International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Student Placement Prediction Using Various Machine Learning Techniques

Milind Ruparel*[1], Dr. Priya Swaminarayan[2]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: When it comes to helping students achieve their goals, campus placement stands out as a crucial factor in every educational institution's consideration. Each student shares the common objective of graduating from college with a job offer in hand. To address this, a predictive model has been developed for this study, aimed at determining a student's likelihood of securing placement. The main objective of this research is to analyze historical data from the previous academic year, forecast placement opportunities for current students, and support efforts to increase the percentage of successful placements within institutions. The study also aims to propose a recommendation system that predicts whether an existing student will be placed. Four distinct machine learning classification algorithms have been utilized for this purpose: the K-Nearest Neighbors (KNN) algorithm, logistic regression algorithm, random forest algorithm, and Support Vector Machine (SVM) algorithm. These algorithms independently predict outcomes, and their efficiency is evaluated based on the dataset used. The ranking of efficiency is determined by the dataset's characteristics. This research contributes to identifying students with academic potential, enabling them to focus on and improve both their technical and social skills, thereby enhancing their chances of success in securing placement.

Keywords: *Student Placement, Categorical Encoding, Machine Learning,* Feature Importance, Voting Classification.

1. Introduction

In the realm of higher education, one of the pivotal metrics for evaluating a university's efficacy is its placement rates. These rates serve as a tangible indicator of the institution's ability to prepare and propel its students into the professional sphere successfully. Prospective students, too, keenly scrutinize these statistics provided by admissions officers, as they seek assurances regarding their prospects upon graduation. Consequently, the focus of this paper revolves around the intricate task of forecasting and analyzing the demand for college placements, a pursuit that not only contributes to the developmental trajectory of academic institutions but also significantly impacts the career trajectories of their graduates.

To achieve a nuanced understanding of placement dynamics, sophisticated classification algorithms such as K-Nearest Neighbors (KNN), Logistic Regression, Decision Trees, and Support Vector Machines (SVM) have been deployed within the Placement Prediction system. These algorithms play a pivotal role in generating predictions pertaining to the likelihood of an undergraduate securing employment with a prospective organization during campus recruitment drives. Central to this predictive framework is the comprehensive evaluation of a student's academic journey, encompassing factors such as grades, attendance records, and credit accumulation. Leveraging data from previous academic cycles, these algorithms are meticulously analyzed to refine their predictive accuracy [1].

The landscape of higher education has witnessed a surge in the proliferation of universities and colleges, each equipped with dedicated placement offices committed to facilitating graduates in securing fulfilling and remunerative careers. Notably, enhancing students' competitiveness in the job market stands as a paramount objective for numerous academic institutions. However, as the standards of educational entities ascend, the task of crafting precise placement prognostications becomes increasingly intricate. Addressing this challenge necessitates the infusion of novel insights into educational processes and entities within management systems, thereby augmenting overall quality.

Machine learning techniques, adept at mining vast repositories of both contemporary and historical data, serve as indispensable tools in this endeavor. By harnessing information gleaned from databases maintained by academic institutions, these techniques enable a comprehensive analysis of placement trends and patterns, empowering institutions to make informed decisions aimed at bolstering students' employability and success in the professional realm [2].

2. Literature Study

Several recent studies have delved into the domain of student placement prediction using machine learning

[1] *Research Scholar, Faculty of Information Technology & Computer Science, Parul University, Vadodara, Gujarat, India*
Email: milind.ruparel@gmail.com
ORCID ID : 0000-0002-2231-2019
[2] *Dean, Faculty of Information Technology & Computer Science, Parul University, Vadodara, Gujarat, India*
Email: priya.swaminarayan@paruluniversity.ac.in
ORCID ID : 0000-0002-2535-2745

techniques. Ambili and Abraham [1] conducted a comprehensive evaluation of employability prediction employing ensemble learning techniques, showcasing the efficacy of such approaches. El Mrabet and Moussa [2] proposed a framework for predicting academic orientation utilizing supervised machine learning methods, highlighting the potential of these techniques in educational settings.

In a study by Baffa, Miyim, and Dauda [4], machine learning algorithms were explored for predicting students' employability, contributing to the growing body of research in this area. Similarly, Parida, Kumarpatra, and Mohanty [8] investigated the prediction of employment recommendations using machine learning procedures and geo-location-based recommender systems.

