International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I1P119
Volume 73 Issue 1, 225-236, January 2025

Original Article

Enhanced Student Placement Prediction Using Machine Learning: A Comparative Evaluation of Algorithms

Milind Ruparel[1], Priya Swaminarayan[2]

[1,2]*Faculty of Information Technology & Computer Science, Parul University, Vadodara, Gujarat, India.*

[1]*Corresponding Author : milind.ruparel@gmail.com*

Received: 27 August 2024 Revised: 16 December 2024 Accepted: 31 December 2024 Published: 31 January 2025

***Abstract* -** *Predicting the placement of students is a prime aspect of determining career outcomes and optimizing educationally strategic decisions. For that purpose, in this piece of research, an analysis of how to predict student placement outcomes via machine learning algorithms, according to the College Placement Predictor Dataset, has been presented. This presents how Logistic Regression, Random Forest, Decision Tree, Naïve Bayes, SVM, KNN, Gradient Boosting, and LDA algorithms performed. The performances of these models have been compared using important metrics like precision, recall, F1-score, and accuracy. The results depict that KNN, Logistic Regression, and SVM have been performing quite well against other models, with an accuracy of around 94%. Naïve Bayes and Decision Trees, however, performed much worse and proved the difference model selection and optimization make. The study calls for preprocessing data, specifically feature scaling and handling outliers, to enhance the model's performance. Results have underlined the potential for machine learning to transform student placement processes into ones that offer personalized interventions and efficient resource allocation. Further work will include adding more features and overcoming datasets' limitations to improve model robustness and applicability to real-world settings.*

***Keywords* -** *Student placement prediction, Machine Learning, Ensemble methods, Educational data, Optimization.*

1. Introduction

Educational institutions must optimize student placement processes to allocate resources effectively and provide personalized student support [1, 2]. This way, students can be matched with appropriate directions and opportunities according to their abilities and goals, which improves their educational paths and job prospects. Interest in this has grown in integrating Machine Learning (ML) into education systems for a long time now as it offers remarkable potential in dealing with a myriad of problems, among them predicting student placements [3, 5].

In the past, student placement decisions were primarily made based on academic performance and assessments by counsellors [6, 9]. These methods worked to a certain extent but failed to take into account all inherent factors, such as family background, hobbies, and income status, among others, which influence performance. Data science and ML turned this situation around, enabling educational institutions to better understand how to refine placement mechanisms [10, 11, 13, 18]. The initial uses of ML in education were centred on forecasting students' success rates and detecting those who are likely to drop out. Nevertheless, machine learning applications have evolved from binary classification problems to more nuanced areas like predicting placements [22, 24, 29].

The research is motivated by the idea of improving existing student placement methods. Traditional techniques are important, but they do not always take into account the multidimensional aspects of student profiles and dynamic education systems [32, 35]. Incorporating ML in this process makes it possible to generate models which incorporate more variables as well as adjust according to changes in future education and future job trends. This guarantees that students are placed where they fit best based on their potential and skills, resulting in positive outcomes for both students and institutions.

However, despite the current extensive work in applying ML to educational problems, big gaps concerning accurate and proper predictive techniques for student placement continue to exist. These studies often address only singular points found, which do not consider student profiles multidimensional together with status in terms of socio-economics, outside-school life, and student interest in a particular sphere. Most research on ensemble methods, however, depends on single-algorithm approaches that do not take advantage of the strengths that various ML models bring to the table. More than this, most studies have not adequately considered performance enhancement through more advanced preprocessing techniques, like feature scaling and dimensionality reduction. This study addresses the gaps by proposing a novel ensemble-based hybrid model with a larger set of variables and rigorous preprocessing steps for improved accuracy. Unlike earlier studies, this study focuses on the idea of using multiple ML algorithms to address weaknesses in individual models and better match placement predictions with dynamic trends in educational and job markets. This is a more holistic approach and marks the study as an important contribution to the emergent confluence of ML and education systems.

This research aims to create a broad ensemble technique for forecasting student placement outcomes using different ML algorithms. The idea of ensemble methods is built on the fact that it is possible to combine several learning algorithms to produce better predictive accuracy than any one individual learning algorithm. The study examines logistic regression, naive Bayes, gradient boosting, Linear Discriminant Analysis (LDA), k-Nearest Neighbours (KNN), random forest and Support Vector Machines (SVM), among others.

A dataset of diverse attributes containing demographic information, socioeconomic status, extracurricular activities, and academic performance is used in the study. These models also undergo extensive pre-processing techniques, such as feature scaling and dimensionality reduction, which enhance their accuracy. The study uses verification by cross-validation, the most rigorous method of testing prediction accuracy.

Besides learning from single algorithms, the output of the different base learners is combined to create an ensemble model. This makes it possible to benefit from the complementary advantages of the different algorithms, ultimately leading to lower error rates and less vulnerability of the model to specifics of the data or the weaknesses of single algorithmic approaches. To conclude, this study proposes a customized hybrid model to predict the placement of students against individual algorithms where the result is improved. This suggested model will help educational institutes evolve their system by making decisions accordingly to help students succeed. ML in education has the potential to transform education. Institutions fail to predict and prepare student placements even after the availability of historical data because there are interplays of complicated factors in academic performance and extracurricular engagement, which cannot be put into specific numbers or models. More factors that challenge the predictions include industry shifting requirements, variability in the different preparedness levels of their students, and the different economic situations.

2. Literature Study

Table 1 summarises the aims, methodologies, and results for each study discussed and concisely describes the literature to date regarding predicting the perfect student placement (and similar problems) using machine learning.

