IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Enhancing Student Career Outcomes With Predictive Analytics

Revathi Talari[1,a)], Pavan Kumar Potla[2,b)], Taslem Shaik[3,c)], Soniya Ubbala[4,d)], Venkata Sairam Goriparthi[5,e)]

[1]Assistant Professor, Department of Computer Science and Engineering, NRI Institute of Technology, Agiripalli- 521212, Vijayawada, Andhra Pradesh, India

[2,3,4,5]B. Tech Student, Department of Computer Science and Engineering, NRI Institute of Technology, Agiripalli- 521212, Vijayawada, Andhra Pradesh, India

Abstract: Student career prediction play an important role in assessing employability trends and enhancing career guidance in schools. This study applies advanced machine learning techniques like Logistic Regression, D T, R.F, SVM, and Neural Networks to predict student placement results from academic records, technical competencies, soft competencies, internships, and extracurricular activities. The system improves accuracy using data-driven information rather than traditional, human-centered evaluation methods. By automating the process, it minimizes human intervention, eliminates mistakes, and gives fair and impartial appraisals, providing equal opportunities for all students. The models are validated with accuracy, precision, recall, and F1-score to determine the best approach, with Decision Tree model achieving a staggering 91.25% accuracy, 89.19% precision, 89.73% recall, and 89.46% F1-score. It also helps students in identifying areas that require improvement on their part so that they may improve their technical as well as soft skills and prepare themselves for the job. Its scalability and flexibility allow it to be easily applied in various institutions, courses, and departments. It closes the gap between academia and industry, and thus maximizes placement plans for students, reduces unemployment, and allows institutions to tailor training courses to enhance the employability of students.

Keywords: Student Placement Prediction, M.L, Classification Models, Decision Tree, R.F, Naive Bayes, Employment Prediction, Data-Driven Approach, Career Development.

INTRODUCTION

Student placement is responsible for making or breaking the reputation and success of an educational institution. With rising competition in the job market, universities and colleges try to make their placement services better by determining the chances of students getting job offers. The conventional methods of predicting placement depend on subjective factors and manual judgment, which is not always precise. To overcome this problem, M.L models are employed to examine past student data and forecast placement results more accurately.

This model emphasizes creating a student placement prediction system based on M.L algorithms like D.T, R.F, and Naive Bayes. Based on different academic and demographic characteristics, the model will classify students into two groups "Placed" and "Not Placed." The dataset considers parameters like academic marks, past placement history, extracurricular activities, and technical skills to give a complete analysis. The system takes advantage of previous trends to forecast a student's chances of placement, allowing universities to make informed decisions.

The M.L model adopts a systematic process, such as data preprocessing, feature selection, model training, and model evaluation. Different classification metrics like accuracy, precision, recall, and F1-score are employed to measure the performance of the models. The Decision Tree and Random Forest models have been found to be highly accurate in predicting placement outcomes, providing consistent results. Random Forest's ensemble learning technique assists in minimizing overfitting and enhancing generalization, which makes it a robust choice for real-world applications.

LITERATURE SURVEY

Pal, A. K., & Pal, S. [1] (2013) In this study, EDM is employed to mine student data and predict placement results. With several classification algorithms executed using WEKA software, the research evaluates students' academic performance for training and placement. The proposed model indicates correlations between academic achievement and campus selection success. Out of the algorithms experimented with, Naive Bayes Classifier was the most accurate (86.15%), followed by Multilayer Perceptron (80%). The findings indicate the effectiveness of data mining techniques in educational**,** decision-making.

Meher, K., Gaikwad, R. L., Sangoi, J. A., Khot, A., Rana, M., Sall, S. and Patil, M. [2] (2024) study on Student Placement Prediction and Analysis for Improved Career and Institution Development. The article gives detailed insight regarding how student placement influences career and institution development. They used M.L models for forecasting the outcome of the input placements of the students based on their qualifications, age, and work experience. Three algorithms, namely Decision Tree, Naive Bayes, and R.F were utilized to study the pattern of the placements in the study. During the first experiment, the accuracy of the classification was 84.2 percent, which showed that the model was robust. The model's accuracy increased to 92.1 percent in the second experiment using the same data features.

