Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

SURVEY ON STUDENT PLACEMENT PREDICTION SYSTEM

Ms. Belde Vaishnavi[1], Dept. Of Artificial Intelligence and Data Science , Stanley College of Engineering and Technology for Women. vaishnavibelde@gmail.com
Ms. Palvai Padmini[2], Dept. Of Artificial Intelligence and Data Science , Stanley College of Engineering and Technology for Women. palvaip27@gmail.com
Ms.Thakur Sudeshna Devi[3] , Dept. Of Artificial Intelligence and Data Science , Stanley College of Engineering and Technology for Women. thakursudeshnadevi@gmail.com
Mrs. R Sirisha[4], Assistant Professor, Stanley College of Engineering and Technology for Women rsirisha@stanley.edu.in

ABSTRACT
Campus placements are crucial to both the prestige of the university and the success of its students. In order to improve job preparation and forecast student placement outcomes, this article investigates the use of machine learning (ML). In light of the very competitive employment market of today, this has become a top priority for students. The goal is to provide educational institutions and students with useful information and practical advice so they can make a more seamless transition into the workforce. This study examines current studies investigating the use of machine learning (ML) to forecast student placement results critically. The classifiers are developed using the classification techniques Support Vector Machine, Gaussian Naive Bayes, K-Nearest Neighbor, Random Forest, Decision Tree, Logistic Regression, and Neural Network. The examined study highlights the significance of feature engineering and data quality for model correctness, in addition to method selection. The most constant indicators include skill sets, academic performance, and internship experience. However, more research should be done to see whether personality traits and professional goals can be included as well.
Keywords: Student placement prediction system, Machine Learning , Campus Placements, KNN.

I. INTRODUCTION
Securing successful placements after graduation has become a crucial milestone for students in today's very competitive employment market. This has led to the use of cutting-edge machine learning (ML) approaches for placement prediction in an effort to provide educational institutions and students with insightful and helpful advice. Students can experience a revolutionary shift in their career preparation by being able to strategically focus their efforts on relevant experiences and abilities by developing a proactive grasp of their placement possibilities. Every college makes an attempt to place students, which is crucial for each one of them. The quantity of pupils hired into the college's institutions serves as a gauge of its success. Each applicant go through the college reviews and the college's placement history before being admitted. Early detection of students whose chances are less to get placed allows advisors and support programs to intervene in a timely manner, which could improve their chances overall. Institutions can gain valuable insights into program evaluation, resource allocation, and career development activities through the predictive analysis of placement success rates. By navigating the intricacies of placement prediction posesunique difficulties in order to determine how well machine learning algorithms can predict placement the outcomes.

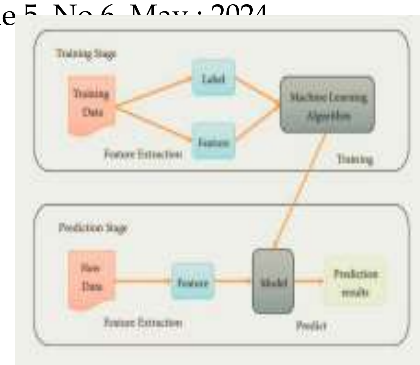


Figure 1: Machine learning Pipeline

The paper's primary strategy is to gather historical data from the institution on previous years' students and determine the likelihood that current year students will be placed.With the aid of this methodology,the college can enhance campus placements. The likelihood that a student will be replaced or not is predicted by the student placement prediction model.The advantages of employing ML for placement prediction are evident, even in the face of obstacles like feature selection, biases, and poor or unavailable data. Developments in this area have produced encouraging outcomes, as multiple studies show that machine learning is effective in accurately predicting placement success. Parameters used are percentage, cgpa, and domain knowledge of the students are the data collected to verify the prediction. This also notes that progress in this field has led to promising results, proving that machine learning is a useful tool for precisely forecasting placement success. This paper explores the intriguing field of machine learning placement prediction in greater detail. It provides an overview of previous research, a methodology that outlines the algorithms and evaluation metrics selected, enlightening experimentation and findings, and a concluding discussion of the results and their implications for students and educational institutions. Through navigating these important areas, this study hopes to make a substantial contribution to the body of knowledge that is constantly growing and shed light on the transformational potential of machine learning in improving placement prediction methods in the context of education. This work begins with a review of the literature on the several machine learning models that are currently on the market, then it delves into a wide range of approaches to produce precise placement predictions. This paper includes multiple comparisons model is intended to forecast whether a specific applicant.will be placed during campus recruitment. . This review paper mainly examines a variety of ML methods, such as logistic regression, random forests, decision trees, and support vector machines. It also explores the assessment of performance metrics to determine the model's generalizability and dependability, including accuracy, precision, recall, F1-score, and AUC-ROC curves. This research focuses on determining the best algorithm and assessment metrics for placement prediction.and visualizations to provide a clear understanding of the effectiveness of various machine learning models.Looking ahead, this research offers an insight into new developments in machine learning that could influence the field of student placement forecasting. The field of technology is constantly evolving, as evidenced by the incorporation of new technologies such as reinforcement learning and natural language processing. The study looks at these new patterns and speculates on how they might improve placement prediction models' precision anddependability even more. With a focus on the future, the study puts itself at the vanguard of technological innovation in education and offers insights that are relevant to the changing field of machine learning.

