### **PREDICTION OF EMPLOYABILITY OF**

### **GRADUATES USING ENSEMBLE LEARNING**

**A PROJECT REPORT**

***Submitted by***

## **SRIHARSHA ITHA[Reg No: RA1911003010225]**

## **AAKASH PANDEY [Reg No: RA1911003010252]**

*Under the guidance of*

## **Dr. G. ABIRAMI**

(Assistant Professor, Department of Computing Technologies) *in partial fulfillment for the award of the degree*

*of*

### **BACHELOR OF TECHNOLOGY**

### in

## **COMPUTER SCIENCE AND ENGINEERING**

Of

**FACULTY OF ENGINEERING AND TECHNOLOGY**



S.R.M. Nagar, Kattankulathur, Kancheepuram District

**October 2022**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

(Under Section 3 of UGC Act, 1956)

**BONAFIDE CERTIFICATE**

### Certified that 18CSP109L project report titled “PREDICTION OF EMPLOYABILITY GRADUATES USING ENSEMBLE LEARNING” is the bonafide work of “**SRIHARSHA ITHA** [Reg No: RA1911003010225] and **AAKASH PANDEY** [Reg No:

### RA1911003010252]”, who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE**

Dr. G. Abirami Assistant Professor

Dept. of Computing Technologies

**SIGNATURE**

Dr. P. Madhavan Associate Professor

Dept. of Computing Technologies

**SIGNATURE**

Dr. M. Pushpalatha

Professor

# **ACKNOWLEDGEMENTS**

We express our humble gratitude to **Dr. C.** **Muthamizhchelvan**, Vice Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support. We would also like to thank Dr. T.V. Gopal, Dean CET, SRM Institute of Science and Technology, for his continued support and encouragement. We extend our sincere thanks to **Dr. Revathi Venkataraman**, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her invaluable support. We wish to thank **Dr. M. Pushpalatha**, Professor & Head, Department of Computing Technologies, SRM Institute of Science and Technology, for her valuable suggestions and encouragement throughout the period of the project work.

We are extremely grateful to our Panel Head, **Dr. P. Madhavan**, Associate Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for his inputs during the project reviews. We register our immeasurable thanks to our Faculty Advisor, **Dr. P. Madhavan**, Associate Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr. G. Abirami**, Assistant Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for providing me an opportunity to pursue my project under her mentorship. She provided me the freedom and support to explore the research topics of my interest.

We sincerely thank staff and students of the Computing Technologies Department, SRM Institute of Science and Technology, for their help during my project work. Finally, we would like to thank my parents, our family members and our friends for their unconditional love, constant support and encouragement.

Sriharsha Itha Aakash Pandey

iii

# **ABSTRACT**

### Graduates` employability is a very big worry for the universities and predicting their employability can help Institutions improve their placement ratio and ranking. If graduates are unemployed or unable to find work, then they will be are often more likely to face many difficulties like difficulty in finding future job opportunities and less income etc. Economy is another thing it affects because if people don’t have jobs they would have less money to spend which basically leads to a weaker economy. Comparatively unemployed don’t consume as much as those who are employed, so having more unemployment can very much hurt the economy. The proposed system focuses on predicting employability and identify weak areas of graduates so that they can work on their weak areas and be more likely to receive a job, lessen number of unemployed students after finishing their graduation and Improving Institution curriculum to build new technical skills, both for educating, training and reskilling current and future graduates.

### The project has considered various Supervised Machine Learning algorithms like random forest, decision tree, xgboost, svm and logistic regression in order to predict employment using several parameters whose values are analyzed. For the proposed experimental measures, we have collected the dataset from the Kaggle website which contains the records of people collected by Indian Universities.

### Using earlier results taken from other reference papers, we found out logistic regression gives us highest accuracy compared to other supervised model to predict employability for graduates also taking precision and recall into consideration we found out logistic regression best suites our requirement. Many test cases have been made for the logistic regression model in order to check and predict the future issues.

# **TABLE OF CONTENTS**

**ACKNOWLEDGEMENTS** iii

**ABSTRACT** iv

**1) INTRODUCTION 2**

1.1 What is employment?

1.2 Reasons for unemployment

1.3 Why prediction of employability?

**2) LITERATURE REVIEW 6**

2.1 [Research Wor](file:///C:\Users\aakash\Projects\Minor%20Project\aakash-harsha-final-review-doc.docx#_TOC_250006)k

**3 SYSTEM ARCHITECTURE AND DESIGN 10**

[3.1 Proposed Work](file:///C:\Users\aakash\Projects\Minor%20Project\aakash-harsha-final-review-doc.docx#_TOC_250005)

3.2 [Block diagram for employability prediction](file:///C:\Users\aakash\Projects\Minor%20Project\aakash-harsha-final-review-doc.docx#_TOC_250004)

[3.3 Real time working](file:///C:\Users\aakash\Projects\Minor%20Project\aakash-harsha-final-review-doc.docx#_TOC_250003)

3.4 Analysis of design

**4) METHODOLOGY 16**

4.1 [System Design](file:///C:\Users\aakash\Projects\Minor%20Project\aakash-harsha-final-review-doc.docx#_TOC_250001)

4.2 Dataset

4.3 Pre-Processing

4.4 Data Analysis

4.5 Separation of data into two parts

4.6 Training

4.7 Model Evaluation and Accuracy Score

4.8 Predictive Analysis

* 1. Implementation

**5) Coding and Testing 25**

5.1 Coding and screenshots

**6) Results and Observation 28**

6.1 Prediction of Employability

6.2 Confusion matrix and classification report .

6.3 Performance and accuracy measures

* 1. Summary of prediction model

**7) CONCLUSION 31**

**8) REFERENCS 32**

# **CHAPTER 1**

**INTRODUCTION**

# **What is Employment?**

Employment is a contract between two entities who manage work performance for their benefit. One party, the employer (which may be a corporation, non-profit organization, cooperative, or other organization), will pay the other worker under the contract in return for performing an assigned job. Depending on the type of work the worker does, the prevailing conditions of the sector, and the bargaining power between the parties, it may be paid as an hourly, piecework, or annual salary. Employees in some sectors may receive tips, bonuses or stock options. Depending on the type of employment, workers may receive benefits in addition to their wages eg: health insurance, housing, LTA etc. Employment is usually regulated by labor laws, organizations or legal contracts.

