ORIGINAL RESEARCH

SAMRIDDHI Volume 12, Special Issue 2, 2020 Print ISSN: 2229-7111 Online ISSN: 2454–5767

Student Placement Prediction System using Machine Learning

Varsha K. Harihar*, D. G. Bhalke

Department of Electronics and Telecommunication Engineering, All India Shree Shivaji Memorial Society College of Engineering, Pune, Maharashtra, India

Abstract

This paper presents an overview of the machine learning techniques that can be used to predict a student's placement performance. The ability to predicting the performance of a student is a very essential task of all educational institutions. Since this is the task of predicting student's placement of undergraduate students. This paper can be used to predict the probability of an undergraduate student getting placed by applying different machine learning algorithms. In this system, multilayer perceptron (MLP), logistic model tree (LMT), sequential minimal optimization (SMO), simple logistic, and logistic classifiers are applied to predict student performance. These classifiers independently predict the results and then compare the accuracy of the algorithms, which is based on the data set. After performing analysis on different matrices (time taken to build classifier, correctly classified instances, root mean squared error, incorrectly classified instances, precision, recall, F-measure, ROC area) by different machine learning algorithms, we are able to find which algorithm is performing better than other on the student data set, so that we are able to make a guideline for future improvement of student placement performance in education.

Keywords: Classification, LMT, Machine learning, MLP, Prediction, Simple logistic, SMO.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2020); DOI: 10.18090/samriddhi.v12iS2.17

Introduction

Students studying in the final year of an engineering college start feeling the pressure of placement. They feel the need to know where they stand and how they can improve their skills to chances of getting a job. The main objective to build the system can be helpful to increase progress in student performance. So, students can analyze where they need to improve to secure a good placement in the near future. Placement holds the most important part of a student's life. This system will have a major impact on the performance of the students, resulting in good placements. This paper presents an overview of the machine learning techniques that can be used to predict a student's placement performance. The ability to predicting the performance of a student is a very essential task of all educational institutions. This paper can be used to predict the probability of final year students for the placement. This system uses the machine learning classifier—simple logistic, logistic, locally weighted learning (LWL), Naive Bayes (NB), logistic model tree (LMT), sequential minimal optimization (SMO) for support vector machine (SVM), and deep learning classifier multilayer perceptron (MLP) for this work. This system can be helpful to increasing probability of students getting placed. A model is constructed by applying the classifiers on the dataset. The ten-fold crossvalidation is accomplished. The parameters, like accuracy, precision recall, kappa statistics, ROC area, true positive rate (TPR), and false positive rate (FPR) are considered for evolution of classifier. In this paper, the prediction of student performance is done by applying algorithms and classification techniques in Waikato Environment for Knowledge Analysis (WEKA) tool. By applying algorithm and techniques on student's data, we can obtain the good student performance for the placement. This system will help to improve student's performance and probability of students placement. All these will help to improve quality of institute.

Corresponding Author: Varsha Kailas Harihar, Department of Electronics and Telecommunication Engineering, All India Shree Shivaji Memorial Society College of Engineering, Pune, Maharashtra, India, e-mail: varshaharihar12@gmail.com

How to cite his article: Harihar, V.K., & Bhalke, D.G. (2020). Student Placement Prediction System using Machine Learning. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 12(SI-2), 85-91.

Source of support: Nil

Conflict of interest: None

Literature Review

Shreyas Harinath *et al*.[1] proposed two different machine learning classification algorithms, *viz*., Naive Baye's classifier,