Sah and Singh [9] and Sah and Singh [10] conducted studies on student career prediction using machine learning, emphasizing the relevance and utility of predictive models in guiding students towards successful career paths. Maurya [11] explored the prediction of students' careers using machine learning algorithms, providing insights into the predictive capabilities of these techniques.

Valte et al. [13] delved into placement prediction using machine learning, contributing to the discourse on predictive modeling for student placements. Pandey and Maurya [14] focused on career prediction classifiers based on academic performance and skills, highlighting the multifaceted aspects considered in predictive modeling.

Moreover, Mani [24] assessed the employability of students using data mining techniques, shedding light on the application of these methods in evaluating students' readiness for the job market. Al-dossari and Alkahlifah [26] developed a machine learning approach for career path choice for information technology graduates, showcasing the versatility of machine learning in career guidance.

Rajashekar [21] explored a campus placement prediction system using a bagging approach, contributing to the advancement of predictive modeling in educational settings. Hinton, Vinyals, and Dean [33] discussed the distillation of knowledge in neural networks, providing insights into knowledge transfer and optimization in machine learning models.

Other studies such as Ambili and Abraham [1], El Mrabet and Moussa [2], and Baffa, Miyim, and Dauda [4] have contributed significantly to understanding and improving student placement prediction through machine learning techniques. Parida, Kumarpatra, and Mohanty [8], Sah and Singh [9], Sah and Singh [10], and Maurya [11] have provided valuable insights into career prediction and employability assessment using machine learning.

Furthermore, research by Valte et al. [13], Pandey and Maurya [14], Mani [24], Al-dossari and Alkahlifah [26], Rajashekar [21], and Hinton, Vinyals, and Dean [33] has expanded the scope of machine learning applications in student placement, career guidance, and knowledge distillation.

These studies collectively underscore the growing interest and advancements in the application of machine learning techniques for student placement prediction and career guidance, offering valuable contributions to academia and industry alike.

3. Proposed Methodology

The flow diagram you provided outlines the steps involved in building and evaluating a predictive model for student placement using machine learning techniques. Let's break down each step-in detail:

3.1. Input Data:

This step involves gathering the dataset containing information about students and their placement outcomes. The dataset typically includes features such as academic performance, skills, attendance, internship experiences, etc., along with the target variable indicating whether a student was placed or not.

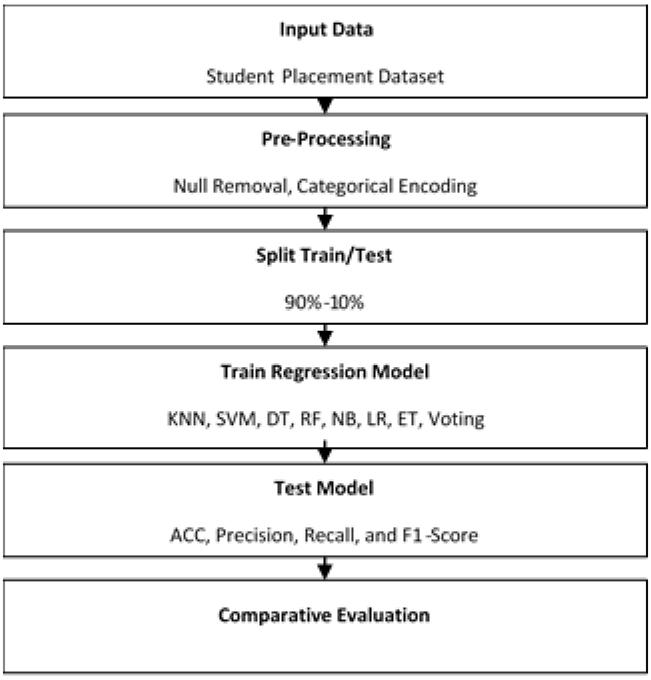

Fig. 1. Proposed Flow

3.2. Pre-Processing:

Null Removal: In this step, any missing or null values in the dataset are handled. This can be done by either removing rows with missing values or imputing them using techniques like mean, median, or mode imputation.

Categorical Encoding: Since machine learning algorithms require numerical inputs, categorical variables in the dataset (e.g., gender, department) are encoded into numerical format using techniques like one-hot encoding or label encoding.

3.3. Split Train/Test:

The pre-processed dataset is divided into two sets: a training set (90% of the data) and a test set (10% of the data). The training set is used to train the machine learning models, while the test set is used to evaluate their performance.

