ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 1| January 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

A COMPREHENSIVE EVALUATION OF EMPLOYABILITY PREDICTION USING ENSEMBLE LEARNING TECHNIQUES

Ambili P S[1], Biku Abraham[2]
[1]School of CSA, REVA University, Bangalore, INDIA ORCID: 0000-0002-7082-1205
[2]Department of CA, Saintgits College of Engineering, Kottayam, INDIA ORCID: 0000-0003-1054-9622

ABSTRACT

The Quality Education plan reflected in the Goal four of the United Nations 2030 Agenda for Sustainable Development stresses the essential function of Education in developing an equitable and just society and reaching complete human potential. The present-day Technical Higher Education system focuses only on the skill sets a student acquires on completion of the programme. The visibility of private HEI is made through the national and international rankings they have and the stakeholders are concerned about the campus placements of their wards. The IT firms chooses HEI s for campus drives by their visibility and they need skilled workforce with multidisciplinary abilities. This study by collecting information from the student database, employing ensemble learning techniques, focuses on such aspects makes an efficient prediction on factors that contribute to high quality learning and improved possibility of campus placement. The proposed model extracts data from dataset with 20 attributes. This enhanced predictive approach based on stacking ensemble learning approach predicts the campus placement chances to an accuracy of 90.21%.

KEYWORDS: *quality education, multidisciplinary abilities, placement prediction, ensemble learning, stacking ensemble.*

I. INTRODUCTION

The Goal eight of United Nations 2030 agenda for sustainable development, decent work and economic growth, can be achieved through the goal four Quality Education. As higher education systems grow and diversify, this education sector has been identified as one of the promising areas for private and foreign investments. The need for a skilled workforce with multidisciplinary abilities across the sciences will be increasingly in demand for campus placements and is a challenge to the institutions in the current scenario. The introduction of new generation courses with learner-centric, discussion-based, flexible and industry-oriented curriculum has now become a practice of private educational institutions to ensure learning outcomes and employability to its graduates. The rationale of this work is an attempt to provide tools/guidance to technical HEIs and curriculum developers on how to structure programmes based on student placement possibilities and thus achieve goal eight of sustainability development using performance-based assessments.

The work helps in boosting the institutions to device fruitful steps to irradicate unemployment among the technically qualified youth. The observe allows the institutions, students, curriculum designers and the college through enhancing their knowledge on how a great deal the scholars are getting to know and what sort of they're at risk of employment in campus. There have been plenty of studies on education data mining and few on student employability prediction but the prediction with ensemble learning techniques are less presented [2,3,12,13].

Ensemble Learning Methods

Ensemble styles that train numerous students (the learned model can be known as a postulation) and furthermore consolidate them for use, with Boosting and Bagging as agents, are a sort of cutting-edge proficiency approach. Gathering styles train numerous students to break a similar issue. As opposed to normal education approaches which attempt to build one student from preparing information, group styles attempt to develop a bunch of students and join them. Troupe proficiency is additionally called commission-grounded education or learning various classifier frameworks [6]. Group is essentially a style that help the delicacy of powerless students (likewise applied to as base students) to solid students, which can make genuinely exact guesses. It joins two or further calculations of practically equivalent to or various sorts called base students. This makes a stronger framework, which consolidates the guesses from every one of the base students to get our last " exact" and more uncertain biased choice. Fig1 shows a typical ensemble skeleton.

There are three garments of early grants that prompted the ongoing area of group styles; that is, consolidating classifiers, outfits of feeble students and admixture of specialists.

1. Joining classifiers was significantly concentrated on in the example acknowledgment local area. Experimenters in this string for the most part work serious areas of strength for on and attempt to plan significant joining rules to get more grounded consolidated classifiers.
2. Gatherings of powerless students was significantly concentrated on in the AI people group. Requests in this field much of the time work on feeble students and attempt to plan significant calculations to support the presentation from frail to solid.
3. Admixture of specialists was considerably concentrated on in the brain organizations' local area. Experimenters for the most part think about a pinnacle and-overcome

methodology, attempt to gain proficiency with an admixture of parametric models deliberately and use consolidating rules to come by a general outcome.

