A Comparative Study On Students Placement Performance Using Data Mining Algorithms

N. Premalatha, Dr. S. Sujatha

Abstract: In recent days, the prediction of unemployment becomes a major and critical issue since it helps the government to take decision and policies that can improve the rate of employment. The prediction of unemployment offers various giants to learn about the upcoming trends related to economics. Forecasting of unemployment receives huge attention from many organizations, governments, research institutes and also research scholars. Many methods have been applied to predict/forecast students' placement performance. This paper discusses the comparative analysis on students' placement performance using different types of data mining algorithms and also describes the processes involved in the educational data mining.

Index Terms: Accuracy, Classification, Educational Data Mining, Machine Learning, Placement, Prediction, Unemployment.

1 INTRODUCTION

THE unemployment of students/graduates arises due to many factors such as their performance in academics, analytical skills, communication skills, CGPA, backlogs and many more. Many researchers follow the methodology of Work Integrated Learning (WIL) which helps in students' professional development and improving the skills in various aspects. Data Mining Techniques (DMT) is used for extracting useful information from a pool of available data. The complexity of the particular problem increases when the total number of data increases. Many researchers including schools, colleges and business firms rely on the data mining techniques for obtaining a better solution for the dataset available with them. In this article, the comparative analysis with respect to students' placement is chosen and the results obtained with the help of various techniques have been tabulated for better understanding. This article might help the researchers to have a clear idea about the employment statistics of students from various parts of the world.

2 STATISTICS OF UNEMPLOYMENT IN INDIA

The unemployment rate in India is quite higher this year when compared to previous years. Many factors act as a key token for this issue. The data to be presented here is very alarming and it needs to be given the utmost care in such a way that the rate of unemployment has to be decreased in the near future. According to the International Labour Organization (ILO), India will again see its unemployment rate at 3.5 per cent in 2018 and 2019, the same which was seen in 2017 and 2016.According to the latest data, there will be 18.9 million jobless people in India next year a little more than 18.6 million for 2018 [1]. Few reasons are the absence of quality institutions, students lacking in technical aptitude and skills, creating jobs to match upto the employment rate, frequent shift in their mindset and Indian students opting for international institutions. The rate of unemployment in India for the year 2017-18 & 2018-19 is graphically represented in figure 1. It is observed that the rate of unemployment for the year 2018-19 is higher when compared to the year 2017-18.

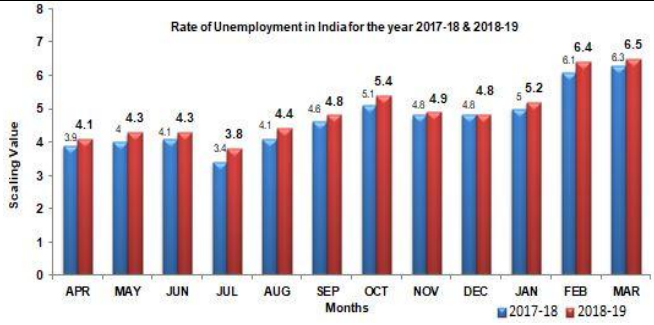

Fig.1. Rate of Unemployment in India for the year 2017-18 & 2018-19

3 EDUCATIONAL DATA MINING

EDM can give the answer for many questions from the patterns obtained from the dataset. Educational data mining can be defined as "An emerging discipline concerned with developing methods for exploring the unique types of data that come from educational settings and using those methods to better understand students, and the settings which they learn in" [2]. But the development of data mining and analytics in the field of Education was fairly late, compared to other fields. However, the challenge for EDM of online learning is due to its specific features on data. While many types of data have sequential aspects, the distribution of educational data over time has unique characteristics; for instance, a skill may be encountered many times during a school year, but separated over time and in the context of quite different activities. Research in education has resulted in several new pedagogical improvements. Computer-based technologies have transformed the way we live and learn. Today, the use of data collected through these technologies is supporting a second-round of transformation in all areas and learning with different achievements. Data mining is a powerful new technology with great potential to help Schools and Universities focus on the most important information in the data they have collected about the behavior of their

