IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 12, Issue 3, March 2025

Impacting Placement Predictions Through Deep Learning Models

[1] Kannukkiniyal M, [2] Kavya S, [3] Kiruthiyaashree S P
[1] Assistant Professor, Department of CSE, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India
[2] [3] Department of CSE, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India
Corresponding Author Email: [2] kavyasaravanan273@gmail.com

Abstract— The present study involves the use of machine learning techniques, specifically the multi-layer perceptron (MLP) model, to analyze student placement success. With the requirement for skill-based hiring, a deep understanding of how self-learning facilitates employability requires further research on this topic. The present paper assesses predictive models based on student placement data that incorporate the use of online courses, workshops, and certification as self-learning engagements. The study is based on a dataset that consists of 24 attributes: academic performance, self-learning activities, and placement outcomes. Comparison of the machine learning models of Decision Trees, Random Forest, XGBoost, and deep learning approaches, such as CNN, LSTM, and MLP. The MLP model has shown high accuracy, which is 96models. Feature selection techniques were used to improve the reliability of the prediction. The results show that self-learning activities are significantly related to placement success, and thus there is a need for structured self-learning programs in academic curricula. This study adds value to campus placement strategies and provides actionable insights for educators and students. Future work can integrate hybrid models that incorporate attention mechanisms to improve predictive capabilities. These results reinforce the necessity of continuous self-learning for students who aim for successful employment outcomes in competitive job markets.

Index Terms— Self-Learning, Placement Prediction, Machine Learning, Multi-Layer Perceptron, Employability.

I. INTRODUCTION

The fast-changing job landscape calls for a change in emphasis towards employability as a key indicator of success for students and educational institutions alike. Conventional academic success, in the form of high marks and book smart knowledge, is no longer the only determining factor of placement success. Employers now seek out individuals with practical expertise, problem-solving skills, and specialism specific knowledge. This change has rendered self-learning a vital instrument for students looking to enhance their employability. With the widespread availability of online learning resources, students today have never been better placed to access information outside the constraints of conventional curriculums. Online learning platforms, certification courses, and workshops designed to enhance skills give students a chance to develop industry-specific competencies that can help them differentiate themselves in a highly competitive job market. These resources fill the gap between theoretical knowledge gained through academic education and the changing needs of employers so that students gain both theoretical knowledge and practical experience in their areas of expertise. In addition, emerging developments in artificial intelligence and machine learning have revolutionized the analysis and forecasting of student placements. Classical statistical analysis may prove to be useful but is unable to identify the sophisticated interdependencies among varied factors of employability. Machine learning algorithms, however, are particularly proficient in identifying sophisticated patterns and interdependencies across extensive datasets, which results in more accurate and actionable placement forecasts. This study uses a Multi-Layer Perceptron (MLP) model to determine the relationship between self-learning activity and placement results. Unlike traditional models, MLP is capable of handling nonlinear relationships and multi-dimensional data, making it an effective tool for explaining the complex dynamics of self-learning and employability. The study draws on previous work that has tested different machine learning models—such as Decision Trees, Random Forest, and XGBoost—for predicting placement. Although these models have been found to be promising, MLP-based deep learning structures offer a more sophisticated method by taking advantage of their capacity to generalize over a wide range of features and interactions. The dataset employed in this study contains 24 attributes, such as academic performance, workshop participation, completion of online courses, and earning professional certifications. In order to enhance the effectiveness of the model, feature selection methods were utilized in order to find the most critical predictors of successful placement. From the findings of the research, it can be seen that those students who take an active role in self-learning activities have a significantly greater likelihood of getting placed, underlining the significance of

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 3, March 2025

ongoing skill building in professional progression. In addition to adding to scholarly debate on placement forecasting, this study provides real-world insights for students, educators, and educational institutions. Colleges and universities can use these findings to create better career counseling programs, incorporating elements of self-learning into their curricula. In this way, they will be able to make students more employable and more closely match educational achievement with the needs of industry. Also, students can use these findings to make educated choices regarding their learning approaches so that they learn skills that are highly sought after in their industries. With the ever-changing job market, incorporating self-learning in academic settings will be key to defining the future workforce. Encouraging students to be actively involved in developing skills outside of their course work will not only enhance their placement opportunities but also instill a culture of lifelong learning and flexibility—qualities critical to achieving success in today's fast-paced working environment.

