A Review: Recruitment Prediction Analysis Of Undergraduate Engineering Students Using Data Mining Techniques

Vandana Mulye[*1], Dr. Atul Newase[#2]

[*1]P.hD. Scholar (Dept. of Comp. Science & Application, Dr. A.P.J. Abdul Kalam University, Indore M.P.), India
[#2]Research Guide (Dept. of Comp. Science & Application, Dr. A.P.J. Abdul Kalam University, Indore M.P.), India

Abstract - *At present, the Recruitment of Engineering students is the biggest challenging task and problem in India Recruitment is an aspiration of each engineering student. After studying hard, spending more money and time, every student eagerly waits for recruitment, but at last, most of the students don't get recruited. The present situation of recruitment of engineering students is very disgraceful. Although there is a huge amount of well-reputed, equipped, good infrastructure engineering institute is placed in India, but they are unable to provide recruitment to each student. Recruitment is the right of each engineering student. To overcome such type of situation, the prediction method of this research will help students as well as Institutions. Hence this prediction method will help to engineer Institution to identify the main qualities which are essential to get recruitment. Prior identification of student's eligibility can help to engineer institutions to upgrade their student's qualities to get recruited as well as students also. This paper presents a review of various studies made by different investigators, researchers on recruitment prediction analysis of students using data mining techniques.*

Keywords — *Classification, Data mining, Logistic Regression, Machine Learning, Prediction.*

I. INTRODUCTION

 In the era of a dynamically changing competitive world, everyone seeks to get at the top position and wills to have a good livelihood. To achieve this, every youth wants to get a noble earning source and fortified life. Engineering is a noble profession, and to achieve this profession, every student wants to score top, spend more money to get admission in a reputed, highly equipped, and good infrastructure Institution. Even low scores students also spend more money to get Admission to a top institution.

To overcome such type of problem, this research will be very helpful. This research particularly focuses on the main attributes which are must necessary in a student to get recruitment. In this research, data is collected from B.Tech. Final year engineering students. After data collection, data mining techniques ID3 decision tree algorithm is applied on student's dataset mainly on nine attributes like as Academic grade, Practical Knowledge, Skilled Certificates, Project accomplishment, Subject Knowledge in written exam and Interview, Fear in Written exam and interview, Communication skill in an interview, Confidence in the interview, Practical interest for recruitment prediction. After performing the decision tree algorithm and relevant formulas on the student's dataset, this research originates that information gain for the 'skilled certificate' attribute is topmost among all other attributes and is the root node in the decision tree.

Hence this research concludes that to acquire recruitment, every student must have one skilled certificate, alongside the skilled certificate, a student should have a mark above 80% "A" grade, excellent practical knowledge, subject knowledge should be more than 80%, good communication skill and confidence in an interview, project accomplishment should be accomplished during course, the student may be or not have fear during written test and interview. This research will be beneficial for students and engineering institutions. This research will also be helpful to decrease the unemployment problem in our country. This research paper proposes a review of various recruitment prediction methods and student's performance.

II. RELATED STUDIES

K. Sripath Roy et al.(2018) [1] proposes the advanced machine learning algorithm for a student's career prediction. They used three algorithms, SVM, XG Boost, Decision Tree, on a dataset. Among all three algorithms, they found that the SVM algorithm provided more correctness with 90.3%, and then XG Boost provided correctness with 88.33%. They required many parameters for student's career prediction like as student's academic scores in various subjects, specializations, programming, and analytical capabilities, memory, personal details like relationship,

interests, sports, competitions, hackathons, workshops, certifications, books interested and many more. Finally, they developed a web application to enter the parameter and show the result of the prediction. In the future, highly advanced upgraded applications can be developed.

Pothuganti Manvitha and Neelam Swaroopa (2019) [2] proposes a supervised machine learning technique for the prediction of campus placement. This research applied the decision tree and random forest algorithm to the student's dataset. The algorithms are applied on previous year data sets of the students, and the parameters used to construct the model and parameters used for this research is an academic history of the student like overall percentage, backlogs, credits.

