Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Prediction Of Undergraduate Students' Career UsingVarious Machine Learning And Ensemble Learning Algorithms

Akanksha Pandey[1*] , L S Maurya[2]

[1*]Research Scholar Of M.Tech (Computer Science & Engineering)Shri Ram Murti Smarak College of Engineering and Technology, 13 km, Bareilly Nainital Road, Ram Murti Puram, Bareilly-243202, (U.P)-INDIA.

[2]Professor (Department of Computer Science & Engineering) Shri Ram Murti Smarak College of Engineering and Technology,13 km, Bareilly-Nainital Road, Ram Murti Puram, Bareilly-243202, (U.P)-INDIA.

Abstract

In the current scenario, the students need to identify their area of interest in an academic field so that they can opt for the right career courses they have an interest in and are capable of going through. The students have to go through many available options to draw the career path. In this paper, we are predicting the career an engineering student can select after their graduation using machine learning classification techniques. We will be describing the machine learning classification techniques that will help students to support their decision-making. The algorithmic methods and machine learning algorithms are presented here. We will be discussing our classification in machine learning algorithms to predict the career options for engineering students. The objective of the research is to find the factors that can affect students' decision to choose the right career path using machine learning techniques.

Keyword : Artificial Intelligence, Ensemble Method, Machine learning, Python, Supervised Learning.

I. INTRODUCTION

Engineering is one of the best career streams apart from medicine, which most students are opting for, some due to interest while some due to parental pressure, as it is the most defined career option in the world. There are engineering students who come out of college every year. Many students choose their stream after their graduation. Opting for the right carrier has becomea complex science nowadays as there are multiple career options and job competitions in the market. Researchers have even suggested machine learning classification technology to explore the right career option.

With the lack of proper guidance from professional services, students face problems in choosing the right career path. Many times they mismatch their career path in terms of their personality skills and interest. Students are even forced to opt for a career stream in engineering as pressure from family and the greed for high pay. The students in the past who have passed engineeringand started working for MNC but still lack interest and skills, make them unhappy. The upcoming generation, hereby, has now started to opt for the streams that interest them.

Machine learning technique for career guidance has been developed for engineering graduates who have completed their graduation or in the last semester and are still confused about which part of the field they should opt for. It’s a big challenge for those students to make the correct decision regarding the career they opt for as their complete future depends upon this. Thus, we have considered other aspects which will help us to opt for the right career path not just based on the academic score but also on the basis of their personality which is important for making their decision.

The general objective of this research is to develop a classification model for predicting career options after engineering undergraduate students. The purpose of this study is to govern whether the student’s academic performance is determined by aptitude or personality and to develop a model to analyze student’s performance. Machine learning is the best technique to identify and analyze the data and use them in order to perform the predictions automatically. Processes for a large amount of data are analyzed to discover patterns and rules after data. These generated rules and patterns can be easily processed and defined by a computer to characterize the new data. It is an automatic process that helps to improve the updated data. Therefore, as a result, it helps the students to improve their learning activities as well as it helps them to analyze the career path for their future.

To calculate educational performance at the global level for students, classification techniques are applied. The prediction model in terms of student-related variables is assessed.

In this paper, the future career options for students are predicted based on their skills, interest, hobbies, links, etc. the rules learned are portrayed in the context of the decision tree. In this work, the ensemble learning algorithm is preferred to improve prediction accuracy. The application of algorithms of Machine Learning involves analysis of data, visualization of data, performance prediction, providing feedback and recommendations, and grouping students. For this, pre-processing of the data of student performance is done, to extract features and for selection. It also involves data cleaning, tokenization, sentimental analysis, and removal of words. At the end of the preceding processes, the final performance of the student is obtained that will help them to analyze the best career options they should opt for after their graduation in engineering.

Section II addresses literature review. Section III represents proposed work. Section IV discusses research design and methodology. Section V describes the results and discussions of our research work. Section VI shows the conclusions of the entire research. Finally, Section VII lists the references.

II. LITERATURE REVIEW

In this section, we will review some papers in the correlated area.

Iqbal et al have discussed various machine learning techniques to predict grades after students in various courses. Models such as matrix factorization, classification, and regression are used to analyze the collected data from ITU, Pakistan. They have evaluated performance by usingmachine learning techniques and it has been found that RPM is the best among various machine learning techniques (Iqbal, 2017).

Vaidu et al has implemented machine learning techniques based on student performance to predict their employability skills. They have used KNN and Naïve Bayes models in order to classify the students into numerous groups. The result of the prediction of the employability of the students from the KNN algorithm is 95.33% accurate and that is all for the Naïve Bayes is 67.67% accurate (Vaidu, 2017).

In order to predict our future performance after students, Byung-Hak et al have used a Grit Net algorithm based on deep learning. As per the logistic regression, Grit Net gives more accurate results according to this research paper. They have taken data from the Udacity Nanodegree Programs (Schaar, 2017).

Jie et al how also proposed a machine learning approach to predict student performance in degree programs. In this investigation, the past, as well as present performance of the students, is evaluated. It uses a bi-layered structure that compromises multiple phase predictors along with a data-driven approach that is based on the efficient factors to base predict. This research paper has shown that the proposed method gives a more accurate result as compared to the benchmark approaches.

The machine learning algorithms examined by Pojon Murat are used to predict student performance. Pojon Murat has used three different algorithms like Linear regression, Naïve Bayes classification, and decision tree on two separate data sets, that is, Roberson and anotherone is featured engineering version. As per the result, Naïve Bayes is the best technique used for the first data set as it gives an accuracy of 98% while Decision Tree is the best technique for the second database as it gives the accuracy of 78% (Singh, M., 2013).

Singh et al have used some machine learning techniques to predict the academic performances of the students' subjects wise in their engineering field. To analyze the scores of the subject on the basis of the previous semester they predict the success scores of the students in the ongoing courses. For this purpose, decision tree classifier and Naive Bayesian techniques have been used and it has shown that decision tree gives the more accurate result as compared to Naïve Bayes (Bendengnu Ksung, 2018).

By using the machine learning techniques like Support Vector Machine, Random Forest, Gradient Boosting, and Naïve Bayes, Pushpa S et al predicts the student's performance whether they will fail or get a pass in the previous semester. As per the prediction, the accuracy rate of Random Forest is higher as compared to other algorithms, that is, 89.06% (Pushpa, 2017).

Bendengnu Ksung et al have used the DNN model, that is, Deep Neural Network to predict student performances. The research paper by Bendengnu Ksung et al compares the DNN with machine learning algorithms like Naïve Bayes, ANN, and Decision Tree. According to this, DNN achieves 84.3% accuracy which is better as compared to Machine Learning Techniques (Gerritsen, 2017).