Irene Treesa Jose et al. [3] (2020) pioneered the creation of a placement predictor. A predictor is a device that can determine the probability of a student placement in a company. Various determinants are used to construct it, such as university data, placement test scores, etc. One of the issues was the choice of the most appropriate classification algorithm, given that the model is training on data of previous students. The study compared four machine learning models, which were KNN, SVM, Logistic Regression, and Random Forest. The final result was the following: SVM accuracy rate was 100%; Logistic Regression came second at 97.59 percent. The finding may be useful for choosing the most suitable algorithm for the effectiveness of the placement prediction system.

P. Manimaran, G. Priyadharshini, and N. Yamuna Devi [4] (2022) investigates predicting student placement eligibility using data mining techniques. The authors built a predictive model to classify students into dream companies, super dream companies, or mass recruiter companies based on their prior academic and extracurricular achievements. The authors applied machine learning algorithms (e.g., KNN, Logistic Regression, Random Forest, and SVM) to examine historical applicant data to predict current student's placement opportunities. The final accuracy from the test data set for the Random Forest model was reported as 73.49%.

Ruparel and Swaminarayan [5] (2025) experimented with a number of M.L techniques in predicting outcomes from student placement. Models that were considered included Logistic Regression, Random Forest, Decision Tree, Naive Bayes, SVM, KNN, Gradient Boosting and LDA. The authors deemed KNN, SVM and Logistic Regression models to be at the top of the performance pyramid among the models used. The accuracy of the top three models was the same at 94%. The performance of the Decision Tree model and Naive Bayes was also worthy of note and it is important that when developing such a model that the model is selected carefully and that there is potential for improved performance on the model in the future as indicated by the authors of the report.

Maurya, Hussain, and Singh [6] (2022) deals with the supervised M.L classification models for predicting student’s placements vis-a-vis their skills in aptitude, coding, communication, and technical skills. The authors appraised the effectiveness of the following algorithms in predicting placement results: SVM, Gaussian Naive Bayes, KNN, R.F, Decision Tree, SGD, and Logistic Regression. The dataset is applied on random forest which gives 86% result and the decision tree gives 84% accuracy.

Imtiyaz Ahmad Magray and Gurinder Kaur Sodhi [7] (2024). The objective was to gain knowledge of advanced data analysis, pre-processing, and model selection. The learning model consists of the following algorithms:

Decision Trees, R.F, XGBoost, and KNN. As a preventive measure against overfitting, Randomized Search was chosen as the optimizer for the XGBoost algorithm. The model accuracies are as follows: Decision Tree 87.74%, Random Forest and XGBoost 87.60%, KNN 85.18%.

Patel, N. K. M., Goutham, N. M., Inzamam, K. A., Kandi, S. V. and Sharan, V. R. V. [8] (2022), Assistance to student success via campus recruitment are assisted by machine learning technique to aid in interacting and predicting from large data sets. They found that by employing the J48 algorithm, prediction accuracies of 97.27% were achieved compared to Naïve Bayes and multilayer perceptron at 85.92% and 94.94%, respectively. It also provides evidence of accuracy in the evidence-based practice to prepare students for placement.

According to Khandale, Bhoite, [9] (2019), the student placement system of predictor maximizes institute's reputation and significantly reduces the work of Training and Placement Office. Machine Learning can predict student's placement of the future based on the available data and past placement records. The model will improve its accuracy through proper data pre-processing and selecting correct features. Five classifiers namely Logistic Regression, SVM, KNN, Decision Tree and Random Forest as well as ensemble methods such as Bagging method, Boosting method, and Voting Classifier were deployed. The accuracy of 78% was obtained using XGBoost and AdaBoost.