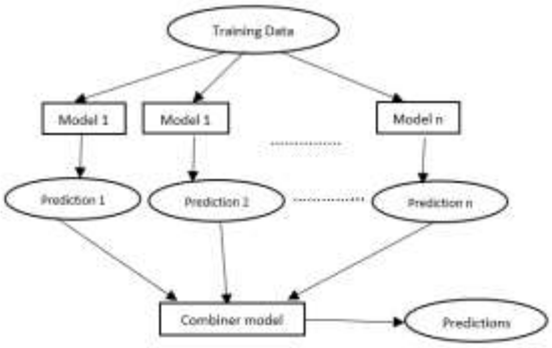

Fig 1: General Ensemble Armature

Many recent research are employing ensemble machine learning techniques and are reporting encouraging results for a variety of applications. Inspired by these successful outcomes of ensemble combinational techniques, the authors attempt to explore ensemble learning methods for employability prediction. Despite much research consideration of applying ensemble approach for many prediction and recommendation systems, there is a still limited amount of attention focusing on employability predictions by considering different categories of data.

Relevance of Work

On thorough evaluation of the similar methods employed, the authors noticed a degradation in the prediction process since the attributes contributing to the intelligent quotient level of students were never considered. The authors identified demographic data as well as the mentor index play a vital in contributing to the attitude of the person. Apart from technology skills such attributes are necessary for such prediction systems. The authors attempted to employ various ML algorithms and then applied the ensemble methodology for getting the accurate result. The results at various stages are analyzed with actual data available and the true positive cases are proved.

Analysis

Precisely, the authors attempt to address the following open research queries:

(i) Are ensemble techniques effective for learning variety of attributes and making effective predictions.

(ii) How to upgrade and adapt existing machine learning models with ensemble techniques for the dataset.

From the experimental studies conducted, the authors were able to attain some encouraging results.

As a summary, the authors were able to introduce the efficient set of attributes to be selected for predicting employability using ensemble techniques which can be considered as the major contributions in this work.

II. LITERATURE REVIEW

The aim of this research work is based on the high potentiality that can be applied in machine learning techniques for predicting the performance of graduates based on several factors. The research suggests that applying different machine learning and to the top ensemble models can produce better results. Multilevel classification proved to be one of the best approaches for classification and these approaches construct ensemble members using bagging scheme [7]. Many studies included the characteristics of students like behavioural features, academic and demographic performances in the training data set for supervised machine learning techniques. The study involves the comparison of many machine learning algorithms and it is observed that logistic regression can give better accuracy. The performances of the algorithms were evaluated by AUC and RUC, which shows the real accuracy of the algorithm [8]. Support Vector Machine (SVM) classifier also featured as one of the best algorithm for prediction analysis [19].

On the dataset, the study performed classification using Decision Tree (ID3), Naive Bayes, K-Nearest Neighbor and Support vector machines to predict student's academic performance. After implementing the ensemble method, it is found that the accuracy of the model increased. The researchers also used Bagging, Boosting, and Voting Algorithm that are the common ensemble methods [18]. The authors suggested an ensemble algorithm The authors cautioned an ensemble set of rules “Roughly Balanced (RB) Bagging” the use of sampling approach to enhance the unique bagging set of rules for facts units with skewed magnificence distributions. The variety of samples in the biggest and smallest lessons are different, however they're correctly balanced whilst averaged over all the subsets, which helps the method of bagging in a extra suitable way [9].

Ensemble mastering strategies, have validated incredible capacities to enhance the prediction accuracy of base mastering algorithms. The authors suggest the speculation which offers development in accuracy of multi method strategies to ensemble mastering is because of a boom withinside the variety of ensemble individuals which are formed. Multi method ensemble mastering strategies are extra correct than their issue ensemble mastering strategies [21].