To predict the student's performance, Gerritsen L. et al used data from the Learning Management System in reference to educational data by using Neural Networks. For this paper, aMoodle Log dataset is considered that has a file that contains 4601 students' information. In this paper, the performance of Neural Networks is compared with six classifiers named as, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Support Vector Machine, Logistic Regression,

and Random Forest. According to this paper, Neural Network is more accurate as compared tothe other six classifiers (Hernandez, 2018).

In order to predict the University's dropout student list, Martin S. et al have used four techniques of Machine Learning to analyze the performance. These four techniques include Random Forest, Support Vector Machines, Logistics Regression, and Neural Networks. For this research paper, the dataset of Instituto Tecnológico de Costa Rica (ITCR) students is used who have enrolled between 2011 and 2016. Among these four algorithms, the Random Forest algorithm is preferred to be best to predict University dropouts.

Ahmad F. Subahi (2018) has used the new artificial Neural Networks to predict the career path of the students based on the dataset. He proposed the data-driven system to collect the data to predict all the future career paths available.

Machine learning algorithms such as the SVM decision tree and XG boost are used by K. Sripath Roy (2018) to create a model of student career predictions. Among these algorithms, the Support Vector Machine gives the more accurate result that is 90.3%.

Mubarak Al Baraka Umar Jesus is the artificial neural network technology to predict a student's academic performance. In this study, the model of neural networks is created which predicts a student's GPA by using their personal information, place of resistance, and academic information. According to this model, the accuracy of prediction is 73.68%.

III. PROPOSED WORK

There are intense websites and web applications that help students in order to know their suitable career path but the drawback of this system is that they only use personality traits to predict the career which might not give a consistent result. Similarly, numerous websites are there over the Internet that suggests students to opt for a career as per their interest. These systems do not have the capacity to understand whether the student can survive in that particular field or not.

Beth Dietz-Uhler & Janet E. Hurn's paper suggests a need for learning analytics to predict and improve the student performance to enlighten the importance of student's interest, trends ability etc (UD Beth, 2013). Lokesh Katore, Jayant Umale, Bhakti Ratnaparkhi's paper predicted the different classifiers have different accuracy in order to predict a career of a student (KS Lokesh, 2015).[10]

So, here we are proposing a model for career prediction of students

Let's have a look at these machine learning algorithms that help students to predict careers after their graduation in engineering.

A. **Decision Tree:** It is a supervised learning technique used for Classification as well as Regression problems. Though it can be used for both but most commonly it is used for Classification problems. Structurally it is a tree classifier where the features of dataset are represented by internal nodes, decision rules are represented by its branches and outcome is denoted by each leaf nodes.

There are two nodes in decision tree, namely decision nodes and leaf nodes. Decision nodes make any decision and have multiple branches, while the leaf nodes are basically the output of the decision taken by the decision nodes and they don't have any branches.

B. **Random Forest:** Random forest is a supervised learning process used for classification and regression. It builds decision trees based upon the samples in classification. In regression, the treesare built as per the average of the samples. One of the key features of this algorithm has to be the ability of handling data sets having continuous variables for the cases of regression and categoricalvariables for classification. Having said that, this algorithm gives better results for classification problems as compared to regression (Torgo, 1996).

To be able to understand the working of random forest methods, a concept called the ensemble technique should be known. Ensemble is basically combining two or more models. Hence, predictions are made based upon a set of models, instead of individual models. Ensemble utilizes two types of methods:

- **Bagging:** As mentioned above, a different training subset is generated from the sample trainingdata and majority voting determines the final output data. Random forest uses this method for classification.
- **Boosting:** In this method of ensemble, the final model created gives the maximum accuracy. Thisis done by generating a sequential model upon combining weak learners with strong learners. Examples of this include ADA BOOST, XGBOOST.

Having understood the ensemble methods, let's look at the random forest technique, which uses the firstmethod, that is, bagging. The steps involved in random forest algorithm are as follows:

Step 1: x number of random records are taken up from a data set with y number of records.
Step 2: Individual decision trees are built for each and every data set sample.
Step 3: An output is generated from all individual decision trees.
Step 4: The majority voting for classification and average of the samples for regression decides the finaloutput data.

C. **Voting Classifier :** A voting classifier is a ML model that is trained on an ensemble of different models which predicts an output (class) depending on the highest probability of the selectedclass as output. It predicts the output class based on maximum majority votes by simply collecting the results of each classifier passed to voting classifier. We create a model that is trained on the different models and predict outputs on the basis of their combined majority instead of creating thosemodels separately and finding accuracy for each of them.

D. **Gradient boosting :** Gradient boosting gives prediction model in the scheme of an ensemble of weak prediction models involving decision trees. The resulting algorithm is known as gradient boosted trees when the weak learner is a decision tree. It is built in a stage wise manner just like in other boosting methods , but it allows optimization of a random differentiable loss function. Gradientboosting basically combines different weak "learners" into a single strong learner . It is easiest to understand in the least-squares regression setting. The gradient boosting is a combination of gradient descent and boosting.

E. **XG Boost classifier :** XG Boost is an algorithm which has recently gained popularity inapplied machine learning for tabular data. XG Boost is designed for performance and speed byapplication of gradient boosted decision trees. It belongs to a collection of tools of Distributed

machine learning community. XG Boost is a software library can be accessed from different interfaces after downloading and installing on your machine. The main interfaces supported by XGBoost are:

- C++
- Command Line Interface (CLI).
- R interface
- Python interface
- Java and JVM languages like Scala and platforms like Hadoop
- Julia

F. **Ada Boost classifier : A**daBoost algorithm uses a short decision tree as weak learners which are added in a sequence to the ensemble and it was one of the very first successful boosting approaches. A weak learner is a simple model with some skill on the dataset and boosting is a class of ensemble ML algorithm which involves combination of predictions by many weak learners. Eachsubsequent model tries to correct the predictions made by the preceding model. This is accomplishedby by weighting the training dataset to focus on samples that previous models predicted incorrectly.

