International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

To Study the Use of Machine Learning Algorithms for Employability Predictions of Undergraduate Students

Seema S. Patil[1], Dr. P. P. Patil[2]

[1]Assistant Professor, Maharshi Karve Mahila Mahavidyalay Satara, Maharashtra, India
[2]Associate Professor, Bharati Vidyapeeth (Deemed to be University) Pune Y. M. Institute of Management Karad, Maharashtra, India

Abstract
Predictions is a speculation about something which may happen in future. Predictions is not based on previous data, experience or knowledge but it is important to make right decisions for future. In order to get advantage of predictions and to automate the prediction process, machines are trained to make predictions and such field comes under machine learning. Various fields such as crime prediction, natural calamities, health care and weather forecasting are some of the applications of prediction. In Higher education institutions, Training and Placement officers who work as a human expertise to identify a skilled undergraduate for employment based on various factors. The researchers have applied various regression and classification machine Learning algorithm to calculate Employability score and make predictions about the employability of a BCA undergraduate Students of Mahila Mahavidyalaya, Satara. The objective of paper is to calculate employability score (numeric value) and predict employability (categorical value). This paper presents the use of various machine Learning algorithm and compares their performance to predict the employability of BCA undergraduates. A dataset of undergraduates was tested and results were discussed.

Keywords: Employability Prediction, Linear Regression, Multiple linear regression, Logistic Regression, Decision tree

Introduction:

After graduation, to get a job in a reputed company is a dream of every undergraduate and every institute tries to place maximum candidates through institute placement cell. but it requires continuous efforts on to maintain good academic performance as well as various soft skills. If institutes and undergraduates go through the process of Training-Assessment-Evaluation iteratively, it will provide benefits to both. In this paper data of 100 undergraduate BCA students were tested using various machine Learning algorithm. To calculate Employability score, regression algorithms are used and for employability predictions, classification algorithms are used. The dataset consists of scores of various parameters like Academic Performance (AP), Aptitude Skills (AS), Technical Skills (TS), Communication Skills (CS), Project(P), Internships (I) and certifications(C) etc. The data within dataset send as input to regression algorithm then algorithm calculates employability score or predict employability using classification algorithms . If Employability status is ‘Employable’ means student get ready to get job and If Employability status is

'Not Employable' means students' needs to make improvement in performance or need to work on weak skills. Due to Categorized output, institutes TPO able to know in advance among 100 BCA undergraduate students how many students will be employable.

In today's competitive job market, many students complete their academic education but still face difficulty obtaining employment due to a lack of required skills. Educational institutions often find it challenging to evaluate whether students possess the necessary abilities demanded by industry.

Traditional evaluation methods mainly focus on academic performance, which does not fully represent a student's technical skills, aptitude, communication ability, internship experience, and project work. As a result, it becomes difficult to accurately assess a student's employability.

Therefore, there is a need for an intelligent system that can analyze multiple student performance parameters and predict their employability level.

The Employability Prediction System using Machine Learning aims to analyze various factors such as academic performance, aptitude skills, technical skills, communication skills, internship performance, project work, and certifications to predict a student's employability score or employability category. This system helps students identify skill gaps and improve their readiness for employment.

Literature Review:

G.GAUTAMI et.al(2020) reveals that Linear regression is one of supervised machine learning tool which helps to predict future trend of student placement based on collected dataset of final year students .Results of this prediction will help students to better understand their weak areas and working on these areas will help students to achieve higher number of placements in an institution. The study suggest that educational institutions plays a key role in placements to prepare and guide the students through the process of placements by providing them needed training. With the proper placement training, students learn how to put forth their knowledge and abilities in the right way to fetch the best of jobs.students graph based on placement practice tests will be generated on the final year student dataset which help students to realize fields like aptitude, reasoning, verbal, technical, communication skills they should work upon more.

