Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System

Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda

Abstract*: Campus placement plays a vital role in every educational institution in helping students to achieve their goals. Data mining classification can be used as a useful tool for extracting the associated information from the large scale student dataset. Data mining methods have been used broadly in the area of the education system which involves various methods and approach for discovering knowledge. In this paper, a predictive model is designed which can predict the category of placements (dream companies, super dream companies and mass recruiter companies) in which students are eligible by considering their past performance in academics and other curricular activities. The model will also suggest further skills required for future recruitments which may help the students for placement preparation. The paper also provides real-time experimental results and findings along with performance measures used for model validation which helps in achieving the milestone of outcome-based education (OBE) in educational institutes as it is given utmost importance in present scenario to ensure better placement prospects in students, which would in turn help the students for carrier building.

Keywords*: Classification, Data Mining, Outcome-based education, Placement Prediction.

I. INTRODUCTION

An OBE curriculum starts with a clear picture of what students should attain in accordance with the curriculum designed, tutoring methods adopted and graduate attributes to be met. The final attainment achieved will ultimately make sure that learning happens and course outcomes are attained. At this moment, OBE is being adopted at a faster pace at engineering institutions all over India. In India, it is considered to be a giant leap forward for improvement in technical education which would help Indian Engineers to compete globally. OBE, which is a student-centred instruction system, focuses on student performance through gauged outcomes as knowledge gained, skills inculcated and attitudes perceived. The institution is given the right to decide on the assessment method for candidates during the program. Various assessment tools for gauging Course Outcomes include class tests, assignments, quiz, project work, labs, presentations, mid-semester and semester-end examinations, employer/alumni feedback which could be incorporated in educational institutions to meet the objectives of OBE. Even though the adoption of OBE at engineering institutions would be a great initiation for higher education in India, but the real success lies in the effective implementation and rigorous accreditation procedures to be met in order to guarantee that quality education is continued. Therefore OBE will help in filling gaps in academics to match with industry standards; thereby helping students in achieving better job prospects. In educational institutions, the student's placement plays a key role in up-lifting institutional standards. Student's academic performance and their academic skills are strongly influenced by placements. To attain high-quality placements, students should be adapted with qualities like problem-solving skills, sincerity and hard work, teamwork, and multitasking. It will be a boon to all students if these qualities are compiled in advance before the commencements of placement drives. Considering the above inputs a model can be proposed which may predict the outcome of the student's placements option, based on their past performance in academics thus bringing the above concept to reality.

At present, a huge amount of data is compiled and stored in educational institutions related to student enrollment, progress reports, examination results and many more. Educational data mining (EDM) is an evolving discipline, which is concerned with data set exploration from various sources and developing methods. Many of the techniques that we use in EDM come from computer science in the field of machine learning and data mining; where computers are used to analyze a huge amount of data. An important technique used in EDM is prediction modelling, in which a model could be developed which infers from one specific aspect of the data i.e. predicted variable which would help in future event prediction. In this paper, a model is proposed which could help the students in predicting the placement category by considering their past academic performance. The model also provides the necessary suggestions and guidelines required to improve the student's skills for the further recruitment process.

Revised Manuscript Received on September 20, 2019.

***** Correspondence Author

Abhishek S. Rao*, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: abhishekrao@nitte.edu.in

Aruna Kumar S V, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: arunkumarsv@nitte.edu.in

Pranav Jogi, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: jogi.pranav17@gmail.com

Chinthan Bhat K, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: bhatkchinthan@gmail.com

Kuladeep Kumar B, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: deepsalyan@gmail.com

Prashanth Gouda, Dept. of Information Science & Engineering, NMAM Institute of Technology, Nitte, India. Email: prashantgoud98@gmail.com

Hence the proposed model is a classification problem, which predicts the categorical label such that the students are eligible to get placed in Dream/Super Dream/MRC. To build the model, we first need to collect the data for prediction. The required data were collected from five engineering departments. Academic details of five years were consolidated and used in the present study for training the model. The consolidated data had various information regarding technical events participated, internships underwent, certifications courses completed, and CGPA attained. Collected data were then synchronized to compute the features for model prediction. Later, machine learning algorithms such as KNN (K-Nearest Neighbors), SVM (Support Vector Machine) and ANN (Artificial Neural Network) were applied to identify patterns in the data that are associated to the prediction model. Once the model is created, it was assessed using various performance metrics such as accuracy, precision, sensitivity, F1-score, and Area under the ROC Curve (AUC).

Rest of the paper is organized as follows. Section II gives an overview of the literature, highlighting on various data mining approaches, classification algorithms, applications of machine learning techniques, model design, performance analysis and various case studies with respect to student placement prediction. Section III gives an overview of the system architecture. Section IV highlights on experimental evaluation and results obtained for various classification algorithms along with the performance metrics. Section V gives the comparative study of various performance metrics used for model validation in order to identify a suitable algorithm for the given dataset. Section VI concludes the paper along with the recommendation.

II. LITERATURE REVIEW

In this paper, the author emphasized on the mapping of students performance pattern using a data mining approach with the help of a case study in Indonesia. K-means clustering was used for revealing the hidden patterns from the data. A total of 306 valid cases were used for the study. Various attributes selected were GPA, Lecture grades, and Lab grade. The student's performance was classified into three categories i.e. low, average and smart [1].

This paper highlights on a database application for huge data management in colleges. Decision Tree and association rules algorithm were used for data pre-processing, in which student's scores were analyzed using decision tree algorithm and consumption, psychological status was analyzed using association rule algorithms. Feature selection method was used to select the best feature among the given dataset. Attributes used for the study were data analysis from library records, consumption records, student score, and psychological tests. Mean phenotypic value and analysis were applied to analyze the reasons for poor performance and psychological status testing model methods were used to determine physiological diseases in students. The performance was classified as poor, and excellent. Finally, the designed system could automatically show the results using the data mining algorithms chosen for the study [2].