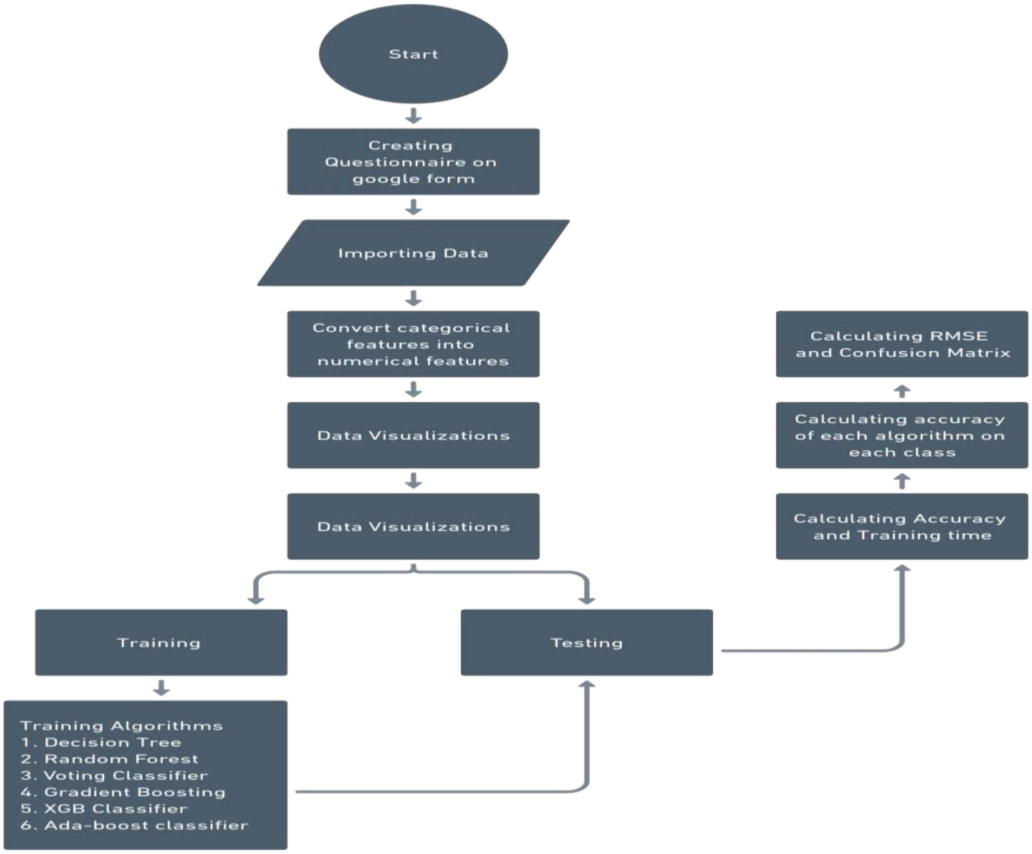

Start
Creating Questionnaire on google form
Importing Data
Convert categorical features into numerical features
Data Visualizations
Data Visualizations
Training
Testing
Training Algorithms
1. Decision Tree
2. Random Forest
3. Voting Classifier
4. Gradient Boosting
5. XGB Classifier
6. Ada-boost classifier
Calculating Accuracy and Training time
Calculating accuracy of each algorithm on each class
Calculating RMSE and Confusion Matrix

Fig 1. Proposed Flow chart

IV. RESEARCH DESIGN AND METHODOLOGY

A. Data collection

A questionnaire containing 10 questions is Designed to collect data. It contains 8 columns as input variables and an another column for career option selection is created which is dependent on each inputvariable. Using the questionnaire a total of 330 observations were collected.

Questionnaire columns are-

1. User name
2. What is your percentage in class 10?
3. What is your percentage in class 12/Diploma?
4. What is your marks percentage in B.Tech till date?
5. Rate your aptitude skill
6. Rate your communication skill
7. Rate your technical skill
8. Rate your management skill
9. Rate your general knowledge
10. Which career option will you prefer after graduation?

B. Data preprocessing

Data preprocessing consists of a set of techniques to transform an unstructured data into structured data.Designed questionnaire contains raw values where we have to preprocess the raw data before using machine learning classification algorithms.

C. Encoding categorical features

Collected data contains only categorical features. First we have to encode each feature beforing training our data because machine learning algorithms perform well with numerical data. For independent features,we encoded ratings value of the questionnaire as Excellent - 5, Very good - 4, good - 3. Average - 2, Poor
- 1 and for dependent feature, we used a python library known as label encoder for encoding this feature.After encoding dependent feature, the assigned number of each class are classified as - Gov Job - 0, M.Tech/ME/MS -1, MBA - 2, Others - 3, Prvt Job - 4.

D. Data Visualizations

Data visualization is creating graphical pictures to represent the data. Data visualization helps to tellstories to the users by organizing the data into a form easier to understand.

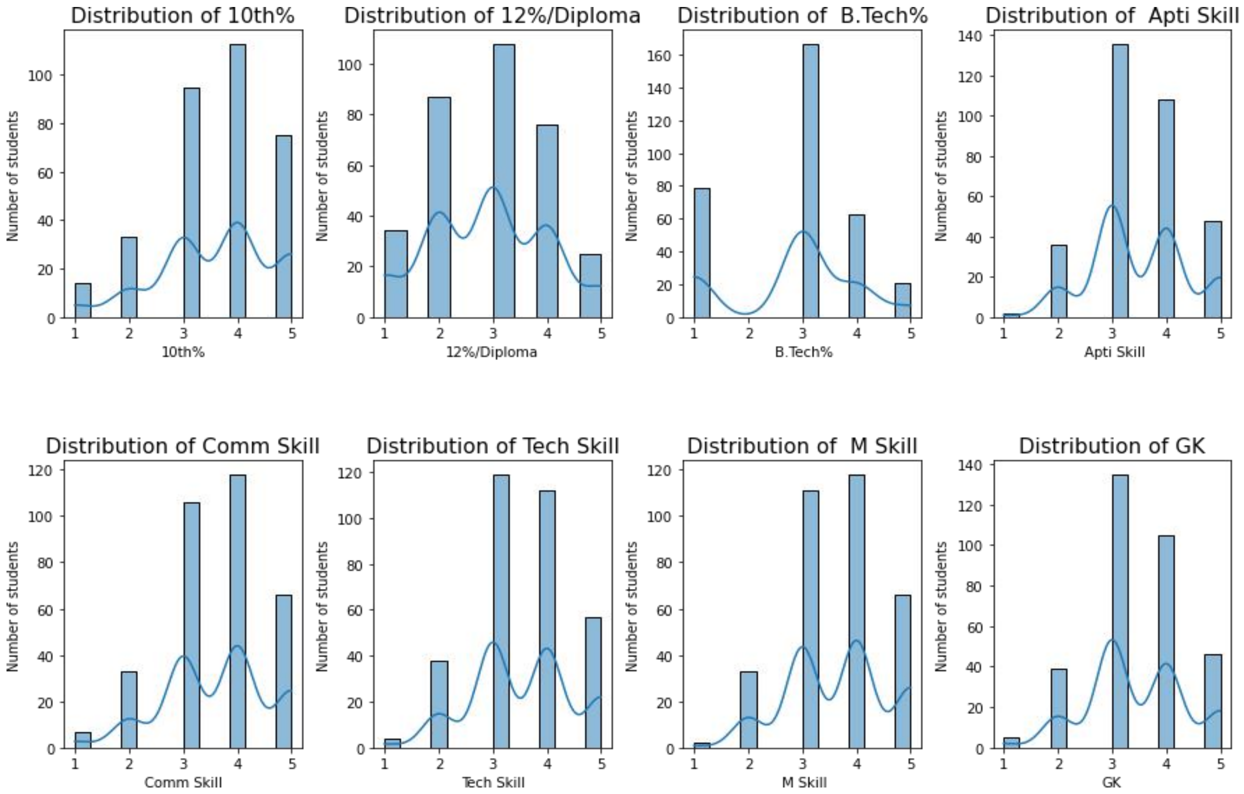
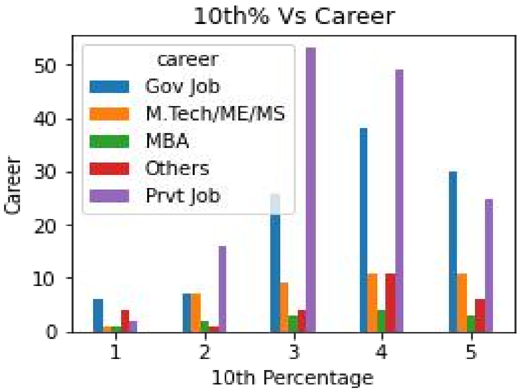