K. Sripath Roy et al.(2018) proposes the advanced machine learning algorithm for a student's career prediction. They used three algorithms, SVM, XG Boost, Decision Tree, on a dataset. Among all three algorithms, they found that the SVM algorithm provided more correctness with 90.3%, and then XG Boost provided correctness with 88.33%. They required many parameters for student's career prediction like as student's academic scores in various subjects, specializations, programming, and analytical capabilities, memory, personal details like relationship, interests, sports, competitions, hackathons, workshops, certifications. Finally, a web application is developed to show the result of the prediction.

Pothuganti Manvitha and Neelam Swaroopa (2019) proposes a supervised machine learning technique for the prediction of campus placement. The decision tree and random forest algorithms are applied on previous year data sets of the students, and the parameters used to construct the model and parameters used for this research is an academic history of the student like overall percentage, backlogs, credits. The correctness gained after examination for the Decision tree is 84%, and for the Random Forest is 86%. The productivity of the two algorithms are likened in terms of correctness. but the Random Forest algorithm is superior to predict the placement results.

Shreyas Harinath et al.'s (2019) research proposes machine learning techniques to predict the placement status of the student. In this research, two different machine learning classification algorithms are used, namely the Naive Bayes Classifier and K Nearest Neighbors [KNN] algorithm for the prediction of a

student's placement position. Depending on provided dataset, these algorithms predict the outcomes uniquely and then differentiate the effectiveness of the algorithms, which is dependent on the dataset. In the forthcoming, this research will focus on adding some more attributes for better prediction of placement.

Vinutha K (2020) proposes different Machine Learning Algorithms which predict employability of engineering graduates depending on student's academic performance and employability skills. This research used different machine learning algorithms like Logistic Regression, Decision tree, k-nearest neighbor, Support Vector Machine and Naïve Bayes, ANN(Artificial Neural Network) to build a model for prediction of employability of engineering graduate students. The dataset used to create a model for this research is USN(Unique Student Number). Name, First to eight semester marks, Online courses completed, Internship, Technical papers presented. This research found that the highest correctness of 87.42% with Artificial Neural, 85.2% correctness with Logistic Regression classifier, and 84.21% correctness with Naïve Bayes classifier.

C. Jayasree & K. K. Baseer (2018) proposes some data mining and Machine Learning Techniques for Prediction Student Performance Improve their Employability and reducing dropout rates. This research work shows that the academic performances of the students are dependent on their past performances. To predict the performance of the student's author considered various datasets like previous grades, research work time, parent's status, GPA, school support, higher education, internet usage, travel time, etc. This research applied a few algorithms (linear regression, K-means clustering, and neural networks using Weka and Azure tools to student datasets. Machine learning studio using azure has been the best tool for real-time applications and can prove to be a powerful tool in academia. The performance of k-means and neural networks is very effective.

S. Celine et al. (2015) proposes a model to predict the employability of candidates using Logistic Regression . The study aims to automatize the employment procedure to predict the possibility of employability. This study apply Logistic Regression machine learning technique to predict employability. This research used four parameters as Aptitude skills (β1), Communication skills(β2), Technical skills(β3), and Personality skills(β4) for the prediction of employability. In the future, this study can be simply converted into another area of prediction like health care, weather forecasting, natural calamities, crime prediction, etc., where the outcomes will be a contrast in nature by varying the self-regulating variables.

V.Rameshet al. (2015) paper proposes the use of different data mining techniques to examine the performance of Computer Course students. This research uses different data mining techniques and presents the uses of classification techniques as a predictive tool. Different classification techniques are applied to student data, and finally, this research found that the Multilayer Perception algorithm is most suitable for predicting the student's performance. Multilayer Perception algorithm provides 87% prediction, which is larger than other algorithms. The student's data parameters used for this research are English, Maths, Programming language, Practical marks. This research tried to use a classification technique to predict the student's performance and also measure the performance of Naïve Bayes Simple, Multilayer Perception, SMO, J48, and REP Tree data mining techniques.

