I. J. Education and Management Engineering, **2025, 1, 10-24**
Published Online on February 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2025.01.02

Identifying Patterns and Trends in Campus Placement Data Using Machine Learning

Raghavendra C. K.*
Department of Computer Science and Engineering, B N M Institute of Technology, Bangalore,560070, India
E-mail: raghav.ck.clk@gmail.com
ORCID iD: https://orcid.org/0000-0002-0321-0674
*Corresponding Author

Smaran N. G.
Department of Computer Science and Engineering, B N M Institute of Technology, Bangalore, 560070, India
E-mail: smaran.ng@gmail.com
ORCID iD: https://orcid.org/0009-0004-8147-8416

Spandana A. P.
Department of Computer Science and Engineering, B N M Institute of Technology, Bangalore, 560070, India
E-mail: ap.spandana@gmail.com
ORCID iD: https://orcid.org/0009-0001-3005-8591

Vijay D.
Department of Computer Science and Engineering, B N M Institute of Technology, Bangalore, 560070, India
E-mail: vijaydragh@gmail.com
ORCID iD: https://orcid.org/0009-0007-8812-4292

Vishruth M. V.
Department of Computer Science and Engineering, B N M Institute of Technology, Bangalore, 560070, India
E-mail: vishruth.mv10@gmail.com
ORCID iD: https://orcid.org/0009-0006-8948-4145

Received: 20 April, 2024; Revised: 23 May, 2024; Accepted: 11 July, 2024; Published: 08 February, 2025

Abstract: This research delves into the utilization of machine learning algorithms to address the urgent challenge of assisting students in navigating a highly competitive job market. Recognizing the limitations of conventional methods in delivering effective guidance for securing job opportunities, there is a growing imperative to integrate advanced technology. Our model using Machine Learning (ML) algorithms offers customized solutions and emphasizes the algorithms that exhibit the highest effectiveness within this context. In the contemporary employment, achieving success extends beyond mere academic credentials, necessitating a holistic grasp of industry trends and in-demand skills. Through the application of machine learning, a fresh approach is presented, encompassing the gathering, and preprocessing of diverse data that encompasses skill proficiencies. This data forms the bedrock upon which ML algorithms operate, predicting and enhancing students' likelihood of securing favorable job placements. The proposed work focuses on the careful selection of suitable machine learning algorithms, with special attention given to classification techniques such as Linear Regression, Random Forest, Decision Tree Classifier, K-nearest neighbors Classifier, and ensembled models. By meticulous evaluation and Ensemble Technique, these algorithms unearth intricate patterns within the data, deciphering the multifaceted factors influencing job placement outcomes. By deconstructing the performance of each algorithm, the report provides valuable insights into their strengths and potential synergies.

Index Terms: Machine learning, In-demand skills, Ensemble Technique, Trends in Placement

1. Introduction

In the dynamic environment of modern education and career development, campus placements have become a key factor in determining the career paths of graduates. This project utilizes data science and machine learning to revolutionize how educational institutions prepare students for their professional journeys. By examining campus placement data, we aim to identify patterns and trends that can enhance decision-making, refine curricula, and better align student skills with the needs of the job market. Our methodology includes thorough data collection, cleaning, and organization to uncover important relationships and changes in recruitment practices. By employing machine learning models, we can forecast future placement trends, offering valuable insights to students, educators, and institutions alike. This research extends beyond mere data analysis; it aims to empower stakeholders to make informed decisions that lead to more personalized education and a closer match with industry requirements. The ultimate and final goal of this study is to build a supportive ecosystem where graduates possess not only academic expertise but also the skills most valued by employers. By pin- pointing the industries and skill sets that are in demand, we hope to contribute to a more successful and rewarding futurefor all involved.

2. Literature Review

Jeevalatha et al., 2014 study investigates the use of decision tree algorithms for predicting student placements based on academic performance. The research indicates that decision trees can achieve good accuracy in predicting placement outcomes. However, the study lacks comparison with other algorithms, which limits the overall assessment of its effectiveness.[1] Maurya et al., 2021 research explores the application of various machine learning algorithms, such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest (RF), for predicting student placements based on academic performance. The study found that RF performed the best with high accuracy. Despite these findings, there is limited exploration of feature engineering, which could enhance the predictive power of the models.[2] Sheetal & Bakare, 2016 paper combines fuzzy logic with KNN to predict student placements, considering factors like academic performance, extracurricular activities, and communication skills. The study reports improved accuracy when using this combined approach compared to KNN alone. However, the research emphasizes the need for larger datasets to validate the model's effectiveness.[3] Ahmed et al., 2018 proposes a performance-based placement prediction system utilizing historical placement data. The study achieves promising results, but the methodology lacks transparency as the specific machine learning algorithm used is not disclosed, which raises concerns about the reproducibility and reliabilityof the findings.[4]

Ishizue et al., 2018 research extends beyond academic metrics to include factors such as class attitude and psycho-logical scales for predicting placement success. The study employs machine learning models, but the specific algorithms and their accuracy are not reported, necessitating a detailed performance evaluation to validate the approach.[5] Manikandan et al., 2018 paper uses data mining techniques to predict campus placement performance based on academic and co-curricular activities. While the study reports satisfactory accuracy, it lacks specific details about the algorithms used, making it difficult to assess the model's robustness and applicability.[6] Rathore & Jayanthi, 2017 employs a fuzzy infer- ence system (FIS) to predict student performance in placement training. The study achieves good accuracy but does not compare the FIS with other machine learning approaches, which limits the understanding of its relative effectiveness.[7] Patel & Tamrakar, 2017 study explores the use of data mining techniques for campus placement prediction in higher education. However, the research does not provide details on the specific algorithms used or their accuracy, which leavesthe effectiveness of the proposed approach unclear.[8]

Goyal & Sharma, 2017 paper presents a decision support system for predicting student placements using data min-ing techniques. Similar to Patel & Tamrakar (2017), this study does not specify the algorithms or provide performance metrics, which reduces the transparency and usefulness of the findings.[9] Surya et al., 2022 research employs supervised machine learning algorithms, including Logistic Regression (LR) and Decision Trees (DT), for predicting student place-ments. The study reports good accuracy, but the specific algorithms and their comparative performance are not mentioned, highlighting a need for in-depth analysis and feature engineering.[10] Nagamani et al., 2020 focuses on using supervised machine learning algorithms to predict and analyze student placements with the goal of improving educational standards. The study achieves reasonable accuracy, but it does not disclose the specific algorithms used, which limits the clarity of the research.[11] Thakar & Mehta, 2017 research combines clustering and classification techniques to enhance the accuracy of student placement predictions. While the study reports improved performance, it lacks detailed information aboutthe specific algorithms used, making it challenging to evaluate the model's overall effectiveness.[12]

Casuat & Festijo, 2019 explores machine learning approaches for predicting student employability. The study reports promising results, but the lack of details regarding the specific algorithms and accuracy metrics limits the comprehensive- ness of the research.[13] Bai & Hira, 2021 propose a hybrid deep belief network model to predict student employability. The study reports improved accuracy compared to traditional machine learning models, but further comparisons with other deep learning approaches are needed to fully assess the model's performance.[14]