and K-Nearest Neighbors (KNN) algorithm. These algorithms independently predict the placement of students using machine learning. Results were then comparing the efficiency of the algorithms.[2-4] This model helps to improve their technical skills. Neelam Swaroopa *et al*.[5] proposed model is to predict the same with an algorithm. Data were collected from the same institution for which the placement prediction is done, and also suitable data preprocessing methods were applied. This model is also compared with other traditional classification algorithms, such as, decision tree and random forest with respect to the accuracy, precision, and recall. From the results obtained, it is found that the proposed algorithm performs significantly better in comparison with the other algorithms mentioned.[5-13] Tansen Patel *et al*.[14] has presented a paper on various clustering techniques and compares the result of placement prediction for higher education. K. Prasada Rao *et al*.[2] proposed system classification techniques are used on student database to predict the learning behavior. This study helps to identify the slow learner, rectify the failures early. This paper also compares the performance of J48, Naive Bayes, and random forest algorithm. Karishma B. Bhegade *et al*.[3] presented work on classification methods to propose a well-behaved carrier for the student. Undisciplined and violent student affects their carrier. Classification rules produced by the decision tree are famous due to easy interpretation. Various kinds of classifiers are tried for calculation of accuracy, as well as, performance. M. Usha Rani *et al*.[4] proposed a mode using EDM. Classification was done in order to predict students in different class categories, like high, medium and low. The classifiers used are SVM. Keshav Kumar *et al*.[6] proposed work for the development of placement predictor system (PPS) using a logistic regression model. Lakshmipriya. K *et al*.[7] presented a model to analyze the data mining methodologies to study the behavior of graduated students in terms of their career interests, by analyzing the corresponding database, and as a matter of fact, information, like mobile phone usage, internet connectivity, subject assignments, study timings, and grade. Jai Ruby *et al*.[8] presented work on the prediction of student academic performance using various classification algorithms. All the classification algorithms MLP, ID3, J48, REPTree, NBTree, SimpleCart, and Decision Table considered in the study show prediction accuracy above 68% for the student dataset. D. Ganesh Gopal *et al*.[12] has proposed model sum of difference method has been used to achieve the goal and extract the patterns from the given dataset. Praveen Rani *et al*.[13] presented a study by using simple K-mean clustering and J48 classification algorithm for student's segregation. Animesh Giri *et al*.[15] proposed a placement prediction system that predicts the probability of an undergrad student getting placed in an IT company by applying the machine learning model of K-nearest neighbor's classification.[14,15] Ranjita Kumari Dash *et al*.[16] proposed a paper which is a selection of best classifiers from the different dataset using WEKA. In a previous study, the placement prediction system will predict the probability of an undergraduate student getting placed in a different company by applying the machine learning model of K-nearest neighbor classification. They are also comparing the results of the same against the results obtained from other models like logistic regression and support vector machine.[16-21] By analyzing the above studies, the proposed technique is efficient to predict student's placement performance. Here, this system can be predicting the placement of students by using a machine learning technique and also deep learning technique, which gives better accuracy for this dataset.[21-26]

Methodology

Proposed Technique and Attribute Selection

Placement prediction in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. In this paper, the focus on machine learning techniques and also deep learning techniques to predict placement status. The student data provided as text input through excel to weka. Attribute selection on the student's database is given in Table 1. The data set used in this study contains 1,000 records are collected from the training and placement department of AISSMSCOE college. The data set can be used to predict and improve the performance of students by using different classification techniques. Figure 1 shows, the methodology of the student's placement prediction system.

Table 1: List of attributes

S. No.	*Attributes*	*Type of data*
1	Branch	Nominal
2	Gender	Nominal
3	10th%	Numeric
4	10th board	Nominal
5	12th%	Numeric
6	12th board	Nominal
7	Diploma%	Numeric
8	FE-I SEM%	Numeric
9	FE-II SEM%	Numeric
10	SE-I SEM%	Numeric
11	SE-II SEM%	Numeric
12	TE-I SEM%	Numeric
13	TE-II SEM%	Numeric
14	Agg. engineering	Numeric
15	Live backlogs	Numeric
16	Year down	Numeric
17	Education gap	Numeric
18	Company selected	Nominal
19	Placed/not placed	Nominal