3.4. Train Regression Model:

In this step, various regression models are trained using the training data. The models listed in the diagram include: K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and Extra Trees (ET) are common.

Voting: An ensemble technique where multiple models are combined to make predictions.

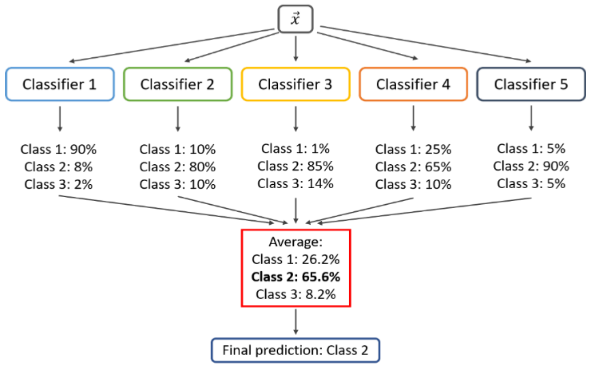

Fig. 2. Voting Classification

3.5. Test Model:

After training, each model is tested using the test dataset to assess its performance. The evaluation metrics used in this step include Accuracy (ACC), Precision, Recall, and F1-Score. These metrics help measure how well the model predicts student placement outcomes.

3.6. Comparative Evaluation:

Finally, the performance of each model is compared based on the evaluation metrics. This comparative evaluation helps determine which model(s) perform best in predicting student placement. Models with higher accuracy, precision, recall, and F1-score are considered more effective in this context. Overall, this flow diagram represents a structured approach to building and evaluating predictive models for student placement, starting from data preprocessing to model training and comparative assessment. It emphasizes the importance of using multiple models and evaluation metrics to ensure robust and accurate predictions.

4. Experiment and Results

	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status
0	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.0	Mkt&HR	58.80	Placed
1	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed
2	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.0	Mkt&Fin	57.80	Placed
3	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.0	Mkt&HR	59.43	Not Placed
4	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.50	Placed
...	...	...	...	...	...	...	...	...	...	...	...	...	...
210	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.0	Mkt&Fin	74.49	Placed
211	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.0	Mkt&Fin	53.62	Placed
212	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.0	Mkt&Fin	69.72	Placed
213	F	74.00	Others	66.00	Others	Commerce	58.00	Comm&Mgmt	No	70.0	Mkt&HR	60.23	Placed
214	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.0	Mkt&HR	60.22	Not Placed

215 rows × 13 columns

Fig. 3. Data reading

	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status
0	1	67.00	1	91.00	1	1	58.00	2	0	55.0	1	58.80	Placed
1	1	79.33	0	78.33	1	2	77.48	2	1	86.5	0	66.28	Placed
2	1	65.00	0	68.00	0	0	64.00	0	0	75.0	0	57.80	Placed
3	1	56.00	0	52.00	0	2	52.00	2	0	66.0	1	59.43	Not Placed
4	1	85.80	0	73.60	0	1	73.30	0	0	96.8	0	55.50	Placed
...	...	...	...	...	...	...	...	...	...	...	...	...	...
210	1	80.60	1	82.00	1	1	77.60	0	0	91.0	0	74.49	Placed
211	1	58.00	1	60.00	1	2	72.00	2	0	74.0	0	53.62	Placed
212	1	67.00	1	67.00	1	1	73.00	0	1	59.0	0	69.72	Placed
213	0	74.00	1	66.00	1	1	58.00	0	0	70.0	1	60.23	Placed
214	1	62.00	0	58.00	1	2	53.00	0	0	89.0	1	60.22	Not Placed