Samrat Singh and Dr. Vikesh Kumar (2013) uses various data mining techniques to evaluate academic performance of Engineering students for Recruitment purposes. In this research, six classification techniques are used on student data that is Bayes Net, Naïve Bayes, Multilayer Perceptron, IB1, Decision Table, and PART Classification method. As per the pilot outcome, they observed that IB1 Classifier is the most competent technique for such type of student's data. In this research, six parameters are taken into

consideration like as student's Name, Branch, passing percentage (%) of 10th class, passing percentage (%) of 12thclass and Graduation passing percentage (%) and Final Grade for analysis. For future studies, this research will be beneficial for educational organizations, colleges, universities, and industrial organizations or corporate sectors. In the future, they can use other data mining techniques like clustering, Prediction and Association rules, etc., on student data.

T.Malathi et al. (2015) proposes the classification data mining techniques to analyze the student's recruitment process like Selected, Waiting, and Not Selected. A classification algorithm is used to classify the students based on their talent. This research will support the Lecturers to select the students and organize them for the recruitment process. This research is very helpful to improve the student's performance and cut down the failure rate. In order to improve the studying capability of students, this research will provide useful information to lecturers.

Siddu P. Algur et al. (2016) proposes classification models which uses two algorithms Random Tree and J48. These two models are used to predict the recruitment status of a student on their academic performance. The academic performance is measured by different attributes like examination score, communication skill, and placement preparation hours breaks taken during the course, extracurricular activities, cultural activities, and the number of industrial visits for recruitment prediction. Between these two classification models, the Random Tree classification model is discovered good as compared to the J48 classification model. The accuracy of the Random Tree classification model provides correctness 85%, and the correctness of the J48 classification model is detected 74%. In forthcoming to progress the prediction/classification correctness by applying some other data mining techniques such as KNearest Neighbor classification technique, Naive Bayesian classification techniques, etc.

Keno C. Piad et al. (2016) compared and implements classification algorithms on IT employability dataset to predict the employability of IT graduate students. In this research work, the researchers compare five algorithms under classification technique on nine variables of employability dataset. The result shows logistic regression with an accuracy of 78.4 is implemented. Based on logistic regression analysis, three academic variables directly affect; IT_Core, IT_Professional, and Gender identified as significant predictors for employability. The research work can utilize data to generate more rules and predict more accurately IT employability. Thus, other classification algorithms maybe explore and generate new data models for better prediction.

Tripti Mishra et al. (2016) proposes different classifier and develop employability model formed on the appropriate classifier to predict the employability of MCA (Master of Computer Applications) students. For the prediction of employability, this research used various classification techniques like Bayesian methods, Multilayer Perceptron and Sequential Minimal Optimization (SMO), Ensemble Methods, and Decision Trees and also observed the most suitable algorithm related to the question of this research. The parameters used for this research are academic performance, socioeconomic conditions, job skills, and emotional skills, etc. Measuring the performance of all the algorithms they obtain J48 (a cropped C4. 5 decision tree) algorithm of WEKA is most applicable among all the algorithms to predict the employability of (MCA) students and identify the algorithm which is best suited for this problem. In the forthcoming, this research work will consider the students of B. Sc. and B. E. also.

Nor Azziaty Abdul Rahman et al.(2017) proposes six classification models, namely K-Nearest Neighbor, Naive Bayes, Decision Tree, Neural Network, Logistic Regression, and Support Vector Machine to acquire a suitable model to predict employment status among fresh graduates. Supervised and unsupervised Machine Learning Algorithms were used in this research that as K-Nearest Neighbour,

Naïve Bayes, Decision Tree, Neural Network, Logistic Regression, and Support Vector Machine. The parameter they used for this research is Gender, Program Academic. In forthcoming, there is a necessity for more parameters like grade subjects taken during the study period, the results of the oral test, and work status.