Saidani et al., 2022 study focuses on predicting student employability based on internship experiences using gradientboosting models. The research achieves good accuracy and provides valuable insights, but it would benefit from a more extensive exploration of feature importance.[15] Hariharan et al., 2022 paper utilizes various machine learning techniques to predict student placements. While the study reports promising results, it lacks detailed comparisons between the different algorithms used, which would provide a better understanding of the most effective approaches.[16] Manvitha & Swaroopa, 2019 research explores the use of supervised machine learning techniques for predicting student placements. The study achieves reasonable accuracy, but it lacks a thorough analysis and comparison of the algorithms employed, which would strengthen the findings.[17]

N. Kumar et al., 2020 focuses on predicting campus placements using machine learning techniques. The study reports reasonable accuracy but lacks detailed analysis and comparisons between different algorithms, which would provide a more comprehensive understanding of the model's effectiveness.[18] Bhoite et al., 2023 employ probabilistic machine learning models for placement prediction. The study achieves promising results, but it lacks detailed comparisons with other machine learning approaches, which would strengthen the conclusions drawn.[19] Basha et al., 2023 research com- bines exploratory analysis with machine learning techniques to predict student placements. The study reports good ac- curacy but could benefit from a more thorough exploration of feature importance to enhance the interpretability of the results.[20]

Table 1. Summary of Papers of Literature Survey

Paper author name	Undisclosed	Accuracy	Research Gap
Jeevalatha et al., 2014	ML	Good	Lack of comparison with other algorithms
Maurya et al., 2021	Data Mining	High	Limited exploration of feature engineering
Sheetal & Bakare, 2016	Fuzzy Inference System	Improved	Requires larger datasets for validation
Ahmed et al., 2018	Data Mining	Promising	Lack of transparency in methodology
Ishizue et al., 2018	Data Mining	Not reported	Needs detailed performance evaluation
Manikandan et al., 2018	LR, DT	Satisfactory	Lack of specific algorithm and accuracy details
Rathore & Jayanthi, 2017	Supervised ML	Good	Limited comparison with other ML techniques
Patel & Tamrakar, 2017	Clustering & Classification	Not reported	Lack of specific algorithm and accuracy details
Goyal & Sharma, 2017	ML	Not reported	Lack of detailed performance evaluation
Surya et al., 2022	Hybrid Deep Belief Network	Good	Requires in-depth analysis and feature engineering
Nagamani et al., 2020	Gradient Boosting	Reasonable	Lack of specific algorithm details
Thakar & Mehta, 2017	Various ML	Improved	Requires larger datasets for validation
Casuat & Festijo, 2019	Supervised ML	Promising	Needs detailed performance evaluation
Bai & Hira, 2021	ML	Improved	Requires comparison with other deep learning models
Saidani et al., 2022	Survey	Good	Limited exploration of feature importance
Hariharan et al., 2022	Modified Model	Promising	Lack of detailed comparisons
Manvitha & Swaroopa, 2019	Various ML	Reasonable	Requires in-depth analysis and feature engineering
N. Kumar et al., 2020	Various ML	Reasonable	Lack of detailed analysis and comparisons
Bhoite et al., 2023	Data Mining	Promising	Lack of detailed comparisons
Basha et al., 2023	Fuzzy Inference System	Good	Limited exploration of feature importance

3. Methodology

The research was diligently undertaken through a framework of testing and meticulous data processing. This approach aimed to ensure the integrity and robustness of the developed solution. The project's scope was methodically divided into distinct and coherent modules, each serving a specific purpose within the broader context. The following modules encapsulated the project's sequential progression:

- Data description
- Multi-model accuracy determination and model selection
- Data processing
- User interface

3.1. Data description

The dataset used in this study is derived from Kaggle, a prominent platform for public datasets, and consists of 2982 instances with 10 distinct attributes. These attributes reflect various aspects of the interview process, such as ’GENERAL APPEARANCE’, ’MANNER OF SPEAKING’, ’PHYSICAL CONDITION’, ’MENTAL ALERTNESS’, ’SELF-CONFIDENCE’, ’ABILITY TO PRESENT IDEAS’, ’COMMUNICATION SKILLS’, and ’STUDENT PER-FORMANCE RATING’. The ’CLASS’ attributes serves as an indicator of employability, providing a label that categorizes students based on their potential to secure employment. Together, these attributes offer a holistic view of the factors thatinfluence students’ performance during interviews.

Before the dataset could be analyzed, it was necessary to perform a comprehensive preprocessing process. The first step was to address any missing values, which can cause biases and inaccuracies in the results if not properly managed. For numerical attributes, techniques like mean or median imputation were applied, where missing values were replaced by the mean or median of that specific attribute’s existing values. This approach helps maintain the overall distribution of the data. For categorical attributes, mode imputation was used, replacing missing values with the most common category within the attribute, thus preserving the original distribution. The next critical step was the identification and treatment of outliers. Outliers can distort the results of an analysis, especially in algorithms that are sensitive to extreme values. Statistical methods such as Z-score analysis, which calculates how far a data point is from the mean in terms of standard deviations, and Interquartile Range (IQR) filtering, which identifies outliers based on their distance from the first and third quartiles, were employed. Depending on the severity and impact of the outliers, they were either adjusted to fall within an acceptable range or removed entirely.

Data normalization was also a vital part of preprocessing, given the diverse scales of the attributes in the dataset. For example, while some attributes might range from 1 to 10, others could be binary or categorical. To ensure that no single attribute dominated the analysis, normalization techniques were applied. Min-Max scaling, which adjusts the data to a specific range (typically 0 to 1), was used to standardize the attributes. Additionally, Z-score normalization, which converts data into a distribution with a mean of 0 and a standard deviation of 1, was utilized where appropriate. This step is particularly crucial for machine learning models that rely on distance calculations, such as k-nearest neighbors andsupport vector machines.

In summary, these preprocessing steps—managing missing values, addressing outliers, and normalizing the data—were essential to ensure the dataset’s quality and reliability. By carefully preparing the data, the study set the foundation for accurate and insightful analyses, leading to more reliable predictions and conclusions.

	GENERAL APPEARANCE	MANNER OF SPEAKING	PHYSICAL CONDITION	MENTAL ALERTNESS	SELF-CONFIDENCE	ABILITY TO PRESENT IDEAS	COMMUNICATION SKILLS	Student Performance Rating
count	2982.000000	2982.000000	2982.000000	2982.000000	2982.000000	2982.000000	2982.000000	2982.000000
mean	4.246814	3.884641	3.972166	3.962777	3.910798	3.813883	3.525486	4.610664
std	0.678501	0.757013	0.744135	0.781982	0.807602	0.739390	0.743881	0.692845
min	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	3.000000
25%	4.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	3.000000	5.000000
75%	5.000000	4.000000	5.000000	5.000000	5.000000	4.000000	4.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Fig. 1. Description of dataset.

3.2. Multi-model accuracy determination and model selection

Algorithm Selection is a critical phase where the most suitable ML algorithms, such as Logistic Regression, Gaussian Naive Bayes, Random Forest, Support Vector Machine, and K-Nearest Neighbors, are considered based on the unique characteristics and requirements of the data. This step ensures that the chosen algorithms align with the specific nuancesof predicting college students’ placements using academic performance data.