In this paper, the author has examined the potency of applying machine learning techniques for a stronger prediction of successful graduation in students. More than 30,000 student's observations were considered for the study. Random forest with genetic algorithm was successfully used for feature selection. Predictive strength was assessed by matthews correlation coefficient. Attributes used for the study were GPA, retention and graduation [3].

In this paper, the author designed a classification based algorithm (CBA) for predicting slow learners among students. WEKA (Open source tool) was used for filtration of desired potential variables. Student academic record dataset was applied on various CBA like multilayer perceptron, Naïve Bayes, SMO, J48 and REPTree and comparative study was done. Paper highlighted the importance of predictive model for slow learners. 152 records were taken for analysis in this research. Attributes selected were student's gender, type of board, the medium of instruction, Private tuition, area of school level, a computer at school, grade attendance and many more. Multilayer perceptron classifier performed the best with 75% accuracy [4].

This paper highlights on the importance of campus placement in student's life and how the prediction of the campus placement can be achieved with the help of data mining algorithms like fuzzy logic and K-Nearest Neighbors (KNN). Results indicated that the accuracy and execution time of KNN was 97.33% and 13458 (msec) when compared to fuzzy logic with accuracy 92.67% and execution time of 450 (msec) [5].

In this paper, the author designed a data model for better management of student's performance, using senior student's dataset. J48 algorithm was used and the accuracy of the model was compared with other data mining algorithms. High accuracy of 97.27% was achieved by J48 algorithm when compared to Naïve Bayes and multilayer perceptron i.e. 85.92% and 94.94% respectively. Selected features for the study were CGPA, arrears, attendance, PUC marks, Engineering cut off, Medium of education and type of board. Performance classified notations given were good average and best [6].

This paper illustrates that without examinations, machine learning algorithms could also be applied to analyze student's placement examination based on features like psychological scale, programming tasks and the student answered questionnaires. Decision tree classification model with decision tree gave an F-measure of 0.912 when 9 explanatory variables were used. Whereas the best-ranking model with SVM rank has a normalized discounted cumulated gain of 0.962 when 20 explanatory variables were used [7].

This paper emphasizes the use of DM methods in the field of education. A TPO management system was designed which could successfully detect eligible students for campus drive. Decision tree C4.5 algorithm was applied for company's past year data and current requirement, which would indirectly be useful to students since the model would send a notification to eligible candidates thereby helping them to know if they are eligible for it.

This would help them prepare well in advance for the campus drive. The attributes used for the study were academic history like percentage marks, skill set, programming skill, communication skill, analytical skill and teamwork [8].

In this paper, the author has highlighted on the model which could predict the recruitment in an organization using the ID3 decision tree algorithm to effectively select candidates in a cost-effective manner within less time. Qualitative and quantitative attributes selected were years of experience, employment status, current salary, level of education, whether from top tier school, internships and research papers [9].

This paper highlights a case study in Ethiopia were DM techniques were used for classifying students into different university based on their entrance exam. Students who had higher exam score, females from developing regions were given the highest priority to get placed in their first choice department. WEKA Data mining tool was used and various algorithms like (J48, Naïve Bayes, and Random Forest) were used for prediction model building. The analysis result showed that random forest algorithms with 22 attributes performed the best with an accuracy of 99.33% [10].

In this paper, the author has developed a student's academic performance prediction model for BE and M. Tech students of CS and EC department using two classification methods namely decision tree and fuzzy genetic algorithm. The attributes used were internal marks which include attendance, assignment marks, an average of two sessional, admission score which included a weighted score of 10^{th} and 12^{th}, and exam score. The developed model would early predict student performance for each subject so that lecturers could classify students and take remedial actions for helping students to improve in their performance for better results. The system is more transparent as students can view their academic details and other updates online. Industries which have to tie-up with the college can also search students based on their needs [11].

In this paper, the author had conducted studies on designing student placement prediction system using fuzzy interference. The system helped in managing huge data and classifying it further into two classes i.e. placement and non-placement category. Various variables considered for the study were 10^{th} and 12^{th} marks, B. Tech and M. Tech CGPA, and backlogs (if any). Finally, the proposed design could predict if the student is eligible for placement or not [12].

In this paper, the author had designed a model which could predict the student performance analysis using a decision tree classifier. Classification technique was used for skill analysis based on various attributes related to performance and student activity. Results could predict the improvement area needed and also help in making wise decisions to prevent academic loss. Improvement area which student must work on for goal fulfilment is also emphasized by this technique [13].

In this paper, the author had designed a model and did a comparative study to predict specializations which suits the candidate based on the entered information. Accuracy, precision and truth positive rate were considered for comparing between algorithm results. The present study would indirectly help students in selecting the best-suited specialization for better placement possibility. Attributes used for the study were a name, category, age, sector rank, address, and gender branch. Clustering model Viz Rock algorithm gave an accuracy of 84.3% when compared with FP growth (81%) and Naïve Bayes (77%). Hence the clustering model proved to be the best in predicting the correct specialization [14].

This paper emphasizes on the use of classification algorithms for student data by predicting placement prospect in various multinational companies so that students who are eligible can prepare accordingly. Various attributes like roll number, SSLC and HSC marks, UG and PG marks, programming languages are known and company information known to students was considered for prediction. Performance comparison for various classification algorithms like J48, REPTree, Naïve Bayes, BayesNet and MLP were carried out. Results indicated that the classification accuracy got was better for Tier-1 companies than Tier-2 class companies [15].

In this paper, the author had conducted studies on the use of data mining techniques for campus placement prediction and use of WEKA software for design and implementation. Various parameters which could be considered for calculating student performance are the academic performance, communication skills, technical skills, vocational training and projects. Various clustering algorithms like simple k-mean, Farthest-first traversal, filtered clustering, hierarchical clustering were used for model building. It was seen that the time is taken for building simple k-mean, Farthest-first traversal and filtered clustering was just 0.02sec in comparison with hierarchical clustering (0.09 sec) and density-based clustering (0.08 sec) [16].