In 2017, Thangavel, Bharathi, and Sankar [10] created a recommendation system that incorporated M.L that capitalized on student's past performance information ultimately leading to placement outcome being categorized into five categories AI, Dream Company, Core Company, Mass Recruiters, Not Eligible and Not Interested. The recommendation model was beneficial to the placement cells by determining which student selection would be most beneficial for information seeking improvement in skills. The recommendation model also provides feedback to students to indicate the likelihood or chances of being placed with respect to the required improvement skills. Specifically, the recommendation model provides feedback in soft and interpersonal skills, as well as technical skill. The recommendation model was ultimately beneficial in assisting an institution in improving the training provided to the students involved to help improve placement with a 71.66%.

Traditional Machine Learning Models for Placement Prediction

Classic machine learning methods for student placement prediction involve statistical and rule-based models that consider attributes like academic record, technical competencies, and extracurricular activities. Logistic Regression is popular because it is simple and can provide estimates of placement probability but does not handle complicated, non-linear patterns in data. K-Nearest Neighbors (KNN) makes placements based on similarity to the previously placed candidates, providing reasonably good accuracy at the cost of becoming computationally intensive with increases in dataset sizes. Decision Trees give a hierarchal mode of placement, but they easily overfit the data unless sufficient pruning is provided. Random Forest is an ensemble technique involving decision trees with better prediction values and less likelihood of overfitting but heavier computational requirements.

Navie Bayes, a probabilistic classifier, is computationally efficient for small datasets but relies on feature independence, reducing its performance for placement prediction. Support Vector Machines (SVM) are effective in high-dimensional spaces and are capable of processing non-linearly separable data, but they require high computational power for large dataset. Though these classic models are interpretable and computationally efficient, they tend not to encode complex relationships between placement factors. Therefore, newer methods like deep learning and ensemble methods have come into favor because they offer higher accuracy by identifying complex data patterns effectively.

Proposed System

The system proposed here is a M.L based Placement Prediction Model which attempts to do accurate and effective prediction on student's placement opportunity. Unlike the conventional systems which are purely academic records driven, this system integrates several parameters like software skills, extracurricular, work experiences, certificates among other and benchmarks them against the current market conditions. The decision trees, random forests, and logistic regression machine learning algorithms were employed in the model because they are capable of recognizing and accurately predicting placements. The major advantage with this system is that it is data-driven in nature. It captures and analyzes student data like scores at different academic levels, attendance, projects, internships, soft skills, among others. Empirical model building is used to analyze the resulting data and determine the relevant factors that lead to successful placements. The process requires the elimination of human prejudices and standardizes the evaluation for all students, allowing for objective predictions. The system further makes suggestions on what a particular learner should specialize in based on what and when the learner performs best. In case a student has a decreased likelihood of placement, the system recommends what corrections are required, i.e., taking extra courses, certifications, or improving communication skills. This encourages students to become proactive in enhancing employability prior to placement drives.

The advised system incorporates a real-time teacher, student, and placement officer inter-face. It provides forecasts regarding placements, industry information, and employment market trends. With the aid of this information, universities and colleges are able to modify training programs to coincide with industry expectations, therefore raising the overall employment rate. In this way, companies also benefit from the system by receiving suitable candidates in a shorter time, thus reducing the cost of the placement process. The recommended system represents, in conclusion, a more intelligent, open, and streamlined approach to predicting placement. It not only lowers the degree of error in predictions, but also equips students with the tools to fulfill the parameters for their employment. By closing the gap between educational institutions and the labor market, it enhances the standard of placements and assists institutions in maximally increasing the placement rate.

