with higher mean SHAP importance have a greater influence on the model's decision-making process. The results of the feature importance ranking can be seen in Figure 3. This figure shows that the features with the most significant influence on the model's predictions are Computer System Administration, Computer Organization, Information Systems, Entrepreneurship, Professional Ethics, Work Reporting, Web Programming, Artificial Intelligence, GPA, and Final Project I. These features have high Shapley values, indicating that their changes significantly impact the model's decisions.

In addition, other features that also had a substantial impact include Introduction to Information Technology, Software Engineering II, Basic Programming, Pancasila, Final Project II, Computer Networks, Data Mining, Advanced Computer Networks, Internship Duration, and Mobile Device Programming. Although their contribution is lower compared to the top features, these features still play an essential role in determining the model's predictions. Thus, the results of this SHAP analysis provide insights into the most influential factors in the model, as well as how each feature contributes to the prediction decision.

V. CONCLUSION

This research has evaluated the performance of various machine learning models for classification based on accuracy, precision, recall, and F1 scores. The results showed consistent and high predictive performance across multiple models, with decision tree, random forest, XGBoost, LightGBM, CatBoost, and GBM achieving identical scores of 92% for precision, 92% for recall, and 92% for F1 score, while their accuracy remained at 85%. On the other hand, logistic regression showed a higher recall (100%), resulting in an F1 score of 96% and an overall accuracy of 92%, demonstrating its effectiveness in capturing all relevant instances without missing any.

Additionally, SHAP analysis enhances the model interpretability by identifying the most influential features in the prediction process. Global SHAP analysis ranked features based on mean SHAP importance, revealing that Computer System Administration, Computer Organization, Information Systems, Entrepreneurship, Professional Ethics, Work Reporting, Web Programming, Artificial Intelligence, GPA, and Final Project I impact the model's predictions most. Lower but significant contributors include Introduction to Information Technology, Software Engineering II, Basic Programming, Pancasila, Final Project II, Computer Networks, Data Mining, Advanced Computer Networks, Internship Duration, and Mobile Device Programming. This insight provides a deeper understanding of how specific academic and professional attributes contribute to predictive outcomes.

The findings highlight the importance of integrating explainable artificial intelligence (AI) techniques, such as SHAP in predictive modelling, ensuring transparency and interpretability in decision-making. Future research can explore feature selection techniques, larger datasets, and model optimization to enhance predictive accuracy. Additionally, investigating the temporal development of academic and professional variables can provide deeper insights into predicting long-term career success. Continuous improvement of machine learning models and explanation methods will be crucial in enhancing the reliability and application of predictive analytics in education and employment.

CONFLICTS OF INTEREST

The author declares that during the research and writing of this scientific article titled "Interpretable Machine Learning for Job Placement Prediction: A SHAP-Based Feature Analysis," the author has no conflicts of interest with any party.