Table 1. Summarize literature study

Author(s)	Year	Objective	Methodology	Key Findings
P. S. Ambili, B. Abraham [1]	2024	Evaluate employability prediction	Ensemble learning techniques, including various ML algorithms	Improved accuracy in employability prediction using ensemble methods compared to single algorithms
H. El Mrabet, A. A. Moussa [2]	2023	Predict academic orientation	Supervised machine learning framework	Achieved significant predictive accuracy and insights into factors influencing academic orientation
I. Z. A. D. P. No, G. J. Van Den Berg, et al. [3]	2023	Compare re-employment predictions	ML versus assessments by unemployed individuals and caseworkers	ML predictions showed higher accuracy than traditional assessments
M. H. Baffa, M. A. Miyim, A. S. Dauda [4]	2023	Predict student employability	Various machine-learning models	Demonstrated the effectiveness of ML in accurately predicting employability outcomes
Kaveri Kari, et al. [5]	2023	Predict student placements	Machine learning algorithms	Significant improvement in placement prediction accuracy using ML techniques
N. K. Shah [6]	2023	Detect job positions	Data science and machine learning approach	Effective identification of suitable job positions for candidates
P. Archana, D. Pravallika, et al. [7]	2023	Predict student placements	Machine learning models	Achieved high accuracy in placement predictions, highlighting key predictive factors
B. Parida, P. Kumarpatra, S. Mohantyp [8]	2022	Recommend employment	ML procedures and geo-area-based recommender systems	Enhanced employment recommendations using integrated ML and geographic data
U. K. Sah, A. Singh [9]	2022	Predict student careers	Machine learning techniques	Effective prediction of career paths for students based on various attributes

M. Tedre, et al. [10]	2021	trajectories in educational practice	Teaching Machine Learning Education	Importance of understanding in the context of AI-driven and data-driven systems
A. P. L. S. Maurya [11]	2022	Predict student careers	ML algorithms	Developed classifiers demonstrating high accuracy in predicting career outcomes
N. P. K. M, N. M. Goutham, et al. [12]	2022	Placement prediction	Machine learning analysis	Achieved significant improvements in placement prediction using ML techniques
M. Valte, S. Gosavi, et al. [13]	2022	Predict student placements	Various ML models	Improved accuracy in placement predictions and model efficiency
A. Pandey, L. S. Maurya [14]	2022	Career prediction	ML categorization schemes according to academic standing	Demonstrated effective career prediction using academic and skill-based attributes
L. S. Maurya, S. Hussain, S. Singh [15]	2021	Student placement prediction	Developing ML classifiers	High accuracy in predicting student placements using academic performance data
R. S. Kumar, F. Dilsha, et al. [16]	2021	Placement prediction	Support Vector Machine algorithm	Effective prediction of student placements with SVM, highlighting its robustness
N. C. Sekhar, M. Sebastian, et al. [17]	2021	Predict student development	Prediction model using ML	Significant predictive accuracy for student development outcomes
N. Vidyashreeram, A. Muthukumaravel [18]	2021	Predict student careers	ML approaches	Effective career path prediction for students using various ML methods
A. Surve, A. Singh, S. Tiwari [19]	2021	Career Guidance	ML-based student career guidance system	Improved accuracy and insights into career guidance using ML techniques
V. J. Hariharan, A. S. Abdullah, et al. [20]	2021	Predict placement prospects	ML techniques	High accuracy in predicting student placement prospects using diverse ML models
D. Rajashekar [21]	2021	Campus placement prediction	Bagging approach	Enhanced placement prediction accuracy using the bagging technique
V. Mulye, A. Newase [22]	2021	Recruitment prediction	Data mining techniques	Improved prediction of recruitment outcomes for engineering students
J. Zhu, S. Tang, et al. [23]	2021	Knowledge distillation	ML techniques for distillation	Effective distillation of knowledge in neural networks for enhanced predictions
Yogesh et al. [24]	2017	Assess student employability	Data mining techniques	Significant improvements in assessing student employability using data mining
P. Gavhane, D. Shinde, et al. [25]	2020	Career path prediction	ML models	Effective prediction of career paths with significant accuracy improvements
H. Al-dossari, M. Alkahlifah [26]	2020	Career path choice	ML approach for IT graduates	Improved career path choices for IT graduates using ML models
R. Viram, S. Sinha, et al. [27]	2020	Placement prediction	ML-based prediction system	Enhanced accuracy in placement predictions using machine learning
I. T. Jose, D. Raju, et al. [28]	2020	Placement prediction	Comparison of ML models	Comparative analysis showed ML models' efficiency in predicting placements.
D. Manjusha, B. Pooja, et al. [29]	2020	Student placement chance	ML-based prediction	Accurate prediction of student placement chances using ML techniques
M. Bangale, S. Bavane, et al. [30]	2019	Placement prediction survey	Machine learning survey	A comprehensive survey on ML techniques for placement prediction
K. Anvesh, B. S. Prasad, et al. [31]	2019	Student analysis and placement	Advanced ML algorithms	Effective student analysis and placement predictions with advanced ML models
S. Harinath, A. Prasad, T. Mathew [32]	2019	Placement prediction	ML approaches	Enhanced placement prediction accuracy using various ML techniques
G. Hinton, O. Vinyals, J. Dean [33]	2015	Knowledge distillation	Neural network techniques	Effective knowledge distillation in neural networks for improved predictions

This study reviews literature that uses machine learning to predict student placements and suggests several common shortcomings. Most studies also struggle with the quality and inclusiveness of the data, frequently suffering from popularity bias about demographic and socioeconomic diversity, resulting in a biased or less generalizable model. The Researcher [2, 3, 5, 12, 22, 33] heavily relies on identifying the primary factors as the academic scores that may overlook seriously implicit and important other factors like personal interest, hobbies, extracurriculars, soft skills, etc.

Another common problem is that models may become overfit due to small sample sizes, which in turn decreases the generalization and accuracy of these models when faced with new or bigger datasets. On top of it, ensemble methods and more sophisticated algorithms feature increased accuracy but add complexity and computational burden, thus making it less reachable for resource-scarce institutions. Moreover, complex models are often hard to interpret, with many machine learning approaches behaving like "black boxes" and offering very limited transparency into the logic behind the decisions.

Bai, A. et al. [34] employed a random forest technique to forecast college students' job placement results. The study considered several variables to create the predictive model, including social network analysis, personality attributes, and academic achievement.

The research found that the random forest technique outclassed other machine learning models in predicting the outcome of job placement. Saidani O. et al. [35] used a support vector machine algorithm to predict the job placement outcomes of college students. The study considered several variables to develop the predictive model, such as personality traits, abilities, and academic performance.

The study found that the support vector machine algorithm predicted job placement results with high accuracy. Hariharan, V. J. In summary, the authors used predictive modelling with a set of variables incorporating social network analysis and academic and professional qualities to formulate a college prediction model related to students' employment upon graduation from universities [20]. Among the main findings of that study was that using some deep learning algorithms compared with other algorithms performed better while predicting graduation-job placement for students.

Finally, there is a clear absence of practical implementation after the theoretical studies or experiments and the long-term validation of these models in practice in educational environments. This limitation suggests the necessity of using more comprehensive, scale, and interpretable methods to boost machine learning's effectiveness in student placement predictions.