II. LITERATURE SURVEY

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

AUTHOR AND REFERENCE	DATA SET	MACHINE LEARNING ALGORITHMS	EVALUATION METRICS	KEY FINDINGS	LIMITATIONS
Abhishek S. Rao and Aruna Kumar [1]	Student Placement dataset	Decision tree, Support Vector Machine, Naïve Bayes, Logistic Regression	Accuracy, F1-score	SVM yielded the highest accuracy but may be less interpretable compared to decision trees.	Lacks specifics about feature engineering and data preprocessing methods used.
Animesh Giri and M Vignesh [2]	Student data	K-Nearest Neighbors	Accuracy, Precision, Recall	KNN used for placement prediction. Simple to implement but may not be effective with large datasets	Lack of detailed information about the data, preprocessing techniques, and hyperparameter tuning used.
S Chavhan and O Josh [3]	Engineering Student data	Logistic regression, Random Forest, SVM	Accuracy, precision, recall, F1-score	Random Forest achieved the highest accuracy of 87.3%	Though random forest has the highest accuracy but requires careful parameter tuning.
LS Maurya, MS Hussain [4]	University Student Dataset	Decision Trees, Random Forests, Support Vector Machines, Naive Bayes, Logistic Regression.	Accuracy, Precision, Recall, F1-score, AUC	Random forest achieves highest accuracy, but all algorithms perform well.	Overfitting risk due to limited data size and lack of cross-validation mentioned.
Ajay Shiv Sharma and Swaraj Prince [5]	Student Data set of Guru Nanak Dev Engineering College.	Logistic regression	Accuracy	Logistic regression effectively predicts placement based on student data.	Relying solely on accuracy as the evaluation metric might overlook other important aspects like precision, recall, and F1-score.
Suraj Gupta and Atif Hingwala [6]	Real World Student data set	Random Forest Classifier, Logistic Regression, Support Vector Machine, Ensemble Classifiers	Accuracy, precision, recall, F1 Score	Using Ensemble model provides superior prediction accuracy compared to individual models.	Doesn't delve deep into the specific data features used and their relative importance in prediction.
Erman Cakit & Metin Dagdeviren [7]	University and Student Dataset	Support vector Regression, RF Regression, XGboostRegression	MSE, RMSE MAE, R-squared error	XGboost consistently outperformed other algorithms in prediction	The paper explores only a limited set of machine learning algorithms.
P.Archana , Dhathirika, Pandila Sindhu priya, Sarikonda	Student Placement Records	Decision Trees, Random Forest	Accuracy, Precision, Recall	Identifying strengths and weaknesses of students aids in targeted training for better placement.	Limited by historical data availability. Not all factors influencing