- Model training: Once the algorithms are selected, Model Training ensues, involving the use of preprocessed data. The class variable serves as the output, guiding the algorithm to learn and establish connections between the input variables and the target class variable. This training phase equips the models to comprehend the underlying patternsin the data, enhancing their predictive capabilities.
- Model Tuning: Model Tuning becomes pivotal for optimizing performance. Fine-tuning hyperparameters allows for adjustments that can significantly enhance the model’s accuracy. This process involves systematically exploringvarious configurations to achieve the best possible predictive outcomes.
- The subsequent Model Evaluation phase gauges the model’s accuracy by validating it on a distinct dataset from the one used for training. Rigorous evaluation techniques ensure a comprehensive understanding of how well the model generalizes to unseen data, providing insights into its practical effectiveness in predicting student placements basedon academic achievements.

- Finally, Prediction utilizes the trained and optimized model to make future projections about college students' placements for the upcoming academic year. These predictions, rooted in the input variables reflecting academic performance, are crucial for anticipating trends and making informed decisions.

The following machine learning algorithms were considered

1. RandomForestClassifier Explanation: Random forests are an ensemble learning method that constructs a multitude of decision trees during training. The trees are trained independently using random subsets of the data, which introduces diversity in their predictions. The final prediction for a given input is based on the majority vote from all the individual trees. Reason for Selection: Random forests are known for their robustness and ability to handle large datasets with higher dimensionality. They reduce the risk of overfitting by averaging multiple decision trees, which is particularly useful in this context where the goal is to predict student employability based on various attributes.
2. KNeighborsClassifier Explanation: The K-Nearest Neighbors (KNN) algorithm is a simple yet effective instance- based learning method. For a given query point KNN identifies the k nearest points in the training dataset based on a distance metric, typically Euclidean distance. The predicted class $\hat{y}$ is determined by the majority class among these neighbors. Reason for Selection: KNN is intuitive and performs well in scenarios where the decision boundary between classes is non-linear. It was selected to provide a different perspective compared to decision trees and to capitalize on its strength in scenarios with clear class separations.
3. LogisticRegressionCV Explanation: Logistic Regression models the probability of a binary outcome by applying the logistic function to a linear combination of input features. In this case, cross-validation (CV) is used to automat- ically select the regularization parameter, which helps prevent overfitting. Reason for Selection: Logistic regression is a straightforward and interpretable model, making it useful for binary classification problems like predicting em- ployability. It was chosen for its simplicity and the ability to provide probabilities for class membership, which is useful for understanding model confidence.
4. CategoricalNB Explanation: Categorical Naive Bayes applies Bayes' theorem with the assumption that features are conditionally independent given the class label. This algorithm is particularly suited for categorical input features. Reason for Selection: Given that some attributes may be categorical, CategoricalNB is appropriate for this type of data. It is computationally efficient and provides a good baseline for comparison with more complex models.
5. SGDClassifier Explanation: Stochastic Gradient Descent (SGD) is an optimization method that iteratively adjusts model weights based on the gradient of the loss function. The classifier uses a linear combination of features, and SGD allows for the efficient handling of large datasets. Reason for Selection: SGD is chosen for its efficiency in training linear models, especially when working with large-scale data. It also supports a variety of loss functions and penalties, providing flexibility in model design.
6. BernoulliNB Explanation: Bernoulli Naive Bayes is similar to CategoricalNB but is tailored for binary/Boolean features. It assumes each feature follows a Bernoulli distribution and applies Bayes' theorem accordingly. Reason for Selection: BernoulliNB is effective when the features are binary, which can be the case for some derived or transformed attributes in this dataset. It's a good fit for datasets with sparse binary features.
7. DecisionTreeClassifier Explanation: Decision trees split the data into subsets based on feature values, creating a tree-like model of decisions. Each node represents a decision rule, and each leaf node represents a class label. Reason for Selection: Decision trees are highly interpretable and capable of capturing non-linear relationships between features. They were chosen for their ability to easily handle both numerical and categorical data without requiring extensive preprocessing.
8. NuSVC Explanation: Nu-Support Vector Classification (NuSVC) is a variant of Support Vector Machines (SVM) that allows for greater flexibility in controlling the number of support vectors and the margin of separation between classes. It works by finding the hyperplane that best separates the classes while minimizing the number of support vectors. Reason for Selection: SVMs are powerful classifiers, particularly in high-dimensional spaces. NuSVC was chosen for its ability to manage the trade-off between maximizing the margin and minimizing the support vectors, which is useful in complex classification tasks like this one.
9. MultinomialNB Explanation: Multinomial Naive Bayes is an extension of the Naive Bayes algorithm that assumes features follow a multinomial distribution. It is typically used for text classification but can be applied to any scenario where features are counts or frequencies. Reason for Selection: While primarily used in text classification, MultinomialNB was selected to explore its performance on this dataset, particularly if any features represent counts or categorical frequencies.
10. Perceptron Explanation: The Perceptron is a simple type of neural network that makes a binary decision by com- puting a weighted sum of input features and applying a threshold function. Reason for Selection: The Perceptron serves as a fundamental building block for neural networks and was chosen for its simplicity and efficiency in linear classification tasks. It provides a baseline comparison to more complex models.
11. DecisionTreeRegressor Explanation: Decision Tree Regressors are similar to their classification counterparts but are used for predicting continuous outcomes. They split the data based on feature values and predict the

target variable by averaging the target values in the leaf nodes. Reason for Selection: Although not typically used for classification tasks, Decision Tree Regressor was included to explore its potential for predicting numerical aspects of employability that might be derived from classification outputs.

12. VotingClassifier (Ensemble - RandomForest, DecisionTree, KNeighbors) Explanation: The VotingClassifier is an ensemble model that combines predictions from multiple base classifiers (in this case, RandomForestClassifier, DecisionTreeClassifier, and KNeighborsClassifier). The final prediction is determined by majority voting. Reason for Selection: The ensemble approach leverages the strengths of multiple algorithms, reducing the likelihood of overfitting and increasing overall model accuracy. This particular combination was selected for its diversity in base classifiers, ensuring a robust final model.
13. BaggingClassifier (Ensemble - RandomForest) Explanation: Bagging (Bootstrap Aggregating) is an ensemble tech- nique that trains multiple instances of the same classifier (here, RandomForestClassifier) on different subsets of the training data. The final prediction is an aggregate of all individual predictions. Reason for Selection: Bagging was chosen for its ability to reduce variance and prevent overfitting, especially when using models like Random Forest that are already strong on their own. It improves generalization and model stability.
14. VotingClassifier (Ensemble - GradientBoosting, AdaBoost) Explanation: Another ensemble VotingClassifier com- bines Gradient Boosting and AdaBoost, two powerful boosting algorithms. Boosting methods iteratively train models to correct errors made by previous models, thereby enhancing performance. Reason for Selection: Boost- ing algorithms are known for their high accuracy and robustness, particularly in handling complex datasets. This ensemble was selected to combine the strengths of both boosting methods, providing a powerful predictive model.