Madhavi Girase et al. (2018) illustrates performance of Decision Trees as classification techniques to predict the employability of Undergraduate Engineering students. For this purpose, a data set is established with traditional parameters like socioeconomic conditions, academic performance, and some additional emotional skill parameters. This paper works on student data sets like a student's academic performance, additional activities, personality development, technical training.

Azeez Ayofe et al.(2009) examine the skills gaps between the industrial application of Information Technology and university academic programmes (curriculum). It looks at some of the causes, and considers the probable solutions for bridging the gap between them. Author concludes that Education with relevant syllabuses and training in specific areas play crucial roles in achieving rapid changes in updating technical and engineering skills.

V.Manjula et al.(2017) proposed student's employability prediction using Decision Tree algorithm for B.E. students only.he used Waikato Environment for Knowledge Analysis (WEKA), an open-source data mining tool for analysis or prediction.various attributes are taken into consideration to predict student's employability for various categories like BPO only, IT only, BPO & IT and Not Placed.Using decision tree algorithm,he demonstrates relationship between student's marks and employability.

Ankita Kadambande et al.(2017) proposes Predicting academic performance of students to encourage and assist them for quality education and improve educational growth .various data mining techniques are used to analyse student performance with various approaches such as prediction,analysis,visualization etc. In this paper,support vector machine(SVM),a supervised learning technique is used for prediction of student's performance predicted result helps staff to guide students againest academic improvement and also help to TPO to know eligibility and applicability of student for campus interview.

Cherry D. Casuat et al.(2020) compares accuracy of different learning techniques and proposed a system which will use SVM machine learning technique which gives highest accuracy.

Vandana Mulye et al.(2021) proposes comprehensive review of the literature on different recruitment prediction techniques. Academic and recruitment relevant factors are an essential cradle for predicting the upcoming recruitment possibilities. These predictions can help aware students to observe their abilities, and they can recover them before campus recruitment. The recruitment prediction scheme is also very supportive for institutions to increase the number of recruitment of their students as well as upgrade their academic plan and design policies in upcoming years.

S.Celine,M.Maria Dominic,M.Savitha Devi(2020),"Logistic Regression for Employability Prediction", proposed model which will predict the employability using Logistic regression. He used data of 10 undergraduate students of computer science against 4 factors like Aptitude skills, Technical skills, communication skills, personality skills, assign weight to each factor, implement logistic regression method of supervised learning & make prediction of each candidate as employable or not employable as a dictonomus output or binary output.

Machine Learning Algorithms:

1. Random Forest

Random Forest is an ensemble machine learning algorithm that builds multiple Decision Trees and combines their predictions. To predict Employability of undergraduates, various parameters like Academic performance (AP), Aptitude Skills (AS), Technical Skills (TS), Communication Skills (CS), Project(P), Internships(I) and certifications(C) etc. have complex and non-linear relationships with employability. Random Forest works well because it combines the predictions of many decision trees, reducing error and improving accuracy. Instead of relying on one decision tree, Random Forest Creates many decision trees. Each tree is trained on random samples of the dataset. Each tree gives a prediction. The final prediction is the average (regression) or majority vote (classification).This reduces overfitting and improves prediction accuracy.

2. Decision Tree

A Decision Tree is a supervised learning algorithm. A Decision Tree is used for classification and regression. In an Employability Prediction System, it analyzes student attributes (skills, academic performance, etc.) and predicts employability score or category. The algorithm works like a tree structure where

- Root Node → Main decision based on an important feature
- Decision Nodes → Conditions on features
- Leaf Nodes → Final prediction (Employable / Not Employable or score)

3. Linear Regression

Linear Regression is a machine learning algorithm which is easy and simple to implement. It used to predict value based on single independent variables. It handles linear relationship between input parameters(skill parameters) and employability easily.