The correctness gained after examination for the Decision tree is 84%, and for the Random Forest is 86%. The random forest presented the finest outcomes. The productivity of the two methods is likened in terms of correctness. Henceforth, the Random Forest algorithm is superior to predict the placement results.

Shreyas Harinath et al.'s 2019 [3] research proposes machine learning techniques to predict the placement status of the student. In this research, two different machine learning classification algorithms are used, namely the Naive Bayes Classifier and K Nearest Neighbors [KNN] algorithm for the prediction of a student's placement position. The algorithms ponder the factors such as USN, Tenth, and PUC/Diploma results, CGPA, Technical and Aptitude Skills. These algorithms individualistically predict the outcomes and then differentiate the effectiveness of the algorithms, which is dependent on the dataset. In the forthcoming, this research will focus on adding some more attributes for better prediction of placement.

Vinutha K (2020) [4] proposes different Machine Learning Algorithms for the prediction of employability of engineering graduates depended on student's academic performance and also employability skills. This research used different machine learning algorithms like Logistic Regression, Decision tree, k-nearest neighbor, Support Vector Machine and Naïve Bayes,

ANN(Artificial Neural Network) to build a model for prediction of employability of engineering graduate students. The dataset used to create a model for this research is USN(Unique Student Number). Name, First to eight-semester marks, Online courses completed, Internship, Technical papers presented. This research found that the highest correctness of 87.42% with Artificial Neural, 85.2% correctness with Logistic Regression classifier, and 84.21% correctness with Naïve Bayes classifier. Further, this research will be upgraded with student's empirical datasets.

C. Jayasree & K. K. Baseer (2018) [5] proposes some data mining and Machine Learning Techniques for Prediction

Student Performance Improve their Employability and reducing dropout rates. This research work shows that the academic performances of the students are dependent on their past performances. To predict the performance of the student's author considered various datasets like previous grades, research work time, parent's status, GPA, school support, higher education, internet usage, travel time, etc. This research applied a few algorithms (linear regression, K-means clustering, and neural networks using Weka and Azure tools to student datasets. Machine learning studio using azure has been the best tool for real-time applications and can prove to be a powerful tool in academia. The performance of k-means and neural networks is very effective.

S. Celine et al. (2015) [6] proposes a predictive and progressive model to predict the employability of a set of candidates using Logistic Regression. The main goal of this study to automatize the employment procedure, which implements the logistic Regression technique to predict the possibility of employability. This study discovers the application of the Logistic Regression machine learning technique to predict employability. This research used four parameters as Aptitude skills ($\beta1$), Communication skills($\beta2$), Technical skills($\beta3$), and Personality skills($\beta4$) for the prediction of employability. This study tries to demonstrate an application of machine prediction to the employability of a candidate in the employment process. In the future, this study can be simply converted into another area of prediction like health care, weather forecasting, natural calamities, crime prediction, etc., where the outcomes will be a contrast in nature by varying the self-regulating variables.

D. Satish Kumar et al. (2015) [7] proposes a six predictor logistic regression model to predict the probability of MBA student campus placement. It is a classification model. They used six parameters in this model for the prediction of placement possibility. The six parameters are CGPA in UG and PG, Specialization in UG and PG, Soft Skill Score, and Gender. This research used the R software package to examine the data. Predictable Outcomes of the research specify that the probabilities of campus placement are affected by four predictors: CGPA, Specialization in PG, Specialization in UG, and Gender.

V.Rameshet al. (2015) [8] paper proposes the use of different data mining techniques to examine the performance of Computer Course students. This research uses different data mining techniques and presents the uses of classification techniques as a predictive tool. Different classification techniques are applied to student data, and finally, this research found that the Multilayer Perception algorithm is most suitable for predicting the student's performance. Multilayer Perception algorithm provides 87% prediction, which is larger than other algorithms. The student's data parameters used for this research are English, Maths, Programming language, Practical marks. This

research tried to use a classification technique to predict the student's performance and also measure the performance of Naïve Bayes Simple, Multilayer Perception, SMO, J48, and REP Tree data mining techniques.