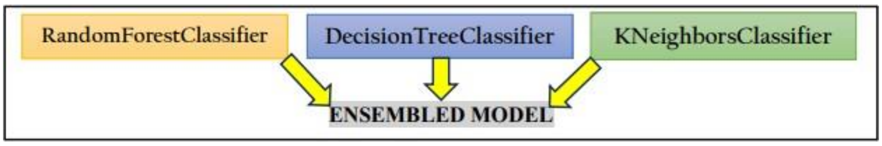

Fig. 2. Ensembled model used in the prediction of employability of a candidate.

3.3. Data processing

The analysis focused on a student cohort participating in a two-round interview process, aimed at gauging employabil- ity. The base model was considered as shown in Fig 3. In which X students appeared for the interview, X1 students were considered to have not qualified for ROUND 2, and X2 students did not qualify in ROUND 2, finally only Y students tohave successfully passed both the rounds.

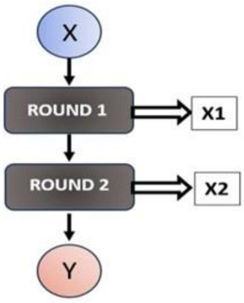

Fig. 3. Representation of the base model.

The initial round, denoted as ROUND 1, involved an assessment of several key skills, including ’GENERAL AP- PEARANCE’, ’MENTAL ALERTNESS’, ’PHYSICAL CONDITION’, and ’SELF-CONFIDENCE’. For the subsequent analysis, exclusively these pertinent attributes were selected, reflecting a judicious feature subset. To establish robust model performance, the dataset was meticulously pre-processed. The data underwent a split into training and testing sub- sets, with a test size parameter set at 0.2. The training phase involved an ensemble model, capitalizing on its amalgamation of diverse learning paradigms. The model was then subjected to testing, generating predictive outcomes for individual stu- dents. These prognostic outputs were systematically categorized into two distinct data frames: one encompassing student deemed employable and the other encapsulating those perceived as less employable.

A pivotal facet of the analysis encompassed a granular exploration of the less employable student cohort. Specifically, an intricate visualization was crafted, juxtaposing the skill proficiency of these students against their corresponding av- erage scores. This graphical representation facilitated nuanced insights into skill-specific performance trends, shedding light on potential areas for improvement. Subsequently, the cadre of students identified as employable progressed to ROUND 2, mirroring the iterative evaluation process. Paralleling the ROUND 1 methodology, the employable students underwent model testing in the context of ROUND 2 with ’STUDENT PERFORMANCE RATING’, ’ABILITY TO PRESENT IDEAS’, ’MANNER OF SPEAKING’, ’COMMUNICATION SKILLS’ considerations. Ultimately, a refined set of employability predictions emerged, crystallizing each student’s vocational prospects.

In summation, the study’s technical narrative underscores a meticulous curation of data preprocessing, ensemble model training, recursive assessment rounds, and skill-centric visualizations. This multifaceted approach delves into the interplay between distinct skill dimensions and employability prognosis, exemplifying a robust and technically rigorous methodology.

3.4. User Interface

The User Interface was developed using Streamlit, with its Python-based framework module to create a visual representation of datasets that were analyzed by using various machine learning algorithms, we have also provided tabs for inputting the placement scores for the model to predict and display the required output. The interface has been designed keeping aesthetic approach as the key element. The introduction page provides the working of the machine learning model which has been trained and tested using a pre-determined data set. The next tab allows user to upload the data of the placement which will be used to calculate the outcome of the targeted class.

4. Results

This research was undertaken to leverage Machine Learning techniques to analyze campus placement data, aiming to unveil underlying patterns and trends that could elucidate the determinants of placement outcomes. To gain a comprehen- sive perspective, the cumulative values of individual skills were aggregated to provide an overarching view of the datasetas shown in Fig 4.

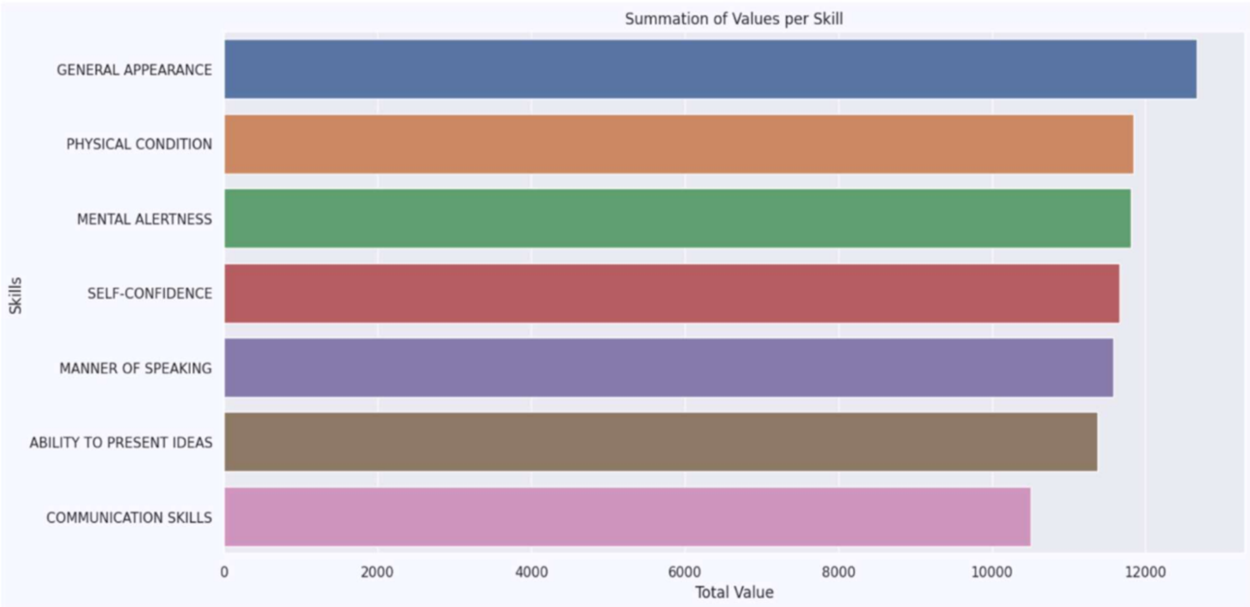

Fig. 4. Summation of values per skills.

To discern the significance of each skill set in influencing students’ employability, a pie chart was generated, as depicted in Fig 5. This visualization depicted the percentage distribution of each skill set for an average employable student. Notably, it became evident that all skills contributed equally to students’ employability. Moreover, the project encom passed the creation of visualizations to highlight students’ performance in different skill categories. Fig 6. showcased the average scores of students across various skill sets. The analysis unveiled commendable performance in terms of general appearance yet indicated a need for improvement in communication skills and overall performance. To delve into interrelationships between different categories, the correlation was explored through (the Seaborn library in Python) correlationmatrix computation. Refer Fig 7. for the same.