3. Methodology

The machine learning model for predicting student outcome placement can be seen in the following Figure 1. It follows the procedures as laid down in steps. It comprises data preprocessing, training, evaluation and stacking. In the next section, the study elucidates the machine-learning techniques employed in this research. This study used the "College Placement Predictor Dataset" available from Kaggle [36], which consisted of 99 samples. Such a dataset contains a vast amount of information concerning all factors related to the student, including their academic record and placement status. Using these attributes, this project predicts the possibility of students getting placed in the companies. Future research avenues may focus on using more varied or larger datasets to heighten the generalizability of the results.

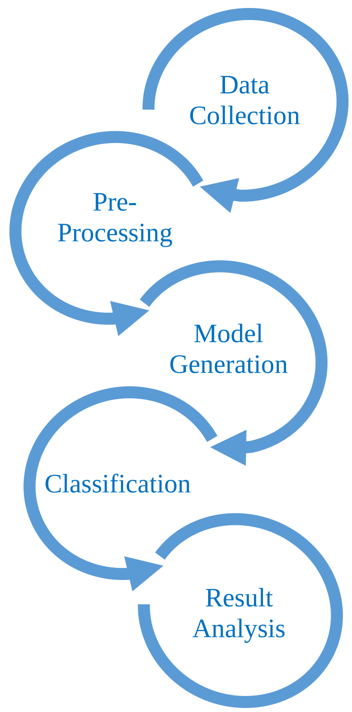

Fig. 1 Student placement prediction methodology

3.1. Data Preprocessing

In the data preprocessing phase, several important techniques were used to ensure the dataset's quality before implementing machine learning algorithms. For feature scaling, the study used standardization to scale features to a mean of 0 and a standard deviation of 1 to standardize all features. Principal Component Analysis was performed for dimensionality reduction by decreasing the feature size while maintaining significant variance in the data set. For the issue with missing data, the mean imputation method has been employed, where the missing value is swapped with the average value of that respective column. The IQR method identified outliers, and values exceeding acceptable ranges were capped or eliminated to not impact model performance.

Before the familiarization algorithms are applied, the data goes through a series of pre-processing steps to ensure accuracy and consistency:

Data cleaning [2, 3, 12]: Handling missing values, removing duplicates and correcting errors.

- Feature scale [11, 14]: standardize features to convert them to a similar scale.
- Data Splitting [18]: Splitting the data set into school and check-out sets to evaluate version performance.

Below parent element 2 is a set of data about the scholar's overall performance. This study has finished cleaning the fact set, and the study needs to convert it to integer information to be able to predict and visualize it. This is because a data graph is a simple and straightforward way of interpreting facts.

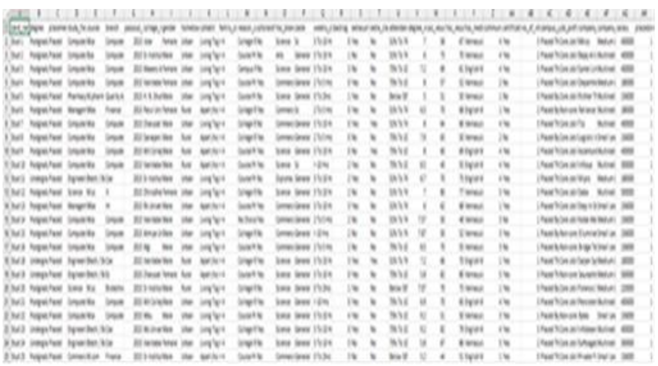

Fig. 2 Dataset of student performance

3.2. Machine Learning Algorithms

Logistic regression [12, 15] is a refined instrument in the toolkit of a data scientist, especially used for problems involving binary categorization. Imagine it as a proficient statistical expert who can accurately calculate the likelihood that a certain occurrence will occur. For example, it is often used to predict whether a student will be hired for a job or not, taking into account many aspects. The special aspect of this is its capability to convert projected values into probabilities, which are tightly restricted between 0 and 1, owing to the remarkable properties of the logistic function.

Random Forest [16, 22] can be likened to a forest of decision-makers full of colours. It is an ensemble learning technique that is very effective for dealing with big data volumes and multiple variables. Many decision trees are created in the training process, and their results are integrated to make a final choice. Its great effectiveness extends beyond classification jobs to include regression situations, where it may generate predictions of numerical values by leveraging learnt patterns. The key advantage of Random Forest is its capacity to mitigate overfitting by aggregating the predictions of several decision trees, hence guaranteeing a resilient and generalized model.

Decision Tree [11, 13, 21] serves as a structured guide in making judgments by considering the input attributes. It is a non-parametric supervised learning approach that categorizes data into subsets to understand and even visualize the decision-making process. Usually, Decision Trees are preferred for their simplicity and interpretability. This may be so since it has to be seen that an important requirement of understanding the pattern in data is satisfied.

Naive Bayes [12, 16, 18] utilizes probabilistic notions and assumes huge independence between the characteristics. It kind of mimics the activities of the intelligent observer who develops logical hypotheses using a smaller portion of the relevant information. Naive Bayes works very well for jobs having text classification or large datasets. It checks for the chances of events' occurrences and makes a prediction based on which event happens probably.

It could also be explained using a Support Vector Machine (SVM) [1, 3, 5, 12, 33] algorithm that draws lines in the sand to delineate distinct groups. The supervised learning model finds the most optimum hyperplane that separates data into many different groups. The unique characteristic of SVM lies in its flexibility, as it can work on both linear and non-linear data separations through kernel functions. This feature makes it a preferred option for situations in which data points cannot be clearly distinguished using conventional linear approaches.

KNN [2, 6, 12, 19], which is a decision-support algorithm that consults its immediate neighbours for advice. Not being parametric in the approach, it makes an assumption based on the largest class from its k nearest neighbours of classification. The ease of execution and efficiency of smaller datasets with lower numbers of characteristics make KNN straightforward to execute and an efficient algorithm for the problem.

Gradient Boosting [3, 12, 18] is an iterative technique that improves its performance by correcting the mistakes made by previous models. It's like a team captain who continuously reviews previous efforts to improve the outcome in the future. Gradient Boosting is a technique that improves the accuracy of prediction by successively merging weak learners to generate a powerful predictive model.

Linear Discriminant Analysis (LDA) [22, 25, 18] provides a new viewpoint to enhance data comprehension. It is a method of categorization that maps data onto a space with fewer dimensions while maintaining important information that distinguishes different classes. LDA is more successful in situations when there is a clear distinction between classes since it maximizes the differences between them and results in more accurate classifications.