# **Reasons for Unemployment**

Unemployment is impacted by the status of the economy, which can greatly decrease job openings. In the developing economy, economic growth is very slow. This slow growth rate does not provide enough employment opportunities for the growing population. The labor supply far outstrips the available employment opportunities. Constant population growth is a serious problem. It is one of the main causes of unemployment. During the period of the plan, the number of unemployed people increased rather than decreased. Lack of scientific and technical research due to high costs. This leads to excessive use of foreign technology and to technological unemployment in the country. Due to low rate of capital accumulation due to low savings and investments. Shortages of capital impede the expansion and growth of businesses and limit their ability to create jobs. In developing countries often the economy is very much dependent on Agriculture but expansion and diversification programs have been plagued by a lack of financial resources.

# .**1.3 Why Employability Prediction?**

By using employability prediction institutes and universities can better equip graduates with relevant skills and technical knowledge required for being able to be employable. It also helps in identifying strength and weak areas of graduates and gives them concrete data to improve upon by making their strong areas stronger and improving their weaker areas graduates become much more like to land their dream jobs. Prediction of Employability does not do help graduates but it also helps universities to improve their job placements and ranking due to this future student will be more compelled to join the university. It will reduce the number of unemployed students after their graduation and hence even help the economy directly. It also helps in Improving universities curriculum to build new skills and competencies for educating training and upskilling current and future graduates. Our project uses various Data Mining (DM) and ML models in order to predict Employability for graduates, Our project tells user that whether a graduate is employable or not employable using several metrics like 10th,12th marks, specialization, work experience, mock interview test scores, masters etc. in that we found that comparing the results from the earlier papers in terms of accuracy, precision and recoil tests logistic regression algorithm is the best fit for our use case. An increasing number of studies are investigating data mining techniques for predicting employability. However, these studies show significant differences in terms of the data used, the methods used, and even the metrics considered for Model assessment. Our project’s aim is to do

early prediction of graduates whether they are employable or not so that they can they improve upon their weak areas this helps in great extent of finding a job

# **CHAPTER 2**

## **LITERATURE REVIEW**

Literature searches are conducted to obtain specific information about the work or projects we have undertaken. This allows you to adopt current methods, techniques, and algorithms implemented by other authors to complete your research and work and obtain useful and accurate predictions. It helps us understand existing trends and technologies and how they continue to work. This gives us a clear idea of the best way to perform our work and obtain results and calculations that are useful for imparting knowledge and understanding to those who are not familiar with it or who are new to it. Literature reviews help us to understand the step-by-step process by which a person performed work and help us to better understand the subject or topic. It gives us deep knowledge and expertise in the areas we wish to investigate and research. We can make further improvements and apply our knowledge and expertise to find better solutions that benefit society and deepen our understanding.

# **2.1 Research Work**

The placement records of the cumulative college are analyzed to predict whether a particular person is placed or not. Placement is very important as it determines future career of graduates. The Placement dataset is considered which consists of the placement records of graduates which are further analyzed. This paper considers different types of data mining and ml models to predict the people who are placed with the as less error rate as possible. After analyzing the dataset and applying the necessary models it comes to notice that 150 people out of 219 are placed companies and the remaining 69 people are not unplaced in any companies. The logistic regression among all the models gives the highest accuracy of 93 percent to predict the placement of graduates with the least error. This paper also considers confusion matrix to find the total accuracy for the parameters.

The ml based prediction and the different data mining and machine learning models used for their prediction. The prediction is considered to predict the placement by using different prediction techniques. After carefully analyzing and evaluating the data the correct and accurate prediction. This paper experiments with several models and techniques taken together instead of considering a single model. The pre-processing is done using various methods such as feature selection, feature extraction which is used with the different prediction and classification techniques for better results and accuracy.

The model with the highest accuracy among all the models is taken to predict the placement because higher the accuracy more accurate the results are with the least possible error. The way to solve it is dividing it 3 phases into where the starting phase is the one which improve quality, analyze the data so that we can get a clear idea about it. Feature detection and feature extraction is used and pre-processing which also help to achieve a higher accuracy by further processing the data which is very useful. The second phase deals with the different algorithm and model training in which different models are considered to predict whether the person is placed or not. The third phase is the model evaluation and accuracy score in which the performance of different models which are computed by previous papers are analyzed to find the model which will give the best accuracy with the least possible error. Higher the accuracy more accurately the model will detect prediction. Attributes that are considered and are of value are 10th,12th marks, specialization, work experience, mock interview test scores, masters are considered. XGBoost, Cat ,LightGBM classification techniques. Maximum accuracy achieved in case of all the models is 79 percent [1].

Decision trees, artificial neural networks(ANN), K nearest neighbors, SVM[2], etc. are considered for prediction of placement. The model with the highest accuracy will be able for the placements with least error’s helps by analyzing placement records. It is used for the early prediction of placement that prevent the unemployment of graduates from increasing. So early detection helps to find graduates their dream job by starting to work on their weakness, technical competencies, and communication skills to best match company skillset [3] . A metric is taken which is used to predict the employment. If the metric has the value 0 then the person is non placed and if the metric has the value 1 then it is placed [4]. Several supervised machine learning techniques are considered. The model with highest score and least error is used to predict the placement most accurately.ID3,JFTtree and BFTtree algorithm is used for classification and validate prediction of placement whether the person is placed or not its usually used on vast dataset often having error rate of maximum 9.35 [5].

The effect of being unemployed is serious it not only affects graduates and effects the economy as well not only that the purchasing power parity also goes down using SVM we are able to achieve 89% accuracy [6].The metric used are value 0 and 1 which stand for non-placed and placed respectively. Major factors for unemployment are rate of growth population, lack of scientific research [7], lack of job creation options etc. Using these metrics, we will be able to use other algorithms over them to find which gives us maximum accuracy and minimum error rate. The modelling techniques used are RMO, Random Forest [8].

Being unemployed not only takes toll on economy but it also reduces morale and self-confidence of unemployed graduate [9].Unemployed metrics can be because not having good score in few or some other metrics due to which he maybe unemployed but its not always the case graduates can keep working on their weak areas still be able to land a successful or dream job, The algorithms which help us in predicting efficiently and successfully can neural network [10].

Applying algorithms for classifying such as SMO, naïve Bayesian, multilayer perception the maximum accuracy obtained is by multilayer perception which is 75% [11]. Algorithms which are Ensemble in nature those are best for classification [12].

It’s not completely graduates responsibility universities and higher institute also play vital role in find proper job opportunities [13].Based on evaluation random forest, naïve Bayesian achieved highest cross validation which is 93% [14].