Samrat Singh and Dr. Vikesh Kumar (2013) [9] propose various data mining techniques to examine the data of Engineering student's academic performance for Recruitment purposes. In this research, six classification techniques are used on student data that is BayesNet, Naïve Bayes, Multilayer Perceptron, IB1, Decision Table, and PART Classification method. As per the pilot outcome, they observed that IB1 Classifier is the most competent technique for such type of student's data. In this research, six parameters are considered for student's data set like as student's Name, Branch, passing percentage (%) of 10th class, passing percentage (%) of 12thclass and passing percentage (%), and Final Grade for analysis. For future studies, this research will be beneficial for educational organizations, colleges, universities, and industrial organizations or corporate sectors. In the future, they can use other data mining techniques like clustering, Prediction and Association rules, etc., on student data.

T. Malathi et al. (2015)[10] proposes the classification data mining techniques to analyze the student's recruitment process like Selected, Waiting, and Not Selected. A classification algorithm is used to classify the students based on their talent. This research will support the Lecturers to select the students and organize them for the recruitment process. This research is very helpful to improve the student's performance and cut down the failure rate. In the future, this research will provide useful information to lecturers to improve the studying capability of students.

Siddu P. Algur et al. (2016) [11] proposes two algorithms-Random Tree and J48 to construct classification models applying the Decision Tree notion. These two models are used to predict the recruitment status of a student on their academic performance. The academic performance is measured by different attributes like examination score, communication skill, and placement preparation hours breaks taken during the course, extracurricular activities, cultural activities, and the number of industrial visits for recruitment prediction. Between these two classification models, the Random Tree classification model is discovered good as compared to the J48 classification model. The accuracy of the Random Tree classification model provides correctness 85%, and the correctness of the J48 classification model is detected 74%. In forthcoming to progress the prediction/classification correctness by applying some other data mining techniques such as K-Nearest Neighbor classification technique, Naive Bayesian classification techniques, etc.

Keno C. Piad et al. (2016) [12] proposes five classification algorithms that were compared on the IT employability dataset to predict the employability of IT graduate students.

This paper predicts the employability of IT graduates using nine variables like as Location, gender, It _core, It_profeesionally Ite-electives, Language _and _humanities, Mathematics_natsci, Social_science_com, Free_electives, etc. . In this research work, the researchers compare five algorithms under classification technique on IT employability dataset. The result shows logistic regression with an accuracy of 78.4 is implemented. Based on logistic regression analysis, three academic variables directly affect; IT_Core, IT_Professional, and Gender identified as significant predictors for employability. The research work can utilize newly acquired data to generate more rules and predict more accurately IT employability. Thus, other classification algorithms maybe explore and generate new data models for better prediction.

Tripti Mishra et al. (2016) [13] proposes different classifier and develop employability model formed on the appropriate classifier to predict the employability of MCA (Master of Computer Applications) students. For the prediction of employability, this research used various classification techniques like Bayesian methods, Multilayer Perceptron and Sequential Minimal Optimization (SMO), Ensemble Methods, and Decision Trees and also observed the most suitable algorithm related to the question of this research. The parameters used for this research are academic performance, socioeconomic conditions, job skills, and emotional skills, etc. Measuring the performance of all the algorithms they obtain J48 (a cropped C4. 5 decision tree) algorithm of WEKA is most applicable among all the algorithms to predict the employability of (MCA) students and identify the algorithm which is best suited for this problem. In the forthcoming, this research work will consider the students of B. Sc. and B. E. also.

Nor Azziaty Abdul Rahman et al.(2017) [14] proposes six classification models, namely K-Nearest Neighbor, Naive Bayes, Decision Tree, Neural Network, Logistic Regression, and Support Vector Machine to acquire a suitable model to predict employment status among fresh graduates in HEI(Higher Education Institutes). Supervised and unsupervised Machine Learning Algorithms were used in this research that as K-Nearest Neighbour, Naïve Bayes, Decision Tree, Neural Network, Logistic Regression, and Support Vector Machine. The parameter they used for this research is Gender, Program Academic. In forthcoming, there is a necessity for more parameters like grade subjects taken during the study period, the results of the oral test, and work status.