- *N.Premalatha is currently working as a Lecturer in the Department of Mathematics and Computer Applications, PSG Polytechnic College, Coimbatore, Tamilnadu, India, PH-9600233352. E-mail: prema.ctlatha@gmail.com*
- *Dr. S. Sujatha is currently working as Associate Professor & Head, Department of Computer Science, Dr. G. R. Damodaran College of Science, Coimbatore, Tamilnadu, India PH-9894434365. E-mail: sujathas05@gmail.com*

students and potential learners. Data mining involves the use of data analysis tools to discover previously unknown, patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms and machine learning methods. These techniques are able to discover information within the data that queries and reports can't effectively reveal. Figure 2 describes the various compositions of EDM.

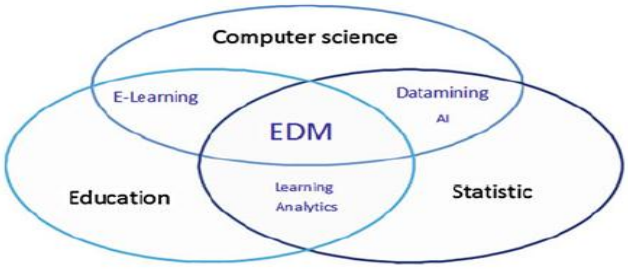

Fig.2. Various Components of EDM

4 Process involved in Educational Data Mining (EDM)

In recent years, many researchers have developed and implemented certain techniques on the available / collected data set from various sources. A general data mining process flow is given in figure 3 [2] for a better understanding.

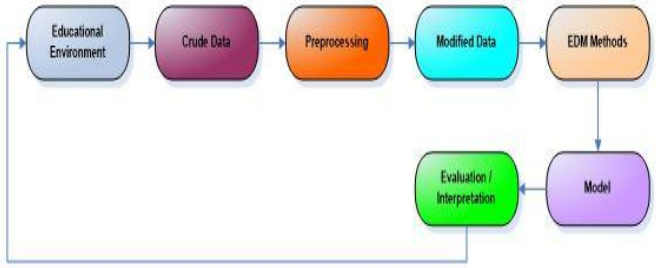

Fig.3. Steps involved in the process of EDM

From figure 3, it is clearly known that the evaluation of a dataset comprises of various steps such as data preprocessing, modification of data, applying various EDM methods, model prediction, evaluation and interpretation. This process might vary based on the type of dataset available. There are two types of datasets: balanced and unbalanced. Balanced data set are those which have an equal number of positives and negatives. The unbalanced dataset is the one which has unequal distribution of positives and negatives. Preprocessing is one of the important processes in the dataset used for removing noisy values and those values can be replaced by either '0' or deleting those particular values.

5 Important parameters to be calculated in the process of EDM:

The following are the important parameters to be calculated in the process of DM and the formulas are as follows:

(i) Sensitivity (SN)
(ii) Specificity (SP)
(iii) Accuracy (ACC)
(iv) Matthews Correlation Coefficient (MCC)

Sensitivity refers to the True Positive rate (TP) and is expressed as

$$SN = TP /(TP + FN) \quad (1)$$

Specificity (SP) refers to the True Negative rate (TN) and is expressed as

$$SP = TN /(TN + FP) \quad (2)$$

Accuracy (ACC) refers to the ratio of the total number of samples classified correctly by the classifier to the total number of samples.

$$ACC = ((TN + TP) /(TN + TP + FP + FN)) * 100 \quad (3)$$

Matthews Correlation Coefficient (MCC) refers to the coefficient between the actual classification and the predicted classification. The value ranges between -1 and 1.

$$MCC = (TP * TN)(FN * FP) / sqrt(TP + FN) * (TN + FP) * (TP + FP) * (TN + FN) \quad (4)$$