REFERENCES

[1] World Economic Forum, "The Future of Jobs Report," 2023. [Online]. Available: https://www.weforum.org/reports/the-future-of-jobs-report-2023
[2] M.H. Baffa, M.A. Miyim, and A.S. Dauda, "Machine learning for predicting students' employability," *UMYU Sci.*, vol. 2, no. 1, pp. 001–009, Mar. 2023, doi: 10.56919/usci.2123_001.
[3] A.A. binti Kahlik dan A.Y.S. Al-Hababi, "Predicting post-internship employability using ensemble machine learning approach," *J. Cogn. Sci. Hum. Dev.*, vol. 10, no. 2, pp. 87–101, Sep. 2024, doi: 10.33736/jcshd.7518.2024.
[4] S. Ramos-Pulido, N. Hernández-Gress, and G. Torres-Delgado, "Analysis of soft skills and job level with data science: A case for graduates of a private university," *Informatics*, vol. 10, no. 1, pp. 1–13, Mar. 2023, doi: 10.3390/informatics10010023.
[5] H. Sahlaoui *et al.,* "Predicting and interpreting student performance using ensemble models and Shapley additive explanations," *IEEE Access*, vol. 9, pp. 152688–152703, Oct. 2021, doi: 10.1109/ACCESS.2021.3124270.
[6] S. Ramos-Pulido, N. Hernández-Gress, and G. Torres-Delgado, "Exploring the relationship between career satisfaction and university learning using data science models," *Informatics*, vol. 11, no. 1, pp. 1–18, Mar. 2024, doi: 10.3390/informatics11010006.
[7] Y. Aswini, J. Jersha, S.B. Chakravarthi, and S. Aditiya B., "Predicting student placement outcomes using machine learning techniques," *Int. J. Nov. Trends Innov. (IJNTI)*, vol. 2, no. 10, pp. 63–69, Oct. 2024.
[8] M.K. Shukla *et al.,* "Students placement prediction model using logistic regression," in *Int. Conf. Innov. Adv. Technol. Eng.*, 2017, pp. 1–4.
[9] M. Kumar *et al.,* "Predicting college students' placements based on academic performance using machine learning approaches," *Int. J. Mod. Educ. Comput. Sci. (IJMECS)*, vol. 15, no. 6, pp. 1–13, Dec. 2023, doi: 10.5815/ijmecs.2023.06.01.
[10] C. Patro and I. Pan, "Decision tree-based classification model to predict student employability," in *Proc. Res. Appl. Artif. Intell.*, 2021, pp. 327–333, doi: 10.1007/978-981-16-1543-6_32.
[11] H.Q. Nguyen *et al.,* "Career path prediction using XGBoost model and students' academic results", *CTU J. Innov. Sustain. Dev.*, vol. 15, no. Special issue: ISDS, pp. 62–75, Oct. 2023, doi: 10.22144/ctujoisd.2023.036
[12] D. Mhamdi *et al.*, "Job recommendation based on recurrent neural network approach," *Procedia Comput. Sci.*, vol. 220, pp. 1039–1043, Mar. 2023, doi: 10.1016/j.procs.2023.03.145.
[13] X. Xue *et al.*, "Convolutional recurrent neural networks with a self-attention mechanism for personnel performance prediction," *Entropy*, vol. 21, no. 12, p. 1–16, Dec. 2019, doi: 10.3390/e21121227.
[14] M. Abdelaal, C. Hammacher, and H. Schöning, "REIN: A comprehensive benchmark framework for data cleaning methods in ML pipelines," in *Proc. 26th Int. Conf. Extending Database Technol. (EDBT 2023),* 2023, pp. 499–511.
[15] R. LaRose and B. Coyle, "Robust data encodings for quantum classifiers," 2020, *arXiv:2003.01695*.
[16] K. Zhang *et al.*, "Description-enhanced label embedding contrastive learning for text classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 10, pp. 14889–14902, Oct. 2024, doi: 10.1109/TNNLS.2023.3282020.
[17] M.M. Suarez-Alvarez, D.-T. Pham, M.Y. Prostov, and Y.I. Prostov, "Statistical approach to normalization of feature vectors and clustering of mixed datasets," in *Proc. R. Soc. A*, 2012, pp. 2630–2651, doi: 10.1098/rspa.2011.0704.

[18] Q.-M. Tan, “Normalization in mathematical simulations,” in *Dimensional Analysis*, Heidelberg, Germany: Springer, 2011, pp. 161–179.

[19] V.L. Miguéis, A. Freitas, P.J.V. Garcia, and A. Silva, “Early segmentation of students according to their academic performance: A predictive modelling approach,” *Decis. Support Syst.*, vol. 115, pp. 36-51, Nov. 2018, doi: 10.1016/j.dss.2018.09.001.

[20] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[21] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *KDD '16: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[22] G. Ke *et al.*, “LightGBM: A highly efficient gradient boosting decision tree,” in *NIPS'17: Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3149 – 3157.

[23] L. Prokhorenkova *et al.*, “CatBoost: Unbiased boosting with categorical features,” in *NIPS'18: Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6639 – 6649.

[24] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Front. Neurorobotics*, vol. 7, pp. 1–21, Dec. 2013, doi: 10.3389/fnbot.2013.00021.

[25] S. Sperandei, “Understanding logistic regression analysis,” *Biochem. Med.*, vol. 24, no. 1, pp. 12–18, Feb. 2014, doi: 10.11613/BM.2014.003.

[26] A.B. Parsa *et al.*, “Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis,” *Accid. Anal. Prev.*, vol. 136, pp. 1–8, Mar. 2020, doi: 10.1016/j.aap.2019.105405.

[27] K.K.P.M. Kannangara, W.-H. Zhou, Z. Ding, and Z. Hong, “Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method,” *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1052–1063, Aug. 2022, doi: 10.1016/j.jrmge.2022.01.002.

[28] M.T. Syamkalla, S. Khomsah, dan Y.S.R. Nur, “Implementasi algoritma CatBoost dan Shapley additive explanations (SHAP) dalam memprediksi popularitas game indie pada platform Steam,” *J. Teknol. Inf. Ilmu Komput.*, vol. 11, no. 4, pp. 777–786, Aug. 2024, doi: 10.25126/jtiik.1148503.

[29] H. Kamel and M.Z. Abdullah, “Distributed denial of service attacks detection for software defined networks based on evolutionary decision tree model,” *Bul. Tek. Elekt. Inform.*vol. 11, no. 4, pp. 2322–2330, Aug. 2022, doi: 10.11591/eei.v11i4.3835.

[30] M.K.M. Almansoori and M. Telek, “Anomaly detection using combination of autoencoder and isolation forest,” in *1st Workshop Intell. Infocommunication Netw. Syst. Serv. (WI2NS2)*, 2023, pp. 25–30, doi: 10.3311/wins2023-005.