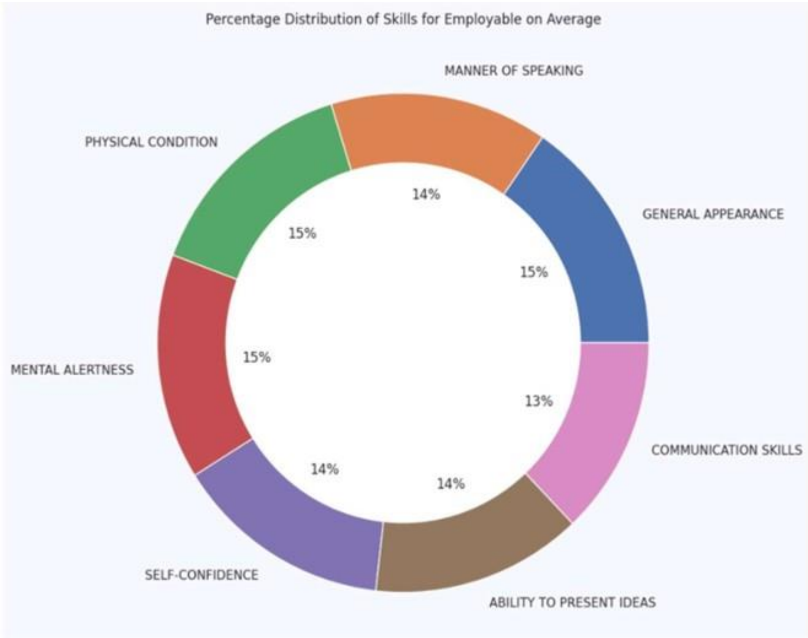

Fig. 5. Percentage distribution of each skill set for an average employable student's.

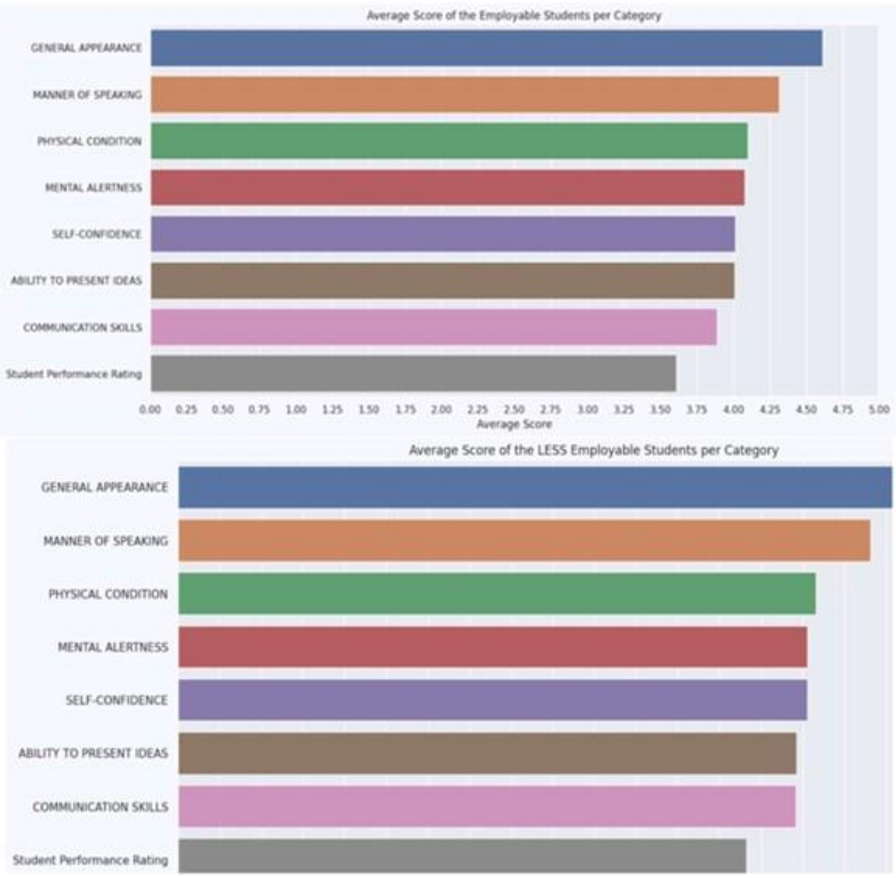

Fig. 6. Average score of employable and less employable students per category.

Subsequently, armed with a deeper understanding of the data, the team proceeded to train diverse Machine Learning models and evaluate their accuracy. The amalgamated model, incorporating RandomForestClassifier, DecisionTreeClassifier, and KNeighborsClassifier, emerged as the top performer with the highest accuracy. This ensemble model, utilizing a VotingClassifier approach, yielded an accuracy of 91.12 percentage. This remarkable outcome underscored a harmonious balance between precision and recall. The model's performance and corresponding accuracy were visually depicted in Fig 8. After preprocessing the data and engineering relevant features, we trained and evaluated a range of models, including Random Forest, Gradient Boosting, Support Vector Machines, and 13 other ML models. Table 2. shows the accuracy ofeach model.

In continuation, the dataset was segregated based on rounds and subjected to individual training. This process resulted in an accuracy of 75.04 percentage for ROUND 1. The model's predictive capabilities were further harnessed to identifythe skill deficiencies of students who didn't qualify in ROUND 1, as illustrated in Fig 9.

Furthermore, students predicted as employable in ROUND 1 were evaluated for employability in ROUND 2, yielding similar predictions. Fig 10. illustrates the average scores per skill category for the less employable students, shedding light on areas where improvements were needed. Finally, the project culminated in an analysis of students who secured jobs, as depicted in Fig 11. This analysis spotlighted the exceptional performance of these students, particularly in presenting their ideas.

Table 2. Multiple models and its accuracy

Sl. No.	Model	Mean square error	Accuracy
1.	RandomForestClassifier	0.09045226130653267	0.9095477386934674
2.	KNeighborsClassifier	0.10552763819095477	0.8944723618090452
3.	LogisticRegressionCV	0.4053601340033501	0.5946398659966499
4.	CategoricalNB	0.4120603015075377	0.5879396984924623
5.	SGDClassifier	0.4120603015075377	0.5879396984924623
6.	BernoulliNB	0.4438860971524288	0.5561139028475712
7.	DecisionTreeClassifier	0.09045226130653267	0.9095477386934674
8.	NuSVC	0.1490787269681742	0.8509212730318257
9.	MultinomialNB	0.4438860971524288	0.5561139028475712
10.	Perceptron	0.5561139028475712	0.4438860971524288
11.	DecisionTreeRegressor	0.0532036362699104	-
12.	VotingClassifier (Ensembled- Random-ForestClassifier, DecisionTreeClassi-fier,KNeighborsClassifier)	-	0.9112227805695142
13.	BaggingClassifier (Ran-domForestClassifier)	-	0.9095477386934674
14.	VotingClassifier (Gradi-entBoostingClassifier, AdaBoostClassifier)	-	0.8174204355108877

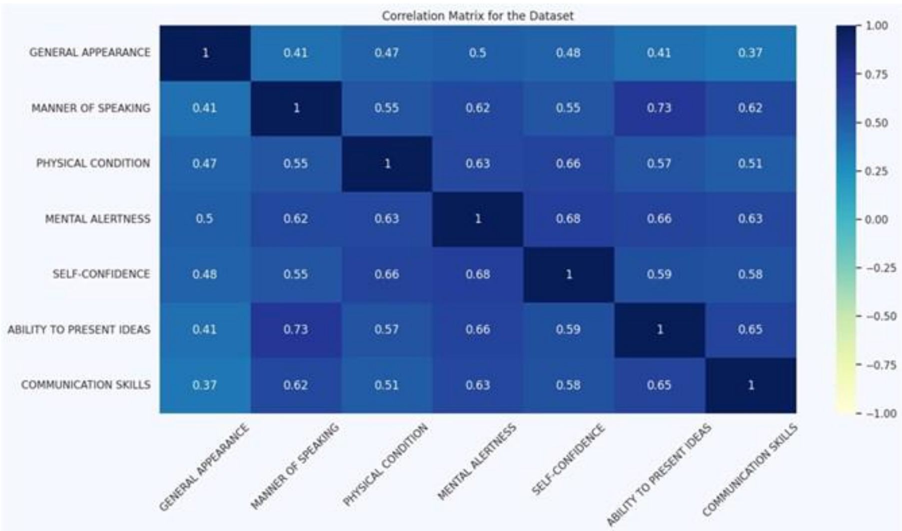

Fig. 7. Correlation of each parameters.