6 Comparative Analysis

This section presents a comparative analysis of various techniques related to Prediction in Placements. Analysis of many research works has been presented in this chapter for a clear understanding of the situation related to placement prediction. Research works published from the year 2011 to 2019 has been taken for a brief analysis. In the year 2011, Kabra and Bickar [3] predicted a model that predicts the academic performance of the students of engineering students. The problem addressed in this paper is related to the enrollment of inferior students in the colleges which leads to the degrading of their academic performance. In order to predict the type of students in prior who are likely to get fail, it will be helpful for the faculty members to take appropriate action against the issue and finally, it will reflect in the good placement of particular students' community. Good placement results in good reputation of the institution where they can attract a huge population of students towards their institution. The authors focused on the classification structure which is used to build a specific structure that is used to create decisions from unseen cases. For the purpose of model creation, training data is applied and for measuring accuracy test data is applied. Among the various classification methodologies, the author implemented the Decision Tree Techniques for prediction of accuracy. The real-time dataset has been

chosen from S.G.R. Education Foundations College of Engineering and Management, which consists of 346 students' data of the year 2009-10 & 2010-110. The accuracy has been calculated by considering the demographic data such as category, gender, 10^{th} marks, 12^{th} marks, address and contact numbers. Post analyzing the data, the accuracy of the model obtained is 60.46% for which the correctly classified instances are 209 and the remaining are incorrectly classified instances. With respect to the three-class prediction, the accuracy obtained is 69.94% for which 242 are the correctly classified instances. Ramesh et al. [4] presented the implementation of Decision Tree, Multi-Layer Perceptron (MLP), SMO, REPTree and Naïve Bayes technique for evaluating the students' performance of an engineering institution. A sample data of about 300 students were collected from the Department of Computer Science and their information related to placement was collected from placement division. The data conversion plays a major role in this methodology. The marks related to seven semesters with respect to the subject (Maths, Programming and Practical Marks) are converted into single Maths Programming and Practical Marks. The cleaning of data is carried out using the cleaning technique. The authors arrived at achieving 87.395% of accuracy with MLP, which is higher when compared with other algorithms reported in the analysis. An Enhanced Model for predicting the placement for the students has been proposed by Ajay Kumar Pal and Saurabh Pal [5] in the year 2013. Nowadays, the students are joining the course mainly to get a secure job. Choosing a career decision with respect to placement plays a vital role in a students' life. This study determines the relationship between the students' achievement and their campus placements. For the purpose of interpretation of potential and useful knowledge, the authors applied the Decision Tree and Naïve Bayes Classifier on the students' dataset collected from the department of MCA for better performance. The comparison results with respect to TP, FP, Precision, Recall and Class are presented for Naïve Bayes, Multi-Layer Perceptron and J48 algorithms. From the results, it is inferred that the highest accuracy obtained is 86.45% and the lowest is 75.38% and the average accuracy is predicted to be 80%. Above all the algorithms reported, Naïve Bayes Classifier performs to be the best. In the year 2014, Dammalapati Rama Krishna et al. [6] formulated the improved version of the Decision Tree Based Algorithm for the purpose of placement prediction among students' community. In this work, the author introduced the concept of attribute selection measure for the purpose of selecting the test attribute at all the nodes of a tree. The test attribute will be