Ensemble mastering first off extracts a fixed of functions with a number of transformations. Based on those found out functions, more than one mastering algorithms are applied to provide susceptible predictive results[16]. Over the preceding years, ensemble mastering has drawn good sized interest withinside the discipline of synthetic intelligence, sample recognition, device mastering, neural community and facts mining. Ensemble learning proved to be efficient and functional in wide area of problem domain and substantial world application [17]. Normally, academic parameters are given more weightage in predicting the academic performance of a student. The study compares the two models: one built using academic parameters only and another using both academic and non-academic (demographic) parameters[1,11].

The research can be useful for predicting student performance and helping educators to make informed decisions by proactively notifying the students [14]. The study used machine learning approaches for predicting students' employability and it gives a promising conclusion that lead to the researchers to be motivated to enhanced the process and to validate the produced predictive model for further study [17].

The study reported all machine learning algorithms that were used and compared to predict employability; reporting that deep learning methods are rarely used although once again identifying the weakness in searching for the optimal parameters [15]. The findings indicate that the understanding of employability is enhanced by considering both structural and individual dimensions [5]. One of the important issues that have be taken into consideration when building classifier ensembles is to select a pool of diverse and complementary individual classifiers for the ensemble [10]. The obtained results of the study reveal that there is a strong relationship between learner's behaviors and their academic achievement [4].

METHODOLOGY

The experimental dataset was the technical student record of the HEI for which the prediction system was built. The dataset consists of 1965 student data who had successfully graduated the institution in three consecutive years. After optimization 20 attributes were selected for prediction of the employability characteristics.

Dataset Characteristics

The data from the data set is categorized primarily to three: Demographic, Scholastic and Co-Scholastic. In the outcome-based education scenario, the Co-Scholastic factors contribute a lot to the achievement of Course Outcome (CO) and Programme Outcome (PO). Measuring of academic performance is a challenging task, since three factors are connected with the performance and prediction- demographic, academic and campus behavioural factors. The authors considered all these factors and the data pertaining to co-scholastic index which contribute directly to the possibility of employment was carefully chosen. Technical skills such as index in product development, paper publication, patent, prize winning in Hackathons are now a scale for measuring the ability of a student. Mentoring plays a key role in the current education state and the index was concluded based on the overall continuous observation and assessment [2,12,22]. It is observed that socio-psychological and educational aspects are more predictive outcome than with medical.

Demographic data which points more to the family characteristics can also be measured as a contributor of scholastic levels as well as attitude of a student. The authors were keen in observing the fact that HR level is an inevitable part of the placement process where attitude of a person plays vital role. The Scholastic factors which are the direct output obtained from evaluation system cannot be replaced in any situation. It is the quantitative measure of eligibility for any selection procedure and has given due importance. Educational data mining techniques are exploring faster, but need more collaborative studies with challenge of minimal attributes [14,20].

The authors carefully applied the Machine Learning algorithms as prediction and classification cases. Machine learning techniques provide accurate predictions on student's assessment marks and achievements. Dimensionality reduction applied in the dataset initially as large number of input features always makes the prediction a challenging task. Stacking ensemble models are found very effective in multi-label ensemble classification problems[18].

Ensemble Technique

Group is the method involved with consolidating forecasts from numerous models to a solitary vaticination to consummate section execution. We join prepared on the equivalent dataset to decide whether they beat the loftiest scoring single model. Every framework is assessed utilizing stacking ensemble. Figure 2 depicts the ensemble methodology used in this study.

Boosting and Bagging

There are two standards of group styles, or at least, successional troupe styles, where the base students are produced successionally, with Boosting as a delegate, and resemblant outfit styles where the base students are created in equal, with Bagging as an agent. The starting incitement of successional styles is to take advantage of the general exhibition can be helped in a remaining of resemblant gathering styles is to take advantage of the freedom between the base students, since the mistake can be decreased decisively by consolidating autonomous base students.