This paper emphasizes the use of data mining for performance prediction in a particular subject by students using a C4.5 decision tree algorithm. The need for the timely accompaniment of students for making wise decisions for preventing academic risks was also highlighted [17].

In this paper, the author designed a placement prediction support system with the help of data mining algorithm. The model developed not only helped in finding placement possibility but also helped in predicting the level of interviews the student may clear. Naïve Bayes and Improved Naïve Bayes were considered for the study. WEKA and NetBeans tools were used for data analytics. Results indicated that Improved Naïve Bayes gave an accuracy of 84.7% when compared to Naïve Bayes (80.96%) when 560 instances dataset were considered for the study [18].

In this paper, the author had conducted a study on student placement prediction using a classification model. Data mining tools (WEKA) was used for analysis. Variables like gender, academic results, seminar performance, lab work, Communication skill, graduation background was considered for the study. Various classification algorithms like Naïve Bayes, MCP and J48 were compared and results indicated that Naïve Bayes was the best-suited algorithm with an accuracy of 86.15% and model building time of 0sec [19].

III. METHODOLOGY

The present study focused on real-time institutional data collected from four major engineering disciplines along with training and placement department. After discussion with the concerned authorities of the departments, various factors which could have a major impact in placement prediction was identified and further selected for the study. Major factors identified for the study were related to academic details, extra-curricular details, MOOC courses, and internships. The collected data were then pre-processed for quality data extraction. Later, best features suitable for the present study were identified and selected for further analysis. Fig.1 displays the system architecture of the proposed model.

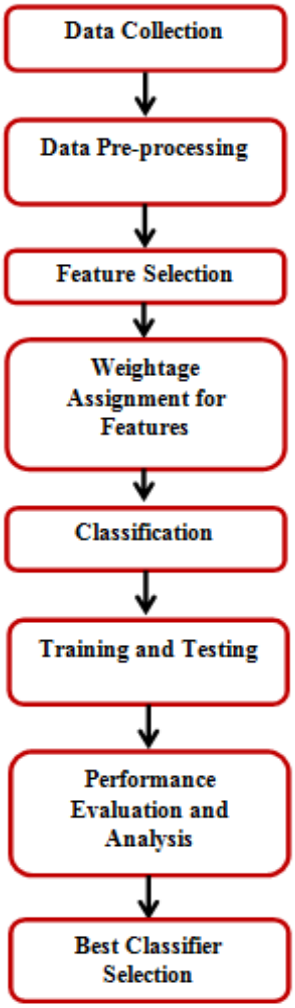


Fig. 1. System Architecture

A. Data Collection

Course grade point average (CGPA) for 4 consecutive years was collected from the academic section. Data related to technical events participated, internships undergone, and certification courses completed were collected from individual departments. As an input for model prediction, information such as skills required for students in getting placed in MRCs, Dream, and Super Dream companies was collected from placement department for helping in placement prediction in a better way.

B. Data Preparation & Pre-processing

Data preparation is a step in a data analysis process in which data from one or more sources is cleaned, transformed and enriched to improve the quality of data prior to its use. The collected data were then pre-processed to fill the missing data and made compatible for further processing.

C. Feature Selection

The data containing eight features was further reduced to four features with the help of the dimensionality reduction approach. This step was carried out with the help of expert advice both from academic and placement departments in order to select the most appropriate features for faster and effective classification of the data.

D. Weightage Assignment for Features

As per the expert's advice, it is evident that students cannot excel in all domains; therefore appropriate weight was assigned for each feature for effective prediction. The weight assigned for internships, extra-curricular details, MOOC courses, and academic performance is 20%, 20%, 20%, and 40% respectively.

E. Classification Methods

The main objective of the classification technique is to assign the pre-defined labels based on the features selected. The proposed model will help in classifying the data into three categories namely Best (Super Dream Company), Good (Dream Company), and Average (MRC). Among the various classification algorithms, proposed work utilizes KNN, SVM and ANN which are considered to be the best-suited algorithms for the present study. A total of 1023 data samples were collected from various departments of the institution.

The Support Vector Machine (SVM) helps in identifying the hyperplane for classifying the data samples. In the case of multiple hyperplanes, the one which has maximum distance from the data points was chosen for better classification. The dimension of the hyperplane depends on the total number of features. In the proposed study, the SVM algorithm used various mathematical functions (kernels) like radial basis function (RBF), linear, sigmoid, and polynomial. Further, experimental investigation on various kernels was applied to the data samples and results were compared based on the performance metrics to identify the best kernel to classify the data. The K-Nearest Neighbor (KNN) algorithm classifies the data point based on the value of 'k'. The 'k' value refers to the number of neighbors. Depending on the value of 'k', the data point is classified based on the voting from each neighbor. Further, an experimental investigation was carried out for multiple neighbors and a comparative study was carried out based on the performance metrics. The Artificial Neural Network (ANN) consists of several nodes. Each node is capable of taking input and performing operations. ANN consists of an input layer, hidden layers and output layer. The output of each node is termed as its activation. An activation function chooses whether a node should be activated or not by calculating the weighted sum. Activation functions such as Rectified linear units (ReLu), Tanh (Hyperbolic tangent), logistic and identity were used in the present study and compared to choose the best activation function for the given dataset.

F. Training and Testing

For better model validation, the dataset in the present study was split into training and testing with the help of SciKit library. Different proportions like 80:20, 70:30 and 60:40 ratios were made and utilized for the study. A proportion of 80:20 signifies that 80% data is considered as training data and rest 20% data is considered as testing data.

G. Performance Evaluation and Analysis

The performance measurement of the model was evaluated with the help of various metrics like accuracy, sensitivity, F1-score and precision. The performance visualization of the multi-class classification problem was analyzed using a graphical plot AUC (**Area under the Curve**) ROC (**Receiver Operating Characteristics**) curve that reveals the analytical ability of a binary classifier system as its discrimination threshold. The ROC curve is generated by plotting the true positive rate against false-positive rates at various threshold rates. The best algorithm based on the performance parameters was selected to predict the placement category of students. Based on the details provided by the students, the placement category could be predicted and the result would be displayed along with the suggestions for further improvement.