The model will be used to predict employability effectively and efficiently. The algos are applied on data set and the relevant attributes are used get the model ready. The best model for such task are Random forest, Decision Tree [15]. This approach considers supervised machine learning techniques. Suggestions are made to improve the accuracy of detecting placed people from non-placed people. Supervised models considered are decision trees, logistic regression [16]. The main focus is to achieve higher accuracy for detecting and predicting placed people who are unplaced. The focus is to get higher prediction accuracy with minimum error. Early detection is key to tackling vulnerabilities. The goal is to achieve accurate predictions with minimal error. This leads to better employment forecasts

Hence also considering all other models and applying best algorithms best suited for our project requirements. The accuracy we got after analysis for the decision tree is 90.69% ​​& for random forest algo it is 91%, while logistic regression has an accuracy of 93.02%. Therefore, from the above analysis and prediction, it is best to use logistic regression algorithm to predict location outcomes.

# **CHAPTER 3**

## **SYSTEM ARCHITECTURE AND DESIGN**

This is used to represent the underlined architecture of the project. Basically, inputs are given to the project, based on which employment opportunities for specific people are determined. The input to this project is a dataset collected from cumulative universities placement it consists of students who are placed and unplaced. These datasets, collected per individual, count the number of values does not present in the dataset, remove irrelevant features, and extract useful features to better understand and analyze the data and from which processed to extract useful results. Achieve higher accuracy. Inputs are sent to various DM & MLmodels to analyze data & accordingly make predictions with some degree of accuracy as to whether a person has Employment or not. Accuracy is very important because the higher the accuracy, the more accurately we can , predict and identify graduates with job. Poor accuracy can cause the model to make incorrect predictions. Accuracy is therefore very important for making accurate and accurate forecasts. The inputs are various data mining and machine learning models such as RF, DT, SVM, XB, LR.

Input data is split into 2 parts. One as training data and one as test data. The training data helps you train your model, and the test data helps you evaluate the trained model to make high accuracy predictions about employability. Testing data provided as inputs to various machine learning models to examine their accuracy in making predictions for predicting employment. Classification reports are based on various parameters such as accuracy, recall, f1 value, and support. All these parameters are analyzed for all different models compared to make predictions. After calculating the accuracy of all models, find the model with the highest accuracy and lowest error to make a prediction. A model with a maximum accuracy of is stored, given random values for all attributes as input, to predict whether a person has gotten placed If we feed the saved model with values for all attributes, we can see that the model correctly predicts placed and unplaced graduates.

# **Proposed Work**

Employment, or Lack of Employment that is Unemployment is arguably one of the most negative economic phenomena as it can affect social cohesion and stability .The project will help us predict the employable people and the non-employable people. From the dataset we will be able to identify the employable graduates and the different parameters are analyzed to come to the outcome of placed or unplaced . In this way we will be able to make people more aware and reduce the chances of being unemployed. The dataset will contain the records of people that are placed or unplaced .

# **Block diagram for employment prediction**

Module one we will be collecting the dataset. The attribute named status is used to tell if the people are placed and not. Outcome value 0 denotes that the student is unplaced and the outcome value 1 denotes that the student is placed. In the 2nd module, we find the number of null/missing values in our dataset. We use clean the dataset by pre processing the data so that we get the highest accuracy possible from our model. We also use data pre-processing to remove any duplicate value and update the null/missing values with the most relevant values. In third Module we will be converting all the string values to numbers.

In the 4th Module we will split the dataset into training and testing data using iloc. 5th Module is the model in which we will feed the training and testing data into our machine learning model to calculate the accuracy and we get accuracy for random values. Sixth Module we use confusion matrix to get a better idea of our model and to get the accuracy.

Block diagrams are used to show the step-by-step process of the steps in action. Our system predicts if a graduate would be employed or not based on the their gender, grades including school, college gpa as well as mock interview % to judge soft skills. This uses various machine-learning algorithm to give the different amount of accuracies and then we compare those accuracies to get the most efficient model possible.

Talking about the dataset, it contains 10th % and board, 12th% and board, graduation gpa, degree name, and internship experience as attributes

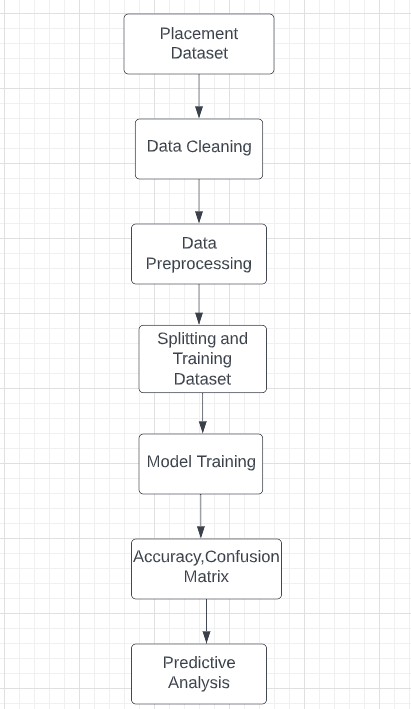


Figure 3.1: Block Diagram of Employability prediction

# **Real time working**

Thinking about this project in real time we will look at the figure and can say that the placement record dataset is taken and fed into employability prediction model. It will have a look at the placement records of the graduates and using metrics and scores in each metric it determines whether the patient is employable or not employable. If the model ‘s metric scores are low then. The model output will be used to detect and predict the placed and unplaced graduates.

If the model predicts unplaced, then the unplaced graduate can improve upon his weak areas and become employable.

# **Analysis of design**

Firstly, the placement dataset is considered which will be used as a input in the project. The dataset consists of the placement details of different people which are collected. The placed details are collected for a set of different attributes whose values are closely analyzed and observed. Based on those values we must come up with a solution to detect and predict placement. Exploring the data helps to extract useful results which can prove beneficial to predict employability in future. Feature extraction helps to find the useful attributes in order to predict the placed. The non-placed records will show unplaced in the dataset whereas the placed records will show placed and will act as outlier because of which we are able to predict such employability records in the dataset. Pre-Processing can be used to find the number of missing values