Madhavi Girase et al. (2018) [15] proposes Decision Trees as classification techniques of data mining to predict the employability of Undergraduate Engineering students. For this purpose, a data set is established with traditional parameters like socioeconomic conditions, academic performance, and some additional emotional skill parameters. This paper works on student data sets like a student's academic performance, additional activities,

personality development, technical training. The result of this research illustrates that the Decision Tree algorithm performed better-applied data mining techniques in this research.

III. TABLE I

Comparatively study

S. N O.	NAM E OF AUT HOR	YEAR OF PUBLI CATIO N	PARAMETER,MET HODS,ALGORITHM S, TOOLS	ATTRIB UTES
1.	K. Sripat h Roy et al.	2018	SVM, XG Boost, Decision Tree	student's academic scores in various subjects, specializa tions, programm ing, and analytical capabilitie s, memory, personal details
2.	P.Ma nv- itha and Neela m Swar oo-pa	2019	decision tree and random forest algorithm	Academic history of the student like overall percentag e, backlogs, credits.
3.	Shrey as Harin ath et al.	2019	Two different machine learning classification algorithms are used:- Naive Bayes Classifier and K Nearest Neighbors [KNN] algorithm.	USN, Tenth, and PUC/Dipl oma results, CGPA, Technical and Aptitude Skills.
4.	Vinut ha K	2020	Various machine learning algorithms like Logistic Regression, Decision tree, k-nearest neighbor, Support	USN(Uni que Student Number). Name, First to
			Vector Machine and Naïve Bayes, ANN(Artificial Neural Network) to build a model.	eight- semester marks, Online courses completed , Internship , Technical papers presented.
5.	C. Jayas ree & K. K. Basee r	2018	Linear regression, K- means clustering, and neural networks using Weka and Azure tools.	previous grades, research work time, parent's status, GPA, school support, higher education, internet usage, travel tim
6.	S. Celin e, et al.	2015	Logistic Regression technique & algorithms	Aptitude skills (β1), Communi cation skills(β2), Technical skills(β3), and Personalit y skills(β4)
7.	D. Satis h Kum ar et al.	2015	Six predictors logistic regression model & R software package.	CGPA in UG and PG, Specializa tion in UG and PG, Soft Skill Score, and Gender.
8.	V.Ra mesh et al.	2015	Naïve Bayes Simple, Multilayer Perception, SMO, J48, and REP Tree data mining	English, Maths, Program ming language, Practical

			techniques.	marks
9.	Samrat Singh and Dr. Vikesh Kumar	2013	BayesNet, Naïve Bayes, Multilayer Perceptron, IB1, Decision Table, and PART Classification method.	student's Name, Branch, passing percentage (%) of 10th class, passing percentage (%) of 12thclass and passing percentage (%) and Final Grade
10.	T. Malathi et al.	2015	Classification algorithm – Decision tree	internal assessment attributes are Here, assignment, quiz, class test marks, lab performance, and external assessment attributes are Previous exam grade and final exam grade.
11	Siddu P. Algur et al.	2016	Two algorithms- Random Tree and J48 to construct classification models	examination score, communication skill, and placement preparation hours breaks taken during the course, extracurricular

				activities, cultural activities, and the number of industrial visits
12.	Keno C. Paid et al	2016	Classification Algorithms	Location, gender, It_core, It_profeesionally Ite-electives, Language _and _humanities, Mathematics_natsci, Social_science_com, Free_electives,
13.	Tripti Mishra et al.	2016	Bayesian methods, Multilayer Perceptron and Sequential Minimal Optimization (SMO), Ensemble Methods, and Decision Trees	socioeconomic conditions, job skills, and emotional skills, etc.
14.	Nor Azziaty Abdul Rahman et al.	2017	K-Nearest Neighbor, Naive Bayes, Decision Tree, Neural Network, Logistic Regression, and Support Vector Machine	Gender, Program Academic.
15.	Madhavi Girase et al.	2018	Decision Trees as classification techniques	socioeconomic conditions, academic performance, and some additional emotional skill parameters.