selected based on the training set information. At this juncture, most of the retrieved values are the frequent value of the decision attribute of Ts. A total data set of 86 students is taken for analysis. The parameter Gain with respect to attitude, technical skills, communication skills, written test, city and grade is calculated. Ajay Shiv Sharma et al. [7] in the year 2014 developed the Placement Predictor System using the logical regression model. The data required for the analysis has been collected from the placement department of Guru Nanak Dev Engineering College (GNDEC). The research and analysis have been carried out with GNU Octave Programming Tool. The type of data used for analysis is in-house placement data of the information technology department. The collected dataset consists of the information related to 10^{th}, 12^{th}, marks from semester I to semester VI, sex and their residential status. The logistic regression model is trained using the collected data. From the analysis, it is concluded that the residential and gender of the candidate plays a major role in the probability of the students getting placed. The testing and the training accuracy of the model were found to be 83.33% and 98.93% respectively. Mosima Anna Masethe and Hlaudi Daniel Masethe [8] in the year 2014 presented a comparative study among various algorithms for the prediction of work-related learning placement using DM algorithms. The mechanism to improve the professional practice along with improving the skills is referred to as Work Integrated Learning (WIL). J48, REPTREE, CART and Naïve Bayes Algorithms are selected for the purpose of comparison. The attributes such as Student number, Gender, SUB_ID, SemGrade, Citizenship and Sponsor are chosen for analysis. The confusion matrix for the aforementioned algorithms has been compared. It is concluded that the developed model fits better in predicting the students eligible for placements. In the year 2014, Jeevalatha et al. [9] implemented the various decision tree algorithms for analyzing the performance of the UG students for placement selection. The authors collected the dataset from Dr.N.G.P Arts and Science College. The three steps involved in the process is the data collection, preprocessing and implementation of various algorithms such as ID3 (Iterative Dichotomiser), CHAID and C4.5 using Rapid Miner tool. The chosen algorithms are validated and the accuracy is determined. Roll number, HSC, UG, Board, Communication, Placed data are the attributes chosen for analysis. Post analysis it is found that the ID3 algorithm performs better when compared with C4.5 and CHAID. The accuracy yielded by the ID3 algorithm is about 95.33%. Ramanathan et al. [10] in the year 2014 implemented the Sum of Difference Method for placement prediction by mining the educational data. The technique used here works on the concept of the similarity measure, which in turn used to find a pattern according to the given attribute. The Sum of Difference method is used to find the pattern for the applied students' data. With respect to the mathematical model, the analysis has been carried out in the C sharp language for predicting the students' placement statistics. The runtime environment used here is the MS Visual Studio 2012. From the analysis, it is inferred clearly that the student will only get placed only when the value goes above 4 and if is below 4, the student is not eligible for placement. In the year 2015, Pratiyush Guleria and Manu Sood [11] presented the application of Bayesian theorem on educational data set for predicting the placement results. The main purpose of using the Naïve Bayes algorithm is that dimensionality of the input to be high and small amount of data. In addition, all the features are conditionally independent for a given class label. The probability of using the Naïve Bayes classifier is framed to determine whether the student will be placed or not. The classification process is carried out using WEKA and the Rapidminer tool is used for developing the visualization techniques for generating results. Praveen Rani and Rajan Vohra [12] developed a comprehensive statistical model for identifying the total number of students eligible for placements, the number of