215 rows × 13 columns

Fig. 4. Data Pre-Process

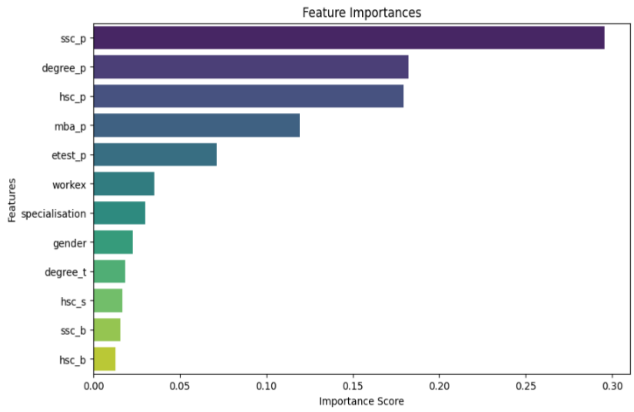

Fig. 5. Feature Importance

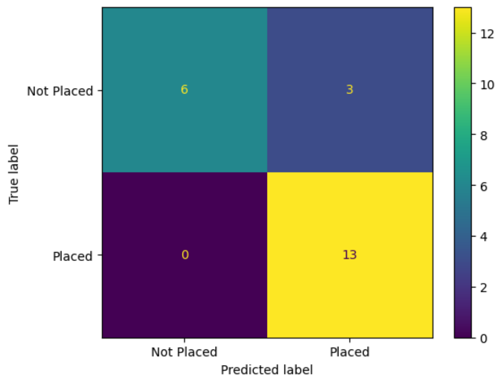

Fig. 6. Confusion Mat of KNN

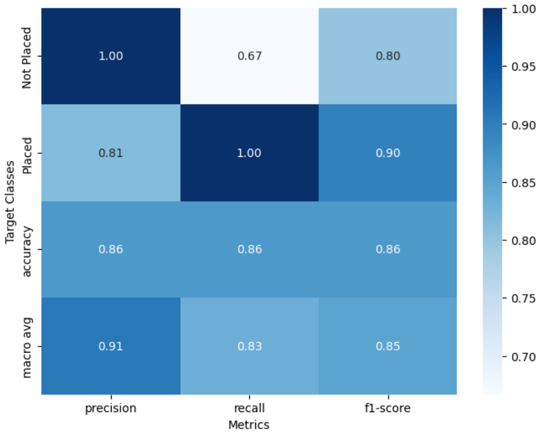

Fig. 7. Classification Report of KNN

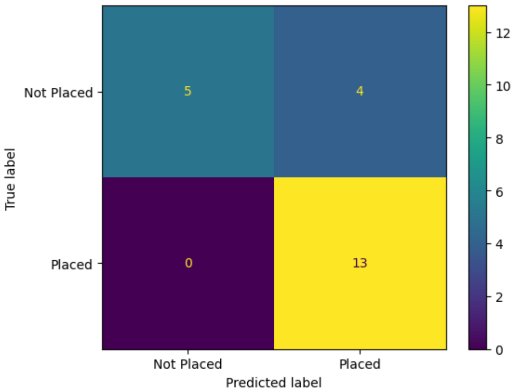

Fig. 8. Confusion Mat of SVM

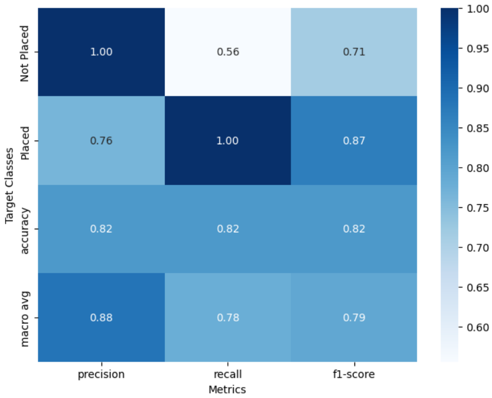

Fig. 9. Classification Report of SVM

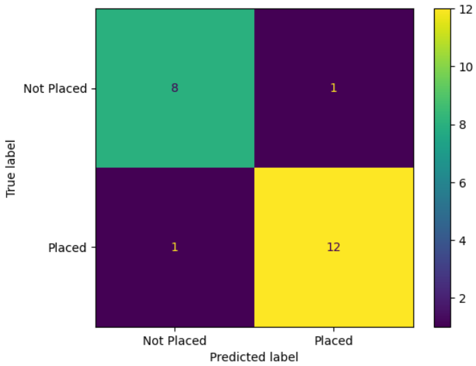

Fig. 10. Confusion Mat of DT

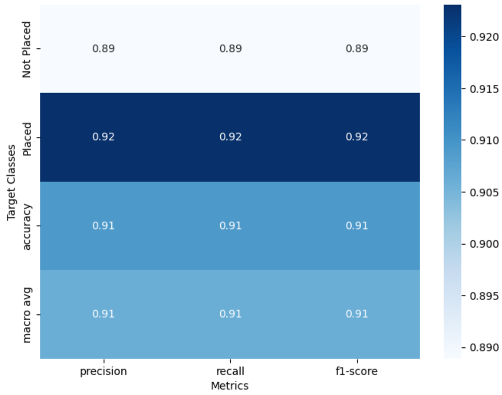