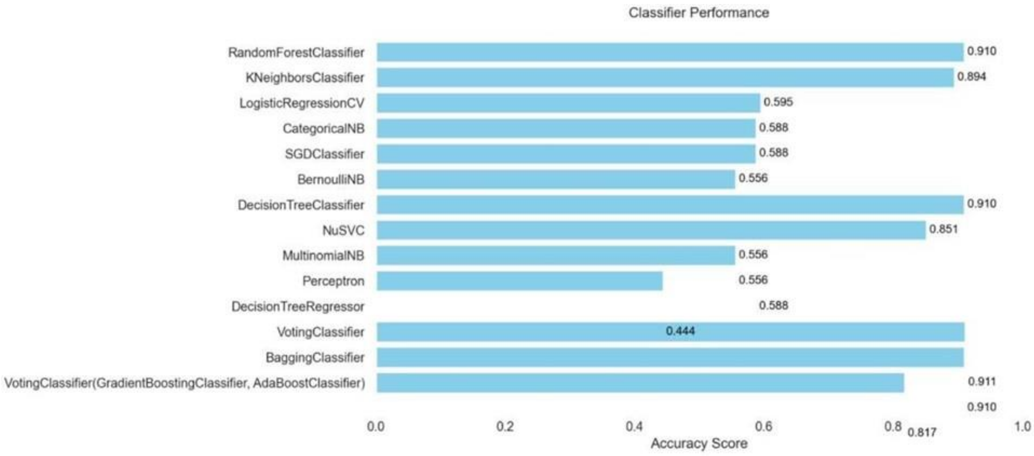

Fig. 8. Model v/s Accuracy graph.

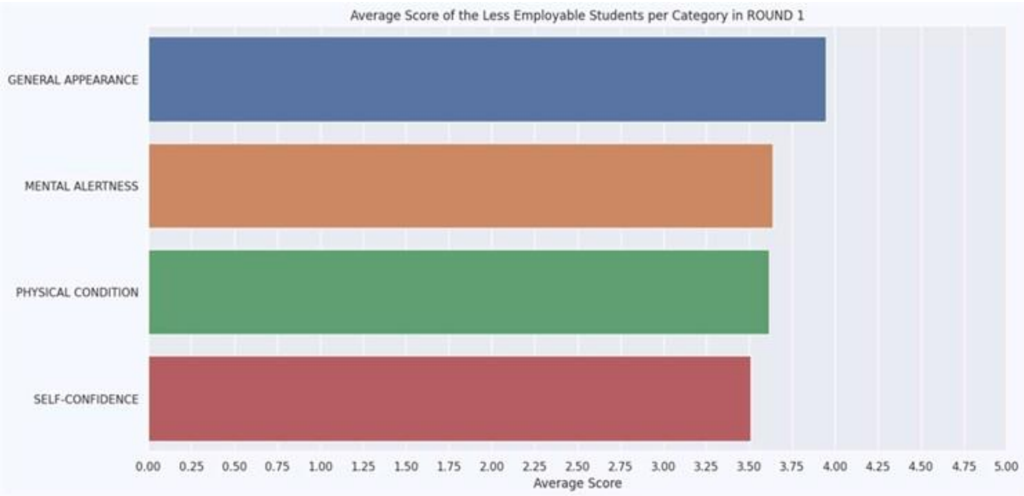

Fig. 9. Average score of less employable students per skill set in ROUND 1.

Fig 12., Fig 13., Fig 14., Fig 15. Show how the model is implemented on a web interface. In essence, the research traversed a comprehensive analytical journey, encompassing skill assessment, correlation analysis, model training, and prediction evaluations, all contributing to a holistic understanding of the factors shaping students' employability outcomes.

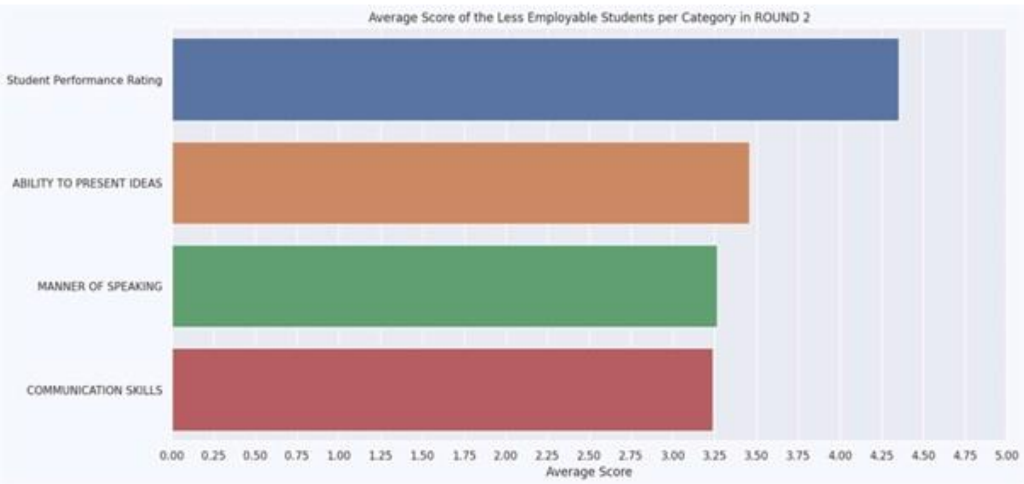

Fig. 10. Average score per category of the less employable students in ROUND 2.

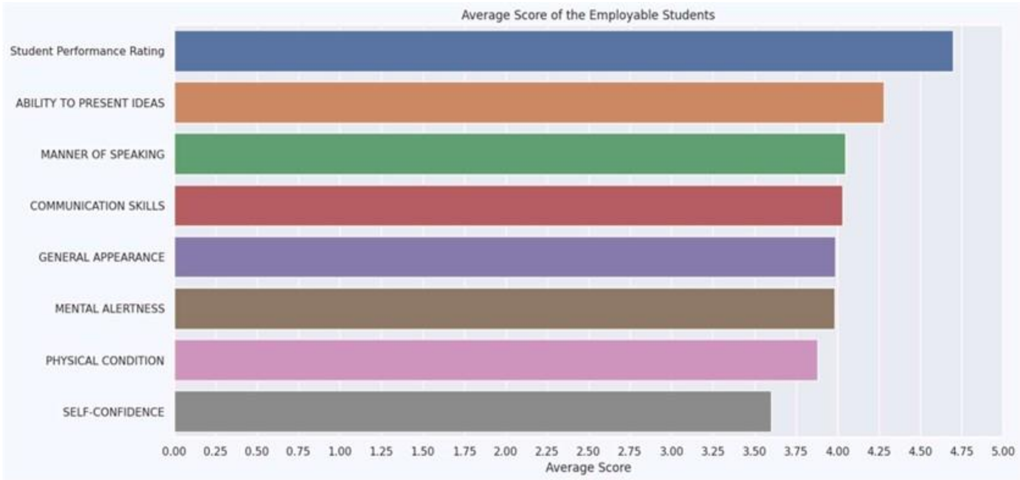

Fig. 11. Average score per category of the employable students.