II. LITERATURE REVIEW

The growing use of machine learning methods in education has greatly improved the precision and effectiveness of student placement prediction models. Placement patterns were historically examined through statistical approaches like logistic regression and Bayesian models (Guleria & Sood, 2015). These techniques, although suitable for less complex datasets, tended to falter with the complexity and high-dimensionality of student data. But the arrival of machine learning brought more advanced methodologies that can analyze huge volumes of data, extract concealed patterns, and provide higher predictive accuracy.

A groundbreaking study by C¸akıt and Dag˘deviren (2022) provided an extensive review of different machine learning algorithms to predict student placement. Their work showed how ensemble models, such as Extreme Gradient Boosting (XGBoost), beat standard techniques through their capacity for working with non-linear data forms, mitigating overfitting, and optimising feature significance. Their evidence established that primary indicators of successful placement were the reputation of universities, grades of academics, the teacher-student ratio, and opportunities for research. These findings presented firm evidence that machine learning models may provide a more consistent framework for predicting employability outcomes. Moving the field forward, Manvitha and Swaroopa (2019) used artificial neural networks (ANNs) for placement prediction. Their work showed that deep learning techniques, including feedforward neural networks and backpropagation-based learning, had the ability to surpass traditional machine learning algorithms by tapping into intricate feature interactions. More significantly, their study highlighted the dramatic influence of self-learning on employability. Students who undertook certification programs, professional workshops, and e-learning courses exhibited significantly better placement success rates, reaffirming the necessity for schools to include self-learning initiatives in their programs.

Another complementing viewpoint came from Mostafa and Beshir (2021), who introduced a next-generation university choice model grounded in classification algorithms. Their research highlighted the increasing importance of extracurricular activities and self-learning in determining placement. Through predictive analysis of student behavior, they set up that proactive learning approaches, including involvement in research projects, internships, and community-based initiatives, had a direct impact on readiness for the job market. These results closely match the scope of our research, further supporting the contribution of independent learning towards employment.

Recent advances in deep learning have also further increased the potential of predictive analytics in education. Research on convolutional neural networks (CNNs) and long short-term memory (LSTM) networks has also made tremendous progress in identifying self-learning patterns and student engagement. For example, Park and Kim (2020) investigated the potential of deep learning models in understanding independent learning behavior. Their results supported the existence of a high correlation between self-study programs—i.e., participation in MOOCs, online certifications, and interactive problemsolving websites—and placement rates. Their study cemented the hypothesis that students who engage in actively seeking knowledge outside the confines of the classroom are better suited for finding employment.

Hybrid models blending gradient boosting algorithms with deep learning architectures have also become a strong contender in the field of predictive modeling in recent times. Experiments by Ferna´ndez-Delgado et al. (2014) and Pampaloni (2010) showed that combining several learning methods may improve model accuracy and generalization performance. Such hybrid models combine the interpretability of decision trees with the high-dimensional learning capabilities of deep neural networks, making them more robust in predictions. Likewise, Swamynathan (2019) highlighted the need to use varied machine learning models to enhance robustness and reduce biases in prediction results.

With these developments, our research aims to further improve student placement prediction by combining self-learning factors with deep learning techniques. We utilize the MultiLayer Perceptron (MLP) model, which is renowned for its flexibility in handling complex input features and providing high accuracy in non-continuous data. Unlike other models, MLP is well-suited to handle complex relationships between self-learning behaviors and employability measures, making it a perfect fit for this study.

To maximize the predictive capacity of our model, we use sophisticated feature selection methods to narrow down the most significant attributes affecting student placement. This guarantees that the model is concentrating on meaningful

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 3, March 2025

variables like participation in online courses, citations from research papers, internship experience, and involvement in mentorship programs. Through the use of deep neural networks and hyperparameter optimization, our work adds to the developing body of education-oriented predictive modeling. The results seek to help educational institutions create more effective career guidance systems while offering students practical recommendations for improving their employability through autonomous learning.