IV. EXPERIMENTS AND RESULTS

The experiments were conducted in an attempt to find the best algorithm for the current model. The algorithms selected to classify the data where SVM, KNN and ANN. The data samples of 1023 instances were used to predict the placements category in which student may get selected.

A. Support Vector Machine (SVM) Experimental Results

In this experiment, SVM classifier was utilized on various kernel functions for varied proportions of data and the results of the performance metrics were compared (Table I).

Table- I: Performance Metrics for SVM Classifier

Ratio	Kernel used	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)	AUC
80:20	RBF	95.12	93.3	96.48	94.8	1.00
70:30	RBF	96.42	94.4	97.49	95.87	1.00
60:40	RBF	95.37	92.44	97.29	94.67	1.00
80:20	Linear	71.22	90.83	88.19	88.61	1.00
70:30	Linear	70.68	93.01	88.24	89.78	1.00
60:40	Linear	70.98	96.3	88.57	91.66	1.00
80:20	Polynomial	79.51	95.75	93.59	94.65	1.00
70:30	Polynomial	81.11	95.32	94.89	95.1	1.00
60:40	Polynomial	81.71	95.72	95.12	95.41	1.00
80:20	Sigmoid	70.24	32.44	25.17	28.35	0.06
70:30	Sigmoid	70.68	32.88	24.3	27.95	0.06
60:40	Sigmoid	71.95	33.33	24.07	27.95	0.04

From Table I, it is evident that the best results for SVM algorithm were obtained for Radial Basis Function (RBF) kernel for 70:30 proportions. Least poor performance was seen when the sigmoid kernel was used for the same proportion.

B. Experimental Result on K-Nearest Neighbor (KNN)

In this experiment, KNN classifier was utilized with multiple neighbors for varied proportions of data and the results of the performance metrics were compared (Table II).

Table- II: Performance Metrics for KNN Classifier

Ratio	No. of Neighbors	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)	AUC
80:20	3	98.05	95.39	97.97	96.52	0.99
70:30	3	98.05	94.76	98.31	96.33	0.98
60:40	3	98.54	95.73	98.75	97.11	0.98
80:20	4	97.56	93.3	97.75	95.18	0.99
70:30	4	98.05	94.76	98.31	96.33	0.98
60:40	4	98.54	95.73	98.75	97.11	0.99
80:20	5	98.05	95.39	97.97	96.52	0.99
70:30	5	98.7	97.23	98.61	97.86	0.98
60:40	5	98.54	95.73	98.75	97.11	0.99
80:20	6	97.56	95.16	97.22	96.01	0.99
70:30	6	98.05	95.84	97.93	96.76	0.99
60:40	6	98.29	95.62	98.36	96.85	1.00
80:20	7	97.56	95.16	97.22	96.01	0.99
70:30	7	98.37	95.99	98.46	97.66	0.99
60:40	7	98.29	95.62	98.36	96.85	1.00
80:20	8	97.56	95.16	97.22	96.01	0.99
70:30	8	98.05	95.84	97.93	96.76	0.99
60:40	8	97.56	95.28	97.25	96.09	1.00

From Table II, it is evident that all the results for the KNN algorithm in all the cases were above 93%. However, when 70% of the data was trained and 30% of the data was tested, considering 7 neighbours, the best result was obtained.

C. Experimental Result on Artificial Neural Networks ANN)

In this experiment, ANN classifier was utilized with various activation functions on the varied proportions of data and the results of the performance metrics were compared (Table III).

Table- III: Performance Metrics for ANN Classifier

Ratio	Activation Function	Accuracy (%)	Sensitivity (%)	Precision (%)	F1-score (%)	AUC
80:20	ReLu	98.54	97.47	98.19	97.78	0.99
70:30	ReLu	99.35	99.7	98.91	99.29	1.00
60:40	ReLu	97.8	97.19	94.34	95.7	1.00
80:20	Logistic	94.15	91.22	95.79	93.14	0.99
70:30	Logistic	88.27	63.24	60.08	61.55	0.99
60:40	Logistic	72.2	33.33	24.07	27.95	0.61
80:20	Tanh	99.02	99.55	98.41	98.96	0.99
70:30	Tanh	98.7	97.23	98.61	97.86	0.99
60:40	Tanh	98.78	97.64	97.93	97.77	1.00
80:20	Identity	94.63	91.44	96.82	93.64	0.99
70:30	Identity	95.44	93.33	96.44	93.41	0.99
60:40	Identity	94.15	92.22	93.41	95.52	0.99

From Table III, it is evident that the results for the ANN algorithm were poor when 60% of the data was trained and 40% of the data was tested using logistic activation function. The best result was obtained when 80% of the data is trained and 20% of the data is tested using Tanh activation function.

V. COMPARATIVE STUDY OF PERFORMANCE METRICS

This section will give a brief summary of various classification algorithms utilized for varied proportions along with comparative results for performance metrics (Fig 2 - 7).