Fig. 12. Summation of values per skills represented in the web interface.

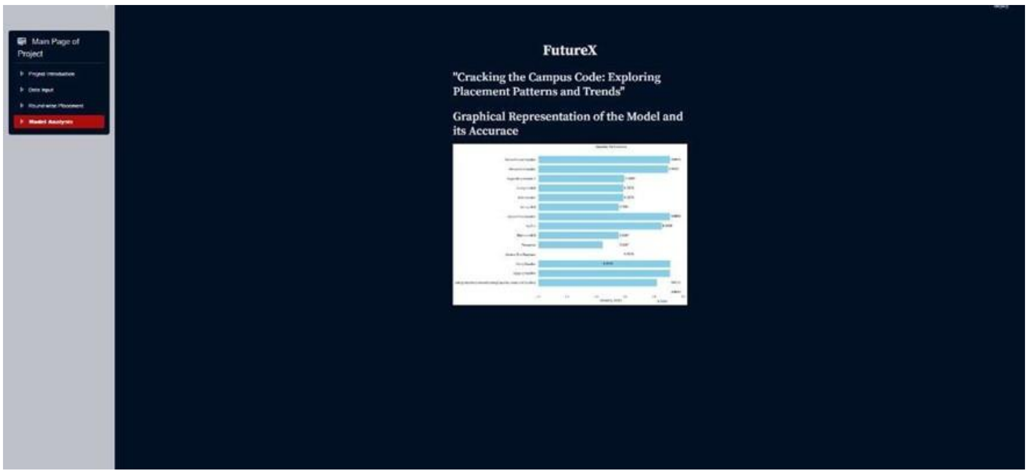

Fig. 13. Graphical representation of model used v/s its accuracy in the web interface.

Fig. 14. Input values web interface for placement prediction.

Fig. 15. Graphical representation of the features affecting employability of students in the web interface.

5. Practical Applications of the Proposed Model

The findings of this research have the potential to significantly improve campus placement processes within educational institutions. By applying the insights derived from data analysis and machine learning, institutions can implement strategies that enhance student outcomes, better align academic programs with industry needs, and streamline their placement procedures. Here is how these insights can be practically applied: This research can pinpoint which skills are most sought after by employers, enabling institutions to update their curricula to address these gaps. For example, if students with strong communication or specific technical skills are shown to be more employable, the institution can enhance focus on these areas across all courses. Understanding which fields are in high demand allows institutions to develop specialized tracks or concentrations that equip students with relevant, marketable skills. For instance, if fields like data analytics or cybersecurity are in high demand, tailored programs can be introduced to prepare students for these sectors.

The predictive models can be used to offer personalized career advice, assessing individual students' strengths and areas for improvement. Career counselors can then recommend specific courses, internships, or extracurricular activities that will enhance the students' employability. By identifying students who may struggle to find employment, institutions can offer additional support, such as mentoring or access to specialized career services, to improve these students' job prospects. By understanding which industries or companies are most likely to recruit graduates, the institution's placement cell can build stronger partnerships with these employers, organize focused recruitment drives, and prepare students for the specific needs of these industries. Insights from the research can guide the design of more effective placement preparation programs. For instance, if self-confidence or presentation skills are shown to correlate with higher placement rates, the institution can emphasize these areas in workshops or training sessions.

The institution can use research insights to establish feedback mechanisms, allowing them to monitor and refine their educational and placement strategies over time. This approach ensures that their methods remain relevant and effective in the changing job market. Institutions can use the findings to establish benchmarks for student performance and employability. By tracking these benchmarks, they can assess the effectiveness of their curricula and placement strategies and make necessary adjustments. Institutions can use data-driven success stories to enhance their reputation among prospective students and employers. These stories can be featured in marketing materials to attract new students and strengthen industry partnerships. By demonstrating a clear understanding of industry needs and a commitment to student employability, institutions can attract collaborations with leading companies, secure funding, and create more opportunities for student internships and placements.

Predictive models can provide insights into future employment trends, allowing institutions to anticipate job market changes and adjust their strategic plans. This might involve developing new programs in emerging fields or expanding industry partnerships. By identifying the factors that most influence employability, institutions can allocate resources more effectively, investing in programs or services that have the greatest impact on student success. Educational institutions can greatly benefit from applying these research findings to improve their campus placement processes. By integrating data- driven insights into their strategies, institutions can better prepare students for successful careers, enhance their academicofferings, and build stronger connections with the job market.

6. Future Research Directions

The dataset used in this study is limited to a specific institution or set of institutions, which may restrict the generalizability of the findings. Future research should consider expanding the dataset to include a wider variety of educational institutions from different regions, academic disciplines, and student backgrounds. This broader dataset could reveal more general patterns and trends, as well as institution-specific or regional differences in employability factors. While this study focused on key interview-related attributes, employability is likely influenced by a broader set of factors, including participation in extracurricular activities, work experience, social skills, and personal attributes. Future studies should incorporate these additional variables to gain a more comprehensive understanding of what drives successful placement outcomes, leading to more holistic models that better reflect real-world employability. Machine learning models can sometimes perpetuate biases found in the training data, potentially leading to unfair outcomes for certain groups of students. To combat this, future research should prioritize identifying and mitigating biases within the models. This could involve using fairness-aware algorithms or applying post-processing techniques that ensure equitable predictions across diverse student populations. The current research primarily focuses on predicting students' immediate employability after graduation. However, to fully understand the impact of education on career success, future research should consider a longitudinal approach. Tracking graduates over several years would provide insights into long-term career progression, stability, and satisfaction, offering a more comprehensive view of the factors that contribute to sustained employability.

While this study suggests several ways educational institutions could utilize the findings, these strategies have yet to be tested in real-world settings. Future research should focus on implementing these recommendations in actual educational environments and assessing their effectiveness. This would involve piloting curriculum changes, enhanced career advising, and other interventions to measure their impact on student outcomes. Although the study explored various machine learning models, there are many advanced techniques, such as deep learning or ensemble methods, that remain untested. Future research could explore these more sophisticated models to determine if they provide better predictive accuracy or additional insights. Additionally, examining the interpretability of these models will be essential to ensure that they are both powerful and understandable. The job market is constantly evolving, with new industries and roles emerging over time. Future research should analyze how these industry trends influence employability and skill requirements. By combining labor market data with student placement information, researchers could develop models that not only predict current employability but also anticipate future job market demands, helping institutions better prepare students for evolving career opportunities.