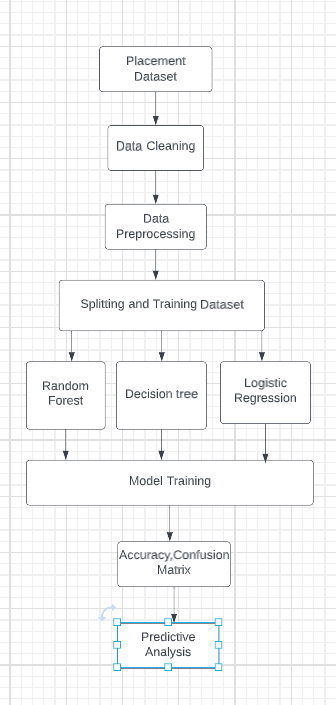


Figure 3.2: Working

in each column of the dataset because any missing values can affect the accuracy with which we are predicting employability. After that we will be dividing or splitting dataset into 2 parts. Train data is there for the training of the different machine learning models that we have considered like Logistic Regression, Random Forest & DT in order to predict employability. Testing data is used to evaluate the trained methods and find the model with the highest accuracy and performance measures that can predict placement correctly and easily without any error. The models can be differentiated on various different measures. The models which will have these parameters better than the rest of the models will be the most appropriate model that we have to select in order to predict placement and that model will be most suitable, correct and accurate in order to predict and detect placement. The model with the highest accuracy is of lot of importance because we will use that model as a base model going forward for detecting and predicting placement. Model that has the maximum accuracy is saved and the placement data of different people can be considered on that for prediction and detection of placement. model having the highest accuracy and the least possible error will benefit to predict the people who are placed and unplaced accurately. The random value of all the attributes are given to the saved model and we see that our model predicts the employable and unemployable people correctly. The saved model will give only one out of the two outputs that is 0 and 1 based on the placement data input. If the model gives 0 as the output for a particular placement data then the person is unplaced but if the model gives 1 as the output then the person is placed. So the saved model can be used to predict the employable and unemployable records from the dataset.

**CHAPTER 4**

**METHODOLOGY**

We will be discussing our project methodology here. We have 6 modules in our project. In the first one we will be collecting a dataset which will be having a lot of features including gender, 10th percentage, 10th board, 12th percentage, 12th board, degree %, degree name, work experience, mock test %, specialization, masters, status, salary, etc.

The attribute named status is used to tell if the people are placed and not. Outcome value 0 denotes that the student is unplaced and the outcome value 1 denotes that the student is placed. In the Second module, we calculate the missing or null values in our dataset. We clean the dataset so that we get the highest accuracy possible from our model. We also use data pre-processing to get rid of duplicate or irrelevant values and update the missing values with the most relevant values. For example, s\_no was an irrelevant column for us as it has no impact on placement. Another column that we will be dropping would be salary because while it can be relevant in other projects but we just want to find the employability of graduates what salary they get is not relevant to us.

In third Module we will be converting all the string values to numbers. Now before moving forward, we need to pre-process and transform our data. For that, we will use *astype()* to change the datatype to *category*. After this we apply cat codes to convert text values to numeric values.

In the Fourth Module we will split the dataset into test and train data using iloc. Fifth Module is the model in which we will feed the training and testing data into our machine learning model to calculate the accuracy and we find out accuracy for random values. We also compare few different algorithms to see which model will give the best accuracy to predict the employment likeliness of a graduate student. We will use different models to calculate the accuracy and choose the best model that is going to have the maximum accuracy.

In sixth Module we use confusion matrix to get a better idea of our model and to get the accuracy with two rows and two columns containing false positives, false negatives, true positive and true negative

**4.1 System Design**

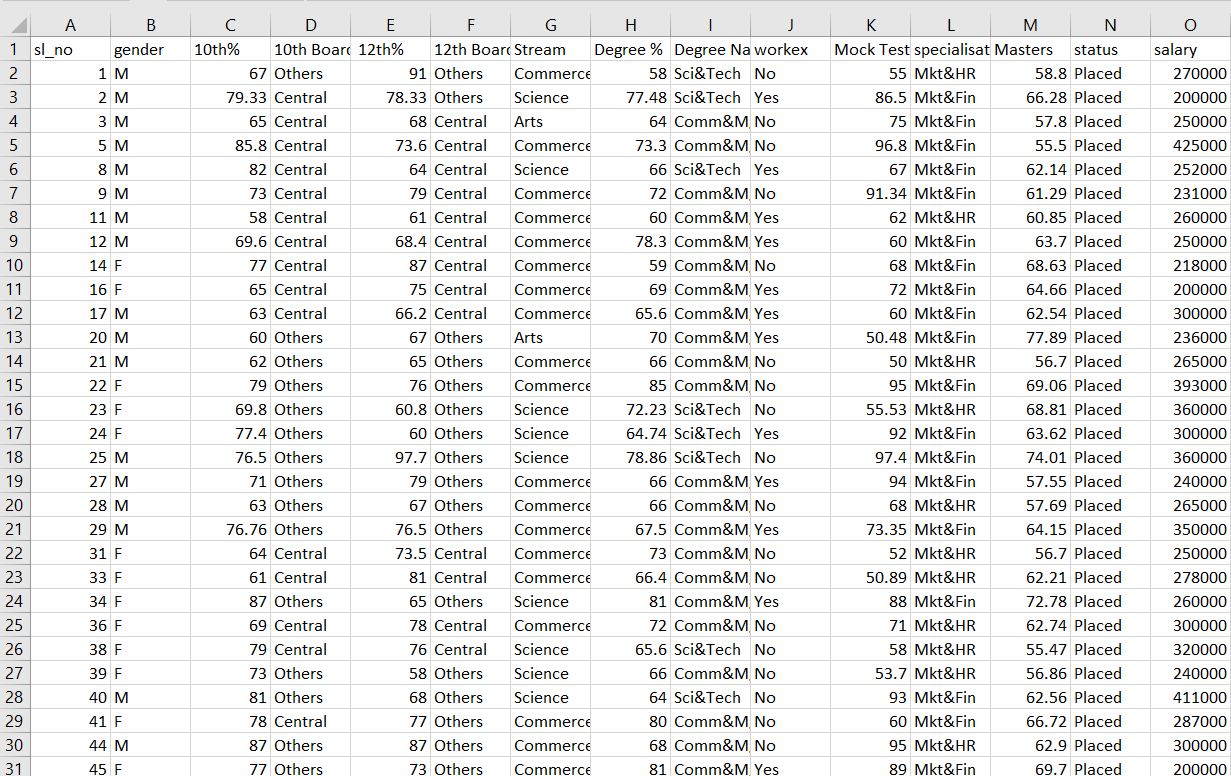
For backend we will be using jupyter notebook and python. In our system we will be using a lot of ml & data mining models and makes model selection according to which one gets the best accuracy. Looking at existing papers we were able to get much higher accuracy than them because we used a characteristic called mock interview percentage which helps us get an idea of the graduate’s soft skills like fluency, confidence and body language which are very important factors in interviews. You could theoretically have very high technical knowledge but if you are unable to communicate that knowledge chances of getting through interviews are low, especially HR interviews. While 10th, 12th percentage can be used to get a small idea of the student’s past performance it’s mostly useless but at the same time very low marks in 10th and 12th can end up in either not being eligible for a lot of companies or it can make getting short listed very tough but once you go for tests and interviews it’s mostly an irrelevant metric.