Fig. 11. Classification Report of DT

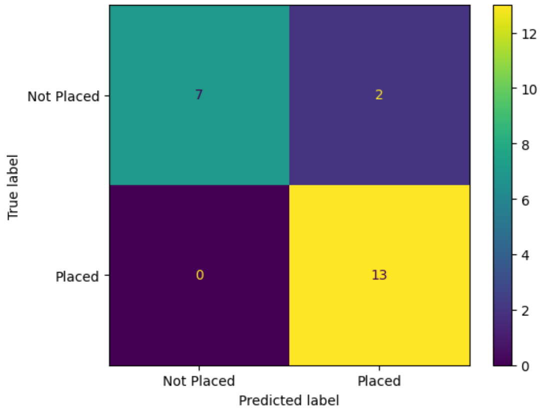

Fig. 12. Confusion Mat of RF

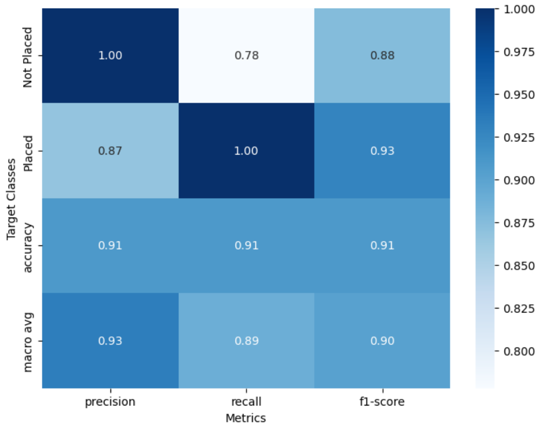

Fig. 13. Classification Report of RF

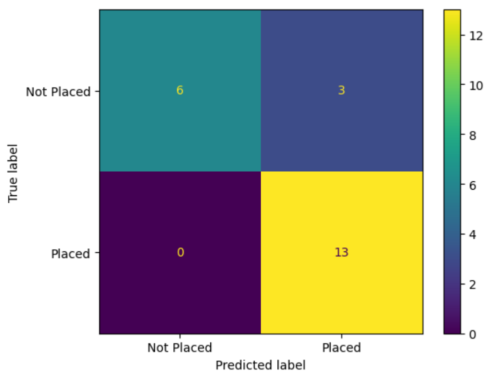

Fig. 14. Confusion Mat of NB

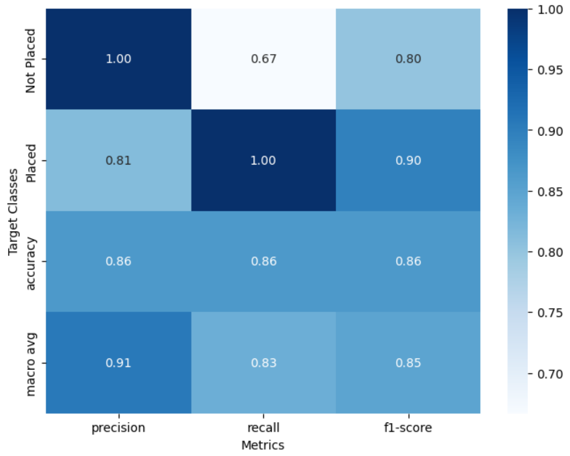

Fig. 15. Classification Report of NB

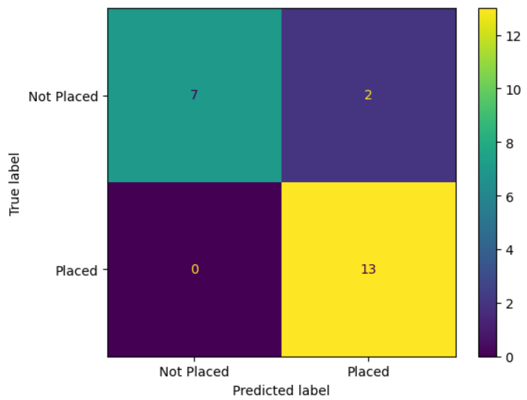

Fig. 16. Confusion Mat of ET

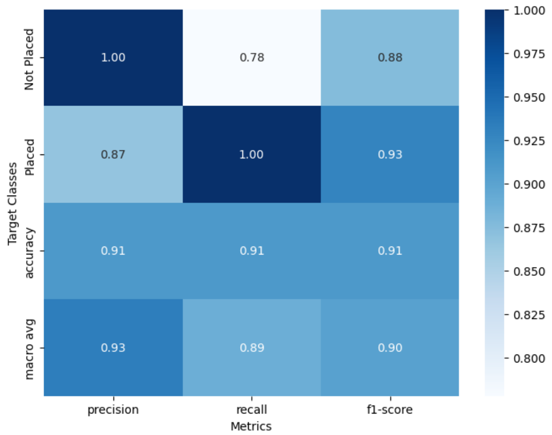

Fig. 17. Extra-Tree Classification Parameters

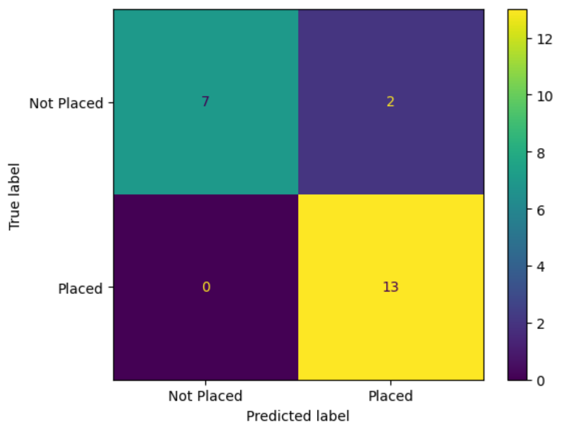