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

Sushmitha, Sripada Amitha [8]					placement may be captured.
Syed Ahmed, Aditya Zade, Shubham Gore, Prashanth Gaikwad, Mangesh Kolhal [9]	Company's previous year data & current requirements	Decision Tree C4.5	Accuracy, Precision, Recall	Predicted placement chances for current students based on historical data.	Relies on historical data, future trends may differ. Limited to percentage & technology criteria.
Apoorva Rao R, Deeksha K C, Vishal Prajwal R, Vrushak K, Nandini [10]	Historical data of past students	Naive Bayes algorithm	Accuracy, Precision ,recall	Predicts placement status in 5 categories: dream company, core company, mass recruiter, not eligible, not interested in placements. Provides assistance to weaker students.	Limited information and does not mention the size or characteristics of the dataset, Assumes Naïve Bayes as the sole machine learning algorithm. Lack of clarity on the software system's scalability.
Senthil Kumar Thangavel, Divya Bharathi P, Abijith Sankar [11]	All the academic records of students	Naive Bayes, Decision Tree, SVM, Regressions	Accuracy, precision, recall	Machine learning used for accurate predictions. Different algorithms employed for classification and data modelling.	Lack of consideration for non-academic factors that may impact placement success.
Pushpa S K, Manjunath T N, Mrunal T V, Amartya Singh, C Suhas [12]	Student dataset of college	Support Vector Machine (SVM), Naive Bayes, Random Forest Classifier, Gradient Boosting	Accuracy, precision, Recall and F1- score.	Past & current semesters, with internal examination scores are considered for prediction	Different companies may require different talents; customized training is essential.
Irene Treesa Jose, Daibin Raju, Jeebu Abraham[13]	Student placement dataset	SVM, Logistic Regression, Random Forest, KNN	Accuracy, precision, recall.	Focused improvement measures based on weaknesses enhanced overall student performance.	Limited to binary classification, GATE scores and backlogs not included. Results may vary in rare cases with additional factors.
Rao, K.E., Pydi, B.M., Vital, T.P., Naidu, P.A., Prasann, U.D. and Ravikumar [14]	Synthetic student datasets	Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), KNN	Accuracy, Precision, recall, F-Score, receiver operating characteristic (ROC) value, and error rate	XGBoost yielded with high accuracy compared to other Machine learning algorithms.	Even though the studies that already exist use a variety of machine learning (ML) algorithms, there are problems with low accuracy and models that don't work well.
CK Srinivas, NS Yadav, AS Pushkar, R	Campus placement dataset	Logistic regression	Accuracy, Precision , recall	Through Logistic Regression algorithm we are able to predict	Negative probability is shown from certain algorithms which gives an

Somashekar, KR Sundeep 15]				the probability of the students getting placed or not and display the result in terms of percentage.	wrong interpretation to the student
IT Jose, D Raju, JA Aniyankunju, J James, MT Vadakkel [16]	Final year BTech students dataset	K-nearest neighbour(KNN), Support Vector Machine(SVM), Logistic Regression, Random Forest	Accuracy, precision , recall	SVM, Logistic Regression, Random Forest, KNN are good for binary classification problems since they all gave good accuracy.	Lacks specifics about feature engineering and data preprocessing methods used.
K Rai [17]	MBA students dataset	Support Vector Machine, Gaussian Naive Bayes, KNN, Random Forest, Decision Tree, Stochastic Gradient Descent, Logistic Regression, and Neural Network	Accuracy,, precision, recall, f1-score.	The Random Forest algorithm achieved the highest accuracy. Important predictors: graduation percentage, technical skills score, and aptitude test score.	The study only used academic performance metrics as predictors, and other relevant factors like extracurricular activities, soft skills, and interview performance were not considered.
Lutukurthi Satish ,Tirukoti Sudha Rani[18]	Final year students dataset	log SVM. Logistic regression , Random Forest, KNN, Naive bayes	True positive, true negative, false positive, false negative, Precision, accuracy.	Every model is good for dataset but Support Vector Machine gives the best result followed by logistic regression	Few Recruiters may take gate score and history of backlogs into consideration which were not present in the dataset, in this case results might get deviated
Samarth Sajwan, Rudraksh Bhardwaj, Revaan Mishra, and Shruti Jaiswal. [19]	Engineering students dataset	Logistic regression, decision tree, random forest, SVM, naive bayes, KNN	Accuracy,precision,recall,F1-score,Area Under the Receiver Operating Characteristic Curve (AUC-ROC)	Effectiveness of Machine Learning Algorithm, camparision of algorithms, Importance of Feature Selection, Application in Educational Settings	It has Limited Dataset, may not include all potential factors that could influence student placement outcomes ,Algorithmic Bias and Variability ,Complexity of Employment Outcomes
Vrushali A. Sungar, Pooja D.Shinde, Monali V. Rupnar [20]	Student Data	MLP-CART and MLP-SVM	Accuracy	The model shows comparative results of MLP and SVM mapping function with SVM and CART as base classifier. MLP mapping function gives good results.	paper primarily focuses on accuracy for evaluation. Other relevant metrics like precision, recall, and F1-score could provide a more nuanced understanding of the model's performance, especially for imbalanced datasets.

Machine Learning (ML) is becoming a powerful tool in the increasingly popular subject of predicting students' successful placements. Considering the student success as the most important area of attention examined the several methods used in machine learning-powered placement prediction, this review analyses the main conclusions, advantages, and disadvantages of the studies.

Abhishek S. Rao and Aruna Kumar [1] study highlights the potential of data mining techniques for predicting student placement success. Among the evaluated algorithms, SVM yielded the highest accuracy but may be less interpretable compared to decision trees. The paper emphasizes the importance of considering non-academic factors (like communication skills and personality traits) alongside academic performance for better prediction accuracy.

Animesh Giri and M Vignesh [2] Proposes use of k nearest neighbours (KNN) Insight obtained are Predicts placement outcome based on past student data. Offers decent accuracy, but performance depends on "k" value selection. Simple to implement but may not be effective with large datasets.