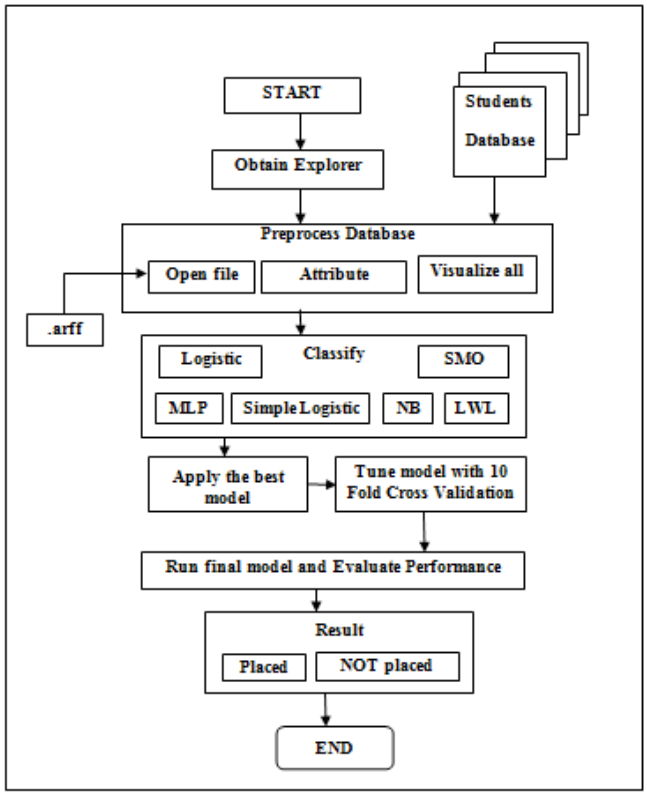

Figure 1: Methodology of the system

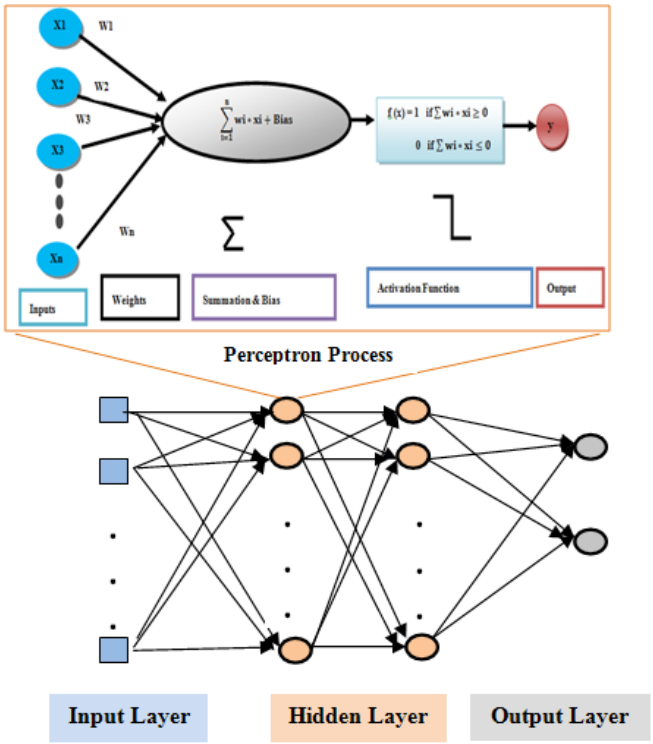

Figure 2: Multilayer perceptron

Classifier

There are various methods through which predict future or unknown events by analyzing current datasets. There are various machine learning techniques through which one can make a prediction model. This work has used the models are NB, SMO, MLP, LMT, simple logistic, logistic, and LWL algorithms.

Naive Bayes Classifier

The Naive Bayes classifier is very effective in many real data applications. This is an easy technique for building a classifier model. The Naive Bayes classification technique is based on the "Bayes Theorem" with an assumption of independence among predictors. The creative Naive Bayesian technique is based on the conditional probability and the maximum likelihood incidence. The formula for Baye's theorem is given as:

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$

Where P(A/B) is posterior probability, P(B/A) is likelihood probability, P(A) is prior probability, and P(B) is a marginal probability.

Logistic Model Tree (LMT)

Classifier for building "logistic model trees," which are classification trees with logistic regression functions at the leaves. The algorithm can deal with binary and multi-class target variables, numeric, and nominal attributes.

Simple Logistic

Classifier for building linear logistic regression models. LogitBoost with simple regression functions as base learners is used for fitting the logistic models. The optimal number of LogitBoost iterations to perform is cross-validated, which leads to automatic attribute selection.

$$s(x) = \frac{1}{1+e^{-x}}$$

Multilayer Perceptron (MLP)

The MLP is also called artificial neural networks or simply neural networks. It is a network of perceptron arranged in the feed forward network. Multilayer perceptron, there can more than one hidden layer (combinations of neurons). The multi-layer perceptron algorithms support both regression and classification problems. This network includes the non-linear activation function. The multi-layer networks are shown in Figure 2.

The training steps of multilayer perceptron is as follows:

1. Forward pass:

a. Calculate the weighted sum of the input to the weights in the hidden layer is given by,

Net input = (w1*x1 + w2*x2 + …..+ wn*xn) + b

$$= \sum_{i=1}^{n} w_n * x_n + b$$

b. A typical activation function used is the Sigmoid equation is,

$$f(x) = \frac{1}{1+e^{-Net\ input}}$$

2. Calculate error

3. $$E = \frac{1}{2}\sum(Actual\ Output - Predicted\ Output)^2$$

4. Backward pass:

Update the weights, $w_{new} = w_{old} - \eta \frac{\partial E}{\partial t}$

Where, is the learning rate. The above process is repeated until the convergence.

