Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7922919

ENHANCED ENSEMBLE VOTING BASED MACHINE LEARNING TECHNIQUE FOR STUDENT CAMPUS PLACEMENT PREDICTION

B.Kalaiselvi[1] Dr.S.Geetha[2]
[1.] Research Scholar, Department of Computer Science, Government Arts College, Udumalpet, Tamilnadu
[2.] Assistant Professor, Department of Computer Science, Government Arts and Science College for Women, Puliakulam, Coimbatore.

Abstract: Academic performances and campus placements decides the student's carrier and reputation of the educational institutions. Today various methodologies are available to predict or forecast the students' performance in academic and placement interviews. In the educational data mining (EDM), apart from data mining techniques, Machine Learning (ML) techniques needed to achieve high standard of prediction results to enhance the quality of the education, institution and student performance. An ensemble model from ML helps to combine different models and provides improved results. In this work, Multilayer Perception (MLP), Naïve Bayes tree (NBT), Logistic model tree (LMT) and J48 classifiers are utilized in ensemble voting process to predict the students' placement possibility. In the first ensemble voting model, BayesNet and J48 classifiers are combined using its mean value of the probability with combination rule and the second ensemble voting model, MLP and LMT are combined using its mean value of the probability, obtained attributed and instances are passed to NBTree for further process. The MLP and LMT based ensemble voting model produced better accuracy (92.8%) than first model (91%) with least error prediction rate.
Keywords: Campus Placement, Ensemble Voting, Logistic Model Tree, NBTree

1. INTRODUCTION

Today, Indian academic educational institutions have huge competitions in admissions and providing more placements than other institutions. In this competitive environment, higher educational institutions needed some advanced techniques and strategies to provide better environment for the students to learn well and show good performance in academic and placements. EDM with advanced ML techniques helps to the higher educational institutions to improvise the teaching strategies, obtaining more students in admission, identifying the student's skills, learning difficulties, enhance the knowledge base and increase the placement count. EDM with ML techniques analyse the student data, extract the hidden knowledge which are required to the educational institutions.

Extracted and obtained knowledge from the student data helps to academic staffs and placement trainers to improvise the teaching strategies, concentrate the non-performing students and make all to score high marks in the academic and getting more offers in the campus interviews with good package from the leading companies.

This work concentrates to satisfy the following three objectives:

1. To analyse the ensemble voting based ML models for student data analysis

2. To compare the performance of the ensemble voting models in EDM and determine the opt model for placement prediction
3. To predict the first and second year student's placement status

In this work, real time student data is used from various departments with academic performance, medium and type of school in the school level, first graduate in the family, area of residence, marks obtained in the end semester exams and activities in the class rooms and placement training related attribute data.

II. LITERATURE SURVEY

Aman et al [1] proposed LMT predictive model for choice selection for further studies with real data set with academic demographic and socio-economic futures based attributes obtained from Univ. of Peshawar. J48 and Random Forest are compared with LMT model. The proposed LMT model achieved 83.1% of accuracy.

LH Son et al [2] proposed MANFIS with RS for student performance prediction process for real student data set from VNU Univ. of Sci. and also 3 educational data sets obtained from KDD data sets. Experimentally validated with other fuzzy and tree based models and obtained good accuracy than fuzzy and tree based models.

Shreyas et al [3] utilised Naive Bayes and KNN ML models for placement prediction in current student data set based on the historical student data set. Passed out student data set with placement status is used as the training data set for the proposed ML models.

Sultana et al.[4] proposed CNN based deep learning model for student performance prediction with train and model of the data set. CNN based deep learning model produced 97.5% of accuracy and it shows better accuracy than other models.

T. Jose et al.[5] analysed SVM, LR, KNN and Random Forest for placement prediction process and compares the performances measures and accuracy. Placement training scores with quantitative aptitude, reasoning, verbal, technical programming and academic CGPA, backlogs and certification details are used as the parameters.