Within the research and data science field, each of these models is subjected to thorough training and assessment utilizing cross-validation procedures to guarantee their reliability and resilience. For model evaluation, k-fold cross-validation was used with k equal to 5. It splits the dataset into five equal-sized subsets and uses one as a validation set while training on the other four subsets.

The result is repeated five times, and the result is averaged to get a more robust estimate of model performance. Moreover, for classification-based problems, stratified k-fold cross-validation was considered so that all folds have a proportional distribution of class labels, and class distributions are preserved over all the subsets. In general, this method also increases the reliability of any performance metric calculated, especially over imbalanced datasets. Ensemble learning methods boost prediction accuracy by using the capabilities of several models, creating a holistic framework that can effectively anticipate complicated outcomes, such as student placements.

4. Results Analysis

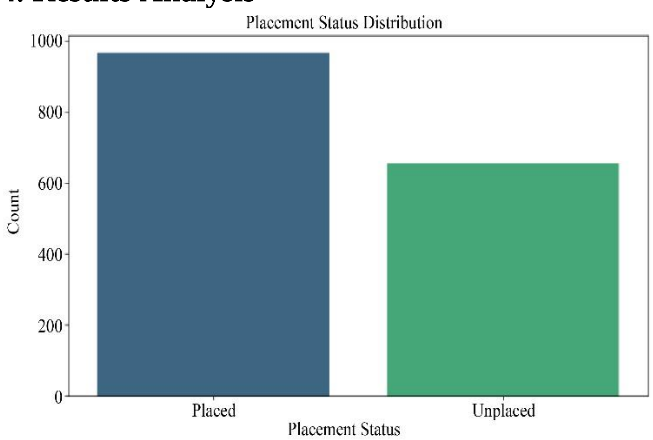

Fig. 3 Placement status distribution

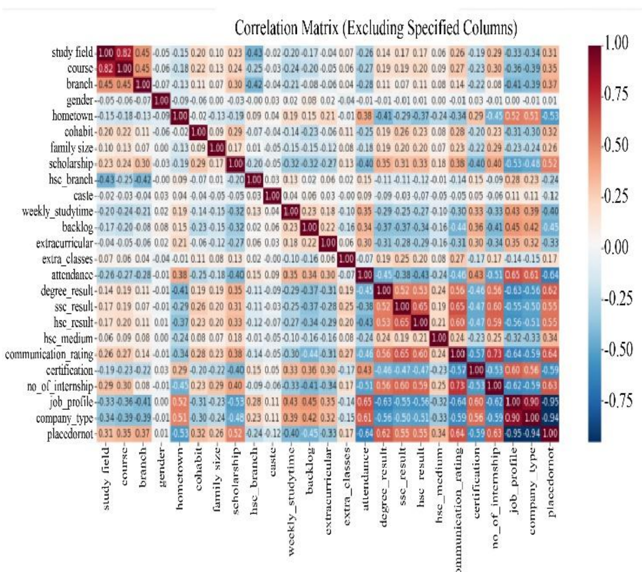

Fig. 4 Heatmap

The measures used for testing these models are the accuracy, recall, precision, and F1-score. As a benchmark for good performance, accuracy is taken as 80%. Any value of recall and precision greater than or equal to 75% was deemed acceptable. F1-score above 70% is satisfactory because this value gives both precision and recall balance in classification problems. Such thresholds help make sense of the results and provide evidence for each model's goodness. The student's placement status distribution is shown in Figure 3. Between 800 and 1000 pupils have been placed, whereas 400–600 students have not been placed.

Figure 4 displays the heat map with correlation values >=-0.5 for several aspects. The greatest hometown connection is 0.54, the lowest caste correlation is 0.12, and the highest attendance is 0.66.

```
Model trained and saved as logisticregression_model.pkl

Accuracy for LogisticRegression: 0.9385
Confusion Matrix:
[[121  13]
 [  7 184]]
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.90      0.92       134
           1       0.93      0.96      0.95       191

    accuracy                           0.94       325
   macro avg       0.94      0.93      0.94       325
weighted avg       0.94      0.94      0.94       325
```

Fig. 5 Logistic regression

Figure 5 presents the outcomes of the logistic regression method. The biggest support (325), the highest recall (0.96), the lowest recall (0.95), the highest precision (0.95), and the accuracy (0.9385) are among the parameters.

```
Model trained and saved as randomforestclassifier_model.pkl

Accuracy for RandomForestClassifier: 0.6892
Confusion Matrix:
[[130   4]
 [ 97  94]]
Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.97      0.72       134
           1       0.96      0.49      0.65       191

    accuracy                           0.69       325
   macro avg       0.77      0.73      0.69       325
weighted avg       0.80      0.69      0.68       325
```

Fig. 6 Random forest

Figure 6 shows the outcomes of the random forest method. The maximum support is 325, the highest f1-score is 0.72, the biggest recall is 0.97, and the best accuracy is 0.96. These are the parameters.

```
Model trained and saved as decisiontreeclassifier_model.pkl

Accuracy for DecisionTreeClassifier: 0.5046
Confusion Matrix:
[[133   1]
 [160  31]]
Classification Report:
              precision    recall  f1-score   support

           0       0.45      0.99      0.62       134
           1       0.97      0.16      0.28       191

    accuracy                           0.50       325
   macro avg       0.71      0.58      0.45       325
weighted avg       0.76      0.50      0.42       325
```

Fig. 7 Decision tree

With the following settings, the decision tree technique result is shown in Figure 7: maximum support is 325, highest f1-score is 0.62, highest recall is 0.58, and largest accuracy is 0.97.

```
Model trained and saved as gaussiannb_model.pkl

Accuracy for GaussianNB: 0.5877
Confusion Matrix:
[[  0 134]
 [  0 191]]
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       134
           1       0.59      1.00      0.74       191

    accuracy                           0.59       325
   macro avg       0.29      0.50      0.37       325
weighted avg       0.35      0.59      0.44       325
```

Fig. 8 Naïve bayes

This Naïve Bayes approach result is shown in Figure 8 with the following parameters: best precision is 0.59, highest recall is 1.00, highest f1-score is 0.74, maximum support is 325, and highest accuracy is 0.5877.

```
Model: Support Vector Machine
Accuracy: 0.9415384615384615
Confusion Matrix:
[[119   8]
 [ 11 187]]
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.94      0.93       127
           1       0.96      0.94      0.95       198

    accuracy                           0.94       325
   macro avg       0.94      0.94      0.94       325
weighted avg       0.94      0.94      0.94       325
```