After performing classification, the analysis was done on different parameters, i.e., time taken to build a classifier, correctly classified instances, root mean squared error, incorrectly classified instances, precision, recall, F-measure, and ROC area. The results of all these parameters were obtained from the confusion matrix. The confusion matrix tells us about the performance of the classification algorithm. From this, the true positive rate which is termed as sensitivity, false-positive rate, or type-1 error, the true negative rate also called specificity and false-negative rate or type-2 error is evaluated. The confusion matrix is shown in Table 2.

Tool and Techniques

For the purposes of this study, WEKA software packages and MS Excel can be used. WEKA is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. WEKA is free software available under the GNU General Public License. The WEKA workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset. In WEKA, datasets should be formatted to the ARFF format. The WEKA is a widely used software for classification and regression tasks. WEKA GUI chooser depicts in Figure 3.

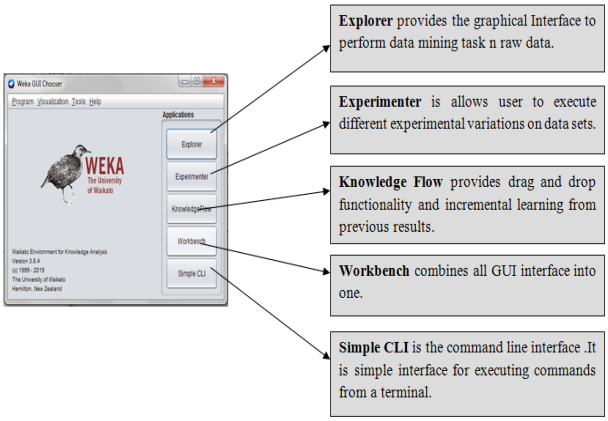

Figure 3: WEKA GUI choose

Table 2: Confusion matrix

Model		Predicted	
		No	*Yes*
Actual	No	TN	FP
	Yes	FN	TP

Implementation of the System

For this system, the student placement.arff file was created in WEKA. This file was loaded into WEKA explorer. The classify panel enables the user to apply classification and regression algorithms to the dataset, to evaluate the accuracy of the resulting predictive model, and to visualize the predictions. The algorithm used for classification is MLP, Naive Bayes, SMO, simple logistic, logistic, and LWL. Under the "Test options," the ten-fold cross-validation is selected as this evaluation approach. Since there is no separate evaluation data set required to apply. This predictive model provides a way to predict whether a new student will be placed or not placed in an organization. The system Output in WEKA shown in Figure 4.

Results and Discussion

The collected dataset contains the result of 1,000 students for the last two years. The dataset contains a number of instances are 1,000 and 20 attributes. The data file has to be in either in “CSV” format or “ARFF” format. For this experiment, data stored in MS Excel and converted into .arff (Attribute-Relation File Format). This file was given as input to WEKA 3.8.4 tool to obtain results. Data pre-processing is the first step in the evaluation of this project. For this project, choose WEKA Explorer interfaces for the classifying model. Here, choose the MLP classifier.

The evolution result of ten-fold cross-validation on student's data is given in Table 3 and the corresponding graph in Figures 5 and 6. In this work, compare the performances of the LWL, Naïve Bayes, Logistic, SMO, Multilayer Perceptron, and LMT. Different classifiers were chosen in this research work and comparative analysis of their performance was done using the WEKA tool. Educational dataset pre-processed and later fed to different algorithms. Training and testing were performed using ten-fold cross-validation technique. The obtained results from the model were measured in different terms, like accuracy, TPR, FPR, ROC area, precision, recall, and