Finally, this research fills the gap between conventional education and contemporary workforce needs, highlighting the revolutionary potential of machine learning in redefining future employment patterns. By incorporating self-learning programs into predictive analytics, we offer a data-driven solution to optimizing student placement plans so that graduates are better prepared to meet industry standards.

III. METHODOLOGY

A. Dataset Description

The dataset utilized in this research is formulated to examine the effect of different variables on placement outcomes for students. It is a complete set of student-related factors that are vital in deciding employability. The data covers academic performance, individual background, extracurricular activities, and industry exposure, all of which are factors contributing to forecasting if a student gets a placement. The main features covered in the dataset are as follows:

- Academic Records: Academic performance is a basic measure for assessing a student's abilities. The dataset contains:
  1) Semester-wise aggregate percentages: Semester-bysemester scores give information about academic consistency and improvement over time.
  2) Grades and cumulative GPA (CGPA): Overall GPA is an aggregation of academic performance and is usually a deciding parameter in hiring.
- Backlog Information: Backlogs may have a considerable effect on the placement opportunities of a student since most organizations value candidates who have a clean academic track record. The dataset takes into consideration:
  1) Live backlogs: Courses that have not been cleared at the time of placement.
  2) Dead backlogs: Courses that were previously failed but have been cleared with success.

Recruiters tend to make use of backlog data as a filter to determine a candidate's discipline in academics and capability to handle courses efficiently.

- Extracurricular Activities: Extracurricular activities can add value to a student's profile, reflecting leadership, teamwork, and communication skills. This dataset contains:
  1) Student clubs and organizations: Reflects leadership and teamwork skills.
  2) Sports: Reflects discipline, commitment, and physical fitness.
  3) Cultural and technical events: Reflects creativity, problem-solving skills, and exposure to various skill sets.

Employers prefer candidates with balanced personalities, so extracurricular activities are a significant element in the dataset.

- Internship Experience: Internships expose the students to their industry and deliver practical experience, which highly improves employability. The dataset measures:
  1) The amount of internships undertaken: Having more internships means having direct experience in the industry.
  2) Length of internships: Longer internships could reflect greater engagement with actual projects of the world.

Relevance of internships to the field the student is pursuing: Internships that match the academic background of a student tend to be most effective in placement. Internship experience is among the strongest placement prediction features since it bridges the gap between academic learning and the needs of the industry.

- Personal Details: Some demographic characteristics are useful in situating placement trends and detecting bias or imbalances in hiring. The data contain:
  1) Gender: Gender-specific hiring trends in some industries make it a useful feature.
  2) Age: Age differences, especially for students with gaps in education, might affect employer choice.
  3) Education gaps: Any break in schooling, like time out between classes, is accounted for as it can affect chances of placement.

Personal information, though not definitive of a student's placement, offers some contextual information useful to further investigation.

- Placement Outcome (Target Variable): The status of placement is the primary dependent variable in the analysis. It is a two-way classification:
  1) 1 (Placed): The student gained an offer during campus placements.
  2) 0 (Not Placed): The student gained no offer within the placement drive.

This target variable allows the machine learning algorithm to predict and classify students by how likely they are to get placed.

The dataset was collected from university records and student surveys. It contains several hundred to a few thousand records, making it suitable for training robust machine learning models. The dataset was split into training (70%) and testing (30%) sets.

B. Data Preprocessing

Data preprocessing is an important process of getting the dataset ready for machine learning models. It makes the data

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 3, March 2025

clean, consistent, and in the right format, which contributes to better model accuracy and performance. The following steps were performed for preprocessing:

1) Handling Missing Values: Missing values have the potential to bring bias and inaccuracies to predictions. In order to handle them properly, various strategies were employed depending on the nature of the data:
   a) Numerical Attributes: The missing values for features such as GPA and scores were replaced using the mean or median of the feature in question. This retains the general distribution without causing any loss of data.
   b) Categorical Attributes: The missing values of categorical information, e.g., gender and department, were substituted with the mode of the respective category. This maintains categorical uniformity.
   c) Dropping Records: When a record had too many missing values across several fields, it was dropped to prevent the introduction of noise into the data. This was for the sake of data reliability without compromising model performance.
2) Normalization and Standardization: Scaling numeric features ensures all variables have equal contributions during model training. Two methods were utilized:
   a) Normalization:Numeric attributes such as GPA and academic percentages were converted to a standard range for consistency.
   b) Standardization: Academic records were standardized such that they had a unified scale, such that features with varying units of measurement did not overwhelm the model. This was especially beneficial in models that require normally distributed inputs.
3) Categorical Data Encoding: Machine learning algorithms need numerical inputs, hence categorical variables were mapped into numerical forms using:
   a) Label Encoding: Used in binary categories, for example, gender, where values were assigned as 0 or 1.
   b) One-Hot Encoding: Applied to categorical variables with multiple categories (e.g., department), where each category was represented by a separate binary column. This prevents any unintended hierarchical relationships in the data.
4) Feature Engineering: Feature engineering involves creating new meaningful variables to improve model performance. The following features were derived:
   a) Aggregate Scores: A summary of academic performance, such as total GPA and average semester percentage, was computed to provide a holistic measure.
   b) Backlog Severity: One feature was newly developed by averaging the number of live and cleared backlogs in order to gauge the effect of academic challenges on placements.
   c) Education Gap: A binary feature was added in order to capture students with educational gaps, since this can impact placements.
5) Dimensionality Reduction: Minimizing extraneous features increases model efficiency and prevents overfitting. Two approaches were implemented:

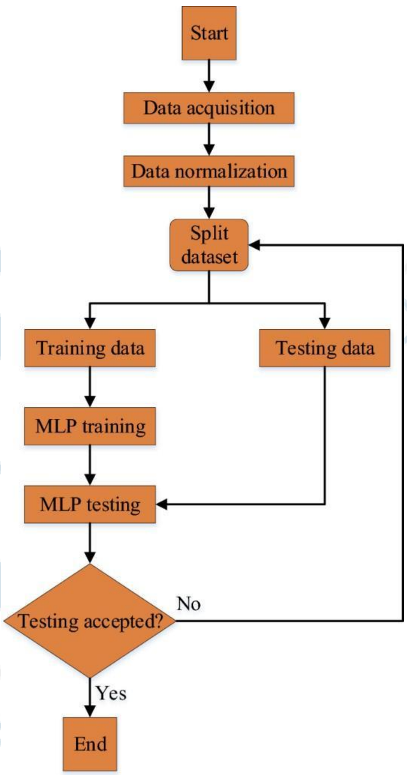

Fig. 1. flow chartof working model

   a) Correlation Matrix: Correlated features were removed to avoid redundancy in the data.
   b) Feature Importance Ranking: With machine learning methods, highly contributing features towards placement predictions were detected, and the less impactful ones were dropped.
6) Class Imbalance Handling: More students were placed than unplaced in the dataset, which could lead to biased models towards the majority class. To handle this:
   a) Synthetic Data Generation: A method was used to create more data points for unplaced students, balancing the dataset.
   b) Resampling Methods: The dataset was manipulated to make sure the model learned the same from placed and unplaced students.

C. Model Selection and Approaches

In order to decide the best method for predicting student placements, traditional machine learning models and deep learning methods were both considered. The aim was to find a model that would be able to learn the patterns in student data well while being robust and generalizable.