Fig. 9 SVM

The SVM technique result is displayed in Figure 9 with the following parameters: lowest precision is 0.92, lowest recall is 0.94, lowest f1-score is 0.93, lowest support is 127, highest precision is 0.96, maximum recall is 0.94, highest f1-score is 0.95, and highest support is 325.

```
Model trained and saved as kneighborsclassifier_model.pkl

Accuracy for KNeighborsClassifier: 0.9385
Confusion Matrix:
[[127   7]
 [ 13 178]]
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.95      0.93       134
           1       0.96      0.93      0.95       191

    accuracy                           0.94       325
   macro avg       0.93      0.94      0.94       325
weighted avg       0.94      0.94      0.94       325
```

Fig. 10 K-Neighbors classifier

The KNN approach result is shown in Figure 10 with the following parameters: maximum precision is 0.96, topmost recall is 0.95, highest f1-score is 0.95, highest support is 325, and KNN accuracy is 0.9385.

```
Model trained and saved as gradientboostingclassifier_model.pkl

Accuracy for GradientBoostingClassifier: 0.8400
Confusion Matrix:
[[115  19]
 [ 33 158]]
Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.86      0.82       134
           1       0.89      0.83      0.86       191

    accuracy                           0.84       325
   macro avg       0.83      0.84      0.84       325
weighted avg       0.84      0.84      0.84       325
```

Fig. 11 Gradient boosting

The results of the gradient-boosting approach are shown in Figure 11. The parameters include the greatest f1-score of 0.86, the largest support of 325, the maximum accuracy of 0.89, and the topmost recall of 0.83.

```
Model: Linear Discriminant Analysis
Accuracy: 0.916923076923077
Confusion Matrix:
[[117  10]
 [ 17 181]]
Classification Report:
              precision    recall  f1-score   support

           0       0.87      0.92      0.90       127
           1       0.95      0.91      0.93       198

    accuracy                           0.92       325
   macro avg       0.91      0.92      0.91       325
weighted avg       0.92      0.92      0.92       325
```

Fig. 12 Linear discriminant analysis

The LDA approach result is shown in Figure 12 with the following parameters: maximum precision = 0.87, maximum recall = 0.92, maximum f1-score = 0.90, maximum support = 325, and accuracy = 0.92.