A. Accuracy Analysis

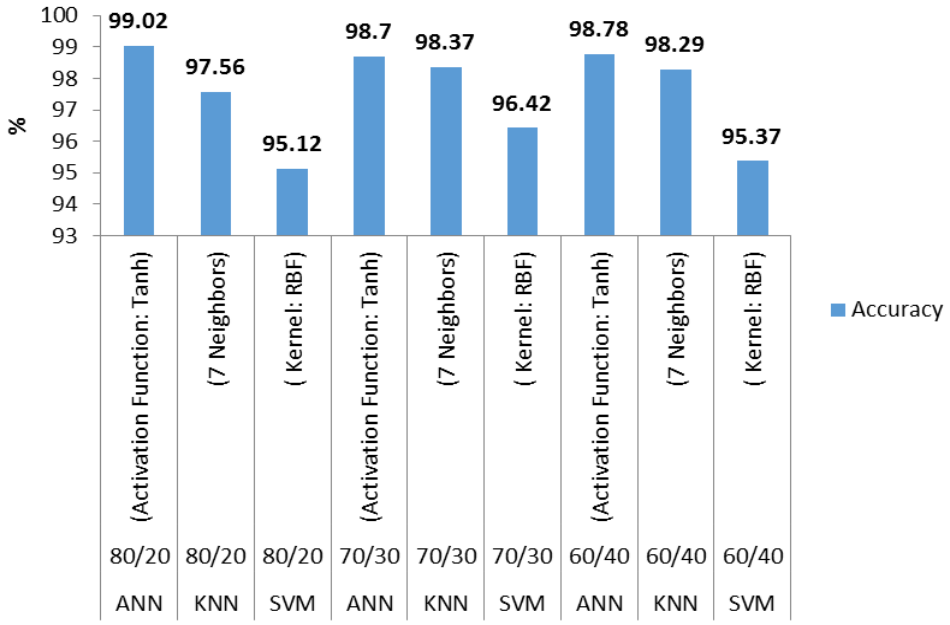


Fig. 2. Comparative study of accuracy analysis of ANN, KNN and SVM algorithms for varied proportions

From Fig. 2, it is evident that the model showed the best accuracy of 99.02% for ANN algorithm for Tanh activation function at 80:20 proportions. The lowest accuracy of 95.12% was obtained for SVM algorithm with RBF kernel for the same proportion.

B. Sensitivity Analysis

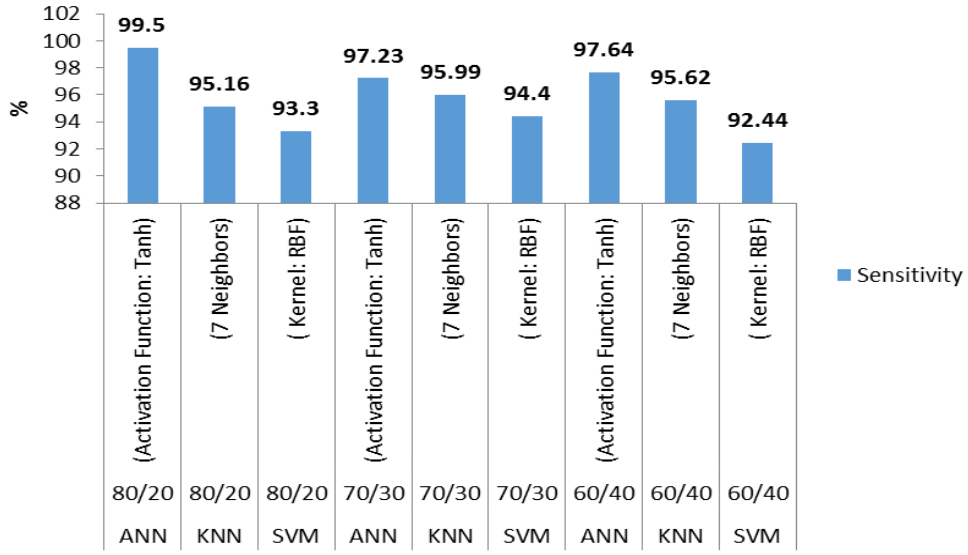


Fig. 3. Comparative study of sensitivity analysis of ANN, KNN and SVM algorithms for varied proportions

From Fig. 3, it is evident that the model showed the best sensitivity of 99.5% for the ANN algorithm for Tanh activation function at 80:20 proportions. The lowest accuracy of 92.44% was obtained for SVM algorithm with RBF kernel for 60:40 proportions.

C. Precision Analysis

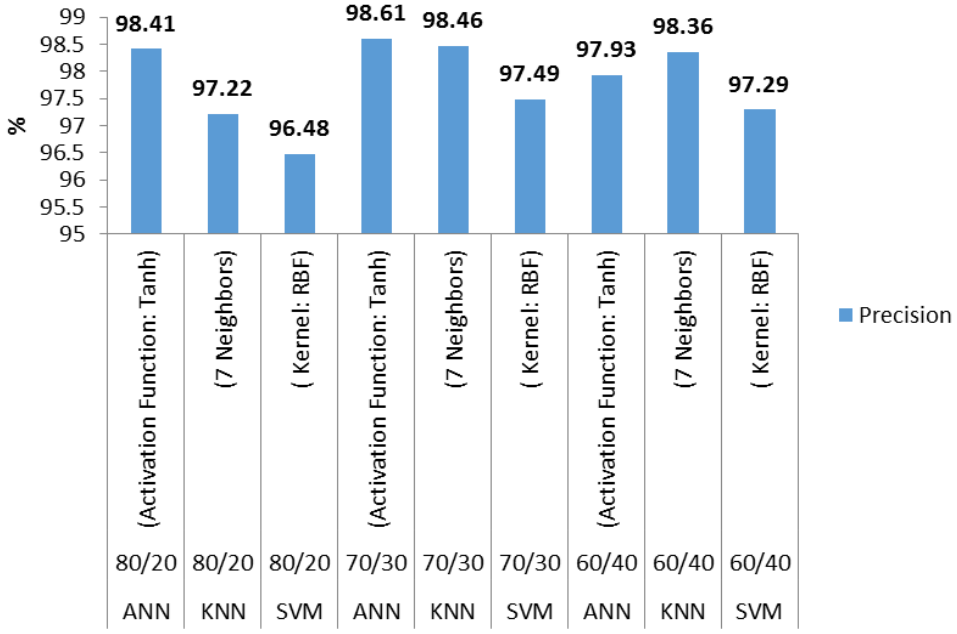


Fig. 4. Comparative study of the precision analysis of ANN, KNN and SVM algorithms for varied proportions

From Fig. 4, it is evident that the model showed the best precision of 98.61% for ANN algorithm for Tanh activation function at 70:30 proportions. The lowest accuracy of 96.48% was obtained for SVM algorithm with RBF kernel for 80:20 proportions.