**4.2 Dataset**

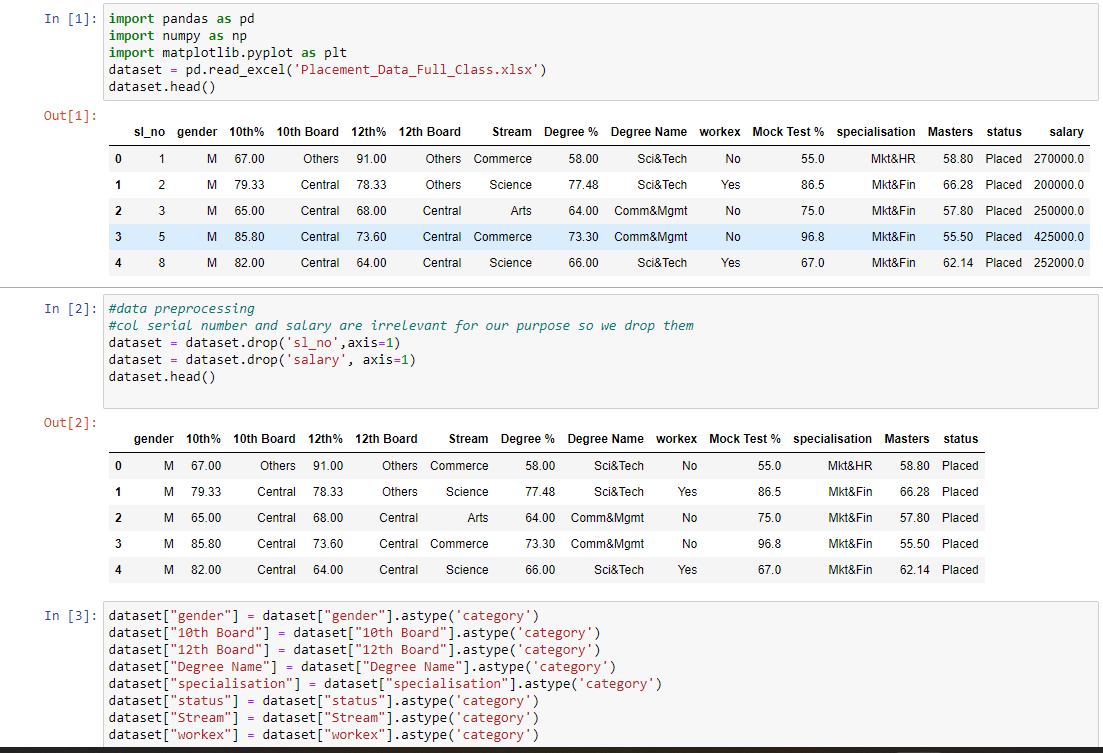
We have taken the dataset from Kaggle which contains all placement data which contains a lot of features including gender, 10th percentage, 10th board, 12th percentage, 12th board, degree %, degree name, work experience, mock test %, specialization, masters, status, salary, etc. The attribute named status is used to tell if the people are placed and not. Outcome value 0 denotes that the student is unplaced and the outcome value 1 denotes that the student is placed. Our dataset contains 200+ records in total.

**4.3 Pre-processing**

Import the required modules. I used pd.read\_csv to read the dataset and removed data columns not used in the model. For example, one of the columns that was removed was Salary. I don't care if the student is employed or not, although it could be an important pillar for other projects. I don't care if they are employed. Before proceeding, we need to preprocess and transform the data. To do this, use the astype() method on some columns and change the data type to categorical. Let's apply code to some of these columns to convert text values ​​to numbers.

 Figure 4.1: Dataset

**4.4 Data Analysis**



**4.5 Separation of data into two parts** Now split the dataset into training and test data. This will be used later to check efficiency.

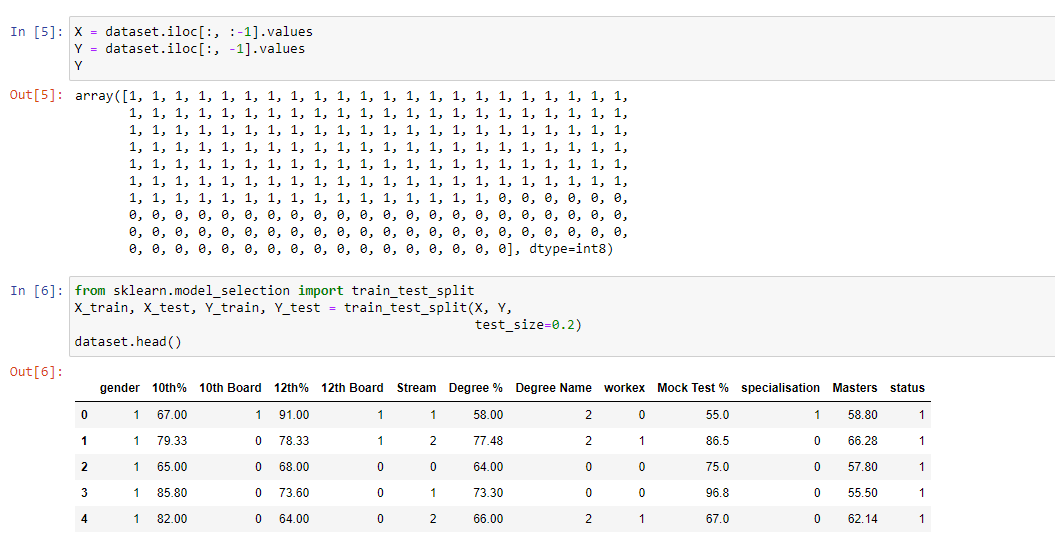
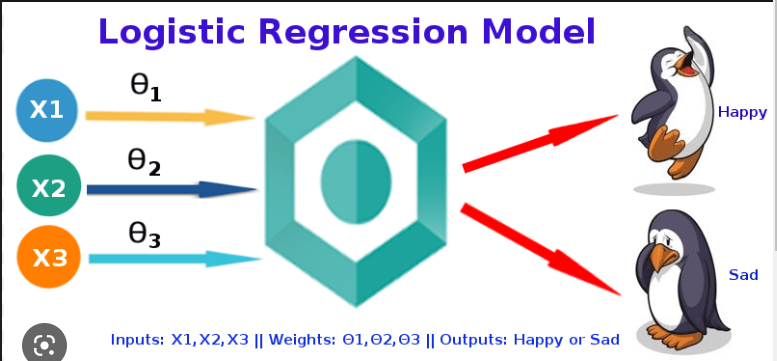