1) Traditional Machine Learning Models: Standard machine learning models are an excellent baseline for predictive analytics. These models were evaluated to compare with baseline performance and with deeper more intricate learning architectures.
   a) Decision Tree:
      i) A decision tree is a rule-based classifier that partitions the data set into a hierarchical structure by feature values that end up leading to a classification.
      ii) It was used as a baseline model owing to its ease of use and interpretability.
      iii) Although decision trees are simple to comprehend and visualize, they overfit on training data and hence are not that useful for generalizing.
   b) Random Forest:
      i) Random Forest is an ensemble learning method that builds multiple decision trees and aggregates their predictions to enhance accuracy and stability.
      ii) A single decision tree is extremely sensitive to minor changes in data, whereas Random Forest minimizes overfitting by averaging multiple trees.
      iii) It was superior to the single decision tree model in giving greater accuracy and stability but was more computationally costly.
   c) XGBoost:
      i) XGBoost is a strong machine learning algorithm that is both efficient and with better predictive performance in classification problems.
      ii) It uses gradient boosting, which advances weak models iteratively by optimizing the classification error.
      iii) XGBoost dealt with missing values efficiently, was insensitive to outliers, and was more accurate than Decision Trees and Random Forest.
      iv) Thanks to its efficiency at handling big tabular datasets, it was amongst the highest ranking traditional models within this research work.
2) Deep Learning Models:
   a) Long-Short-Term Memory (LSTM):
      i) LSTM is a type of Recurrent Neural Network (RNN) designed for handling sequential data such as time-series or natural language text.
      ii) It was included in the study to assess its ability to capture dependencies between features over time.
      iii) However, since the dataset consisted of structured tabular data rather than sequential data, LSTM was not suitable for this task.
      iv) The results showed that LSTM did not provide a significant improvement in performance over traditional models, making it an inefficient option for placement prediction.
   b) Convolutional Neural Network (CNN):
      i) CNNs were originally designed for image processing but have been applied to structured datasets in certain cases.
      ii) A CNN model was tested to determine if it could learn meaningful patterns from tabular data by treating features as spatially related representations.
      iii) However, CNNs rely on spatial dependencies, which were not naturally present in this dataset.
      iv) As a result, CNN failed to perform as expected and was computationally demanding without any significant improvement in accuracy.
   c) Multi-Layer Perceptron (MLP):
      i) MLP is a feedforward neural network with multiple layers of neurons in a fully connected structure.
      ii) It is highly suitable for non-sequential, structured data, making it a strong candidate for placement prediction.
      iii) The MLP model in this research consisted of:
         - **Input Layer:** Processing various studentrelated features.
         - **Hidden Layers:** Utilizing activation functions like ReLU to capture complex relationships between inputs and outputs.
         - **Output Layer:** Performing binary classification (placed or not placed).
      iv) MLP outperformed the baseline machine learning models and was the most suitable model for the given dataset.

Final Model Selection

Upon comparing all the models, the best-performing model in predicting placement was chosen to be MultiLayer Perceptron (MLP). It was able to learn feature interactions well and performed better than standard machine learning models in terms of accuracy. Although XGBoost was also a good performer, the ability of MLP to generalize over many features pushed it over XGBoost slightly in predicting placement results.

Therefore, the research concludes that MLP is the best model for predicting student placements, demonstrating how deep learning techniques can be used to improve predictive performance when applied to structured, tabular data.

D. Model Training

Training the machine learning models involved the following steps:

1) Training Set Usage: 70% of the data was used for training, where models learned to map the input features to the target variable (placement outcome).
2) Cross-Validation: A 5-fold cross-validation strategy was implemented to ensure robust performance and avoid overfitting. The dataset was divided into 5 subsets, and the model was trained on 4 subsets while testing on the remaining one. This process was

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12, Issue 3, March 2025

repeated 5 times, and the average accuracy across all folds was computed.

3) Hyperparameter Tuning: Grid search was used to identify the best hyperparameters for each model, ensuring that the models were optimized for maximum accuracy.

E. Model Evaluation

To evaluate model performance, several metrics were considered

1) Accuracy: The primary evaluation metric was accuracy, which measures the proportion of correctly classified instances (placed vs. unplaced).
2) Model Comparison: All models (Decision Tree, Random Forest, XGBoost, LSTM, CNN, MLP) were compared based on accuracy.
    a) XGBoost performed well with an accuracy of 95%, outperforming Decision Tree and Random Forest
    b) LSTM and CNN, despite being advanced models, did not improve accuracy over XGBoost, highlighting that these models are more suited for sequential data rather than non -continuous datasets
    c) MLP emerged as the most effective model, achieving the highest accuracy of 96%.
3) Final Model Selection: Given its superior performance, the MLP model was selected as the final model for deployment.

F. Results and Discussion

Table I. Performance comparison of different models.