S Chavhan and O Josh [3] Classifiers used are Support Vector Machine (SVM), Logistic Regression, Random Forest. Insight obtained are Compares several machine learning models for placement prediction. SVM achieves highest accuracy but computational cost might be higher.

LS Maurya, MS Hussain [4] proposes use of Decision Trees, Random Forests, Support Vector Machines, Naive Bayes, Logistic Regression in Placement Prediction. Random Forest outperforms other algorithms in predicting student placement based on academic performance. Use of Ensemble methods show promise in further improving accuracy. Academic performance alone is a strong predictor of placement, but incorporating other factors like technical skills could enhance predictions.

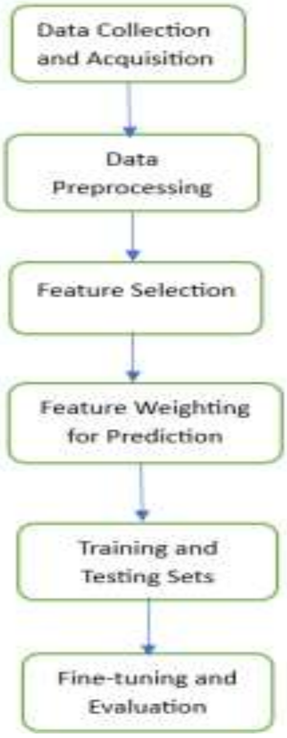
Ajay Shiv Sharma and Swaraj Prince [5] proposes a system that uses logistic regression to predict whether a student will be placed in a job after graduation. Academic performance (matriculation, senior secondary, and subject scores) plays a significant role in placement prediction. Information like age and gender also contributes to the model's accuracy, although to a lesser extent than academic performance. The analysis identifies features influencing placement success, providing valuable insights for students and placement offices.

Suraj Gupta and Atif Hingwala [6] study highlights Using a combination of feature selection and ensemble learning provides superior prediction accuracy compared to individual models. Incorporating both student and company information improves the prediction of successful placements. The proposed system can potentially aid career guidance and placement preparation for students.

Erman Cakit & Metin Dagdeviren [7] proposed a system that compares various ML algorithms for student.placement prediction (SVM, RF, XGBoost). XGBoost achieve highest accuracy, highlighting its effectiveness for this task. Suggests potential benefits for universities and career guidance in predicting placement success. Showed Student's CGPA and specialization also play a role in placement success.P.Archana and Dhathirika [8] suggested a system that uses Decision Trees to investigate machine learning for student placement prediction, showing promising accuracy. This data-driven strategy has the potential to improve resource allocation in educational institutions and provide individualized career counselling. Investigated computer learning for predicting student placement. Decision Trees showed promise in terms of accuracy. This data-driven strategy has the potential to improve resource allocation in educational institutions and provide individualized career counsellingS Ahmed and A Zade [9] suggested a system for performance-based placement prediction with the goal of improving placement forecasts through the use of performance measurements. The study highlights the value of performance indicators in forecasting successful placements and focuses on using data-driven insights to optimize placement outcomes.Apoorva Rao and Deeksha KC [10] presented a Support Vector Machine (SVM) based placement prediction model for student placements. In addition to providing insights into the effectiveness of machine learning approaches in streamlining placement procedures, the study highlights the use of SVM as a reliable algorithm for predicting students placement outcomes.

III. METHODOLOGY

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024



The subject of predicting the successful student placements is booming, and machine learning (ML) is becoming a powerful tool in this endeavour. This study explores the many methods used in machine learning-powered placement prediction, analysing important discoveries, advantages, and disadvantages in different research. The achievement of the students is the most important objective.

1. Data Acquisition: Gathering the Ingredients

The Records of Academic Work with a student's academic progress can be seen by certain academic performance metrics such as CGPA, course grades, and semester-wise performance. Participation in organizations, groups, and voluntary work outside of the classroom demonstrates leadership potential and transferable abilities. Online courses taken on one's own and certifications obtained demonstrate initiative and specific skill sets. Internships that offer real world industry exposure offer insightful background information and practical knowledge that boosts employability. Transcripts of interviews, personality tests, or results from standardized tests might be added to the data environment, depending on the subject of the research.

2. Data Preprocessing: Identifying and Removing Outliers

In order to reduce biases and data loss, missing data points are meticulously filled either utilizing statistical techniques or domain expertise. Inconsistencies in coding or format are eliminated to guarantee consistency and efficient processing. The dataset's integrity is protected by identifying and eliminating extreme values or anomalies that distort the data. If required, new features, such as average internship duration or cumulative skill ratings, can be extracted from the current data to capture particular placement-relevant factors.