Figure 4.6: Separation of data into two parts

**4.6 Training**

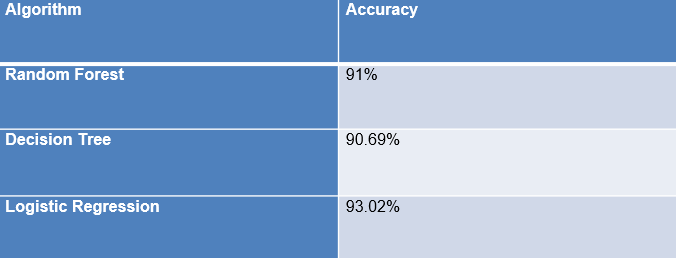
In our model in which we will feed the training and testing data into our ML model to get the result and see how accurate it is then we find out result for random values. We also compare few different algorithms to see which model has the highest accuracy to predict the employment likeliness of a graduate student. We will use many models to calculate the result and choose the best model that is going to have the maximum accuracy. In sixth Module we use confusion matrix to get a better idea of our model and to get the accuracy in % terms

**4.6.2 Logistic regression** Figure 4.7: Logistic Regression

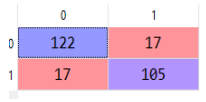
We applied the ML models on our dataset and compared the results Algorithms are applied on the data set and the attributes are used to build the model. The accuracy obtained after analysis for the decision tree is 90.69% ​​and for the random forest it is 91%, while the logistic regression has an accuracy of 93. Therefore, from the above analysis and prediction, it is best to use logistic regression algorithm to predict location outcomes.

**4.7 Model evaluation and Accuracy score**

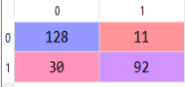
Campus placement is extremely important from the institution's perspective as well as from the student's point of view.

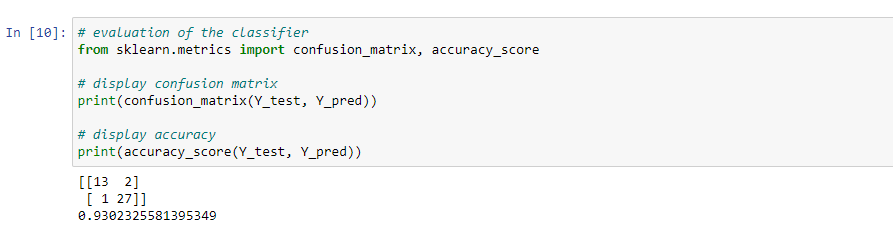


Confusion Matrix for Random Forest:



Confusion Matrix for Decision Tree Algo:





Confusion Matrix for Logistic Regression

**4.8 Predictive Analysis**

We save our model with highest accuracy and feed in data to see the placement probability. After carefully analysing all the different accuracies for all the models we come to the conclusion that logistic regression was giving us the highest accuracy when it came to employability of a graduate, which is why we saved this model. After that we start predicting and if the output comes out to be 0 then that means it’s predicting that the student will not be placed but if it comes out to be 1 then that student will be placed.