Equation for Linear Regression is

$y=\beta0+\beta1X1$

Where y is the dependent variables

X1 is the independent variables

B0 is the intercept

B0,β1are the slopes

4. Multiple Linear Regression-

Multiple Linear Regression (MLR) is a machine learning algorithm used to predict a continuous value based on multiple independent variables. Multiple Linear Regression assumes linear relationship between parameters and employability. It produces a mathematical equation. This algorithm is Easy to interpret.

Equation for multiple linear regression is:

$y=\beta0+\beta1X1+\beta2X2\cdots+\beta nXn$

Where y is the dependent variables

X1,X2,.....Xn are the independent variables

B0 is the intercept

B0,β1,β2…..βn are the slopes

5. Logistic Regression-

Logistic Regression is a classification algorithm that predicts the probability that a student belongs to a particular class (for example, employable or not employable).It uses the sigmoid function to convert calculated values into probabilities between 0 and 1. It collects student skills parameters (academic,

technical skills, communication, etc.) and calculates a weighted sum of the features. Then apply the sigmoid function to obtain probability.
If probability ≥ 0.5 → Employable
If probability < 0.5 → Not Employable

Research Methodology:

In order to predict employability of BCA undergraduate students of Mahila Mahavidyalaya, satara, final year BCA students appear for various soft skills test like AptitudeSkills(AS), TechnicalSkills(TS), CommunicationSkills(CS), Project(P), Internships(I) and certifications(C) etc. By collecting scores of Academic performance and score of various soft skills assessment tests, sample Dataset was prepared as shown in table 1. then according to priority assigned to each parameter, assign weighting factor as shown in table 2. The data in the dataset is analyzed by using the Logistic Regression algorithm and Decision Tree algorithm ,predict employability of each undergraduate BCA student in categorical output format i.e.Employable or Not Employable as shown in Table 3.

Data Analysis And Interpretation:

Table 1: A Sample Dataset consist of score of Academic and soft skills

Student Name	AS	TS	CS	I	P	C	AP
S1	64	34	70	52	72	40	64.4
S2	82	72	75	54	85	80	81.2
S3	44	30	40	52	70	40	63.2
S4	80	73	45	52	70	30	65
S5	82	70	60	52	70	20	70.2
S6	84	67	53	60	73	30	71.2
S7	74	72	45	60	65	40	60.4
S8	90	72	56	55	72	40	71
S9	86	95	58	65	75	40	69.6
S10	52	58	60	60	63	65	71.50

Table 2 : Percentagewise weighting factor of parameter

Sr.No. (i)	Parameter (Pi)	Weighting Factor (Wi)
1	C	5%
2	I	10%
3	P	5%
4	AP	25%
5	AS	15%
6	TS	25%
7	CS	15%

Table 3: Analysis of Data using the Logistic Regression algorithm and predict employability

In study, as we have multiple input independent parameters for calculating employability score, we prefer to use Multiple Linear Regression instead of linear regression. Linear regression handles linear relationship between data within dataset and Multiple Linear Regression handles non-linear relationship. To calculate employability score, multiple independent input parameters like AptitudeSkills(AS), TechnicalSkills(TS), CommunicationSkills(CS), Project(P), Internships(I) and certifications(C) etc. are passed as input to Multiple Linear regression, after implementation we get employability score but calculation of employability score and predicting employability both task are performed by Logistic regression and Decision tree algorithm which help to achieve our objective.so researcher use and implement both algorithms i.e. Logistic regression and Decision tree and compare their performance accuracy as shown in figure 1 and 2.