As machine learning models become more complex, it is important to ensure that their predictions are understandable and transparent. Future research should focus on developing techniques that enhance the interpretability of these models, allowing educators and students to comprehend the factors driving predictions. This could involve using interpretability tools or creating simpler models that balance accuracy with transparency. While this study provides a solid foundation for understanding and improving student employability through data-driven approaches, there is much room for further research. By addressing the limitations and exploring new research avenues, future studies can contribute to a deeper, more nuanced understanding of the factors that influence employability, ultimately benefiting educational institutions andtheir students.

7. Conclusion

In summary, the incorporation of machine learning algorithms into the realm of career guidance has yielded promising outcomes, ushering in a new era of personalized strategies for navigating the intricate job market landscape. Our efforts to harness these algorithms' predictive capabilities have yielded a model that not only enhances job placement predictions but also tailor's recommendations to Placement Officers.Through rigorous experimentation, data preprocessing, and algorithm selection, we've identified an effective model for addressing the challenges of the competitive job market. While this achievement is noteworthy, the journey continues. To maintain relevance in a rapidly evolving landscape, ongoing model refinement, ethical considerations, and bias mitigation are crucial. Ultimately, the successful application of machine learning algorithms in career guidance holds the potential to empower students and Placement Officers with insights and guidance that can significantly influence their path toward successful job placements. The combination of Machine Learning and Python has led to accurate predictions of placement trends in an institution. Students and Placement Officers can use these predictions to make informed decisions about their future career paths.

References

[1] Jeevalatha, T., Ananthi, N., & Kumar, D. S. (2014). Performance analysis of undergraduate student's placement selection using decision tree algorithms. *International Journal of Computer Applications*, 108(15).

[2] Maurya, L. S., Hussain, M. S., & Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. *Applied Artificial Intelligence*, 35(6), 403-420.

[3] Sheetal, M., & Bakare, S. (2016). Prediction of campus placement using data mining algorithm-fuzzy logic and k nearest neighbor. *IJARCCE*, 5(6), 309-312.

[4] Ahmed, S., Zade, A., Gore, S., Gaikwad, P., & Kolhal, M. (2018). Performance Based Placement Prediction System. IJARIIE-ISSN (O), 4(3), 2395-4396.

[5] Ishizue, R., Sakamoto, K., Washizaki, H., & Fukazawa, Y. (2018). Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics. *Research and Practice in Technology Enhanced Learning*, 13, 1-20.

[6] Manikandan, K., Sivakumar, S., & Ashokvel, M. (2018). A Classification Model for Predicting Campus Placement performance Class using Data Mining Technique. *International Journal of Advance Research in Science and Engineering*, 7(6).

[7] Rathore, R. K., & Jayanthi, J. (2017). Student prediction system for placement training using fuzzy inference system. *ICTACT Journal on Soft Computing*, 7(3), 1443-1446.

[8] Patel, T., & Tamrakar, A. (2017). A data mining technique for campus placement prediction in higher education. *Indian J. Sci. Res*, 14(2).

[9] Goyal, J., & Sharma, S. (2017). Placement Prediction Decision Support System using Data Mining. *International Journal of Engineering and Techniques*, 4(2).

[10] Surya, M. S., Kumar, M. S., & Gandhi Mathi, D. (2022). Student Placement Prediction Using Supervised Machine Learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 1352-1355. IEEE.

[11] Nagamani, S., Reddy, K. M., Bhargavi, U., & Kumar, S. R. (2020). Student placement analysis and prediction for improving the education standards by using supervised machine learning algorithms. *J. Crit. Rev*, 7(14), 854-864.

[12] Thakar, P., & Mehta, A. (2017). A unified model of clustering and classification to improve students' employability prediction. *International Journal of Intelligent Systems and Applications*, 9(9), 10.

[13] Casuat, C. D., & Festijo, E. D. (2019). Predicting students' employability using machine learning approach. In *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1-5. IEEE.

[14] Bai, A., & Hira, S. (2021). An intelligent hybrid deep belief network model for predicting students' employability. *Soft Computing*, 25(14), 9241-9254.

[15] Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N., & Alluhaidan, A. S. (2022). Predicting student employability through the internship context using gradient boosting models. *IEEE Access*, 10, 46472-46489.

[16] Hariharan, V. J., Abdullah, S., Rithish, R., Prabakar, V., Suguna, M., Ramakrishnan, M., & Selvakumar, S. (2022). Predicting student's placement prospects using Machine learning Techniques. Available at SSRN 4140544.

[17] Manvitha, P., & Swaroopa, N. (2019). Campus placement prediction using supervised machine learning techniques. *International Journal of Applied Engineering Research*, 14(9), 2188-2191.

[18] Kumar, N., Singh, A. S., T. K., & Rajesh, E. (2020). "Campus Placement Predictive Analysis using Machine Learning," In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India, pp. 214-216, doi: 10.1109/ICACCCN51052.2020.9362836.

[19] Bhoite, Sachin et al. "A Data-Driven Probabilistic Machine Learning Study for Placement Prediction." In *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, 2023,pp. 402-408.

[20] Basha, Md. Shaik Amzad et al. "Unraveling Campus Placement Success Integrating Exploratory Insights with Predictive Machine Learning Models." In *2023 7th International Conference for Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 2023, pp. 1-6.

Authors' Profiles

Raghavendra C. K. is currently working as an Associate Professor in the Department of Computer Science and Engineering at B N M Institute of Technology, Bangalore and obtained his Ph.D degree in the Department of Computer Science and Engineering at SJB Institute of Technology, Bangalore under Visvesvaraya Technological University, Belagavi, India. He obtained his B.E degree in 2009 and M.Tech degree in 2011 from Visvesvaraya Technological University, Belagavi, India. His current research interests are in the area of Data Mining, Web Mining, Big Data Analytics, Recommendation systems Social Networks Analysis and Image Processing.

Smaran Nanjundiah G. is currently pursuing a Bachelor of Engineering in Computer Science and Engineering at BNM Institute of Technology, Bangalore, Karnataka, India. His research interests comprise of Cloud Computing, Software Development and Machine Learning.

Spandana A. P. is currently pursuing a Bachelor of Engineering in Computer Science and Engineering at BNM Institute of Technology, Bangalore, Karnataka, India. Her research interests encompass Technology for Education, Artificial Intelligence and Machine Learning, and Software Development.

Vijay D. is currently pursuing Bachelor of Technology in the field of Computer Science and Engineering in the esteemed institution BNM Institute of Technology an Autonomous Institution Under Vishveshwaraiah Technological University, Belagavi. His Research Interests includes Machine Learning, Artificial Intelligence in variousdomains.

Vishruth M. V. is currently pursuing Bachelor of Engineering in Computer Science and Engineering at BNM Institute of Technology, Bangalore, Karnataka, India. His filed of interests and research includes Machine Learning, Data Science and Software Development.

How to cite this paper: Raghavendra C. K., Smaran N. G., Spandana A. P., Vijay D., Vishruth M. V., " Identifying Patterns and Trends in Campus Placement Data Using Machine Learning", International Journal of Education and Management Engineering (IJEME), Vol.15, No.1, pp. 10-24, 2025. DOI:10.5815/ijeme.2025.01.02