Fig 1. Distribution of numerical features

Above figure (Fig 1). Shows the bar plot of numerical features. In this plot, we can see that most of the class 10th students rated themselves 4 i.e, more than 80% however in 12th class, most of the students rated their score as 3 which is more than 70% and the scores scored by students in B.Tech is relatively less than 12th percentage which concludes that as students reach to higher studies, their scores are gettingreduced. But, many of them have good tech, communication and management skills which are quite relevant in order to succeed in real life.



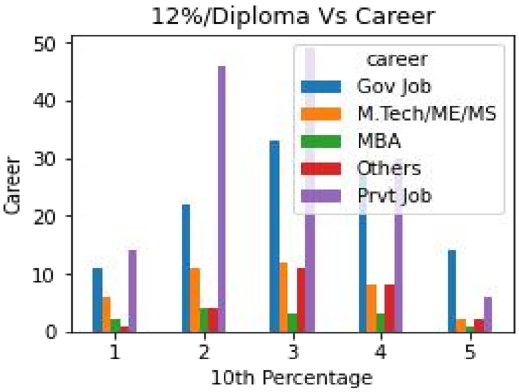
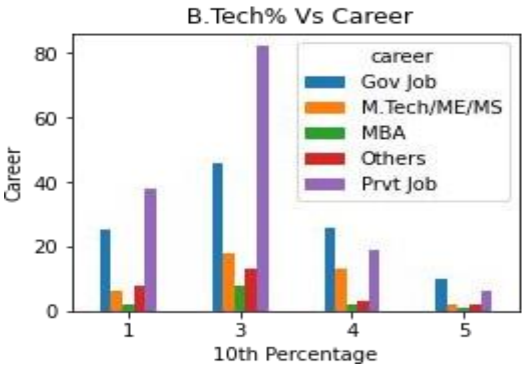


Fig 2 . 10%vs Career

Fig 3. 12th%vs Career



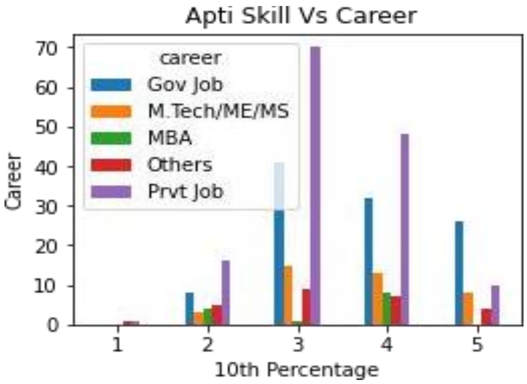
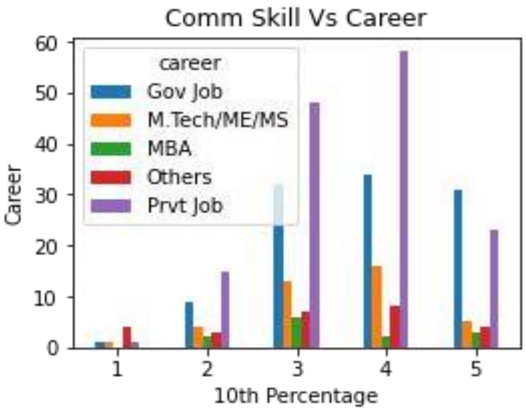


Fig 4. 10%vs Career

Fig 5. 12th%vs Career



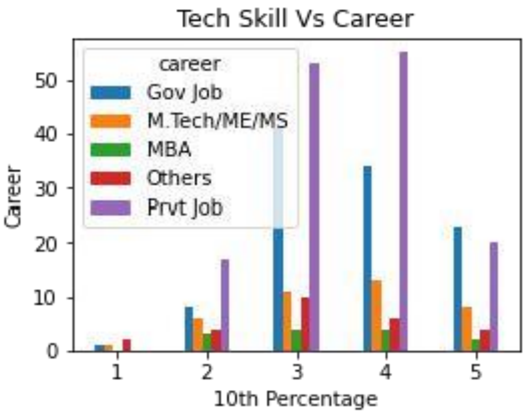


Fig 6 . 10%vs Career

Fig 7. 12th%vs Career

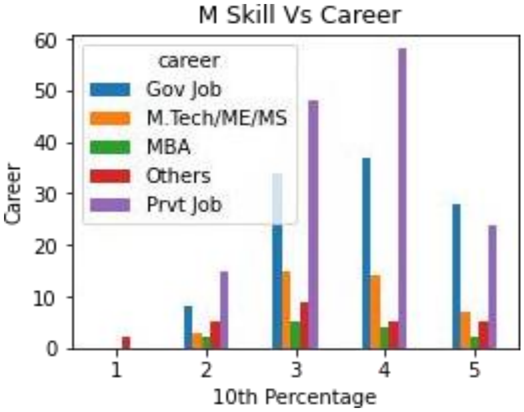


Fig 8 . 10%vs Career

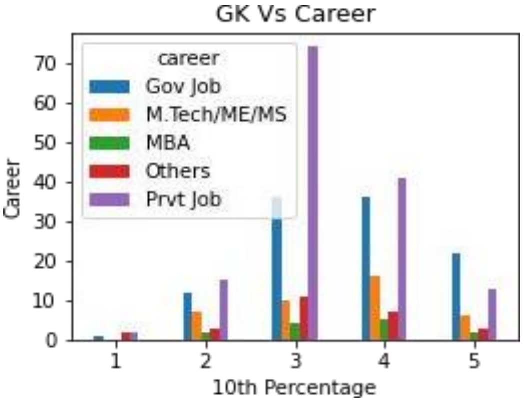


Fig9. 12th%vs Career

Fig 2. to Fig 9 represents that most of the students either scored very well or didn't score much want to do govt jobs or private jobs. Very few of them are interested in doing MBA while M.tech/MS/Me is still a student's choice.