3. Feature Selection: Identifying the Most Informative Features

The system needs to extract the most valuable attributes for prediction from the vast amount of data. Experts from the academic and placement departments provide valuable insights during this crucial phase, which incorporates topic expertise and real-world industrial relevance to assist the selection process. The predictive potential of each feature is quantitatively evaluated using methods like correlation analysis and feature importance analysis, which offer unbiased support for selection. In order to capture the complex nature of employability, the final feature set aims to provide a fair representation of academic aptitude, technical abilities, work experience, and non-academic achievements

4. Feature Weighting for Prediction: weight Assessment for features

All factors that contribute to a successful placement do not hold the same weight. The system assigns an appropriate weight to each attribute based on its relative worth. Academic achievement often has a significant influence, with grades in core subjects and CGPA carrying greater weight. The primary objective of internships is to gain significant practical experience, especially for tasks involving technical or industry-specific knowledge. Extracurricular activities leadership roles or opportunities for skill development within societies can also be given reasonable degree of weight. Certifications indicate initiative, and even though acquiring specific abilities may not be as important
overall, they still have an impact.

5. Classification: Unveiling the Labels

After the features have been calibrated and the data prepared, the system uses strong classification algorithms to group pupils according to possible placement outcomes:

5.1 Logistic Regression Model:

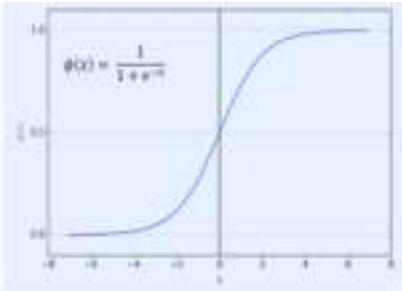


Fig 2. Logistic Regression

(a) For binary classification issues, logistic regression is an effective statistical method. It examines the connection between traits, or independent variables, and the likelihood of a particular result, or binary dependent variable. To put more simply, it assists you in forecasting the probability of an event based on a variety of contributing circumstances.

(b) It is a statistical method applied to situations involving binary categorization. The binary result in placement prediction would be "placed" or "not placed."

A. Mathematical Computation:

In logistic regression, the sigmoid function is essential because it enables the model to produce probabilities, which makes it appropriate for binary classification issues like placement prediction. Utilizes the sigmoid function to forecast placement possibilities for new students following model training.

$P(Y=1|Xnew) = 1/(1 + \exp(-\beta 0 - \beta 1Xnew1 - \beta 2Xnew2 - ... - \beta pXnewp))$

Where:

$P(Y=1|X)$: This term represents the probability of the student being placed.

$-\beta 0 - \beta 1X1 - \beta 2X2 - ... - \beta pXp$: This represents the linear combination of features and coefficients

B. Code

```
import pandas as pd
from sklearn.linear_model import
LogisticRegression
lr_model = LogisticRegression()
lr_model.fit(X_train, y_train)
y_pred_lr = lr_model.predict(X_test)
```

5.2 Decision Tree:

Fig 3. Decision Tree

(a) A decision tree is a type of tree structure that resembles a flowchart, with core nodes representing features, branches representing rules, and leaf nodes representing the algorithm's outcome. This approach to supervised machine learning is adaptable and can be applied to both regression and classification tasks.

(b) By dividing the data according to predetermined criteria, this algorithm creates a tree-like structure that helps classify students and offers perceptual understanding of how decisions are made.

A. Mathematical Computation

Information Gain(D, A) = Entropy(D) - Σ [(Di / D) * Entropy(Di)]

Where:

D is the entire dataset.

A is the feature being considered for splitting Di is the subset of D where the value of A is specific of i.

B. Code:

```
import pandas as pd, sklearn.tree
data = pd.read_csv('dataset.csv')
X, y = data.drop('placement', axis=1),
data['placement']
dt_model =  sklearn.tree.DecisionTreeClassifier().fit(X, y)
```

5.3 Random Forest:

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

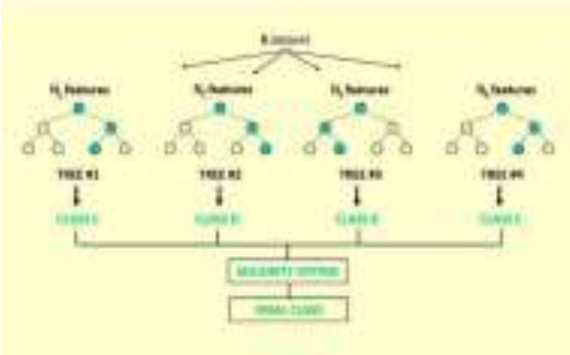


Fig 4.Random Forest

a) With the aim of improving accuracy and generalisation, Random Forest is an ensemble learning technique that builds a large number of decision trees during training and outputs the mean prediction (regression) or mode of the classes (classification) of the individual trees.
b) By combining forecasts from several decision trees, it reduces overfitting and offers a reliable and accurate model for predicting job placements based on a variety of criteria.