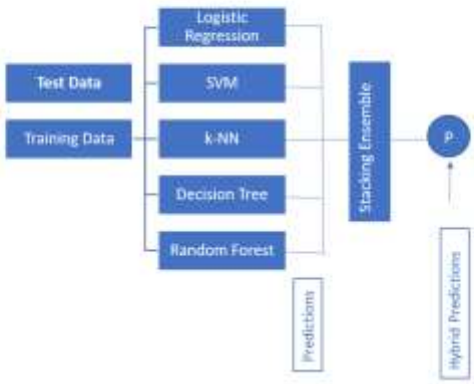

Fig 2: Ensemble Methodology

Stacking

Stacking is a design where a student is prepared to join the singular students are known as the student, or meta-student. We first train the first likewise initiate another informational collection for preparing the meta viewed as info highlights while the first markers are as yet viewed as markers of the new preparation information. A general stacking methodology is represented below.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 1| January 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2023: 8.224 || ISI Value: 1.188

```
Input: Data set D = {(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)};
       First-level learning algorithms 𝔏_1, ..., 𝔏_T;
       Second-level learning algorithm 𝔏.
Process:
   for t = 1, ..., T:        % Train a first-level learner by applying the
      h_t = 𝔏_t(D);          % first-level learning algorithm 𝔏_t
   end
   D' = ∅;                   % Generate a new data set
   for i = 1, ..., m:
      for t = 1, ..., T:
         z_it = h_t(x_i);
      end
      D' = D' ∪ {((z_i1, ..., z_iT), y_i)};
   end
   h' = 𝔏(D');               % Train the second-level learner h' by applying
                             % the second-level learning algorithm 𝔏 to the
                             % new data set D'.
Output: H(x) = h'(h_1(x), ..., h_T(x))
```

III. RESULTS AND DISCUSSION

Prediction Cases and Results

The authors first tried decision tree model on number of straight out highlights present. Truly, one of the significant advantages of this sort of framework is its straightforwardness to comprehend and decipher. This model gave a prediction accuracy rate of 0.8790024. Random Forest is the method of the classes (section) of the singular trees. When applied the model on our dataset, execution contrasting with the previous models, with a delicacy of 0.8905202 is observed. K-Nearest Neighbour to our dataset, models were made with various " k" values fluctuating from 1 to 100 and delicacy of each model being tried by making vaticination on the test information. The authors utilized " knn" work in R for this approach which got back the worth of k = 20 with in general delicacy of 0.8897394. Applying support vector machine model to our dataset, was somewhat trickier since this model requires the dataset to be switched over completely to a configuration of SVM bundle, and direct straightforward scaling to the information (Kumari, Jain, and Pamula 2018). Our initial step is to address every perception in our dataset as a vector of genuine figures, i.e., convert the categorical attributes into numeric information. The prediction rate for single classifiers was accomplished after applying this model to our information, with a delicacy of 0.8848902.

Table1: Accuracy of level 1 classifiers

Model	Accuracy
Decision Tree	0.8790024
Random Forest	0.8905202
K- Nearest Neighbor	0.8897394
Support Vector Machine	0.8848902

Now, the authors present the various results of the study. Figure 3 shows each model's delicacy and a correlation between them.

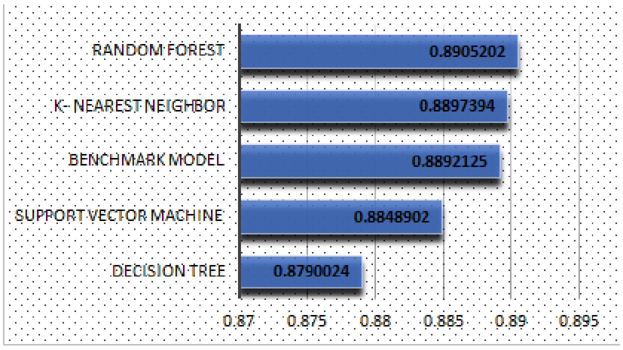

Fig 3. Model Accuracy

Figure 4, clearly represents the accuracy comparisons of various models with the proposed stacking ensemble model.