IV. CONCLUSIONS

A comprehensive review of the literature was implemented formed on different recruitment prediction techniques. After reading various literature on recruitment prediction, understandably, the main cradle for prediction of recruitment is student data relevant to academic and recruitment factors. Academic and recruitment relevant factors are an essential cradle for predicting the upcoming recruitment possibilities. These predictions can help aware students to observe their abilities, and they can recover them before campus recruitment. The recruitment prediction scheme is also very supportive for institutions to increase the number of recruitment of their students as well as upgrade their academic plan and design policies to progress the recruitment status of the Institution for upcoming years.

REFERENCES

[1] K. Sripath Roy, K.Roopkanth, V.Uday Teja, V.Bhavana, J.Priyanka (2018): Student Career Prediction Using Advanced Machine Learning Techniques. International Journal of Engineering and Technology. 7(2.20)(2018) 26-29.

[2] Pothuganti Manvitha, Neelam Swaroopa., Campus Placement Prediction Using Supervised Machine Learning Techniques. International Journal of Applied Engineering Research. ISSN 0973-4562., 14(2019) 2188-2191.

[3] Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A, Tojo Mathew Student placement prediction using machine learning. International Research Journal of Engineering and Technology (IRJET). 06(2019) 4577-4579.

[4] Vinutha K, Yogisha H K., Employability Prediction of Engineering Graduates using Machine Learning Algorithms. International Journal of Recent Technology and Engineering (IJRTE). 8(2020) 4521 -4524.

[5] C. Jayasree, K. K. Baseer., Predicting Student Performance to Improve their Employability by Applying Data Mining and Machine Learning Techniques. International Journal of Computer Sciences and Engineering Open Access. 6(2018) 1291-1308.

[6] S. Celine, M. Maria Dominic, M. Savitha Devi Logistic Regression for Employability Prediction. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 9(2020) 2471 -2478.

[7] D. Satish Kumar, Zailan Bin Siri, D.S. Rao, S. Anusha (2019): Predicting Student's Campus Placement Probability using Binary Logistic Regression. International Journal of Innovative Technology and Exploring Engineering (IJITEE)., 8(2019) 2633 -2635.

[8] V.Ramesh, P.Parkavi, P.Yasodha., Performance Analysis of Data Mining Techniques for Placement Chance Prediction. International Journal of Scientific & Engineering Research. 2(2011) 1-7.

[9] Samrat Singh, Dr. Vikesh Kumar (2013): Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques. IJCSET. 3(2013) 31-37.

[10] T. Malathi, S. Srinivasan, K.R. DilliRani., Prediction of Students Recruitment Process Using Data Mining Techniques with Classification Rules. International Journal of Computer Science and Information Technology Research. 3(2015) 84-87.

[11] Siddu P. Algur, Prashant Bhat, Nitin Kulkarni., Educational Data Mining: Classification Techniques for Recruitment Analysis. I.J. Modern Education and Computer Science. 2(2016) 59-65.

[12] Keno C. Paid, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat., Predicting IT Employability Using Data Mining Techniques. The Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications. (2016) 26-30.

[13] Surendiran,R., and Alagarsamy,K., 2013. "Privacy Conserved Access Control Enforcement in MCC Network with Multilayer Encryption". SSRG International Journal of Engineering Trends and Technology (IJETT), 4(5), pp.2217-2224.

[14] Tripti Mishra, Dharminder Kumar, Sangeeta Gupta., Students' Employability Prediction Model through Data Mining. International Journal of Applied Engineering Research. 11(2016) 2275-2282.

[15] Nor Azziaty Abdul Rahman, Kian Lam Tan1, Chen Kim Lim., Predictive Analysis and Data Mining among the Employment of Fresh Graduate Students in HEI. The 2nd International Conference on Applied Science and Technology (ICAST'17), (2017) 020007-1– 020007-6

[16] Madhavi Girase, Suchita Lad, Prerna Pachpande, Student's Employability Prediction Using Data Mining. International Journal of Scientific & Engineering Research 9(2018) 27-29.