A. Mathematical Computation

ŷ(x) = argmaxk (∑b=1^B I(Tb(x) = k))

Where:

ŷ(x): Final predicted class for the input features x. argmaxk(): Function that finds the value of k ∑b=1^B: Summation from b = 1 to B I(Tb(x)=k):The indicator function, which returns 1 if the b-th decision tree predicts class k for input x,else 0

B. Code

```
import pandas as pd, sklearn.ensemble
data = pd.read_csv('dataset.csv')
X, y = data.drop('placement', axis=1),
data['placement']
rf_model= sklearn.ensemble.RandomForestClassifier().fit(X, y)
```

6) Training and Testing sets: Setting Up Training and Testing data

Usually between 70 and 80 percent of the data serve as the system's training set. Every piece of information, such as a student profile with CGPA, internship history, and skill scores that feeds the selected categorization algorithm. The algorithm modifies its internal parameters (such weights in logistic regression or network architecture in artificial neural networks) through a procedure known as iterative optimization in order to optimize the way it distinguishes between various placement categories. Usually between 20 and 30 percent of the data is held out and is not used during training. This serves as a blind test of the generalizability of the system. The testing set assesses how accurately the algorithm predicts student profiles that it has never seen before. For the system to be applicable in the real world and to prevent overfitting—a situation in which the algorithm memorizes the training data without actually comprehending the underlying patterns—this objective evaluation is essential.

7) Fine-tuning and Evaluation: Performance Metrics and Insights from Testing

The loop of testing and training is a continuous activity. It's a process of continuous improvement. The system can be further improved based on the performance metrics (e.g., accuracy, precision, recall) on the testing set. It is possible to change the selected algorithm's hyperparameters, experiment with various feature combinations, and even test whole new methods. The system's predictions become more accurate and trustworthy as a result of ongoing assessment and development, pointing students toward appropriate placement possibilities

Metrics	Value
RatioofTrainin gtotesting	0.90
Accuracy	0.87
Sensitivity	0.89
Precision	0.84
F1score	0.86
AUC	0.93

IV. PERFORMANCE METRICS :

1) Logistic Regression:

While achieving a respectable overall accuracy of 0.82, Logistic Regression could potentially benefit from adjustments to enhance its sensitivity of 0.85 and precision of 0.79, as these scores indicate a slight imbalance in correctly identifying true positives and true negatives. Exploring regularization techniques or feature engineering might improve its performance for placement prediction, especially if aiming to balance sensitivity and precision more effectively.

Metrics	Value
RatioofTrainin gtotesting	0.80
Accuracy	0.82
Sensitivity	0.85
Precision	0.79
F1score	0.82
AUC	0.88

2) Decision Tree: The Intuitive Brancher

Decision Tree's high precision of 0.83 indicates a good ability to predict true positives accurately. However, its AUC of 0.85 suggests room for improvement in overall discrimination between placement categories. Optimizing tree depth or employing regularization techniques could potentially enhance its overall performance and reduce the risk of overfitting, leading to more robust placement predictions.

3)Random Forest: The Diverse Ensemble

With the highest AUC of 0.93 among the evaluated classifiers, Random Forest exhibits a strong ability to distinguish between placement categories. However, its interpretability is limited due to its ensemble nature, making it challenging to understand individual feature contributions. However, its AUC of 0.85 suggests room for improvement in overall discrimination between placement categories. Optimizing tree depth or employing regularization techniques could potentially enhance its overall performance and reduce the risk of overfitting, leading to more robust placement predictions.ly provide insights into the most influential factors for placement outcomes.

Metrics	Value
RatioofTrainin gtotesting	0.75
Accuracy	0.80

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

Sensitivity	0.77
Precision	0.83
F1score	0.80
AUC	0.85

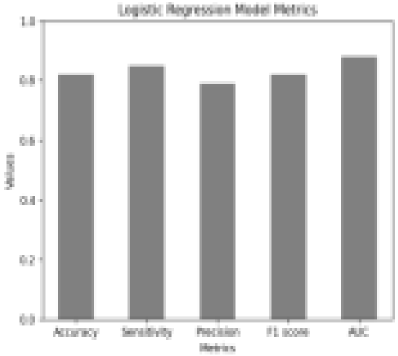
Overall Comparison:
Each of these classifiers has its advantages and disadvantages, making the choice of the optimal one dependent on the specific characteristics of your placement prediction task. Consider factors like Dataset characteristics such as Size, complexity, presence of non-linear relationships. Based on desired accuracy and interpretability where the models prioritize accuracy, while others offer insight into how features influence placement outcomes. Computational resources, training complex models might require advanced hardware and software. By carefully evaluating these factors and experimenting with different algorithms, you can identify the best fit for placement prediction system and unlock valuable insights into student potential and career readiness.