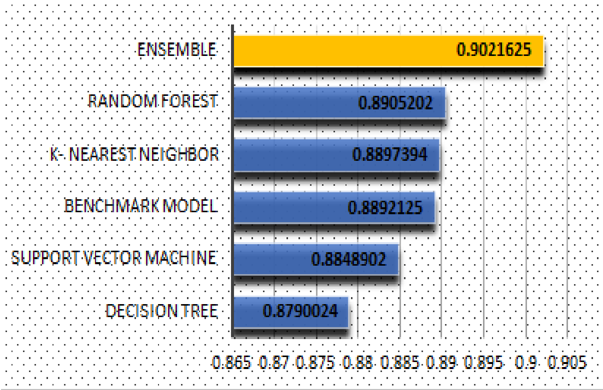

Fig 4. Accuracy comparison with stacking ensemble

IV. CONCLUSION

Predictive models and recommender systems producing accurate suggestions can support learners to grab amazing opportunities. The authors, well experienced academicians in self-financing professional sector, were in constant study of finding a best prediction system for employability which in turn can help the HEIs for the timely educational interventions and equip their stakeholders for suitable employment. This work is an extension of the previous studies made by the same authors by engaging ensemble learning techniques.

For any prediction system, choosing the correct set of attributes which give best result, is a challenge. This study is carefully done by employing different algorithms in Machine Learning and then the stacking ensemble model which combines the predictions and produces more accurate results. To minimize the error possibilities, in prediction, R2, Mean Square Error and Root Mean Square Error scores were studied with two different test data ratios. On comparison with the actual placement data of three consecutive years and the better prediction accuracy from the previous study, the attributes chosen for the datasets can be made optimum. The prospective work may introduce possible consideration of psychological attributes which may help in bringing out the attitude of the ward.

V. REFERENCES

1. *Aggarwal, Deepti, Sonu Mittal, and Vikram Bali. 2021. "Significance of Non-Academic Parameters for Predicting Student Performance Using Ensemble*

Learning Techniques." International Journal of System Dynamics Applications (IJSDA) 10 (3): 38–49.