Fig. 18. Confusion Matrix

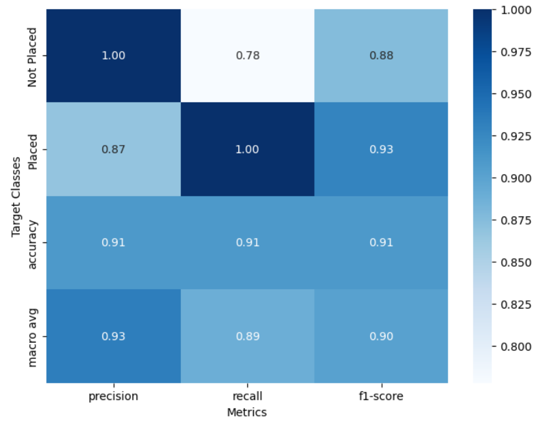

Fig. 19. Linear Regressor Classification Parameters

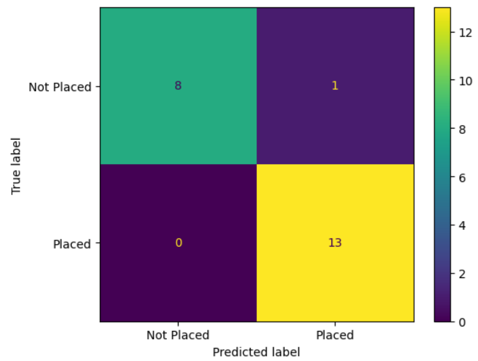

Fig. 20. Voting-Classifier Confusion Matrix

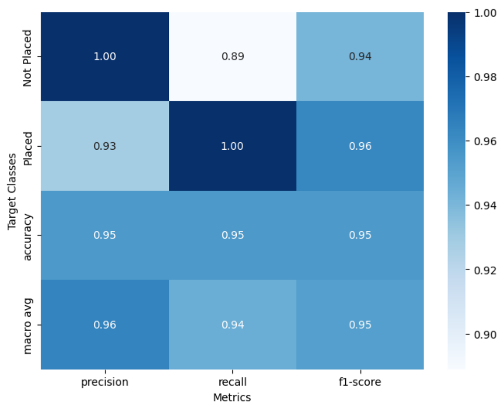

Fig. 21. Voting-Classifier Parameters

Table 1. Analysis of Models

Model	ACC	Precision	Recall	F1-Score
KNN	86%	91%	83%	85%
SVM	82%	88%	78%	79%
DT	91%	91%	91%	91%
RF	91%	93%	89%	90%
NB	86%	91%	83%	85%
ET	91%	93%	89%	90%
LR	91%	93%	93%	93%
Voting	95%	96%	94%	95%