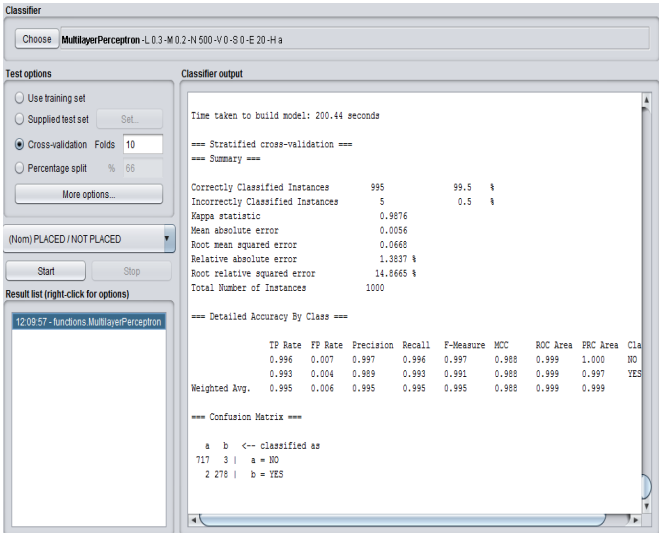

Figure 4: Output in WEKA

Table 3: Evolution of ten-fold cross validation

Model	*MLP*	*LWL*	*SMO*	*NB*	*Logistic*	*Simple logistic*	*LMT*
Accuracy (%)	99.5	99.7	96	69	94.6	96	96
Incorrectly classified instances (%)	0.5	0.3	4	31	5.4	4	4
Kappa statistic	0.987	0.99	0.896	0.273	0.859	0.896	0.896
RMSE	0.071	0.0557	0.2	0.49	0.217	0.218	0.218
TP rate	0.995	0.997	0.96	0.69	0.946	0.96	0.96
FP rate	0.006	0.006	0.101	0.4	0.128	0.101	0.101
Precision	0.995	0.997	0.962	0.709	0.948	0.962	0.962
Recall	0.995	0.997	0.96	0.690	0.946	0.96	0.96
F-measure	0.995	0.997	0.959	0.698	0.944	0.959	0.959
ROC area	0.999	0.993	0.93	0.693	0.962	0.914	0.914
Time to build (s)	116.75	0	0.94	0.1	0.95	0.87	1.48

Table 4: Confusion matrix for the system

		MLP		*LMT*		*LWL*		*Logistic*		*Simple logistic*		*SMO*	
		Predicted											
Model		*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Actual	a	717	3	719	1	719	1	715	5	719	1	719	1
	b	2	278	39	241	2	278	50	230	39	241	39	241

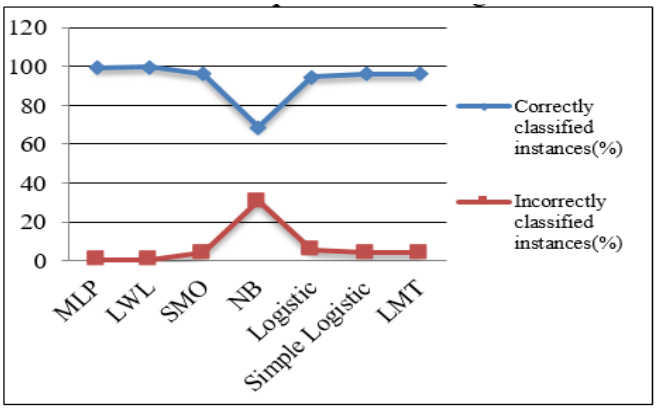

Figure 5: Comparison of classifier

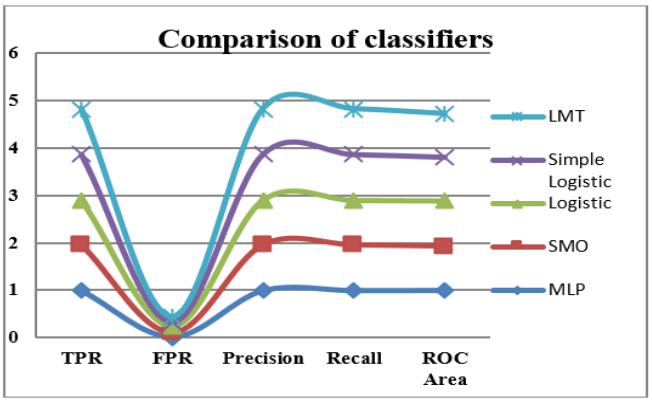

Figure 6: Comparison of classifiers

incorrectly classified instances. The performance is measure in terms of accuracy. The accuracy of the prediction model is defined as the total number of correctly predicted instances. The formula is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100$$

Where, TP, TN, FN, FP represents the number of true positive, true negative, false negative, and false-positive cases.

The confusion matrix for the classifiers is shown in Table 4.