2. *Ambili, P. S. and Biku Abraham. 2022. "A Predictive Model for Student Employability Using Deep Learning Techniques." ECS Transactions 107 (1): 10149, https://doi.org/10.1149/10701.10149ecst.*
3. *Abraham, B., Ambili, P.S. (2023). An Enhanced Career Prospect Prediction System for Non-computer Stream Students in Software Companies. Lecture Notes in Electrical Engineering, vol 984. Springer, Singapore. https://doi.org/10.1007/978-981-19-8493-8_60.*
4. *Amrieh, Elaf Abu, Thair Hamtini, and Ibrahim Aljarah. 2016. "Mining Educational Data to Predict Student's Academic Performance Using Ensemble Methods." International Journal of Database Theory and Application 9 (8): 119–36.*
5. *Berntson, Erik, Magnus Sverke, and Staffan Marklund. 2006. "Predicting Perceived Employability: Human Capital or Labour Market Opportunities?" Economic and Industrial Democracy 27 (2): 223–44.*
6. *Casuat, Cherry D., and Enrique D. Festijo. 2019. "Predicting Students' Employability Using Machine Learning Approach." In 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 1–5. IEEE.*
7. *Dong, Xibin, Zhiwen Yu, Yifan Shi, and Qianli Ma. 2020. "A Survey on Ensemble Learning." Frontiers of Computer Science 14 (2): 241-58.*
8. *Hashim, Ali Salah, Wid Akeel Awadh, and Alaa Khalaf Hamoud. 2020. "Student Performance Prediction Model Based on Supervised Machine Learning Algorithms." In IOP Conference Series: Materials Science and Engineering, 928:032019.*
9. *Hido, Shohei, Hisashi Kashima, and Yutaka Takahashi. 2009. "Roughly Balanced Bagging for Imbalanced Data." Statistical Analysis and Data Mining: The ASA Data Science Journal 2 (5–6): 412–26. https://doi.org/10.1002/sam.10061.*
10. *Krawczyk, Bartosz, Leandro L. Minku, João Gama, Jerzy Stefanowski, and Micha\l Woźniak. 2017. "Ensemble Learning for Data Stream Analysis: A Survey." Information Fusion 37: 132–56.*
11. *Pillai, A.P.S., "AIoMT-Assisted Telemedicine A Case Study of eSanjeevani Telemedicine Service in India", Handbook of Security and Privacy of AI-Enabled Healthcare Systems and Internet of Medical Things, 2023, pp. 445–464, DOI: 10.1201/9781003370321-19.*
12. *Ambili, P. S., and Varghese Paul. "User Span Pattern: A Sequential Pattern Mining Approach for Personalization." International Journal of Applied Engineering Research 11.1 (2016): 621-624.*
13. *Kumari, Pooja, Praphula Kumar Jain, and Rajendra Pamula. 2018. "An Efficient Use of Ensemble Methods to Predict Students Academic Performance." In 2018 4th International Conference RAIT, 1–6. IEEE.*
14. *Lim, Theng-Wai, Kok-Chin Khor, and Keng-Hoong Ng. 2019. "Dimensionality Reduction for Predicting Student Performance in Unbalanced Data Sets." Int. J. Advance Soft Compu. Appl 11 (2).*
15. *Mezhoudi, Nesrine, Rawan Alghamdi, Rim Aljunaid, Gomathi Krichna, and Dilek Düştegör. 2021. "Employability Prediction: A Survey of Current Approaches, Research Challenges and Applications." Journal of Ambient Intelligence and Humanized Computing, 1–17.*
16. *Rincy, Thomas N., and Roopam Gupta. 2020. "Ensemble Learning Techniques and Its Efficiency in Machine Learning: A Survey." In 2nd International Conference on Data, Engineering and Applications (IDEA), 1–6. IEEE.*
17. *Saleem, Farrukh, Zahid Ullah, Bahjat Fakieh, and Faris Kateb. 2021. "Intelligent Decision Support System for Predicting Student's E-Learning Performance Using Ensemble Machine Learning." Mathematics 9 (17): 2078.*
18. *Ambili P S, Agnesh L, & Arun K V. (2023). Siamese Neural Network Model for Recognizing Optically Processed Devanagari Hindi Script. International Journal of Computational Learning & Intelligence, 2(3), 107–113. https://doi.org/10.5281/zenodo.8210372.*
19. *Tripathi, Akarshita, Saumya Yadav, and Rajiv Rajan. 2019. "Naive Bayes Classification Model for the Student Performance Prediction." In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), 1:1548–53. IEEE.*
20. *Wakelam, Edward, Amanda Jefferies, Neil Davey, and Yi Sun. 2020. "The Potential for Student Performance Prediction in Small Cohorts with Minimal Available Attributes." British Journal of Educational Technology 51 (2): 347–70.*
21. *Webb, Geoffrey I., and Zijian Zheng. 2004. "Multistrategy Ensemble Learning: Reducing Error by Combining Ensemble Learning Techniques." IEEE Transactions on Knowledge and Data Engineering 16 (8): 980–91.*
22. *H. V. Ramachandra, A. Ali, P. S. Ambili, S. Thota and P. N. Asha, "An Optimization on Bicluster Algorithm for Gene Expression Data," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/GCAT59970.2023.10353373.*