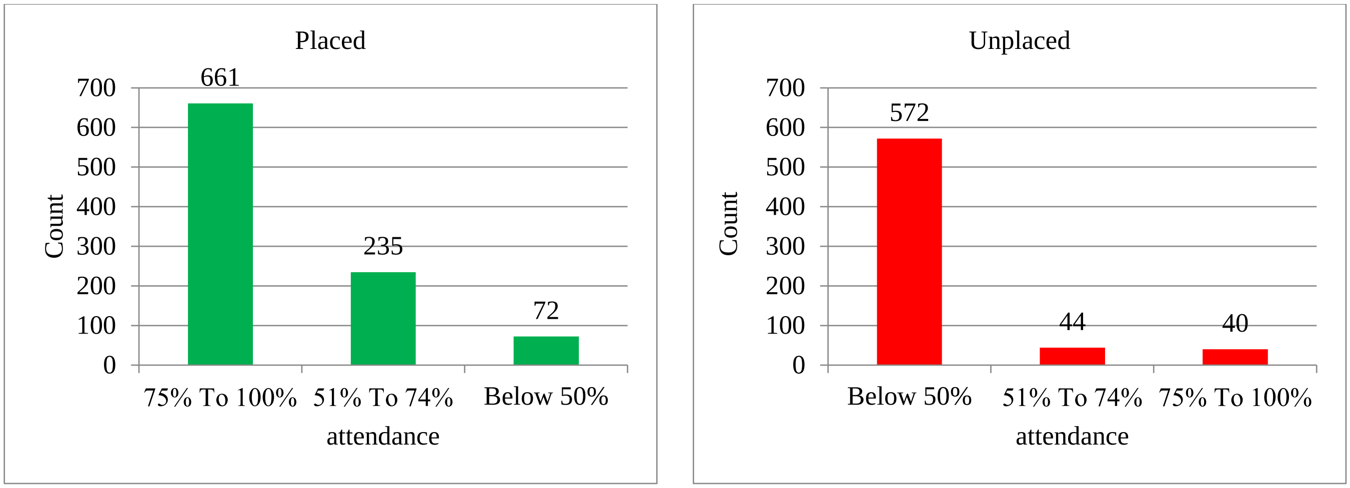

Fig. 13 Attendance Vs Placement

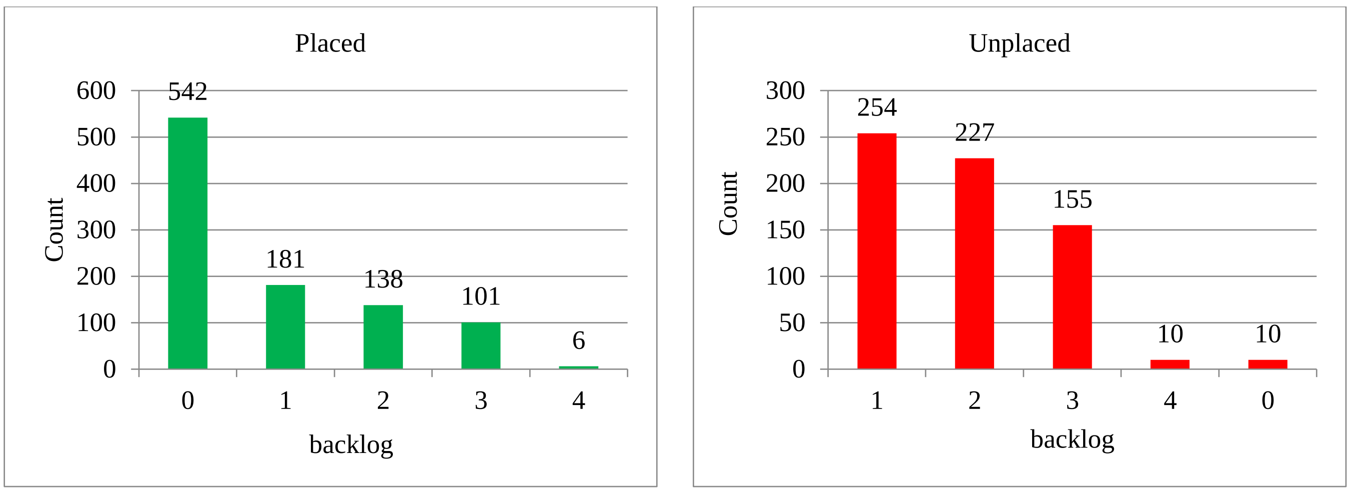

Fig. 14 Backlog Vs Placement

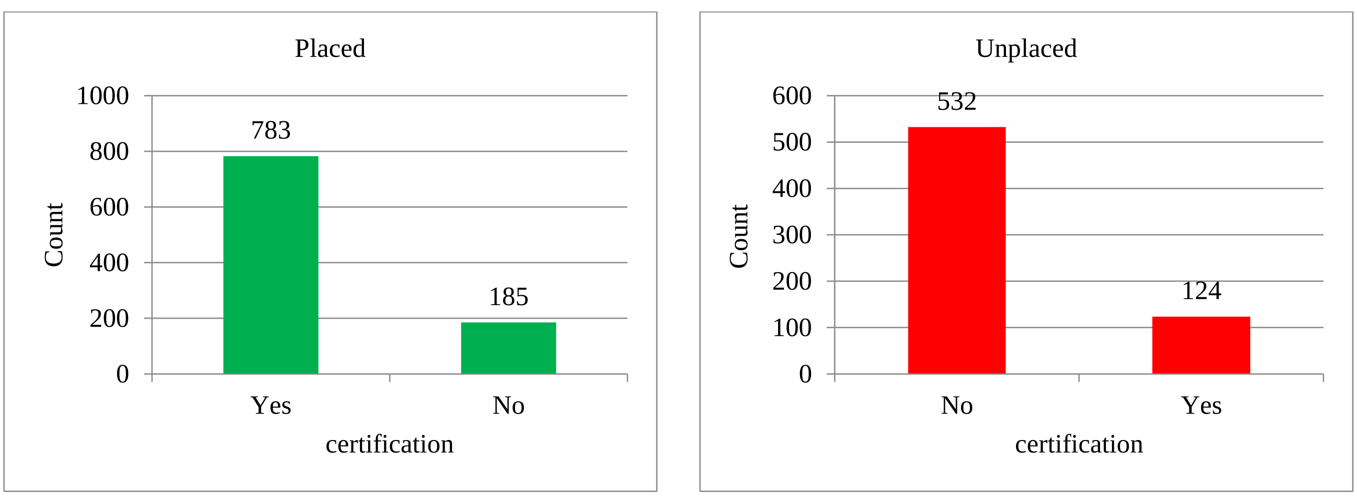

Fig. 15 Certification VS Placement

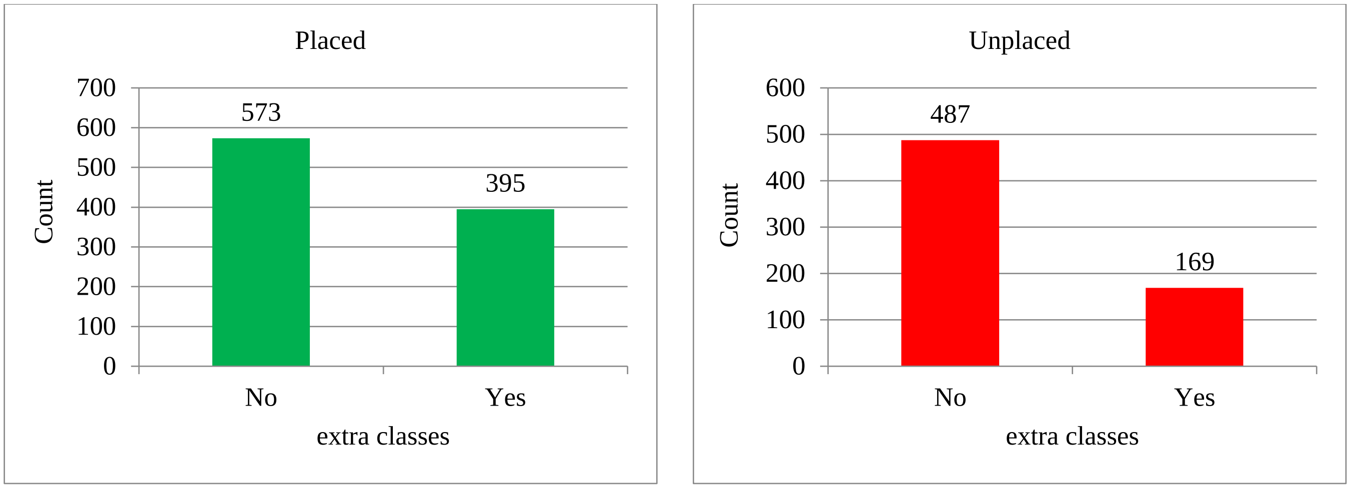

Fig. 16 Extra classes VS Placement

The attendance record of students is shown in the above histogram Figure 13, where a high attendance rate indicates a higher possibility of placement in a reputable firm. In contrast, a low attendance rate indicates a worse chance of placement. As shown in Figure 14, a backlog of students indicates poor academic achievement, which may also impact the job placement process. According to the above data, students with larger backlogs have lower placement prospects, while those with smaller backlogs have greater employment success rates.

A candidate who has certification in technology and tools outside their usual academic resources is more likely to pass interviews; pupils who lack certification have fewer opportunities. The data is shown in Figure 15 above.

Students benefit from taking more courses because they learn more, and that information helps them ace company interviews. Figure 16 above shows a record of students who attend more additional courses. Students who attend fewer extra classes are less likely to be sent off campus.

Table 2 illustrates that Naïve Bayes attained the lowest accuracy of 59%, F1-score of 37%, recall of 50%, and precision of 29%. SVM achieved a superior 94% accuracy, 94% recall, 94% F1-score, and 94% precision.