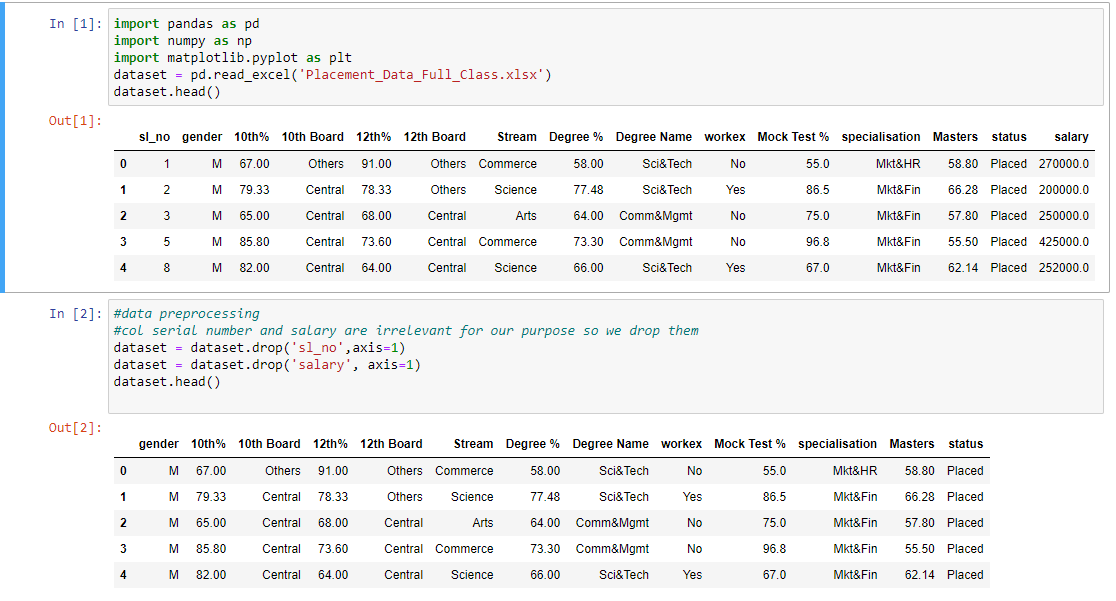
**4.9 Implementation**

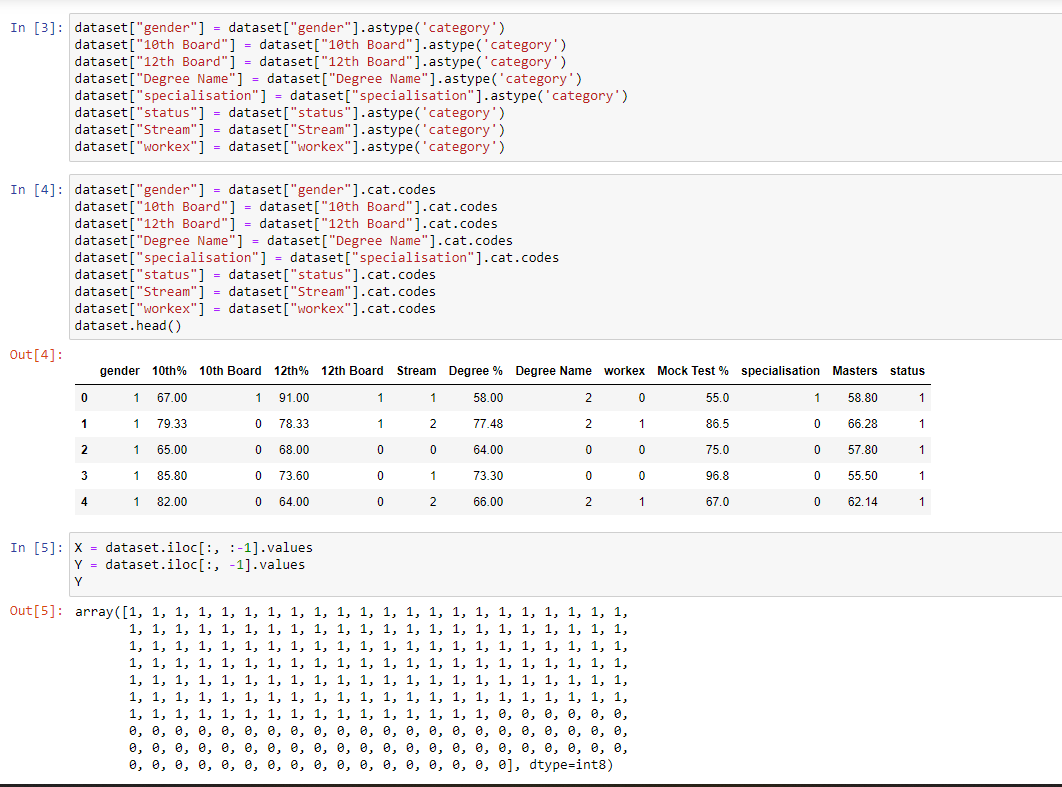
The reason why we are making this project and the reason why we think it’s going to be helpful is that you can find out where you stand placement wise whenever you want and continuously improve your profile and skills so you can be ready when it’s placement season. With the fields we have while filling it you can automatically realise areas for improvement. Some things like 10th% and 12th% are things students would not be able to change now but they can get more internships, better gpa in their current course, and give more mock tests to improve communication skills. Gradually increasing the employability would allow students to gain confidence as well. This will result in more industry ready students which will also be good for the university because these days a big factor while deciding which college to go to is placements so better placement data is going to go a long way for the college as well. The main goal of our project is to predict this accurately and with very few errors. We used many different datasets but this one having mock interview as one of the parameters gave us very good results

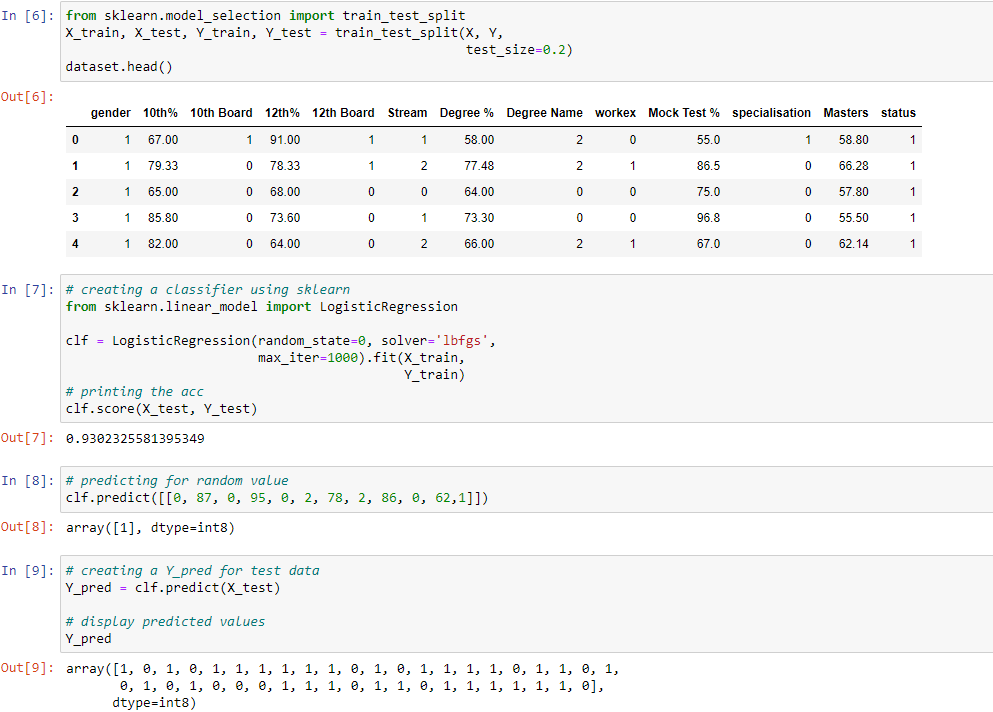
**CHAPTER 5**

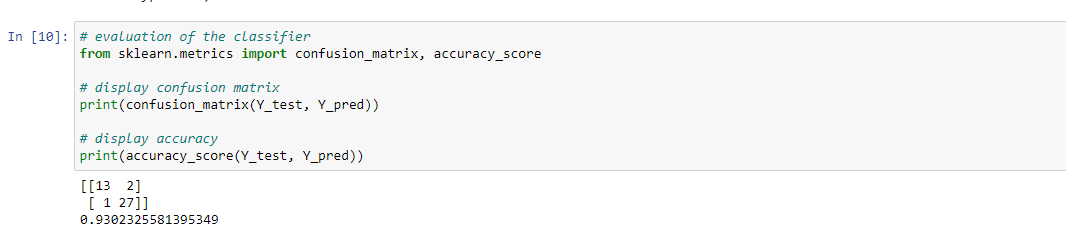
**CODING AND TESTING**

**5.1 Coding and screenshots**









**CHAPTER 6**

**RESULTS AND OBSERVATIONS**

We give inputs to our machine learning model for finding whether a graduate is employable or not. It is very important to be accurate here as careers are at stake which is why we wanted to be very careful with pre processing of data making sure that the columns that are not helpful are dropped. It will allow you to find out where you stand placement wise whenever you want and continuously improve your profile and skills so you can be ready when it’s placement season. With the fields we have while filling it you can automatically realise areas for improvement. Some things like 10th% and 12th% are things students would not be able to change now but they can get more internships, better gpa in their current course, and give more mock tests to improve communication skills. Gradually increasing the employability would allow students to gain confidence as well. This will result in more industry ready students which will also be good for the university because these days a big factor while deciding which college to go to is placements so better placement data is going to go a long way for the college as well. The main goal of our project is to predict this accurately and with very few errors. We used many different datasets but this one having mock interview as one of the parameters gave us very good

**6.1 Prediction of Employability**

Today, the employability of graduates is a major concern for universities, and predicting their employability helps them take timely steps to increase placement rates at institutions. Predict student placement status based on various student attributes using a logistic regression algorithm. Knowing their weaknesses before applying for an internship allows students to work in areas that need improvement to best match the company's skill set.