students to be improved for placements and students not eligible for placements. The authors performed the clustering process for forming major clusters and then by implementing the classification techniques, the efficient model has been predicted. Various parameters like academic performance, technical skills, soft skills, training and projects are considered to capture the desirability and ability of a student for placement. K-Means clustering technique has been employed for clustering process and J48 algorithm has been employed for generating the decision tree. The quantitative analysis of the students according to their academic performance has been revealed using these methods. From the results, cluster 2 has average students and represents about 200 students out of 600 students in the final year of the college, 33 percent of the students those need some improvement in the academic carrier to sit in the placement drive. The cluster 1 has above average students and represents about 230 students and 38 percent of the students those are ready to sit in the placement drive. The cluster 0 has below average students and represents about 170 students and yet only by 28 percent of the students those facing problem even in completing their degrees on time. Mangasuli Sheetal B and Prof. Savita Bakare [13] in the year 2016 suggested a method of fuzzy-logic and KNN for the purpose of measuring the accuracy and the performance levels of the students and final decisions are taken for better prediction in the campus placements. A dataset of 900 students has been collected out of which 600 data are used for the training phase and 300 data are used for the testing phase. In the preprocessing of the data, steps such as cleaning, transformation and integration take place. The data is cleaned by checking manually the attribute entries. Initially, fuzzy logic is implemented for developing the base rules for inferences and post this process, KNN is applied for classification purpose. The obtained results show that the accuracy obtained using the fuzzy logic is about 92.67% and with KNN the accuracy is about 97.33%. With the obtained results, the Knowledge Discovery model is developed. In 2016, Ravina Sangha et al. [14] presented the fuzzy approach for predicting the students' placement eligibility. In this research article, the author implemented the fuzzy approach for extracting the hidden pattern related to multiple factors. The dataset has been chosen from the past records of the enrolled students. The defining of linguistic variables and terms acts as the initial step in fuzzy approach and later the framing of membership function takes place and it is followed by constructing the knowledge base rules. Once it is constructed, the fuzzy value is obtained and it continues to the defuzzification process. As the fuzzy approach gives a variety of possibilities, a best and effective model can be developed. Getaneh Berie Tarekegn and Dr.Vuda Sreenivasarao [15] in 2017 framed an applicability model that includes the evaluation of J48, Naïve Bayes and Random Forest algorithms. The analysis has been carried out with a dataset related to a higher education institution in Ethiopia, Since the number of students getting enrolled into higher education is increasing day by day, an efficient data mining technique is required for the institutes in Ethiopia. The placement with respect to a particular department and choice of interest has been analyzed and the efficiency of the algorithm is also calculated. A total of 1496 instances of placed students data has been collected for analysis and it is received from the assistant registrar office. Upon analysis, it is found that the Random Forest with 22 attributes gives the highest accuracy of 86.564% and 86.5% F-measure with cross-validations and accuracy of 99.33% and 99.3% F-measure with 20% supplied data set. In the year 2017, Ravi Kumar Rathore and Jayanthi [16] presented the Fuzzy Interference System (FIS) that serves many benefits to the educational institutions. This method of implementing the educational dataset provides the opportunity to improve their performance in various aspects. Various classes related to placed students and non-placed students are classified. FIS helps in providing simple recognition in the truth and false values. Using FIS, the membership function can be framed according to the set of defined inputs. The experiment has been carried out with the Fuzzy Interference Tool in MATLAB. A total of 31 students data of M.Tech students have been collected for analysis. Nearly 500 rules for predicting the students' performance for placement training has been created using FIS. The results reveal that if the value is more than 2.1, then the student is eligible for placement and if it is less than 2.1, the students are not eligible. An Improved Decision Tree Algorithm [17] has been framed by Subitha Sivakumar and Rajalakshmi Selvaraj for the prediction of placement statistics in the year 2017. By this method, the relationship between the academic performance of the students and their campus placements is analyzed in various perspectives. Around 600 students from IT and computing are collected for analysis. This improved version of the algorithm involves the calculation of the Normalization Factor (NF) and Association Function (AF) Value. Through this process, a new NF and AF value are calculated and it is taken as the root node for further processing. The final result yields an accuracy of 98.5% with IDT when compared with NB and ANN algorithms. Revathy et al. [18] in the year 2017 suggested an approach that helps in suggesting the company-specific placements. C5.0 algorithm has been implemented to specify in which company the student has to be placed. Post this classifier has been applied to predict a test data set for the purpose of measuring the accuracy. CGPA, various technical skills related to quants, logical and reasoning, marks obtained in programming, marks obtained in various subjects under computer science streams and recruited company are the parameters chosen for analysis. Using C5.0 it is observed that the speed, memory usage and accuracy have been increased when compared with all the other techniques. R tool has been chosen for obtaining the accuracy. In the year 2017, Vadivu and Sornalakshmi [19] presented the application of KNN and Naïve Bayes algorithm for students' employability prediction. The authors use the R tool for analysis. The opportunity for employability prediction is defined as the values Yes/ No. The prediction is measured by their regular performance at the course level. The KNN is used for clustering the similar group of students and then it is passed onto the Naïve Bayes algorithm for predicting whether the student will be employed or unemployed. In 2019, Sonali Rawat [20] gave a comparative study on the predictive analysis for placement of the student. The author had chosen the classification approach as their major course of study and analyzed 20 articles that gave a clear idea about the various techniques used for analyzing the placement