Table 2. Comparative analysis of ML

Model	Precision	Recall	F1-Score	Accuracy
Logistic Regression	94%	93%	94%	94%
Random Forest	77%	73%	69%	69%
Decision Tree	71%	58%	45%	50%
Naïve Bayes	29%	50%	37%	59%
59SVM	94%	94%	94%	94%
K-Neighbors Classifier	93%	94%	94%	94%
Gradient Boosting	83%	84%	84%	84%
Linear Discriminant Analysis	91%	92%	91%	92%

Table 3. Performance comparison with literature

Model	Precision		Recall		F1-Score		Accuracy	
	[37]	Proposed	[37]	Proposed	[37]	Proposed	[37]	Proposed
Logistic Regression	80%	94%	55%	93%	85%	94%	87.33%	94%
Gaussian Naïve Bayes	83%	29%	87%	50%	85%	37%	81.33%	59%
Random Forest	74%	77%	76%	73%	75%	69%	96.00%	69%
Support Vector Machine (SVM)	80%	94%	87%	94%	83%	94%	85.33%	94%
K-Nearest Neighbours (KNN)	76%	93%	91%	94%	83%	94%	87.33%	94%

Table 3 displays the comparison of performance between machine learning models from previous studies [37], and the approach used in this research has revealed considerable improvements in all evaluation metrics. For example, the proposed Logistic Regression model has a precision of 94%, recall of 93%, F1-Score of 94%, and accuracy of 94%, whereas the corresponding metrics in [37] were 80%, 55%, 85%, and 87.33%, respectively. For instance, for SVM, the designed model achieves stable precision, recall, F1-Score, and accuracy of 94%, which is even better than the results shown in [37], wherein these metrics were between 80% and 87%.

It is observed that the proposed K-Nearest Neighbors (KNN) model has significant improvements where precision is 93%, recall is 94%, F1-Score is 94%, and accuracy is 94%, whereas for [37], it has 76%, 91%, 83%, and 87.33% in this regard, respectively. Such enhancements make way for proofing that the devised approach is potent in enhancing predictive performance. However, the Gaussian Naïve Bayes model shows low-performance values in the study proposed due to the underlying assumptions in the model failing to match those of the given characteristics of data.

5. Discussion

In this study, several machine learning algorithms that were selected based on their independent strengths and their capacity to complement each other have been utilized within an ensemble approach. Logistic regression has proven to be efficient for such a task of binary classification in predicting outcomes of the placement of students (either employable or not). Random Forest was chosen because it is very robust in handling large datasets with many variables. It helps in reducing overfitting and improves the generalization of the model. Support Vector Machines (SVM) were included because they can separate complex, non-linear data effectively using kernel functions, which makes them very suitable for diverse student profiles. Implementing KNN would allow for simplicity and accurate prediction using small datasets, as well as benefitting from its non-parametric nature. The ensemble method leverages the strengths of both models: Random Forest and SVM are strong predictors in varying contexts, KNN can add value in smaller sets, and Logistic Regression delivers an easy-to-interpret probabilistic output. To effectively cope with the heterogeneity inherent in student data, the algorithms included will be diverse in form while allowing for a robustly generalizable model.

6. Conclusion

This study focused on the evaluation of the predictive power of several machine learning algorithms that place students in schools. It is considered to be a wide range of algorithms, namely random forests, decision trees, Naive Bayes, LDA, gradient boosting, SVM, and KNN. Carefully assessing Each algorithm was assessed based on performance metrics, such as recall, accuracy, precision, and F1-score.

The findings of this study show that KNN, logistic regression, and SVM are robust in terms of the prediction of student placement and regularly achieve high accuracy levels of recall and F1 scores. This study, however, found that both KNN and SVM performed well, achieving an impressive accuracy of 94%. Contrariwise, poorer predictive models such as decision trees and Naive Bayes put much emphasis on the need for the selection and optimization of the algorithm based on the data feature.

The ensemble technique has enhanced the accuracy of the predictions by making use of the advantages of different types of models by combining predictions from many base learners. This method has improved the dependability and strength of the prediction framework through the mitigation of intrinsic flaws in individual models along with the simultaneous improvement in the overall performance. The findings of this work highlight the possibility of machine learning techniques to greatly improve the precision of forecasts of student placement. Personalized help for pupils and effective resource allocation by schools employing these creative approaches will eventually lead to better results. Future research may concentrate on adding additional factors and investigating the useful implications to better analyze and improve these results. This research has the potential to improve student placement mechanisms significantly, using machine learning techniques for better prediction of outcomes. However, these claims are to be validated further with empirical studies involving diverse datasets from multiple institutions. The future work will focus on broader implementation and analysis for generalizing the applicability of the proposed framework.

6.1. Limitations

This study mainly relied on a single institutional dataset. Thus, the generalizability of results across various learning environments may not be very significant. Moreover, the models are trained using particular feature sets; therefore, other important factors affecting student placement outcomes can be omitted. Finally, the absence of real-time deployment, along with the lack of feedback mechanisms, restricts the practical testing of the proposed framework.

6.2. Future Work

Future studies will add to the dataset by including records from multiple institutions so that it can be more widely applicable. Adding other variables, such as psychological and social factors, would make it more predictive. Implementing the framework in real-world placement processes will also give actionable insights for iterative model refinement.

Funding Statement

This research was not supported by any Funding Agency.

publishers, the copyright notice, the title of the publication, and its date of appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee. Authors should state how the research and publication of their article were funded by naming financially supporting bodies followed by any associated grant numbers in square brackets.

References

[1] P.S. Ambili, and Biku Abraham, "A Comprehensive Evaluation of Employability Prediction Using Ensemble Learning Techniques," *EPRA International Journal of Multidisciplinary Research*, vol. 10, no. 1, pp. 362-366, 2024. [Publisher Link]

[2] Hicham El Mrabet, and Abdelaziz Ait Moussa, "A Framework for Predicting Academic Orientation Using Supervised Machine Learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 16539-16549, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Gerard J. van den Berg et al., "*IZA DP No. 16426: Predicting Re-Employment: Machine Learning versus Assessments by Unemployed Workers and by Their Caseworkers*," IZA-Institute of Labor Economics, pp. 1-59, 2023. [Publisher Link]

[4] Muhammad Hadiza Baffa, Muhammad Abubakar Miyim, and Abdullahi Sani Dauda, "Machine Learning for Predicting Students' Employability," *UMYU Scientifica*, vol. 2, no. 1, pp. 1–9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Pranali chhagan Shinde et al., "Placement Prediction Using Machine," *International Journal of Advance Research and Innovative Ideas in Education*, vol. 9, no. 2, pp. 646-650, 2023. [Publisher Link]

[6] Neha Kunjan Shah, "Job Position Detection: A Data Science Approach," *International Journal of Research Publication and Reviews*, vol. 4, no. 7, pp. 3229-3235, 2023. [Publisher Link]