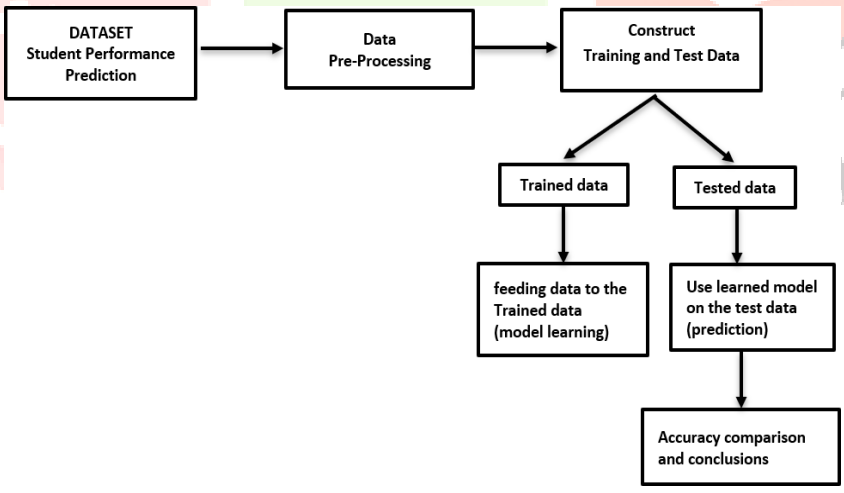


Figure 1: System Architecture for enhancing student career outcomes Machine Learning for enhancing student career outcomes

ML transformed student placement forecasting with evidence-based approaches in evaluating employability on the basis of academic achievement, technical competence, internship, and extracurricular activities. Statistical models were used earlier, while newer ML algorithms like Logistic Regression, Decision Trees, R.F, SVM, and KNN enable better and automated forecasts.

Data Security and Privacy

Privacy and security of data of ML-driven student placement forecasting are top priority to ensure that there is no unauthorized access to academic as well as personal data. Data anonymization, encryption (AES-256), and secure storage are some of the key practices to prevent unauthorized access.RBAC and MFA are additional security practices, whereas HTTPS and VPN ensure safe transportation of data.

Data Collection and Preprocessing

Data collection is done through accumulating student information concerning academic records, technical skills, soft skills, internship experiences, extracurricular activities, and previous placement results from institutional databases, questionnaires, and other publicly available sources. To ensure quality data, I undertook some preprocessing which includes dealing with missing values, removing duplicates, normalizing numbers, and encoding categorical features.

Integration and Real-Time Adaptation

Hybridizing machine learning-based student placement prediction with blockchain increases efficiency, security, and responsiveness. Secure storage and updating of real-time data from educational records, skill tests, and placement results can be done on a decentralized ledger, making it transparent and tamper-proof authenticated. Smart contracts on blockchain enable autonomous placement decisions through immediate qualification verification and matching of students to applicable opportunities. Machine learning algorithms keep learning and adjusting to new data by applying real-time feedback to refine accuracy when forecasting placement trends.

Blockchain Implementation

The integration of blockchain and machine learning for student placement prediction enhances the confidentiality, openness, and integrity of the system. Using several architectures at the same time, the information system implementation complexity becomes too high. Student's academic accomplishments, distributions, and even employment records are in a blockchain. Therefore, it is impossible to delete some information. Let us consider the student placement process. No third parties; the employer and educational institutions can check students' credentials and reduce possible cases of impersonation. The automated placement process is possible via the smart contracts that provide impartial computable job offers based on student credentials and machine learning rules of fairness and transparency.

METHODOLOGY

The student placement forecasting system proposed is systematic in approach, starting with the data collection phase, where students' data such as academic record, technical skills, and soft skills, as well as internship experience, are collected. Thereafter, preprocessing of data is done to remove redundant data, address missing values, convert categorical variables into numerical variables, and perform feature scaling. Different M.L algorithms such as Logistic Regression, Decision Trees, Random Forest, SVM, and Neural Networks are then used and trained using methods such as hyperparameter tuning. The models are tested for accuracy, precision, recall, and F1-score with comparative analysis to find the best algorithm. Placements drivers are used to give insights and suggestions on enhancing employability.