B.Kalaiselvi et al [6], presented ensemble model using AdaBoost classifier with Decision stump, NB Tree and Random Forest classifiers for student placement data analysis and identified AdaBoost with Random Forest classifier produced high accuracy (87.09%) than Decision Stump and NB Tree classifiers. AdaBoost helps to increase the performance of the Random Forest. Without AdaBoost, Random forest produced 79.85% accuracy only.

B.Kalaiselvi et al [7],utilised J48 to classify the student academic data and predict the academic performance in the covid pandemic period. In this model, real time student academic data set is utilised for the classification and end semester exam performance prediction process with 96.42% of accuracy.[8], J48 utilised to predict the student placement possibilities with 87% of accuracy on the full data set. Passed out students data set is utilised in this prediction.

Amirah M S et.al. [9] analyse and suggest the prediction methods for student performance using NB, SVM, DT, KNN and Neural network. Students CGPA and Internal marks based data set employed for this process.

3. METHODOLOGY

In this work, two ensemble voting models are compares based on its performance in placement prediction process in the student data.

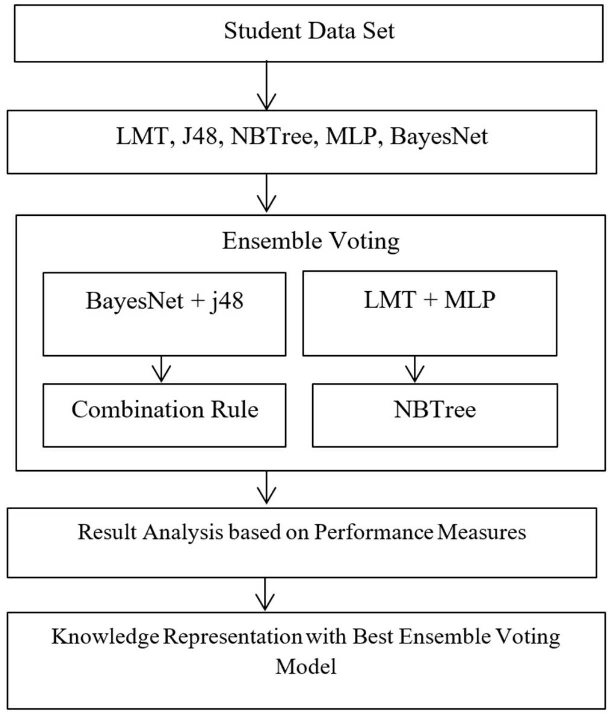


Fig. 3.1 Proposed Methodology

Student data set

The data set is collected directly from the students using Google forms. The collected data from the students are pre-processed with removal of attributes like student name, email id and mobileno,misspelled words correction, handling the missing column attributes in the instance and defining the dependent and independent attributes process.

After pre-processing, student data set is divided into train (75%) and test (25%) data sets. Obtained student training data set is passed to the two ensemble voting based ML models to train and fit the model. The obtained test data set model is used to analyse the fitness of the two ensemble voting based ML models.

After pre-processing, area of living, gender, first graduation in the family, obtained marks in the school level with medium and board of study, obtained marks in the end semester exam, attendance percentage, arrears in the end semester exams, certifications completed, activity in the class room such as seminars, interaction with staffs, placement training and placement status are the key attributes

J48 and BayesNet are used as base classifier in the first ensemble voting model and MLP and LMT are used as the base classifier in the second ensemble voting model. In the both models, mean of the probability obtained from the classifiers are taken for the combination rules. J48+Bayesnet based ensemble voting model output is passed to the combination rule model for further classification and prediction process. In the second model, NBTree is used for further process.

BayesNet

BayesNet is a probability based graphical model which describes local dependencies from the set of attributes. BayesNet is used to calculate the a posteriori (works backwards) probability distribution of nodes (variables) from the given values of other attributes which are observed.