opportunities for the students. This study helped many of the researchers to pursue their research in their specific areas of interest. The author concluded by mentioning that ID3 with accuracy 95.33%, KNN with 97.33%, C4.5 with 88.89%, Naïve Bayes with 86.15 %, Multilayer perception having 87.395% accuracy are the best algorithms for predicting the accuracy. Table 1 presents the observations of various algorithms reported in the literature and the graphical form is represented in figure 4. The value 'NM' in the table 1 represents that the value of accuracy is not discussed in that particular research article.

7 Observations

From table 1, it is clearly observed that many of the researchers carried out their research work by using various data mining techniques and exhibited their performance with respect to its efficiency. It is also inferred that 30% of the authors prefer the Decision Tree and ANN, 20 % of the authors prefer Bayesian Network and C4.5, 10% of the authors prefer MLP and J48 and the remaining 40% of the authors prefer sequential, ID3, CHAID, fuzzy, KNN and sum of difference method for the purpose of prediction. The authors gave their highest priority to Decision Tree Algorithms, KNN, Fuzzy and Naïve Bayes algorithms due to its unique features. Out of the reported literature, the Linear Regression Model gains the highest accuracy of 98.93%, Naïve Bayes Classifier in combination with KNN scores about 97.67% efficiency and ID3 gains an accuracy of 95.33%. On analyzing the Kappa statistic value, it is found that the Decision Tree Algorithm provides 0.6513 and Naïve Bayes Classifier gives 0.5076. The Kappa Statistic value closer to 1 indicates that the model developed is effective for prediction. If the value is less than 0.5 (or) if it 0.5, it indicates that the model has to be evaluated on a larger scale.

8 Conclusion and Future Scope

In this paper, a comparative analysis based on the different types of algorithms used for the purpose of prediction of placements is discussed. When higher education institutions are considered, placement is the most important factor for the attraction of students. As the number of colleges is increasing day by day, people belonging to the entire category are getting admission easily. But their mindset is still running behind the school days. To improve certain factors, a technique like DM is required for extracting some useful information from the dataset available. From the year 2011 till 2019, few remarkable papers have been analyzed and the same is presented here. If the placement of the students' community leads in a successful way, there will be no unemployment of graduates and the same will improve the economy of our country, failing which leads to the concept of unemployment. Based on the analysis carried out in this paper, a deeper analysis will be presented in our future research work with respect to the eradication of unemployment. The solutions for improving the factors will also be discussed improved versions of the algorithms that will help the educational institutions and government authorities to implement advanced techniques and methodologies.

References

[1] https://economictimes.indiatimes.com/jobs/indias-unemployment-rate-hit-6-1-in-2017-2018/articleshow/69595640.cms?from=mdr.

[2] https://link.springer.com/book/10.1007/978-3-319-02738-8.

[3] R.R.Kabra and R.S.Bichkar, "Performance Prediction of Engineering Students using Decision Trees", International Journal of Computer Applications, vol. 36, no.11, pp. 8-12, Dec 2011.

[4] V.Ramesh, P.Parkavi and P.Yasodha, "Performance Analysis of Data Mining Techniques for Placement Chance Prediction", International Journal of Scientific & Engineering Research, vol. 2, no. 8, pp. 1-7, Aug 2011.

[5] Ajay Kumar Pal and Saurabh Pal, "Classification Model of Prediction for Placement of Students", International Journal of Modern Education and Computer Science, vol. 11, pp. 49-56, 2013.