D. F1 Score Analysis

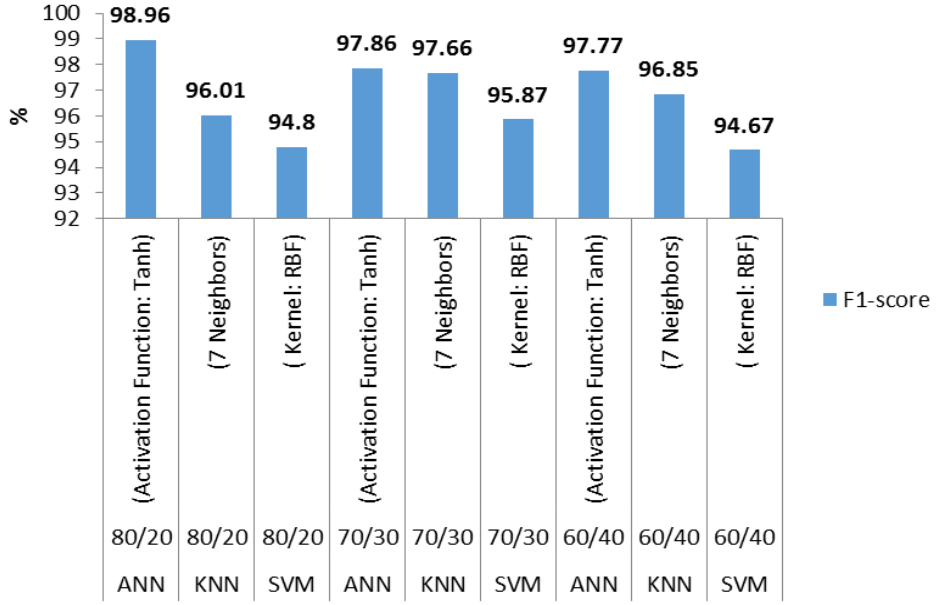


Fig. 5. Comparative study of F1-score analysis of ANN, KNN and SVM algorithms for varied proportions

From Fig. 5, it is evident that the model showed the best F1-score of 98.96% for ANN algorithm for Tanh activation function at 80:20 proportions. The lowest accuracy of 94.67% was obtained for SVM algorithm with RBF kernel for 60:40 proportions.

E. AUC (**Area under the Curve**) ROC (**Receiver Operating Characteristics**) Analysis

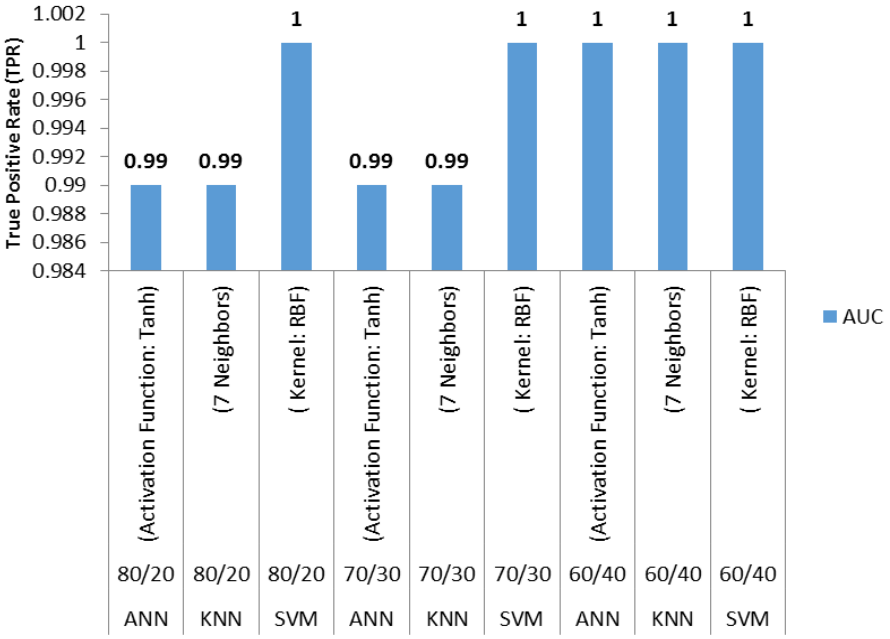
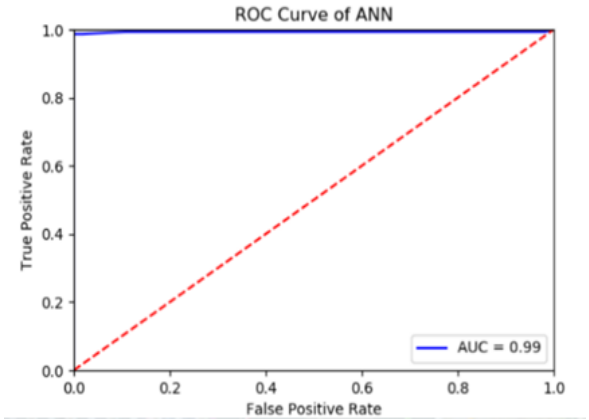
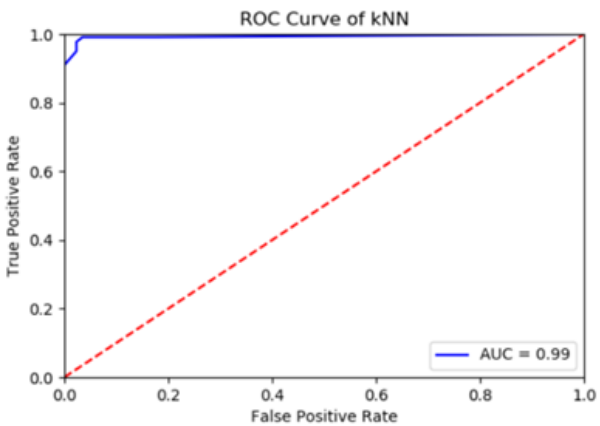


Fig. 6. Comparative study of AUC analysis of ANN, KNN and SVM algorithms for varied proportions