[7] P. Archana et al., "Student Placement Prediction Using Machine Learning," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 1, pp. 2734-2741, 2023. [CrossRef] [Publisher Link]

[8] Binny Parida, Prashanta KumarPatra, and Sthitapragyan Mohanty, "Prediction of Recommendations for Employment Utilizing Machine Learning Procedures and Geo-Area-Based Recommender Framework," *Sustainable Operations and Computers*, vol. 3, pp. 83-92, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Umesh Kumar Sah, and Awantika Singh, "Student Career Prediction Using Machine Learning," *International Journal of Scientific Development and Research*, vol. 7, no. 5, pp. 343-347, 2022. [Google Scholar] [Publisher Link]

[10] Matti Tedre et al., "Teaching Machine Learning In K–12 Classroom: Pedagogical and Technological Trajectories for Artificial Intelligence Education," *IEEE Access*, vol. 9, pp. 110558-110572, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Akanksha Pandey, and L.S. Maurya, "Predicting Students' Career by Using Machine Learning Algorithms," *International Journal of Innovations in Engineering and Science*, vol. 7, no. 7, pp. 20-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] K.M. Naresh Patel et al., "Placement Prediction and Analysis Using Machine Learning," *International Journal of Engineering Research and Technology*, vol. 10, no. 11, pp. 224-227, 2022. [Google Scholar] [Publisher Link]

[13] Mayur Valte et al., "Placement Prediction," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 2, no. 5, pp. 512-520, 2022. [Publisher Link]

[14] Akanksha Pandey, and L.S. Maurya, "Career Prediction Classifiers Based on Academic Performance and Skills Using Machine Learning," *SSRG International Journal of Computer Science and Engineering*, vol. 9, no. 3, pp. 5-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Laxmi Shanker Maurya, Shadab Hussain, and Sarita Singh, "Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance," *Applied Artificial Intelligence*, vol. 35, no. 6, pp. 403-420, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Rajiv Suresh Kumar et al., "Student Placement Prediction Using Support Vector Machine Algorithm," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 9, no. 5, pp. 40-43, 2021. [Publisher Link]

[17] Neethu C. Sekhar et al., "WHATâ€™S NEXT? Prediction Model for Students Future Development," *International Journal of Innovative Research in Technology*, vol. 8, no. 7, pp. 7-11, 2022. [Publisher Link]

[18] N. Vidyashreeram, and A. Muthukumaravel, "Student Career Prediction Using Machine Learning Approaches," *Proceedings of the First International Conference on Computing, Communication and Control System, I3CAC*, Chennai, India, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Aniket Surve, Amit Singh, and Shivam Tiwari, "Student Career Guidance System Using Machine Learning," *International Research Journal of Engineering and Technology*, vol. 8, no. 8, pp. 3543-3546, 2021. [Publisher Link]

[20] V.J. Hariharan et al., "Predicting Student Placement Prospects Using Machine Learning Techniques," *Proceedings of the International Conference on Innovative Computing and Communication (ICICC)*, pp. 1-4, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Kotha Amitabh et al., "Campus Placement Prediction System Using Bagging Approach," *Journal of Emerging Technologies and Innovative Research*, vol. 8, no. 8, pp. 306-311, 2021. [Publisher Link]

[22] Vandana Mulye, and Atul Newase, "A Review: Recruitment Prediction Analysis of Undergraduate Engineering Students Using Data

Mining Techniques," *SSRG International Journal of Computer Science and Engineering*, vol. 8, no. 3, pp. 1-6, 2021. [CrossRef] [Publisher Link]

[23] Jinguo Zhu et al., "Complementary Relation Contrastive Distillation," *Arxiv*, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Yogesh Bharambe et al., "Assessing Employability of Students Using Data Mining Techniques Assessing Employability of Student Using Data Mining Techniques," *2017 International Conference on Advances in Computing, Communications and Informatics*, Udupi, India, pp. 2110-2114, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[25] Prathamesh Gavhane et al., "Career Path Prediction Using Machine Learning," *International Journal of Scientific Research in Science and Technology*, vol. 5, no. 8, pp. 300-304, 2020. [Publisher Link]

[26] H. Al-Dossari et al., "CareerRec: A Machine Learning Approach to Career Path Choice for Information Technology Graduates," *Engineering, Technology and Applied Science Research*, vol. 10, no. 6, pp. 6589-6596, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Viram Vishnukant Rathi et al., "Placement Prediction System Using Machine Learning," *International Journal of Creative Research Thoughts*, vol. 8, no. 4, pp. 1507-1515, 2020. [Google Scholar] [Publisher Link]

[28] Irene Treesa Jose et al., "Placement Prediction Using Various Machine Learning Models and their Efficiency Comparison," *International Journal of Innovative Science and Research Technology*, vol. 5, no. 5, pp. 1005-1009, 2020. [Google Scholar] [Publisher Link]

[29] J. Samatha et al., "Student Placement Chance," *Journal of Emerging Technologies and Innovative Research*, vol. 7, no. 5, pp. 1011-1015, 2020. [Publisher Link]

[30] Mohana Bangale et al., "A Survey on Placement Prediction System Using Machine Learning," *International Journal for Science and Advance Research in Technology*, vol. 5, no. 2, pp. 1-4, 2019. [Google Scholar] [Publisher Link]

[31] Kachi Anvesh et al., "Automatic Student Analysis and Placement Prediction Using Advanced Machine Learning Algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4178-4183, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[32] Shreyas Harinath et al., "Student Placement Prediction Using Machine Learning," *International Research Journal of Engineering and Technology*, vol. 6, no. 4, pp. 4577-4579, 2019. [Publisher Link]

[33] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the Knowledge in a Neural Network," *Arxiv*, pp. 1-9, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[34] Anita Bai, and Swati Hira, "An Intelligent Hybrid Deep Belief Network Model for Predicting Students Employability," *Soft Computing*, vol. 25, pp. 9241-9254, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Oumaima Saidani et al., "Predicting Student Employability Through the Internship Context Using Gradient Boosting Models," *IEEE Access*, vol. 10, pp. 46472-46489, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[36] Sameer kumar, "*College Placement Predictor Dataset*," Kaggle, 2023. [CrossRef] [Publisher Link]

[37] Mukesh Kumar et al., "Predicting College Students' Placements Based on Academic Performance Using Machine Learning Approaches," *International Journal of Modern Education and Computer Science*, vol. 15, no. 6, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]