J48

J48 is an effective tree based ML algorithm for category or sequence based data set. In j48, data set is split into sub sets and utilized divided and conquer strategy for the classification process. Sometimes it will produce empty andinsignificant branches and many fragmentations needed when over fittingproblemare raised.

LMT

In LMT, Logistic Regression (LR) and decision tree based learning system combined together for supervised learning approach. LR function used in the classification trees at the leaves level. LMT algorithm deals with binary or multi-class target attributes / variables.

MLP

Back propagation technique is used in the MLP to classify the instance. The network will be generated by this classifier and it can be modified in the training period. MLP is a fully connected feed forward ANN with input, hidden and output layers.

NBTree

Naive Bayes and decision tree based learning system combined together for supervised learning approach. Naive Bayes used in the classification trees at the leaves level.

Ensemble Voting

It combines the prediction results of the more than one ML model. Softvoting ensemble method is used in this work. Softvoting selects the target label which one has highest weight of probability mean or sum value.

The obtained results from the both models are compared based on the performance measures such as classification accuracy, true and false positive values, true and false negative values and F-Measure values.

4. PERFORMANCE EVALUATION

In this binary class based placement prediction on the student data, F-Measure, Precision and Recall values are analysed. These values are obtained from the true positives, false positives, true negatives and false negatives.

- TP – specifies how many instances accurately estimated as positive
- FP specifies how many instances in accurately estimated as negative.
- TN specifies how many instances accurately estimated as negative and
- FN specifies how many instances inaccurately estimated as negative

5. EXPERIMENTAL RESULTS

This system tested in Core i3 Processor with 8 GB RAM and Python with ML packages utilized to implement the placement prediction process using ensemble voting classifiers.The experimental results first model, J48+Bayesnet based ensemble voting model and second model, MLP + LMT are described in the following table 1.

Table 1 Performance of Measures

Ensemble Voting Model	Accuracy	TP	FP	F1
J48 + BayesNet	91.07	0.91	0.01	0.91
MLP+LMT	92.85	0.93	0.13	0.93

The J48+ BayesNet based ensemble voting model produced 91.07% of accuracy with 0.91 as TP, 0.01 as FP and 0.91 as F-measure values. The MLP + LMT based ensemble voting model produced 92.85% of accuracy with 0.93 as TP, 0.13 as FP and 0.93 as F-measure values. It shows that the MLP and LMT based ensemble voted classifier produced high accuracy than other model in the placement prediction process.

Fig. 5.1 Classification Accuracy

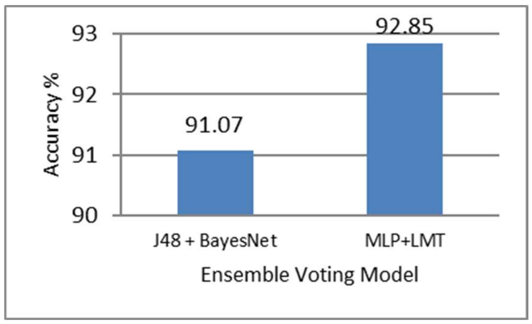
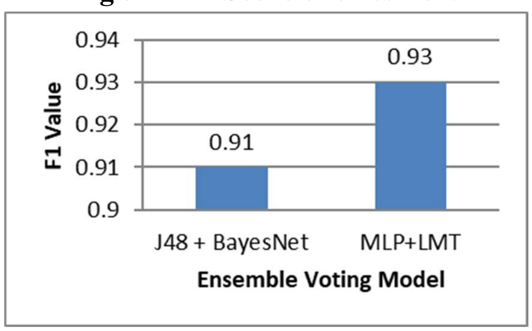


Fig 5.2 F1 –Score of classifiers



The above figures (1 and 2) describe performance of both ensemble voting models.