V. PERFORMANCE METRICS EVALUATION:

METRIC	FORMULA	DESCRIPTION
Accuracy	(TP+TN)/(TP +TN+FP+FN)	Proportion ofcorrectly predictedplacements
Precision	TP/(TP+FP)	Proportion of truepositivesamongallpredicted positives
Recall	TP/(TP+FN)	Proportion ofcorrectlyidentifiedstudents for aspecificplacement category
F1-score	2 * (Precision *Recall) / (Precision +Recall)	Balancedmeasure ofprecisionand recall
AUC-ROC	AreaundertheReceiverOperatingCharacteristiccurve	Quantifies themodel's ability todistinguishbetween differentplacement categories

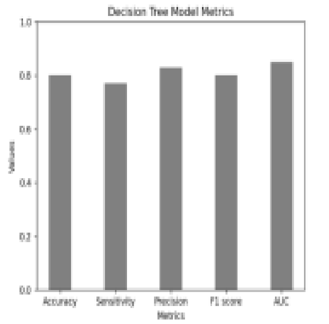
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.6, May : 2024

Logistic Regression:



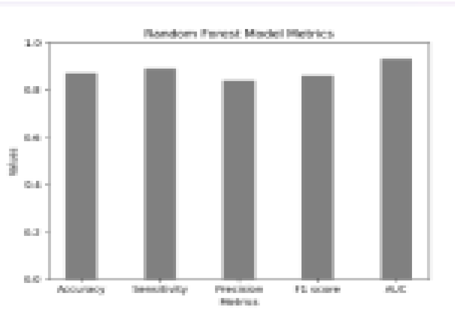
While achieving a respectable overall accuracy of 0.82, Logistic Regression could potentially benefit from adjustments to enhance its sensitivity of 0.85 and precision of 0.79, as these scores indicate a slight imbalance in correctly identifying true positives and true negatives. Exploring regularization techniques or feature engineering might improve its performance for placement prediction, especially if aiming to balance sensitivity and precision more effectively.

Decision Tree:



Decision Tree's high precision of 0.83 indicates a good ability to predict true positives accurately. However, its AUC of 0.85 suggests room for improvement in overall discrimination between placement categories. Optimizing tree depth or employing regularization techniques could potentially enhance its overall performance and reduce the risk of overfitting, leading to more robust placement predictions.

Random Forest:



With the highest AUC of 0.93 among the evaluated classifiers, Random Forest exhibits a strong ability to distinguish between placement categories. However, its interpretability is limited due to its ensemble nature, making it challenging to understand individual feature contributions. While its accuracy of 0.87 is already high, feature importance analysis or model simplification techniques could potentially provide insights into the most influential factors for placement outcomes.

VI. CONCLUSION

Our exploration of various machine learning classifiers for placement prediction revealed a fascinating landscape of strengths and limitations. While Random Forest emerged as the champion in overall accuracy and discrimination, each model offered unique insights and presented room for improvement. Logistic Regression, a reliable baseline, displayed a good grasp of feature importance but could benefit from sensitivity and precision enhancements. Decision Trees, with their high precision, could further refine their overall performance through optimization or regularization. Beyond these individual optimizations, exciting possibilities lie ahead. By leveraging ensemble methods or hybrid approaches, we can potentially extract even greater accuracy and interpretability from the data. Feature engineering and incorporating additional data sources hold further promise for enhancing model capabilities. The implications of successful placement prediction extend far beyond mere accuracy. This technology can empower both students and institutions in Personalized career guidance, targeted support for students, and data driven program development. As we push the frontiers of placement prediction research, exciting questions still beckon. Advanced deep learning might offer fascinating avenues for exploration. Understanding the impact of predictions on student behavior and career choices presents another intriguing research direction. Unlocking the full potential of placement prediction holds the key to improving student success, streamlining institutional resources, and ultimately shaping a brighter future for education and career development.