EXPERIMENTAL RESULTS

Accuracy: Accuracy is the simplest measure to use for classification. Accuracy is the proportion of all correct classifications, both positive and negative. Accuracy is calculated as sum of true negative and true positive to sum of total number of true negative, false negative, false positive, true positive.

$$\text{Accuracy} = (TP + TN) \,/\, (TP + TN + FP + FN)$$

Precision: Precision is an evaluation metric of model performance that tells you how many positive predictions the model creates and are correct. Precision is defined as a ratio of true positive predictions to true positive and false positive predictions.

Precision = $(TP) \ / \ (TP + FN)$

Recall: Recall is the number of true positive divided by the total number of true positive and false negative. It is the number of true positive instances to all actual positive instances.

Recall = $(TP) \ / \ (TP + FN)$

F1-Score: It is a harmonic mean between recall and precision. The value of the same exists in the range [0,1]. It likes to inform us about the accuracy (correctly classifies how many instances) and steadiness (will not leave out any significant number of instances).

F1 Score = $2 * (((precision * recall)/(precision + recall)))$

For comparison purposes, the performance of the Decision Tree model was 91.25% accuracy, 89.19% precision, 89.73% recall, and 89.46% F1-score. While the Decision Tree model's performance is good relative to the baseline model, the Random Forest model reveals a higher degree of accuracy, hence demonstrating the excellence of ensemble learning. Improved performance validates the efficacy of voting among the instances in the bootstrap sampling and bagged training implemented by Random Forest to produce more generalized and less variable predictions of student placement outcome.

Model Performance

This graph indicates accuracy of Random Forest model and number of Trees. X-axis shows the Number of Trees and Y-axis shows the accuracy of Random Forest. The graph begins first at between 0 to 25 trees then the accuracy will increase. The Highest accuracy located at 150 trees then gradually accuracy decreases. The lowest accuracy located at almost 10 trees.