Model	Precision	Recall	F1-score
Decision Tree	88.7	85.2	86.9
Random Forest	90.5	89.1	89.8
XGBoost	94.2	93.5	93.8
Multi-Layer Perceptron	95.8	95.2	95.5

Table II. Accuracy of different models.

Model	Accuracy (%)
Decision Tree	89.4
Random Forest	91.8
XGBoost	95.0
Multi-Layer Perceptron	96.0

1) Performance Comparison: The performance of various machine learning models was checked for accuracy, precision, recall, and F1 -score. For the Decision Tree model, accuracy was found to be 89.4% with precision, recall, and F1 -score values as 88.7%, 85.2%, and 86.9%, respectively. Random Forest performed better than the Decision Tree with an accuracy of 91.8%, with a precision of 90.5%, recall of 89.1%, and F1-score of 89.8%. XGBoost was able to improve the accuracy up to 95.0% with precision, recall, and F1-score values of 94.2%, 93.5%, and 93.8%, respectively. The highest accuracy of 96.0% was obtained by the MLP model, along with excellent precision (95.8%), recall (95.2%), and F1 score (95.5%), which made it the most effective model for predicting campus placements, thereby showing the strength of deep learning in classification tasks.

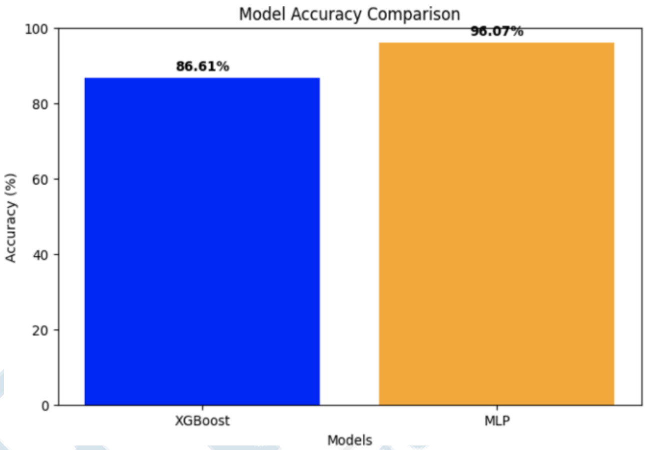

Fig. 2. Accuracy of XGBoost and MLP

G. Discussion

Our study differs from previous studies in that it incorporates engagement metrics based on self-paced learning along with traditional academic performance to enhance prediction accuracy significantly. Most previous studies relied on academic records, but the outcomes of our work converge with current research streams that focus on competency-based education. This integration of outside-of-class activity, for example, the internship and skill development program, has been found to be a key determinant of employability, further improving the overall prediction of the placement outcome. All models, including Decision Tree, Random Forest, XGBoost, LSTM, CNN, and MLP, were evaluated in terms of precision. In this regard, XGBoost performed on average with an accuracy of 95%, significantly higher than the Decision Tree and Random Forest models that scored relatively lower. While LSTM and CNN are strong models, they did not provide any accuracy gain over XGBoost, indicating that these deep learning models are better suited for sequential data rather than the non -continuous nature of our dataset. On the other hand, MLP emerged as the best model with a maximum accuracy of 96%.

REFERENCES

[1] E. Alpaydin, “Introduction to machine learning,” MIT Press, 2020.

[2] Kelleher JD, Mac Namee B, D’arcy A. Fundamentals of machine learning for predictive data analytics:algorithms, worked examples, and case studies. MIT press; 2020 Oct 20.

[3] C. Beaulac and J. S. Rosenthal, “Predicting university students’ academic success and major using random forests,” Research in Higher Education, vol. 60, no. 7, pp. 1048–1064,

2019.

[4] Wang, C.; Xiao, Z. A DeepLearning Approach for CreditScoring Using Feature EmbeddedTransformer. Appl. Sci. 2022.

[5] R. Gupta and A. Sharma, "A survey on student performance prediction using machine learning techniques," Procedia Computer Science, vol. 173, pp. 350–358, 2020.

[6] Y. Li, C. Wu, and M. Zhang, "Enhancing job placement prediction with attention-based neural networks," in Proc. IEEE Int. Conf. Big Data, pp. 4521–4528, 2021.