[6] Dammalapati Rama Krishna, Bode Prasad and Teki Satyanarayana Murthy, "Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm", International Journal of Computer Trends and Technology (IJCTT), vol. 17, no. 4, pp. 190-194, Nov 2014.

[7] Ajay Shiv Sharma, Swaraj Prince and ShubhamKapoor, "PPS - Placement Prediction System using Logistic Regression", Proc. IEEE International Conference on MOOC, Innovation and Technology in Education, pp. 337-341, Dec 2014,

[8] M. A. Masethe and H. D. Masethe, "Prediction of Work Integrated Learning Placement Using Data Mining Algorithms", Proc. of the World Congress on Engineering and Computer Science, vol-I, Oct 2014.

[9] T. Jeevalatha, N. Ananthi and D. Saravana Kumar, "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms", International Journal of Computer Applications, vol. 108, no. 15, Dec 2014.

[10] L. Ramanathan, P. Swarnalatha and D. Ganesh Gopal, "Mining Educational Data for Students' Placement Prediction using Sum of Difference Method", International Journal of Computer Applications, vol. 99, no.18, Aug 2014.

[11] Pratiyush Guleria and Manu Sood, "Predicting Student Placements Using Bayesian Classification", Proc. IEEE Third International Conference of Image Information Processing, pp. 109-112, 2015.

[12] Praveen Rani and Dr. Rajan Vohra, "Generating Placement Intelligence in Higher Education Using Data Mining", International Journal of Computer Science and Information Technologies, vol. 6, no. 3, pp. 2298-2302, 2015.

[13] B. Mangasuli Sheetal and Prof. 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, vol. 5, no. 6, pp. 309-312, June 2016.

[14] Ravina Sangha, Akshay Satras, Lisha Swamy and Gopal Deshmukh, "Student's Placement Eligibility

Prediction using Fuzzy Approach", International Journal of Engineering and Techniques, vol. 2, no. 6, pp. 148-152, Dec 2016.
[15] Getaneh Berie Tarekegn and Dr.Vuda Sreenivasarao, "Application of Data Mining Techniques to Predict Students Placement into Departments", International Journal of Research Studies in Computer Science and Engineering, vol. 3, no. 2, pp. 10-14, 2016.
[16] Ravi Kumar Rathore and J. Jayanthi, "Student Prediction System for Placement Training using Fuzzy Inference System", ICTACT Journal on Soft Computing, vol. 7, no. 3, pp. 1143-1446, Apr 2017.
[17] Subitha Sivakumar and Rajalakshmi Selvaraj, "Adaptive Model for Campus Placement Prediction using Improved Decision Tree", Journal of Engineering and Applied Sciences, vol. 12, no. 22, pp. 6069-6075, 2017.
[18] S. Revathy, G. Roopika, R. Rishitha and P. Revathy, "An Approach to Suggest Company Specific Placement Opportunities Using Data Mining Techniques", International Journal of Computer Science and Mobile Computing, vol. 6, no. 3, pp.196 – 203, March 2017.
[19] G.Vadivu and K.Sornalakshmi, "Applying Machine Learning Algorithms for Student Employability Prediction Using R", International Journal of Pharmaceutical Sciences Review and Research, vol. 43, no. 1, pp. 38-41, April 2017.
[20] Sonali Rawat, "Predictive Analytics for Placement of Student- A Comparative Study", International Research Journal of Engineering and Technology, vol. 06, no. 06, pp. 3423-3429, June 2019.