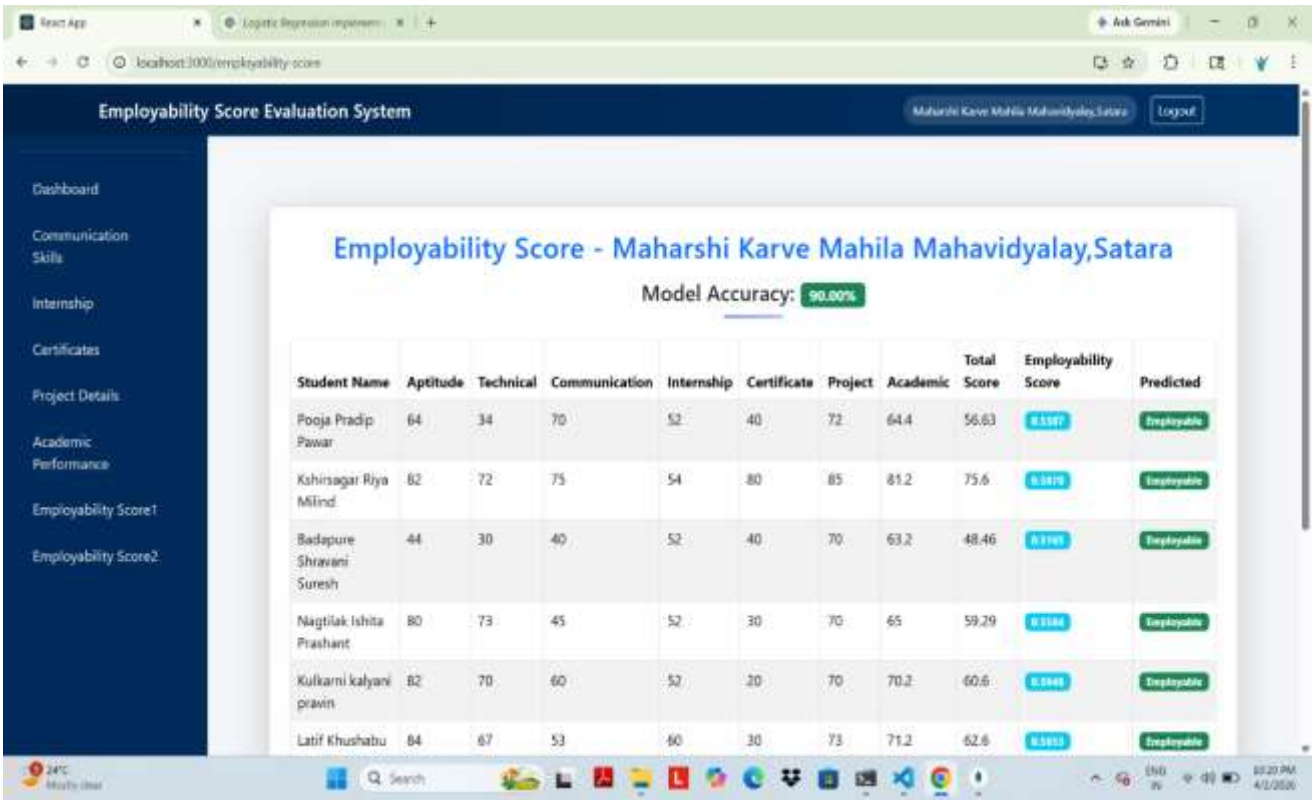

Figure 1:Implementation of ML Model-Logistic Regression Algorithm

Figure 2: Implementation of ML Model- Decision Tree Algorithm

After implementing Logistic Regression and Decision Tree algorithm on data shown in Table 1, results displayed in Figure 1 and Figure 2. As shown in figure 1, implementation of Logistic Regression algorithm gives 90% accuracy and as shown in figure 2, implementation of Decision Tree algorithm gives 100% accuracy.so it is recommended that for precise employability prediction and more accurate results Decision tree algorithm will be preferred.

Conclusion

Many of the previous research papers concentrate on a smaller number of parameters such as CGPA and few soft skills for Employability prediction which leads to less accurate results, but proposed model contains many parameters necessary to predict Employability status which will be more accurate. From above results, researcher conclude that to handle nonlinear relationship among input parameters, decision tree algorithms give more accurate results as compare to Logistic regression. From above results, Institute TPO able to recognize easily that from entire class how many students are Employable and how many students needs some little bit improvement in their performance.