5. Conclusion

This research focused on using various machine learning techniques for student placement prediction. The analysis included models like K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), Extra Trees (ET), Logistic Regression (LR), and a Voting Classifier ensemble. After evaluating these models, it was found that the Voting Classifier had the highest accuracy at 95%. This indicates the effectiveness of combining multiple classifiers for robust predictions. Other models like Decision Trees (DT), Random Forest (RF), Extra Trees (ET), and Logistic Regression (LR) also performed well, consistently achieving accuracies above 90% with strong precision, recall, and F1-scores. While models such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) showed respectable performance, they fell slightly short compared to the top-performing models. Overall, these findings emphasize the importance of leveraging machine learning for student placement prediction. The results can guide educational institutions and placement offices in implementing data-driven strategies to improve students' career outcomes. Future research can explore optimizing model parameters and incorporating advanced ensemble techniques for even better predictions.

6. References and Footnotes

6.1. References

Author contributions

Milind Ruparel: Conceptualization, Methodology, Software, Field study **Dr. Priya Swaminarayan:** Data curation, Writing-Original draft preparation, Software, Validation., Field study **Milind Ruparel:** Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] P. S. Ambili and B. Abraham, "A COMPREHENSIVE EVALUATION OF EMPLOYABILITY PREDICTION USING ENSEMBLE LEARNING TECHNIQUES," EPRA International Journal of Multidisciplinary Research, no. January, pp. 362–366, 2024, doi: 10.36713/epra2013.

[2] H. El Mrabet and A. A. Moussa, "A framework for predicting academic orientation using supervised machine learning," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 12, pp. 16539–16549, 2023, doi: 10.1007/s12652-022-03909-7.

[3] I. Z. A. D. P. No, G. J. Van Den Berg, A. Uhlendorff, G. J. Van Den Berg, G. Stephan, and M. Kunaschk, "DISCUSSION PAPER SERIES Predicting Re-Employment : Machine Learning versus Assessments by Unemployed Workers and by Their Caseworkers Predicting Re-Employment : Machine Learning versus Assessments by Unemployed Workers and by Their Caseworkers," IZA – Institute of Labor Economics, no. 16426, 2023.

[4] M. H. Baffa, M. A. Miyim, and A. S. Dauda, "A periodical of the Faculty of Natural and Applied Sciences , UMYU , Katsina Machine Learning for Predicting Students ' Employability," UMYU Scientifica, vol. 2, no. 1, pp. 1–9, 2023.

[5] B. Pune, "PLACEMENT PREDICTION USING MACHINE," IJARIIE, no. 2, pp. 646–650, 2023.

[6] N. K. Shah, "International Journal of Research Publication and Reviews Job Position Detection : A Data Science Approach," International Journal of Research Publication and Reviews, vol. 4, no. 7, pp. 3229–3235, 2023.

[7] P. Archana, D. Pravallika, P. S. Priya, and S. Sushmitha, "Student Placement Prediction Using Machine Learning," Journal of Survey in Fisheries Sciences, vol. 10, no. 1, pp. 2734–2741, 2023.

[8] B. Parida, P. Kumarpatra, and S. Mohanty, "Prediction of recommendations for employment utilizing machine learning procedures and geo-area-based recommender framework," Sustainable Operations and Computers, vol. 3, no. November 2021, pp. 83–92, 2022, doi: 10.1016/j.susoc.2021.11.001.

[9] U. K. Sah and A. Singh, "Student Career Prediction Using Machine Learning," IJSDR, vol. 7, no. 5, pp. 343–347, 2022.

[10] U. K. Sah and A. Singh, "Review on Student Career Prediction Using Machine Learning," JETIR, vol. 9, no. 4, pp. 262–268, 2022.

[11] A. P. L. S. Maurya, "Predicting Students ' Career by using Machine Learning Algorithms," International Journal of Innovations in Engineering and Science, vol. 7, no. 7, pp. 20–24, 2022.

[12] N. P. K. M, N. M. Goutham, K. A. Inzamam, S. V Kandi, and V. S. V R, "Placement Prediction and Analysis using Machine Learning," IJERT, vol. 10, no. 11, pp. 224–227, 2022.

[13] M. Valte, S. Gosavi, T. Sarode, A. Kate, and P. S. Dhanake, "Placement Prediction," IJARSCT, vol. 2, no. 5, pp. 512–520, 2022, doi: 10.48175/568.