Table 1 Comparison of accuracy with respect to various algorithms ranging from the year 2011-2019

S. No.	Algorithm	Accuracy	Mean Absolute Error	True Positive	False Positive	Precision	Recall	Kappa Statistic
01	Decision Tree Algorithm [3]	69.64	-	-	-	-	-	0.6513
02	Multi-Layer Perceptron [4]	87.395	0.2002	0.788	0.188	0.813	0.788	-
03	Naïve Bayes Classifier [5]	86.45	0.1852	0.906	0.182	0.9	0.818	0.5076
04	Linear Regression Model [7]	98.93	-	-	-	-	-	-
05	J48 [8]	-	0.849	0.758	0.242	0.75	0.75	-
06	ID3 [9]	95.33	-	-	-	-	-	-
07	Sum of Difference Method [10]	-	0.843	-	-	-	-	-
08	Fuzzy Logic & KNN [13]	97.33	-	163	8	-	-	-
09	Random Forest [15]	86.564	-	-	-	-	0.86	-
10	Improved Decision Tree Algorithm [17]	98.5	0.2338	0.818	0.094	0.900	0.818	0.7234
11	C5.0 [18]	75	-	-	-	-	-	-
12	KNN + Naïve Bayes Classifier [19]	97.67	-	-	-	-	-	-

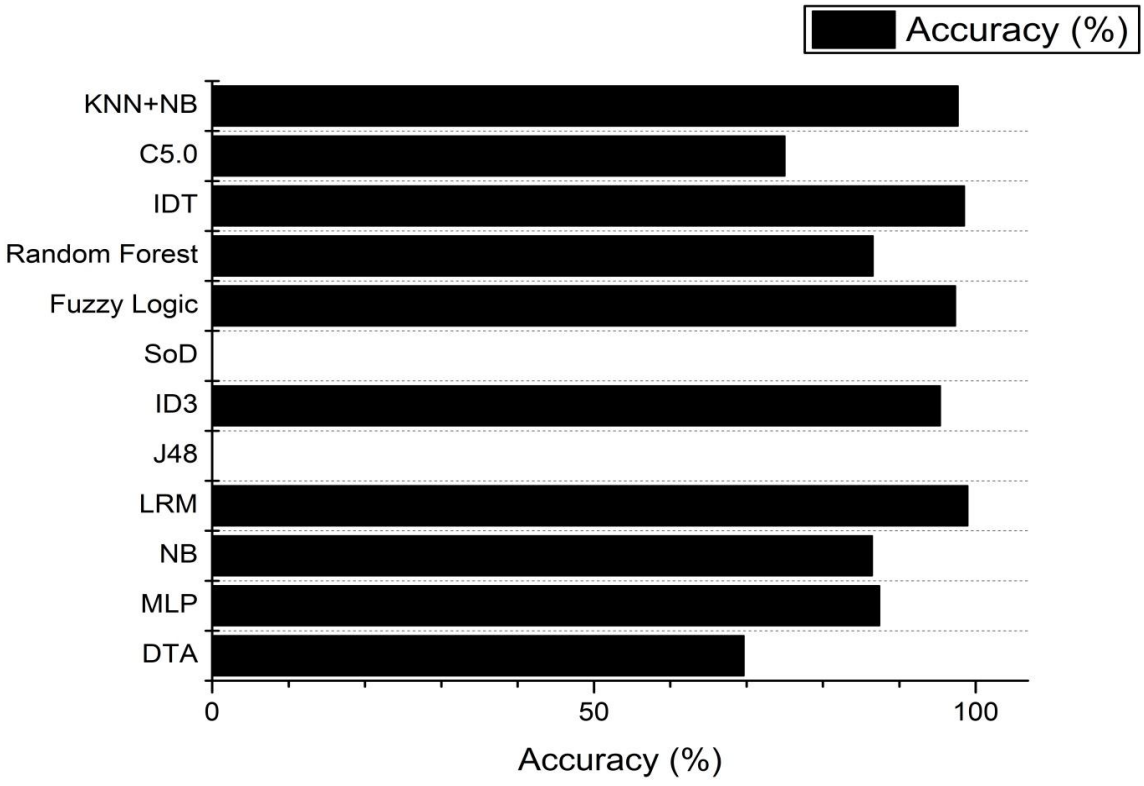

Fig.4. Comparison of Accuracy