Now that we have so many features to deal with, we can ease our work by reducing some features. Even irrelevant features can actually reduce the performance of a machine learning model. For example : KNN algorithm works on nearest neighbor logic, if it find the nearest value from the irrelevant column then ourmodel performance will decrease. So, it's a good practice to reduce the columns.

Let’s see the relation of every column with each other, for that we will use correlation heatmap. Correlation basically tells how one feature is changing with another feature. Its value ranges from -1 to1.If it is +1 then it means one feature is increasing with another feature and if the value is -1 then one feature is decreasing while the other one is increasing.

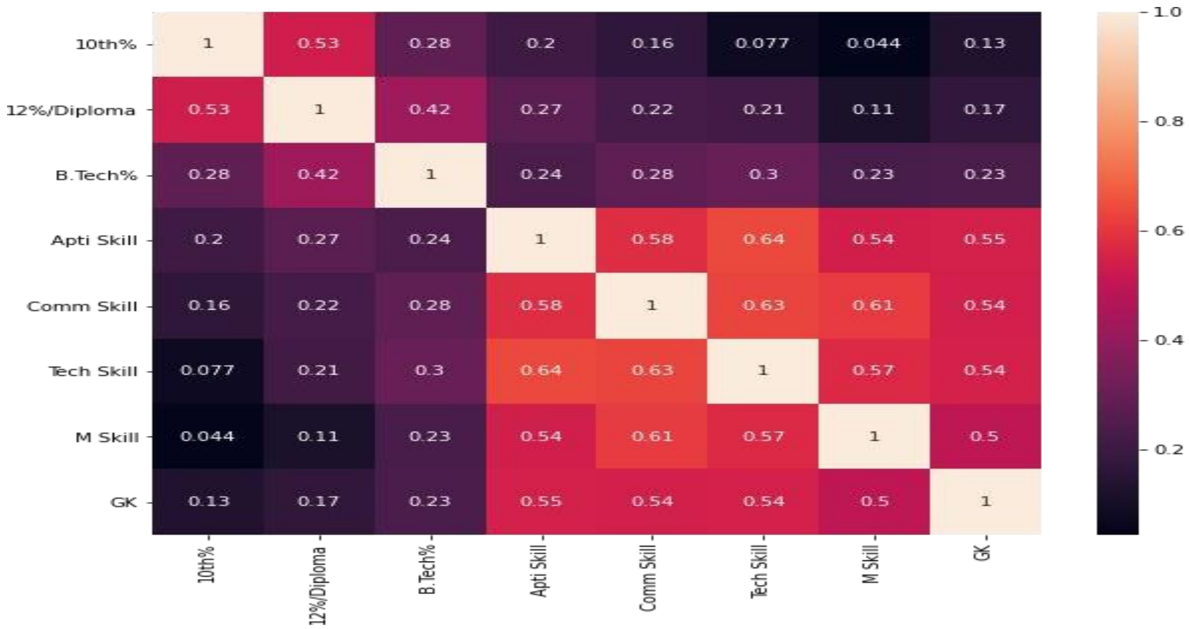


Fig 10. Cor elation Matrix

E. Model Preparation

Training and testing

Since we have a total of 330 observations, we decided to train our model on 80% of data and 20% datafor testing.

F. Classifiers

For this study, we have selected 6 popular machine learning and ensemble learning classificationalgorithms. These are -

- Decision Tree
- Random Forest
- Voting Classifier
- Gradient Boosting
- XG Boost Classifier
- Ada-boost Classifier

V. RESULTS

A. Accuracy

Now that we have prepared our algorithm, the obtained accuracies are shown in Table 1 where accuracyof each algorithm is calculated how much time it took for an algorithm to train itself i.e, execution time. Hyperparameter tuning is done on every algorithm to get the desired accuracy.

S. No.	ALGORITHM	HYPERPARAMETER	EXECUTIONTIME	ACCURACY OBTAINED
1.	Decision Tree	max_depth=5	0.0050 s	40.91%
2.	Random Forest	m_estimators=100	0.1769 s	50.0%
3.	Voting Classifier	voting='hard'	0.2834 s	51.52%
4.	Voting Classifier	Voting='soft'	0.2935 s	53.03%
5.	Gradient Boosting	n_estimators=100, learning_rate= 1	0.4076 s	42.42%
6.	XGB Classifier	n_estimators = 100	0.3142 s	40.91%
7.	AdaBoost Classifie	Default	0.7401 s	33.33%

	r			

Table 1. Calculated accuracy score, Hyperparameter tuning, Execution time and Random state foreach classifier.