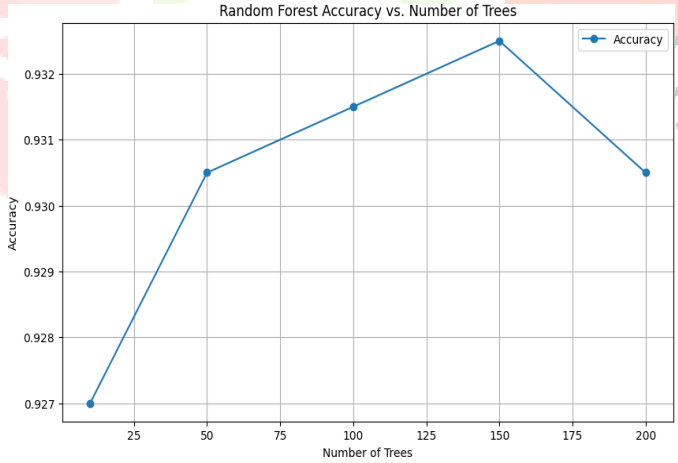


Figure 2: Model Accuracy Comparison

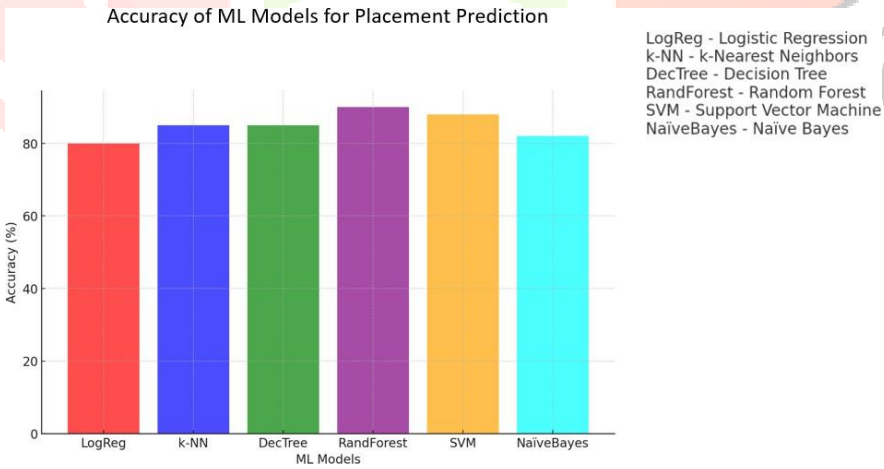
Comparison with Other Models

Placement prediction models have received a great deal of interest from researchers employing machine learning models, as every model has its advantages and disadvantages. Basic models, like logistic regression (Sharma et al., 2014), are simple to employ but need linear data. KNN (Sharma et al., 2021) model does work better in classification, although the expense of resources required to handle big data sets could be exorbitant. Decision trees and random forests (Surya et al., 2022) enhance the model's accuracy based on the feature importance, but they are high variance models and overfit if not regularized.

The Support Vector Machines (SVM) (Aravind et al., 2019) perform well when operating in the high dimensional spaces, but the Naive Bayes learning is rather inexpensive in terms of computational cost, but suffers from the curse of the independence assumption. For instance, the development of deep learning (Maragatham et al., 2024) has fundamentally boosted the level of accuracy to the extent that the algorithms predict more complex patterns as they can better predict things because they can learn to detect patterns more smoothly. For instance, the most divergent algorithms and highest rate of accuracy as well as the most optimal method to place efficiently in employment utilizes the Comprehensive Career Placement Predictor (K. J. et al., 2024).

Table:1 Comparison with Other Models

Model	Accuracy%	Advantages	Limitations
Logistic Regression	**80**	**Simple, interpretable, works for linear data**	**Struggles with complex, nonlinear patterns**
KNN	**85**	**Effective for classification, non Parametric**	**High computational cost for large dataset**
Decision Tree	**85**	**Captures lecture importance, easy to interpret**	**Overfitting without pruning**
Random Forest	**90**	**Handles imbalanced data, reduces Overfitting**	**Computationally intensive**
SVM	**88**	**Works well in high dimensional Space**	**Slow training for large datasets**
Naive Bayes	**82**	**Fast, Works well for smart datasets**	**Assumes feature independence**



CONCLUSION & FUTURE SCOPE

Random Forest model was utilized as part of the Prediction System and it provided more precise placement prediction than the other models, Decision Tree and Naive Bayes. Under the model building strategy, we have already preprocessed the model using the Random Forest algorithm. The model performance was maximized by preprocessing the data, scaling the features, and hyper parameter tuning. We have also examined the Misclassification specifically the False Positives. False Positive indicates that an incorrect variable has been chosen for the model construction and also it's not tuned. The graph of trees vs. accuracy indicates there is a point of diminishing returns in having more trees and the optimal trees to be used in Random Forest needs to be determined. Improved application of the ensemble methods and more sophisticated ensemble methods can yield even improved predictions. Feature engineering is another field, which can be used to make the models even more

accurate. Real placement patterns can also be utilized to make the model more accurate.

There are unlimited uses of the application of machine learning model and real time integration of data in professional predictions for the placement prediction system. It can even extend to developing the system as an end-to-end career counseling system, where the system suggests individual courses, internships, certifications etc. To students who will optimize the chances of being placed in a company. Industry and recruiter partnerships would allow for direct student-to-company mapping, making the recruitment process easier. Cloud and mobile compatibility of the system would enhance its accessibility, XAI would enhance explain ability and this system would be made into a powerful student career development and recruitment automation tool.

REFERENCES

[1] Pal, A. K., & Pal, S. (2013). Classification model of prediction for placement of students. I.J. Modern Education and Computer Science, 2013(11), 49–56. https://doi.org/10.5815/ijmecs.2013.11.07.