**6.2 Confusion matrix and classification report**

We save our model with highest accuracy and feed in data to see the placement probability. After carefully analysing all the different accuracies for all the models we come to the conclusion that logistic regression was giving us the highest accuracy when it came to employability of a graduate, which is why we saved this model. After that we start predicting and if the output comes out to be 0 then that means it’s predicting that the student will not be placed but if it comes out to be 1 then that student will be placed.

**6.3 Performance and accuracy measures**

The on-campus arrangement is extremely important from the institution's perspective as well as from the student's point of view. In this regard, to improve student performance, a work was analyzed and predicted using Decision Tree and Random Forest algorithm and logistic regression classifier algorithms to validate the approach method.

**Random Forest Approach**:

Random forest is a type of ensemble learning in ML. The dataset with all the various attributes will be the input. In RF a lot of decision trees are formed and they all vote the majority vote is displayed as ouput.

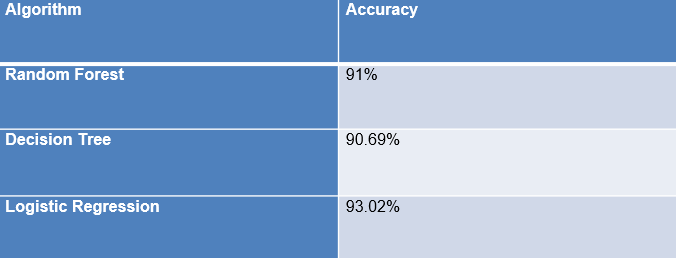
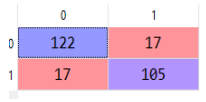


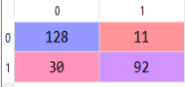
Table 6.1: Performance evaluation of different algorithms

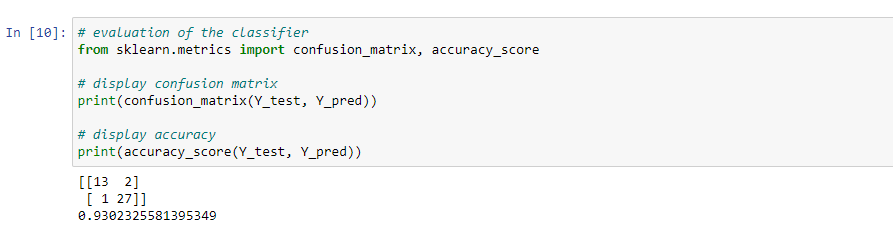
.

Confusion Matrix for Random Forest:



Confusion Matrix for Decision Tree Algo:





**6.5 Summary of prediction models**

Algorithms are applied on the data set and the attributes are used to build the model. The accuracy obtained after analysis for the decision tree is 90.69% ​​and for the random forest it is 91%, while the logistic regression has an accuracy of 93. Therefore, from the above analysis and prediction, it is best to use logistic regression algorithm to predict location outcomes.

**References**

[1] Saidani, Oumaima, et al. "Predicting Student Employability Through the Internship Context Using Gradient Boosting Models." *IEEE Access* 10 (2022): 46472-46489.

[2] Moumen, Aniss, et al. "Machine Learning for Students Employability Prediction." (2022).

[3] Piad, Keno C., et al. "Predicting IT employability using data mining techniques." *2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)*. IEEE, 2016.

[4] Othman, Zalinda, et al. "Classification techniques for predicting graduate employability." *International Journal on Advanced Science, Engineering and Information Technology* 8.4-2 (2018): 1712-1720.

[5] Jantawan, Bangsuk, and Cheng-Fa Tsai. "The application of data mining to build classification model for predicting graduate employment." *arXiv preprint arXiv:1312.7123* (2013).

[6] Premalatha, N., and S. Sujatha. "Optimizable Support Vector Machines To Predict The Students’ Employability Status." *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology* 54.5 (2022): 51-60

[7] Panityakul, Thammarat, et al. "Classification models for employability of statistics and related fields graduates from Thailand universities." *Computer Science* 17.1 (2022): 365-375.

[8] Mishra, Tripti, Dharminder Kumar, and Sangeeta Gupta. "Students’ employability prediction model through data mining." *International journal of applied engineering research* 11.4 (2016): 2275-2282.

[9] Ng, Peggy ML, Tai Ming Wut, and Jason KY Chan. "Enhancing perceived employability through work-integrated learning." *Education+ Training* ahead-of-print (2022).

[10] Ambili, P. S., and Biku Abraham. "A Predictive Model for Student Employability Using Deep Learning Techniques." *ECS Transactions* 107.1 (2022): 10149.

[11] Mohamed, Saouabi, and Ezzati Abdellah. "Data mining approach for employability prediction in Morocco." *Embedded Systems and Artificial Intelligence*. Springer, Singapore, 2020. 723-729.

[12] Repaso, Jennifer Anne A., and Elenita T. Capariño. "Analyzing and predicting career specialization using classification techniques." *International Journal* 9.1.3 (2020).

[13] Shava, Herring. "Linear Predictors of Perceived Graduate Employability among South Africa’s Rural Universities’ Learners during the Covid-19 Pandemic." *International Journal of Learning, Teaching and Educational Research* 21.3 (2022).

[14] Maaliw, Renato R., et al. "Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling." *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2022.

[15] Usita, Maricris M. "Graduates Employability Analysis using Classification Model: A Data Mining Approach." *Journal of Positive School Psychology* 6.3 (2022): 2788-2796.

[16] Srivastava, Vishal, et al. "Modelling student employability on an academic basis: A supervised machine learning approach with R." *Handbook of Research on Innovative Management Using AI in Industry 5.0*. IGI Global, 2022. 179-191.