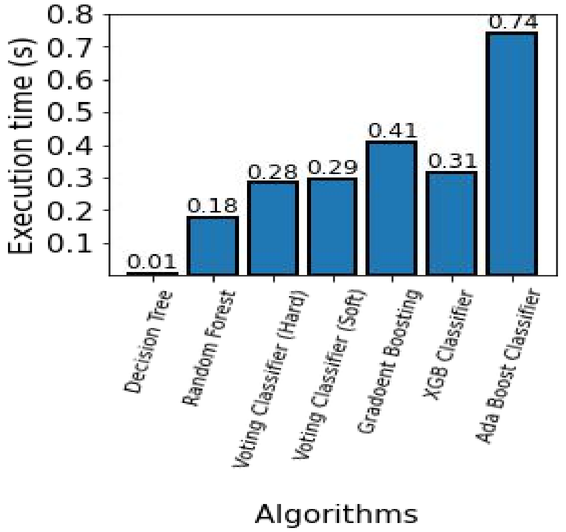


Fig 11 . 10% vs Career

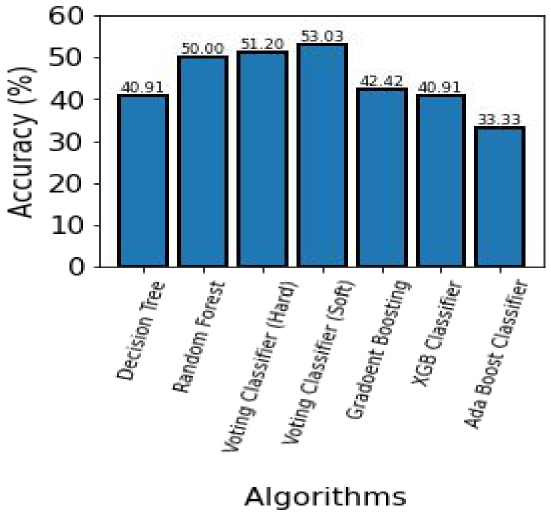


Fig 12. 12th% vs Career

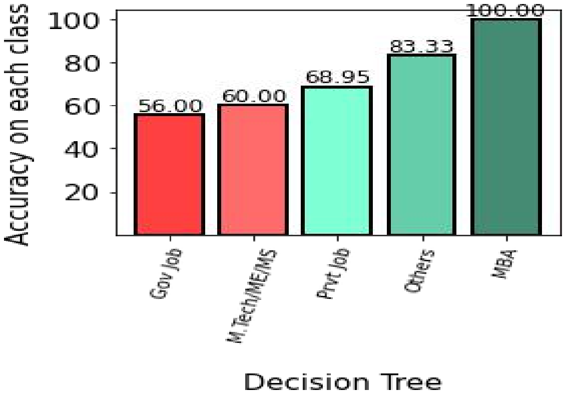
S. No.	ALGORITHM	Govt Job	M.Tech/ME/MS	MBA	Others	Prvt Job
1)	Decision Tree	56.00%	60.00%	100%	83.33%	68.95%
2)	Random Forest	52.00%	40.00%	100%	50.00%	75.86%
3)	Voting Classifier(Hard)	60.00%	80.00%	100%	100%	86.20%
4)	Voting Classifier(Soft)	60.00%	100%	100%	100%	82.75%
5)	Gradient Boosting	52.00%	60.00%	100%	50.00%	62.06%
6)	XGBoost Classifier	52.00%	40.00%	100%	50.00%	62.06%

7)	Adaboost Classifier	60.00%	40.00%	100%	50.00%	62.06%

Table 2. Calculated accuracy score of each Algorithm in each class.

Table 2. Represents the score which actually represents how well our selected classifiers are performing on each class. Each algorithm predicts 100% correct values on MBA class, while very less score is seen on the rest of the class. The reason behind this can be the sample size of each class is relatively different. Voting classifier performs very well with the classes M.Tech/ME/MS, MBA and Others while the overallperformance of XG Boost classifier on each class is relatively lesser than rest of the algorithms.

Below we are visualizing these scores of each algorithm to get better insights of each class.



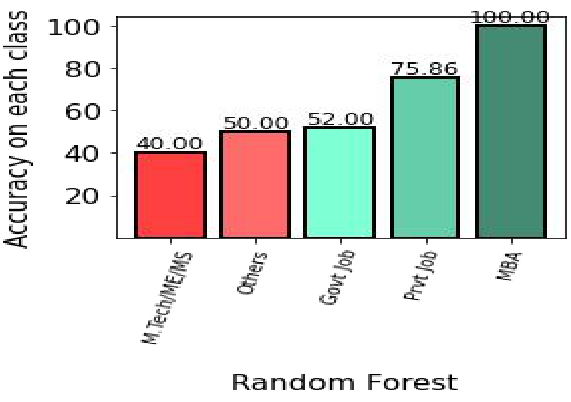
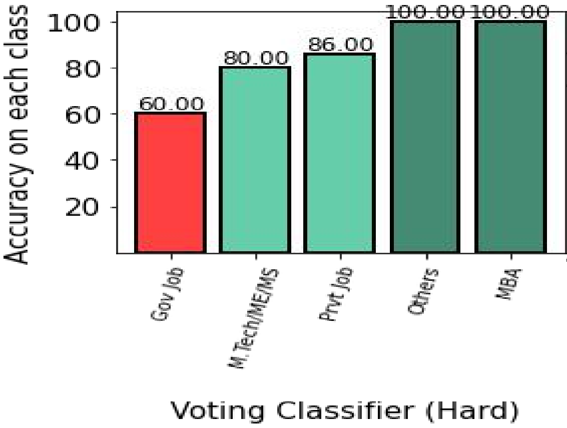


Fig 12 . Decision Tree performance on each class **Fig 13. Random forest performance on each class**



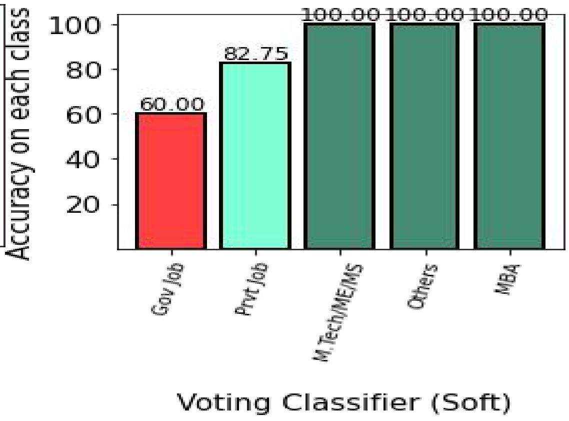
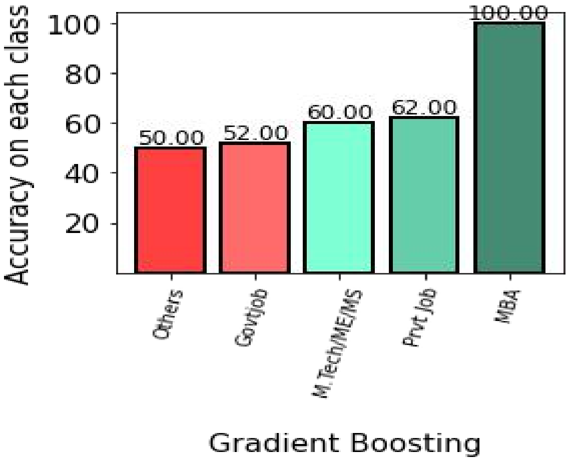