[14] A. Pandey and L. S. Maurya, "Career Prediction Classifiers based on Academic Performance and Skills using Machine Learning," SSRG International Journal of Computer Science and Engineering, vol. 9, no. 3, pp. 5–20, 2022.

[15] L. S. Maurya, S. Hussain, and S. Singh, "Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on," Applied Artificial Intelligence, vol. 35, no. 6, pp. 403–420, 2021, doi: 10.1080/08839514.2021.1901032.

[16] R. S. Kumar, F. Dilsha, A. N. Shilpa, and A. A. Sumayya, "Student Placement Prediction Using Support Vector Machine Algorithm," IJIREEICE, vol. 9, no. 5, pp. 40–43, 2021, doi: 10.17148/IJIREEICE.2021.9507.

[17] N. C. Sekhar, M. Sebastian, N. Suresh, L. Reji, and C. K. Shahad, "WHAT ' S NEXT ? Prediction Model for Students Future Development," National Conference on Smart System and technologies, vol. 8, no. 7, pp. 7–11, 2021.

[18] N. Vidyashreeram and A. Muthukumaravel, "Student Career Prediction Using Machine Learning Approaches," Springer, 2021, doi: 10.4108/eai.7-6-2021.2308642.

[19] A. Surve, A. Singh, and S. Tiwari, "Student Career Guidance System using Machine Learning," IRJET, pp. 3543–3546, 2021.

[20] V. J. Hariharan, A. S. Abdullah, R. Rithish, V. Prabakar, S. Selvakumar, and M. Suguna, "Predicting student ' s placement prospects using Machine learning Tech- niques," SSRG International Journal of Computer Science and Engineering, pp. 2–5, 2021.

[21] D. Rajashekar, "CAMPUS PLACEMENT PREDICTION SYSTEM USING BAGGING APPROACH .," JETIR, vol. 8, no. 8, pp. 306–311, 2021.

[22] V. Mulye and A. Newase, "A Review : Recruitment Prediction Analysis Of Undergraduate Engineering Students Using Data Mining Techniques," SSRG International Journal of Computer Science and Engineering, vol. 8, no. 3, pp. 1–6, 2021, doi: 10.14445/23488387/IJCSE-V8I3P101.

[23] J. Zhu, S. Tang, D. Chen, and S. Yu, "Complementary Relation Contrastive Distillation," arXiv, 2021.

[24] R. Mani, "Assessing employability of students using data mining techniques Assessing Employability of Student using Data Mining Techniques," IEEE, no. October 2020, doi: 10.1109/ICACCI.2017.8126157.

[25] P. Gavhane, D. Shinde, A. Lomte, N. Nattuva, and M. Munjal, "Career Path Prediction Using Machine Learning," IJSRST, vol. 5, no. 8, pp. 300–304, 2020.

[26] H. Al-dossari and M. Alkahlifah, "CareerRec : A Machine Learning Approach to Career Path Choice for Information Technology Graduates," Engineering, Technology & Applied Science Research, vol. 10, no. 6, pp. 6589–6596, 2020.

[27] R. Viram, S. Sinha, B. Tayde, and A. Shinde, "Placement prediction system using machine learning," IJCRT, vol. 8, no. 4, pp. 1507–1515, 2020.

[28] I. T. Jose, D. Raju, J. A. Aniyankunju, J. James, and M. T. Vadakkel, "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1005–1009, 2020.

[29] D. Manjusha, B. Pooja, A. Usha, and B. E. Scholars, "STUDENT PLACEMENT CHANCE," JETIR, vol. 7, no. 5, pp. 1011–1015, 2020.

[30] M. Bangale, S. Bavane, A. Gunjal, R. Dandhare, and S. D. Salunkhe, "A Survey on Placement prediction system using machine learning," IJSART, vol. 5, no. 2, 2019.

[31] K. Anvesh, B. S. Prasad, V. V. Sai, R. Laxman, and B. S. Narayana, "Automatic Student Analysis and Placement Prediction using Advanced Machine Learning Algorithms," IJITEE, vol. 3075, no. 12, pp. 4178–4183, 2019, doi: 10.35940/ijitee.L3664.1081219.

[32] S. Harinath, A. Prasad, and T. Mathew, "Student placement prediction using machine learning," IRJET, pp. 4577–4579, 2019.

[33] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," arXiv, pp. 1–9, 2015, [Online]. Available: http://arxiv.org/abs/1503.02531