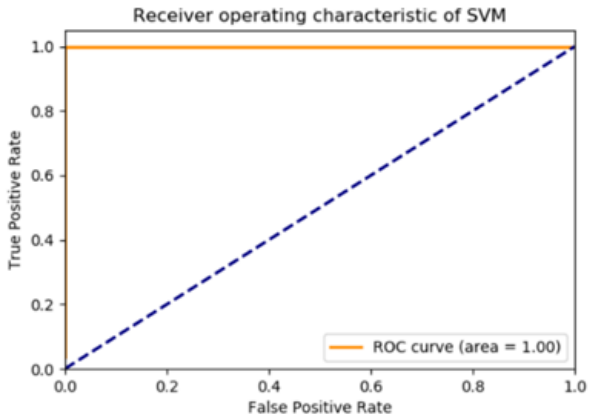
From Fig. 6, it is evident that the model showed a true positive rate (TPR) of above 0.99 in all the cases.



(a)



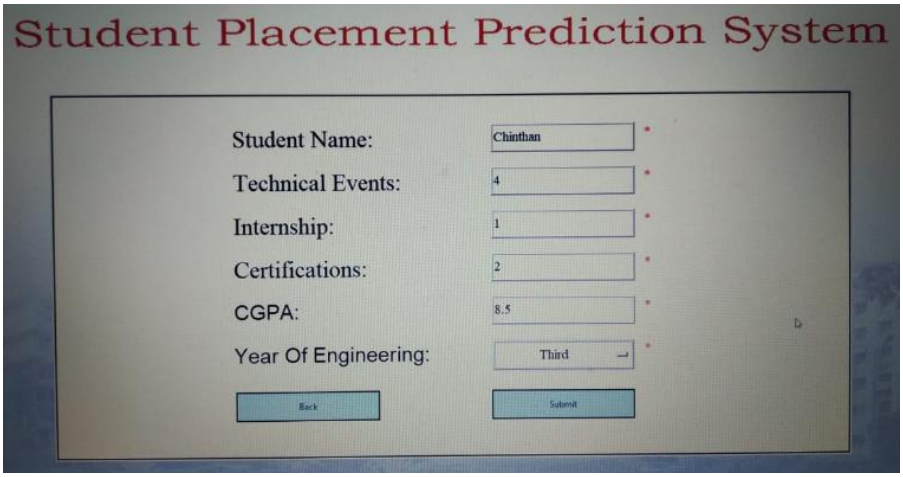
(b)



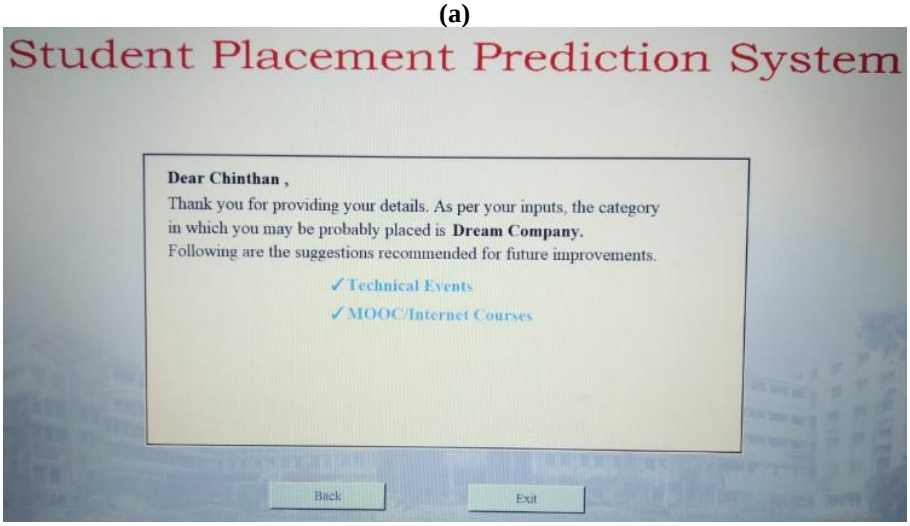
(c)

Fig. 7. ROC curve for (a) ANN (80:20) (Tanh), (b) KNN (70:30) with 7 Neighbors and (c) SVM (70:30) (RBF)

It is evident from Fig. 7, that the true positive rate in all the above cases is much away from the diagonal line with the area under ROC curve values close to one which indicates the model efficiency. From the experimental investigation of the model, it was evident that student data was analysed based on the assigned weight on each given features to predict the placement category in which the student would be eligible based on performance. The model would also suggest further preparations required, which is very much essential to excel in placements and get absorbed in a super dream company. The output of the application shows the student data entry along with predictive results as shown in Fig. 8.



(a)



(b)

Fig. 8. Application output for given student data a) Data Entry, b) Predictive results shown along with suggestions recommended for further improvement

VI. CONCLUSION AND RECOMMENDATION

The proposed work was an attempt to predict student's placement with the help of Data Mining approach. In this work, three best-suited classification algorithms like ANN, KNN and SVM were utilized. From the experimental analysis, it was evident that the results obtained for the ANN algorithm were best compared to other two classifiers when 80% of the data was trained and 20% of the data was tested using Tanh activation function. Similarly for the KNN algorithm when 70% of the data was trained and 30% of the data was tested with 7 neighbours, best results were achieved. Whereas in the case of SVM algorithm best results were got for Radial Basis Function (RBF) kernel for 70:30 proportion. Amongst the three classification algorithms, ANN has given the best results with an accuracy of 99.02% for 80:20 proportions with Tanh activation function. Therefore the predictive model proposed in this study would help the educational institutions to give quality inputs prior to placements in order to help the students to work on their weak areas for getting placed in their dream companies which would indirectly help the institution in achieving the milestone of outcome-based education (OBE) which is the current focus area. From the promising results obtained, it can be concluded that the proposed model could be implemented with the help of other classification algorithms also. The Present study focused on four-year data and was restricted to few major departments with limited features; hence, if the data set could be further strengthened, the model could be successfully utilized for other educational institution related applications like planning of courses and result analysis which is the biggest challenge for any educational institution.