Fig 13 . Voting classifier performance on each class

Fig 14. Voting classifier performanceon each class



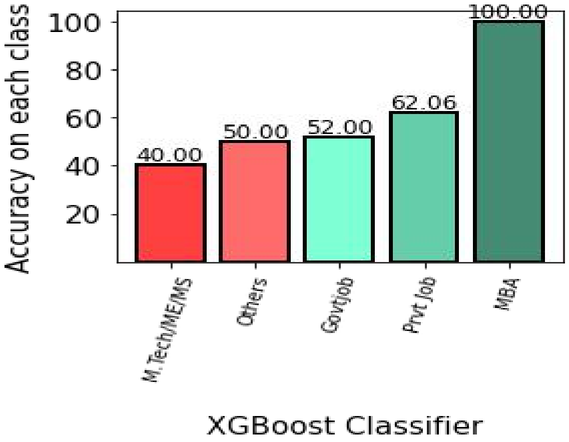


Fig 15 . Gradient boosting performance on each class

Fig 13. XG Boost performance on each class

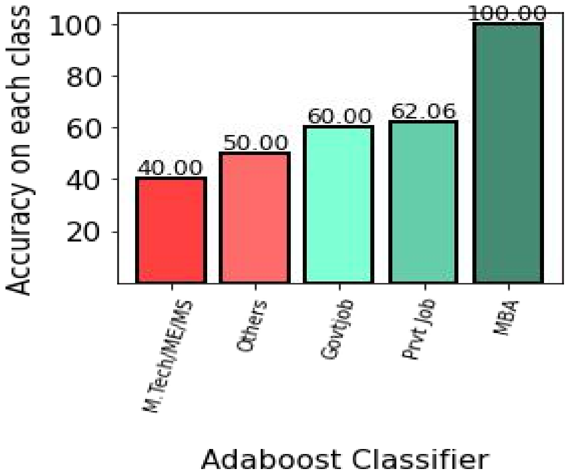
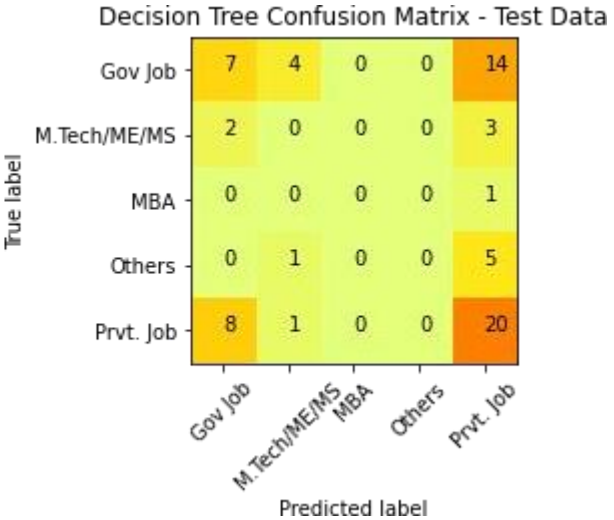
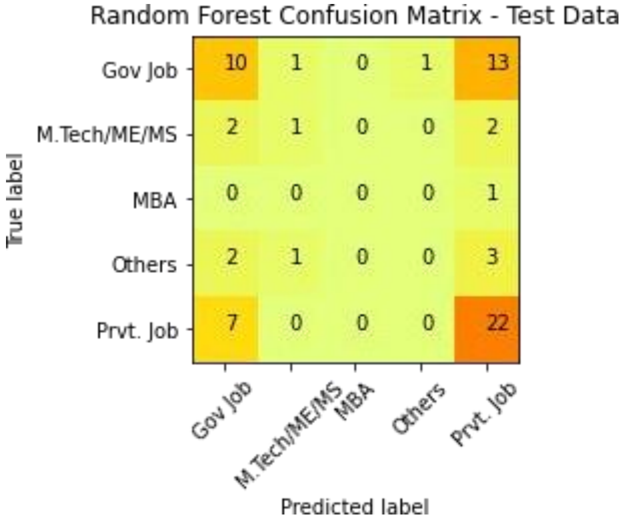
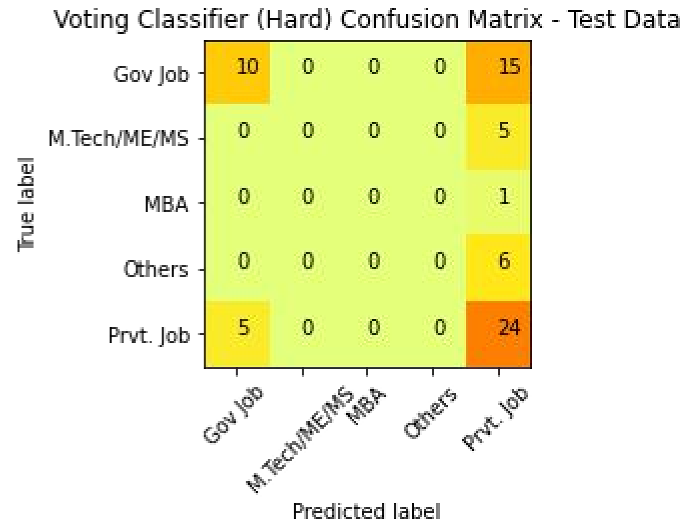
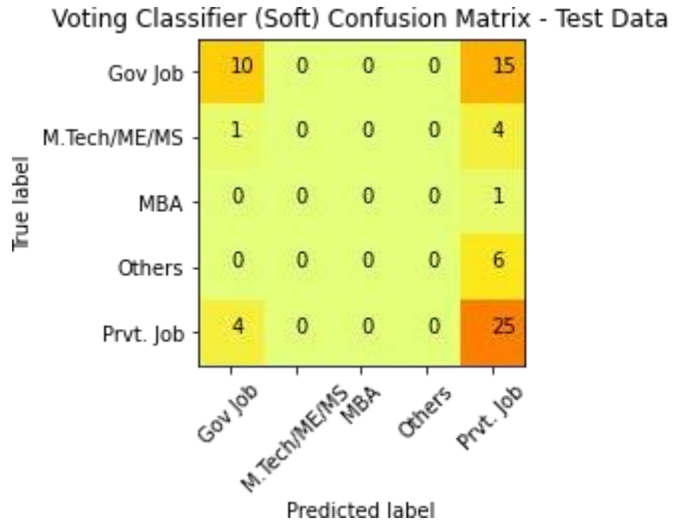
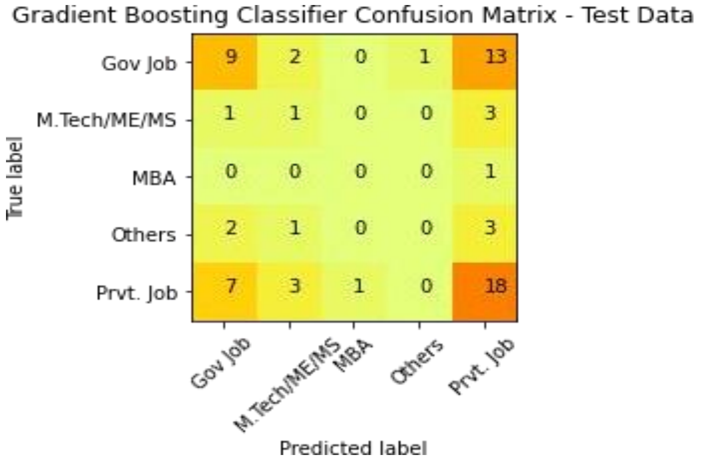