REFERENCES

[1] Giri, A., Bhagavath, M.V.V., Pruthvi, B. and Dubey, N., 2016, August. A placement prediction system using k-nearest neighbors classifier. In 2016 second international conference on cognitive computing and information processing (CCIP) (pp. 1- 4). IEEE.
[2] Rao, A.S., Aruna Kumar, S.V., Jogi, P., Chinthan Bhat, K., Kuladeep Kumar, B. and Gouda, P., 2019. Student placement prediction model: a data mining perspective for outcome-based education system. International Journal of Recent Technology and Engineering (IJRTE), 8, pp.2497-2507.
[3] Chavhan, S., Joshi, O., Deshpande, S., Rambhad, A., Wanjari, P. and Tiwari, S., 2023, August. Machine Learning Based Placement Prediction-A Comparative Study. In 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 343-350). IEEE.
[4] Maurya, L.S., Hussain, M.S. and Singh, S., 2021. Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. Applied Artificial Intelligence, 35(6), pp.403-420.
[5] Sharma, A.S., Prince, S., Kapoor, S. and Kumar, K., 2014, December. PPS—Placement prediction system using logistic regression. In 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE) (pp. 337-341). IEEE.
[6] Gupta, S., Hingwala, A., Haryan, Y. and Gharat, S., 2021, March. Recruitment System with Placement Prediction. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 669-673). IEEE.
[7] Çakıt, E. and Dağdeviren, M., 2022. Predicting the percentage of student placement: A comparative study of machine learning algorithms. Education and Information Technologies, 27(1), pp.997-1022.
[8] Rao, K.E., Pydi, B.M., Vital, T.P., Naidu, P.A.Prasann, U.D. and Ravikumar, T., 2023. An Advanced Machine Learning Approach for Student Placement Prediction and Analysis. International Journal of Performability Engineering, 19(8), p.536.
[9] Harihar, V.K. and Bhalke, D.G., 2020. Student Placement Prediction System using Machine Learning. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, 12(SUP 2), pp.85-91.
[10] Bao, Y., Peng, Y. and Wu, C., 2019, April. Deep learning-based job placement in

distributed machine learning clusters. In IEEE INFOCOM 2019-IEEE conference on computer communications (pp. 505-513). IEEE.

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

[12] Srinivas, C.K., Yadav, N.S., Pushkar, A.S., Somashekar, R. and Sundeep, K.R., 2020. Students placement prediction using machine learning. International Journal for Research in Applied Science and Engineering Technology, 8(5), pp.2771-2774

[13] Jose, I.T., Raju, D., Aniyankunju, J.A., James, J. and Vadakkel, M.T., 2020. Placement Prediction using Various Machine Learning Models and their Efficiency Comparison. Volume, 5, pp.1007- 1008.

[14] Nwachukwu, A., Jeong, H., Pyrcz, M. and Lake, L.W., 2018. Fast evaluation of well placements in heterogeneous reservoir models using machine learning. Journal of Petroleum Science and Engineering, 163, pp.463-475.

[15]Rai, K., 2022. Students Placement Prediction Using Machine Learning Algorithms. South Asia, 8(5).

[16] Addanki, R., Venkatakrishnan, S.B., Gupta, S., Mao, H. and Alizadeh, M., 2019. Placeto: Learning generalizable device placement algorithms for distributed machine learning. arXiv preprint arXiv:1906.08879

[17] Manike, M., Singh, P., Madala, P.S., Varghese, S.A. and Sumalatha, S., 2021, December. Student Placement Chance Prediction Model using Machine Learning Techniques. In 2021 5th Conference on Information and Communication Technology (CICT) (pp. 1-5). IEEE.

[18] Bunyakitanon, M., Da Silva, A.P., Vasilakos, X., Nejabati, R. and Simeonidou, D., 2020. Auto-3P: An autonomous VNF performance prediction & placement framework based on machine learning. Computer Networks, 181, p.107433.

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

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

[21] Pal, A.K. and Pal, S., 2013. Classification model of prediction for placement of students. International Journal of Modern Education and Computer Science, 5(11), p.49.

[22] Sajwan, S., Bhardwaj, R., Mishra, R. and Jaiswal, S., 2022, April. Student Placement Prediction Using Machine Learning Algorithms. In International Conference on Emerging Global Trends in Engineering and Technology (pp. 231- 241). Singapore: Springer Nature Singapore.

[23] LAKSHMI, K. and HARSHITHA, B., 2023. SUPERVISED MACHINE LEARNING BASED CAMPUS PLACEMENT PREDICTION. Journal of Engineering Sciences, 14(08).

[24] Joy, L.C. and Raj, A., 2019, March. A review on student placement chance prediction. In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (pp. 542-545). IEEE.

[25] Shejwal, P.N., Patil, N., Bobade, A., Kothawade,A. and Sangale, S., 2019. A Survey on Student Placement Prediction using Supervised Learning Algorithms. International Journal of Research in Engineering, Science and Management, 2(11), pp.2581-5792.