Table 2 Statistical Measures

Ensemble Voting Model	Kappa	MAE	RMSE	RAE %	RRSE %
J48 + BayesNet	0.65	0.10	0.26	46.5	80.94
MLP+LMT	0.70	0.10	0.23	47.3	69.94

The above table 2 describes statistical performance values of both ensemble voting based ML models. The first model produced 0.65 as Kappa, 0.10 as MAE, 0.26 as RMSE, 46.5% as RAE and 90.94% as RRSE. The second MLP+LMT based model produced 0.70 as kappa, 0.10 as MAE, 0.23 as RMSE, 47.3% of RAE and 69.94% of RRSE. It reveals that the MLP + LMT based ensemble voting ML model produced good performance than J68 + BayesNet model.

CONCLUSION

Ensemble voting based ML model which utilized MLP and LMT classifiers is efficient one for student placement prediction. This MLP and LMT based ensemble voting ML model gives slightly high accuracy and f-measure than other J48 and BayesNet based ensemble ML model.

REFERENCES

1. Aman, Fazal and Ali., 2019. "A Predictive Model For Predicting Students Academic Performance.”, 10th Intl. Conf. on Infor., Intelligence on Info., Intelligence, Sys. andAppli. (IISA), 2019.
2. LH Son and H Fujita, "Neural-fuzzy with representative sets for prediction of student performance", Applied Intelligence, Springer, 2018.
3. Shreyas H, Aksha P, and Suraksha. "Student Placement Prediction using ML", Inl. Research Journal of Engineering and Technology (IRJIET) Volume : 06 Issue: 04 April 2019
4. Sultana, Jabeen, MU Rani, and Farquad. "Student’s performance prediction using deep learning and data mining methods." Int. Journal. Recent Technol. Eng 8.1S4 (2019): 1018-1021.
5. I.T Jose, D Raju, , JA Aniyan, J James, and MT Vadakkel, "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison", IJISRT – May 2020.
6. B.Kalaiselvi and Dr.S.Geetha, "Ensemble Machine Learning AdaBoost with NBTree Model for Placement Data Analysis", International Conference on Intelligent Technologies (CONIT), IEEE-2022.
7. B.Kalaiselvi and Dr.S.Geetha, "Academic performance prediction for first year students in covid pandemic period using data mining technique", Journal of the Maharaja Sayajirao University of Baroda, Volume-54, No.2 (XVI) 2020.
8. B.Kalaiselvi and Dr.S.Geetha, “Analysis of placement performance prediction on students data using machine learning algorithm”, Journal of the Maharaja Sayajirao University of Baroda,Volume-56, No.1(IX) 2022
9. Amirah M S., Wahidah H, &Nuraini A R, " Review on Predicting Student’s Performance Using DM Techniques. Procedia Computer Science, 72, 414–422. http s:// doi.org/10.1016/J.PROCS.2015.12.157
10. Dech T, Dursun D, Phayung M and Nihat K, "Critical Assesmnt of imbalnced class distribution problem : the case of predictingfreshmen student attrition", Expert Systems with Applications,Elsevier 2013
11. Laci, Sergio and Geraldo, Automatic Prediction of Student Performance in STEM_UG programs. ACM, 2016
12. Reddy, C. S., Yookesh, T. L., & Kumar, E. B. (2022). A Study On Convergence Analysis Of Runge-Kutta Fehlberg Method To Solve Fuzzy Delay Differential Equations. JOURNAL OF ALGEBRAIC STATISTICS, 13(2), 2832-2838.
13. Kumar, E. B., & Thiagarasu, V. (2017, October). Color channel extraction in RGB images for segmentation. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 234-239). IEEE.

14. Yookesh, T. L., et al. "Efficiency of iterative filtering method for solving Volterra fuzzy integral equations with a delay and material investigation." Materials today: Proceedings 47 (2021): 6101-6104.

15. Kumar, E. Boopathi, and M. Sundaresan. "Edge detection using trapezoidal membership function based on fuzzy's mamdani inference system." 2014 International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2014.