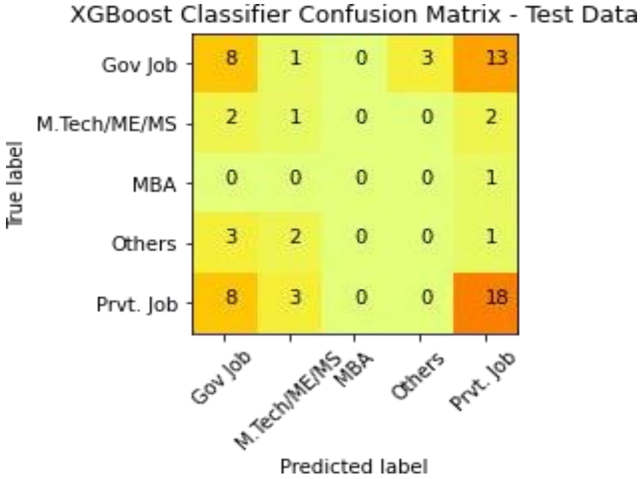
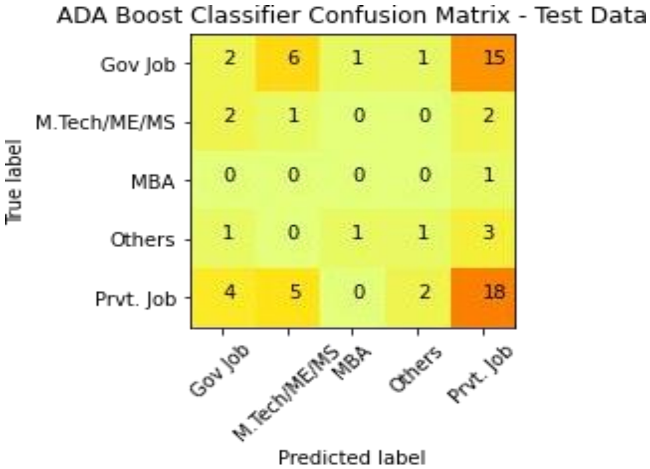
Fig 14 . Gradient boosting performance on each class

B. Confusion Matrix and Heatmap

Confusion matrix is a great technique to check the performance of any classifier. We have a total of 5 classes so the size of the confusion matrix will be 5x5. Graphical representation of confusion matrix i.e, heatmap is obtained and shown below in Table 3 Along with Mean squared value which is used to evaluate the distance between predicted value and actual value and machine learning optimizer tries to fitthe model by reducing this distance. Lesser values of mean square error, better will be the accuracy.

S. No.	ALGORITHM	RMSE	CONFUSION MATRIX
1.	Decision Tree	6.17	
2.	Random Forest	5.74	

3.	Voting Classifier (Hard)	5.68	
4.	Voting Classifier (Soft)	5.32	
5.	Gradient Boosting	6.35	

6.	XG Boost Classifier	6.83	
7.	Adaboost Classifier	6.17	

- **Table 1. Calculated accuracy score, Hyperparameter tuning, Execution time and Random state for eachclassifier.**

VI. CONCLUSIONS

In this study we have created various classification models for prediction of career options for an undergraduate student. Various input features such as students marks percentage in 10th, 12th, B.Tech/Diploma, skills in communication etc. are taken into consideration and output variable was career options a student can choose which were classified as Gov Job, M.Tech/ME/MS, MBA, Others , Prvt Job. We have proposed six most popular machine learning and ensemble learning classification algorithms i.e, decision tree, random forest, gradient descent, voting classifier, xg boost classifier, adaboost classifier . Accuracy of each algorithm is evaluated and the performance sequence of each algorithm is as follows: **Voting classifier (Soft) > Voting classifier (Hard) > Random Forest > Gradient Descent > Decision tree > XG Boost > Ada-boost** classifier which is obvious from Fig. 12. Execution time is also calculated for each algorithm and the sequence of algorithm is as follows: **Decision Tree > Random forest > Voting Classifier (Hard) > Voting Classifier (Soft) > XGB Classifier > Gradient Descent > Ada-boost classifier** (see Fig. 11). It has been seen that Ada-Boost classifier takes a lot of time for training the data and give very low score so ada-boost classifier is not a good fit for this problem. Accuracy score of each algorithm on each class has been evaluated and compared and then we calculated the confusion matrix to check the performance of each algorithm.

VII. REFERENCES

i. Vadivu, G. (2017). K. sornalakshmi,“. Applying machine learning algorithms for student employability prediction using R” International journal of pharmaceutical sciences review andresearch (ISN 0976-044X) March-April, 38-41.

ii. Iqbal, Z., Qadir, J., Mian, A. N., & Kamiran, F. (2017). Machine learning based student gradeprediction: A case study. Ar Xiv preprint arXiv:1708.08744.

iii. Kim, B. H., Vizitei, E., & Ganapathi, V. (2018). Grit Net: Student performance prediction with deeplearning. Ar Xiv preprint arXiv:1804.07405.

iv. Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking andpredicting student performance in degree programs. IEEE Journal of Selected Topics in Signal Processing, 11(5), 742-753.

v. Pojon, M. (2017). Using machine learning to predict student performance (Master's thesis).

vi. Singh, M., & Singh, J. (2013). Machine Learning Techniques for prediction of subject scores: AComparative Study.

vii. Bendangnuksung, P. P. (2018). Students’ performance prediction using deep neuralnetwork. International Journal of Applied Engineering Research, 13(2), 1171-1176.

viii. Pushpa, S. K., Manjunath, T. N., Mrunal, T. V., Singh, A., & Suhas, C. (2017, August). Class resultprediction using machine learning. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon) (pp. 1208-1212). IEEE.

ix. Gerritsen, L. (2017). Predicting student performance with Neural Network. Tilburg University,Netherlands.

x. Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018, July). Perspectives topredict dropout in university students with machine learning. In 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI) (pp. 1-6). IEEE.

xi. García-Peñalvo, F. J., Cruz-Benito, J., Martin-Gonzalez, M., Vazquez-Ingelmo, A., Sánchez-Prieto, J.C., & Theron, R. (2018). Proposing a machine learning approach to analyze and predict employment and its factors.

xii. Chaudhary, D., Prajapati, H., Rathod, R., Patel, P., & Gurjwar, R. (2019). Student Future Prediction Using Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 1104-1108.

xiii. Mostafa, L., & Beshir, S. (2021, June). University selection model using machine learning techniques. In The International Conference on Artificial Intelligence and Computer Vision (pp. 680-688). Springer, Cham.

xiv. Gorad, N., Zalte, I., Nandi, A., & Nayak, D. (2017). Career counselling using data mining. Int. J. Eng.Sci. Comput. (IJESC), 7(4), 10271-10274.

xv. Roopkanth, K., & Bhavana, V. (2018). Student career area prediction using machine learning. IEEE-No.