The comparison of the proposed model with other researchers is shown in Table 5. It indicates the proposed model is better than other models. The proposed model gave an accuracy of 99.5 and 96%. It is much better to predict students' placement performance by using the proposed model. The number of attributes and instances is also greater in the proposed model. This model gives higher accuracy,

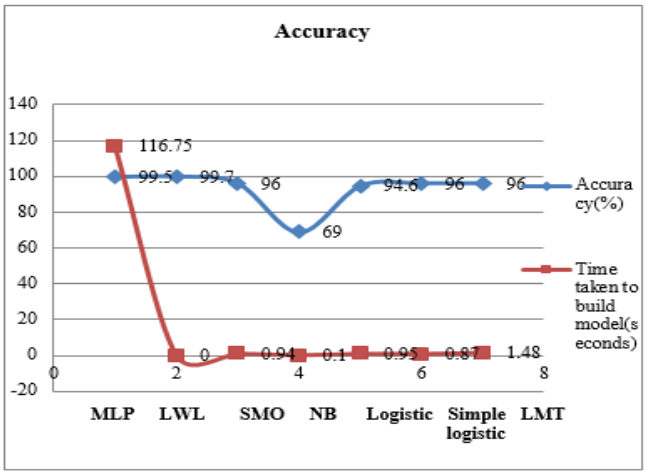

Figure 7: Accuracy of model

Table 5: Performance comparison to other researchers

Ref. No.	*Classifier*	*Total No. of attributes*	*Total No. of instances*	*Test option*	*Time to build*	*Accuracy (%)*
[5]	DT	5	1,000	-	-	86
[8]	MLP	9	165	K-fold cross validation	-	74.8
[23]	J48	12	20	10-fold cross validation	-	79.6
[9]	J48	22	1,496	-	-	88.43
[11]	SOD	11	50	-	-	-
[13]	Clustering	11	600	-	-	-
[14]	Clustering	8	150	-	0.02 s	83
[15]	SVM	10	-	-	-	77.38
[19]	ID3	10	1,342	-	-	95.33
[10]	KNN	8	900	-	13,458 ms	97.33
[22]	DT	9	289	-	0.01 s	71.66
Proposed Model	MLP	20	1,000	10-fold cross validation	116.75 s	99.5

Table 6: Accuracy of model

Classifiers	*MLP*	*LWL*	*SMO*	*NB*	*Logistic*	*Simple logistic*	*LMT*
Accuracy (%)	99.5	99.7	96	69	94.6	96	96
Time taken to build model (in seconds)	116.75	0	0.94	0.1	0.95	0.87	1.48

TPR, precision, recall, F-measure, and ROC area according to Table 3.

The algorithms that gave optimal results are MLP, SMO, simple logistic, LMT with maximum accuracies 99.5%, others all gave 96%. Time taken to build a model for MLP is larger than machine learning algorithms. Students placement prediction system results with accuracy are shown in Table 6.

Conclusion

This paper deals with the prediction of student placement performance using various classification algorithms. This study paper helps the institution to know the placement prediction of the students in advance. This was compared with deep learning technique-MLP and machine learning technique-SMO, simple logistic, in terms of accuracy, precision, true positive rate, FPR, ROC area, recall, incorrectly classified instances, and time taken to build model. The algorithms that gave optimal results are MLP, SMO, simple logistic, LMT with maximum accuracies 99.5%, others all gave up to 96%. The proposed model proved to be the best predicting model for solving placement chance prediction problems compared to all other algorithms.

References

[1] Shreyas Harinath 1, Aksha Prasad 2, Suma H S3," STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING," *International Research Journal of Engineering and Technology* **(***IRJET),* e-ISSN: 2395-0056 ,Volume: 06 Issue: 04 | Apr 2019

[2] K. Prasada Rao, "Predicting Learning Behavior of Students using Classification Techniques,"*International Journal of Computer Applications (0975 – 8887) Volume 139 – No.7,* April 2016

[3] Karishma B. Bhegade, "Student Performance Prediction System with Educational Data **min**ing,"*International Journal of Computer Applications (0975 – 8887) Volume 146 – No.5,* July 2016

[4] Usha Rani, "Student's Performance Prediction using Deep Learning and Data mining Methods," *International Journal of Recent Technology and Engineering* (*IJRTE) ISSN: 2277-3878*, Volume-8, Issue-1S4, June 2019