[2] Meher, K., Gaikwad, R. L., Sangoi, J. A., Khot, A., Rana, M., Sall, S., & Patil, M. (2024). Machine learning prediction techniques for student placement/job role predictions. Nanotechnology Perceptions, 20(6), 1027-1035. https://nano-ntp.com/index.php/nano/article/view/2840.

[3] Jose, I. T., Raju, D. Aniyankunju, J. A., James, J., & Vadakkel, M. T. (2020). Placement prediction using various machine learning models and their efficiency comparison. International Journal of Innovative Science and Research Technology, 5(9), 1005. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/5678.

[4] P. Manimaran, G. Priyadharshini, and N. Yamuna Devi, "Predicting the Eligibility of Placement for Students Using Data Mining," International Journal of Health Sciences, vol. 6, no. S8, pp. 460–467, 2022. [Online]. Available: https://doi.org/10.53730/ijhs.v6nS8.9723.

[5] M. Ruparel and P. Swaminarayan, "Enhanced Student Placement Prediction Using Machine Learning: A Comparative Evaluation of Algorithms," International Journal of Engineering Trends and Technology, vol. 73, no. 1, pp. 225–236, 2025. [Online]. Available: https://doi.org/10.14445/22315381/IJETT-V73I1P119.

[6] L. S. Maurya, M. S. Hussain, and S. Singh, "Machine Learning Classification Models for Student Placement Prediction Based on Skills," International Journal of Artificial Intelligence and Soft Computing, vol. 10, no. 3, pp. 194–207, 2022. [Online]. Available: https://doi.org/10.1504/IJAISC.2022.126337.

[7] Magray, I. A., & Sodhi, G. K. (2024). Predicting student placement in college using machine learning. International Journal for Research in Applied Science & Engineering Technology, 12(3). https://doi.org/10.22214/ijraset.2024.58548

[8] Patel, N. K. M., Goutham, N. M., Inzamam, K. A., Kandi, S. V., & Sharan, V. R. V. (2022). Placement prediction and analysis using machine learning. International Journal of Engineering Research & Technology (IJERT), 10(11). https://doi.org/10.17577/IJERTCONV10IS11051

[9] Khandale, S., & Bhoite, S. (2019, September). Campus placement analyzer: Using supervised machine learning algorithms. International Journal of Computer Applications Technology and Research, 8(9), 358-362. https://doi.org/10.7753/IJCATR0809.1004

[10] Thangavel, S. K., Bharathi, D. P., & Sankar, A. (2017, January). Student placement analyzer: A recommendation system using machine learning. In 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). https://doi.org/10.1109/ICACCS.2017.8014632.

[11] C. K. Raghavendra et al., "Identifying Patterns and Trends in Campus Placement Data Using Machine

Learning," International Journal of Education and Management Engineering, vol. 15, no. 1, pp. 10–24, 2025. [Online]. Available: https://doi.org/10.5815/ijeme.2025.01.02.

[12] M. S. Hassan and A. Goyal, "Predicting Student Placements Using Machine Learning Algorithms," Tuijin Jishu/Journal of Propulsion Technology, vol. 44, no. 4, 2023. [Online]. Available: https://doi.org/10.5281/zenodo.1234569.

[13] I. Portugal, P. Alencar, and D. Cowan, "The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review," arXiv preprint arXiv:1511.05263, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1511.05263.

[14] M. Sweeney et al., "Next-Term Student Performance Prediction: A Recommender Systems Approach," arXiv preprint arXiv:1604.01840, 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1604.01840.

[15] A. K. Pal and S. Pal, "Classification Model of Prediction for Placement of Students," International Journal of Modern Education and Computer Science, vol. 5, no. 11, pp. 49–56, 2013. [Online]. Available: https://doi.org/10.5815/ijmecs.2013.11.06.

[16] A. S. Sharma et al., "PPS - Placement Prediction System Using Logistic Regression," in IEEE International Conference on MOOC, Innovation and Technology in Education (MITE), 2014, pp. 202–207. [Online]. Available: https://doi.org/10.1109/MITE.2014.7020293.