References

1. C. Jayasree, K. K. Baseer., Predicting Student Performance To Improve Their Employability By Applying Data Mining And Machine Learning Techniques. International Journal Of Computer Sciences And Engineering Open Access. 6(2018) 1291-1308.
2. G.Gautami Et.Al(2020),” Prediction Of Student Placement Using Machine Learning Algorithm”, International Journal Of Creative Research Thoughts (Ijcrt), Ijcrt2005188, Volume 8, Issue 5 May 2020 | Issn: 2320-2882,1379-1384

3. K. Sripath Roy, K.Roopkanth, V.Uday Teja, V.Bhavana, J.Priyanka (2018): Student Career Prediction Using Advanced Machine Learning Techniques. International Journal Of Engineering And Technology. 7(2.20)(2018) 26-29.
4. Keno C. Paid, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat., Predicting It Employability Using Data Mining Techniques. The Third International Conference On Digital Information Processing, Data Mining, And Wireless Communications. (2016) 26-30.
5. Madhavi Girase, Suchita Lad, Prerna Pachpande, Student's Employability Prediction Using Data Mining. International Journal Of Scientific & Engineering Research 9(2018) 27-29.
6. Nor Azziaty Abdul Rahman, Kian Lam Tan1, Chen Kim Lim., Predictive Analysis And Data Mining Among The Employment Of Fresh Graduate Students In Hei. The 2nd International Conference On Applied Science And Technology (Icast'17), (2017) 020007-1 020007-6
7. Pothuganti Manvitha And Neelam Swaroopa (2019),"Campus Placement Prediction Using Supervised Machine Learning Techniques ",International Journal Of Applied Engineering Research Issn 0973-4562 Volume 14, Number 9 (2019) Pp. 2188-2191 © Research India Publications. Http://Www.Ripublication.Com
8. S. Celine Et Al. (2015),"Logistic Regression For Employability Prediction",International Journal Of Innovative Technology And Exploring Engineering (Ijitee) Issn: 2278-3075 (Online), Volume-9 Issue-3, January 2020
9. Samrat Singh And Dr. Vikesh Kumar (2013),"Performance Analysis Of Engineering Students For Recruitment Using Classification Data Mining Techniques ",Ijcset |February 2013 | Vol 3, Issue 2, 31-37,Issn:2231-0711
10. Shreyas Harinath Et Al.'S (2019),"Student Placement Prediction Using Machine Learning",International Research Journal Of Engineering And Technology (Irjet) E-Issn: 2395-0056 Volume: 06 Issue: 04 | Apr 2019 Www.Irjet.Net P-Issn: 2395-0072
11. Siddu P. Algur, Prashant Bhat, Nitin Kulkarni., Educational Data Mining: Classification Techniques For Recruitment Analysis. I.J. Modern Education And Computer Science. 2(2016) 59-65.
12. T.Malathi Et Al. (2015),"Prediction Of Students Recruitment Process Using Data Mining Techniques With Classification Rules ",International Journal Of Computer Science And Information Technology Research Issn 2348-120x (Online) Vol. 3, Issue 3, Pp: (84-87), Month: July - September 2015, Available At: Www.Researchpublish.Com
13. Tripti Mishra, Dharminder Kumar, Sangeeta Gupta., Students' Employability Prediction Model Through Data Mining. International Journal Of Applied Engineering Research. 11(2016) 2275-2282.
14. V.Rameshet Al. (2015) ,"Performance Analysis Of Data Mining Techniques For Placement Chance Prediction",International Journal Of Scientific & Engineering Research Volume 2, Issue 8, August-2011 Issn 2229-5518
15. Vinutha K, Yogisha H K., Employability Prediction Of Engineering Graduates Using Machine Learning Algorithms. International Journal Of Recent Technology And Engineering (Ijrte). 8(2020) 4521 -4524.