[5] Neelam Swaroopa," Campus Placement Prediction Using Supervised Machine Learning Techniques,"*International Journal of Applied Engineering Research ISSN* 0973-4562 Volume 14, Number 9, 2019

[6] Keshav Kumar," PPS-Placement Prediction System using Logistic Regression," *IEEE,*2014

[7] Lakshmipriya. K, Dr. Arunesh P.K ," PREDICTING STUDENT PERFORMANCE USING DATA **min**ING CLASSIFICATION TECHNIQUES," *International Journal of Innovation Research in Science and Engineering, vol. No.3,*Feb 2017

[8] Jai Ruby1, Dr. K. David2," Predicting the Performance of Students in Higher Education Using Data **min**ing Classification Algorithms-A Case Study," *International Journal for Research in Applied Science and Engineering* ,Volume 2 Issue XI, November 2014

[9] Getaneh Berie Tarekegn, "Application of Data **min**ing Techniques to Predict Students Placement in to Departments," *International Journal of Research Studies in Computer Science and Engineering* **(***IJRSCSE) Volume 3, Issue 2,* 2016

[10] Mangasuli Sheetal B1, Prof. Savita Bakare2,"Prediction of Campus Placement Using Data mining Algorithm-Fuzzy logic and K nearest neighbor," *International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6,* June 2016

[11] Ramanathan.L, Swarnalatha P,D. Ganesh Gopal, mining Educational Data for Students' Placement Prediction using Sum of Difference Method International Journal of Computer Applications , Volume 99- No.18, August 2014

[12] D. Ganesh Gopal," mining Educational Data for Students' Placement Prediction using Sum of Difference Method," *International Journal of Computer Application Volume 99– No.18,* August 2014

[13] Praveen Rani , Dr. Rajan Vohra, Generating Placement Intelligence in Higher Education Using Data mining International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015

[14] Tansen Patela and Anand Tamrakar,"A DATA minING TECHNIQUES FOR CAMPUS PLACEMENT PREDICTION IN HIGHER EDUCATION," Indian J.Sci.Res. , 2017

[15] Animesh Giri, M Vignesh V Bhagavath, Bysani Pruthvi, Naini Dubey, A Placement Prediction System Using K-Nearest Neighbors Classier2016 second International Conference on Cognitive Computing and Information Processing (CCIP),2016

[16] Ranjita Kumari Dash, "Selection of Best classifier from different dataset Using WEKA," Vol. 2, IJERT, March2013

[17] Mohana Bangale 1, Shubham Bavane ,Akshay Gunjal," Campus A Survey on Placement prediction system using machine learning,"*IJSART -Volume 5Issue2* ,FEBRUARY 2019

[18] Oktariani Nurul Pratiwi ," Predicting Student Placement Class using Data mining," *IEEE International Conference on Teaching, Assessment and Learning for Engineering* **(***TALE),* 2013

[19] Namita Puri 1, Deepali Khot 2, Pratiksha Shinde 3, Kishori Bhoite 4, Prof. Deepali Maste5, \Student Placement Prediction Using ID3 Algorithm," Volume 3 Issue III, March 2015

[20] Senthil Kumar Thangavel ,Divya Bharathi P,Abijith Sankar,, \ Student Placement Analyzer:A Recommendation System Using Machine Learning," 2017 International Conference on Advanced Computing and Communication Systems ,Jan 7 , 2017,INDIA

[21] Oktariani Nurul Pratiwi ," Predicting Student Placement Class using Data mining," *IEEE International Conference on Teaching, Assessment and Learning for Engineering* **(***TALE),* 2013

[22] Ravi Kumar Rathore1 and J. Jayanthi2,"STUDENT PREDICTION SYSTEM FOR PLACEMENT TRAINING USING FUZZY INFERENCE SYSTEM ," *ICTACT JOURNAL ON SOFT COMPUTING*, VOLUME: 07, APRIL 2017

[23] Siddhi Parekh," Results and Placement Analysis and Prediction using Data **min**ing and Dashboard," *International Journal of Computer Applications , Volume 137 – No.13,* March 2016

[24] Ashok M .V," A Placement Prediction System Using K-Nearest Neighbors Classifier," *International Conference on Computational Systems and Information Systems for Sustainable Solutions*, 2016

[25] Weka Tutorials, http://www.technologyforge.netlWekaTutorials

[26] Weka SoftwareTool,Available: htttp://www.cs.waikato.ac.nz/mL/weka/downloading.html