ACKNOWLEDGEMENT

The authors would like to thank the NMAM Institute of Technology, Nitte and the placement department for providing the data for this research. The authors would also like to thank all the department coordinators from ISE/CSE/Mech./ECE for their assistance and valuable advice which helped in carrying out this research work smoothly.

REFERENCES

1. Alfiani, Ardita Permata, and Febriana Ayu Wulandari. "Mapping student's performance based on data mining approach (a case study)." *Agriculture and Agricultural Science Procedia* 3 (2015): 173-177.
2. Liu, Yang, et al. "The Application of Data Mining Techniques in College Students Information System." *2018 International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018)*. Atlantis Press, 2018.
3. Gilbert, Noah. "Predicting Success: an Application of Data Mining Techniques to Student Outcomes." *International Journal of Data Mining & Knowledge Management Process (IJDKP)* 7.2 2017: 1-20.
4. Kaur, Parneet, Manpreet Singh, and Gurpreet Singh Josan. "Classification and prediction based data mining algorithms to predict slow learners in education sector." *Procedia Computer Science* 57 (2015): 500-508.
5. Sheetal, M. B, Savita, Bakare. "Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbor." *International Journal of Advanced Research in Computer and Communication Engineering* 5.6 2016: 309-312.
6. Sumitha, R., E. S. Vinothkumar, and P. Scholar. "Prediction of Students Outcome Using Data Mining Techniques." *Int. J. Sci. Eng. Appl. Sci* 2.6 (2016): 132-139.
7. Ishizue, Ryosuke, et al. "Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics." *Research and Practice in Technology Enhanced Learning* 13.1 (2018): 1-20.
8. Ahmed, S., Zade, A., Gore, S., Gaikwad, P., Kolhal, M. "Performance Based Placement Prediction System." *IJARIIE-ISSN (O)* - 4(3) 2018: 2395-4396.
9. Hazra, S., Sanyal, S. "Recruitment Prediction Using ID3 Decision Tree." *International journal of Advance Engineering and Research Development* 3.10 (2016): 48-52.
10. Tarekegn, G. B., Sreenivasarao, V. "Application of Data Mining Techniques to Predict Students Placement in to Departments." *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)* 3.2 (2016): 10-14.
11. Hamza, H., Indiradevi, S., Kizhakkethottam, J.J. "Student academic performance prediction model using decision tree and fuzzy genetic algorithm." *Procedia Technology* 25 (2016) : 326-332.
12. Rathore, Ravi Kumar, and J. Jayanthi. "Student Prediction System for Placement Training using Fuzzy Inference System." *ICTACT Journal on Soft Computing* 7.3 (2017): 1443-1446.
13. Nichat, Ankita A., and Dr Anjali B. Raut. "Predicting and Analysis of Student Performance Using Decision Tree Technique." *International Journal* 5 (2017) :7319-7328.
14. Suganthi, G., Ashok, M.V. "Unique Data Mining Approach to Predict Placement Chance." *International Journal of Engineering Sciences & Research Technology* 6(3) 2017: 263-269.
15. Manikandan, K., Sivakumar, S., Ashokvel, M. "A Classification Model for Predicting Campus Placement performance Class using Data Mining Technique" *International Journal of Advance Research in Science and Engineering* 7(6) 2018: 29-38.
16. Patel, T., Tamrakar, A. "Data Mining Techniques for Campus Placement Prediction in Higher Education." *Indian J.Sci.Res.* 14 (2) 2017: 467-471.
17. Raut, A. B., Nichat, M. A. A., Students performance prediction using decision tree. *Int. J. Comput. Intell. Res.* 13(7) 2017: 1735–1741.
18. Goyal, J., and Shilpa, S. "Placement Prediction Decision Support System using Data Mining." *International Journal of Engineering and Techniques*, 4(2) 2017: 630-635.
19. Pal, A. K, Saurabh, P. "Classification model of prediction for placement of students." *International Journal of Modern Education and Computer Science* 5(11) (2013): 49-56.

AUTHORS PROFILE

Abhishek S. Rao received his B. E degree in Information Science & Engineering from Canara Engineering College, Bantwal, M.Tech degree in Computer Science and Engineering from NMAM Institute of Technology, Nitte, Visvesvaraya Technological University, India and M.B.A degree in Operations Management from MIT, Pune. His major research interest is in the fields of Data Mining, Machine Learning, Operations Research, and Computer Graphics. He has 2 years of industrial experience and 8 years of teaching experience. He is currently working as Assistant Professor in the Department of Information Science & Engineering at NMAM Institute of Technology, Nitte. He is a member of ISTE and IAENG.

Aruna Kuamr S V received his B.E degree in Computer Science and Engineering from Kuvempu University, M.Tech degree in Software Engineering and Ph.D degree in Computer Science from Visvesvaraya Technological University, India. His area of research includes Machine Learning, Pattern Recognition, Image Processing and Soft Computing.

Pranav Jogi received his B.E degree in Information Science & Engineering from NMAM Institute of Technology, Nitte, Visvesvaraya Technological University, India. His major research interest is in the fields of Data Mining, Machine Learning, and IoT. He is a member of ISTE.

Chinthan Bhat K received his B. E degree in Information Science & Engineering from NMAM Institute of Technology, Nitte, Visvesvaraya Technological University, India. His major research interest is in the fields of Data Mining, Machine Learning, and IoT. He is a member of ISTE.

Kuladeep Kumar B received his B. E degree in Information Science & Engineering from NMAM Institute of Technology, Nitte, Visvesvaraya Technological University, India. His major research interest is in the fields of Data Mining, Machine Learning, and IoT. He is a member of ISTE.

Prashanth Gouda received his B. E degree in Information Science & Engineering from NMAM Institute of Technology, Nitte, Visvesvaraya Technological University, India. His major research interest is in the fields of Data Mining, and Machine Learning. He is a member of ISTE.