

PROFESSIONAL TRAINING REPORT
at
Sathyabama Institute of Science and Technology
(Deemed to be University)

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology Degree in Information Technology

By

D DIYA MARY
REG. NO. 39120026



DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU

APRIL 2022



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC
(Established under Section 3 of UGC Act, 1956)
JEPPIAAR NAGAR, RAJIV GANDHI SALAI
CHENNAI- 600119
www.sathyabama.ac.in



DEPARTMENT OF INFORMATION TECHNOLOGY

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **D DIYA MARY** (Reg. No: 39120026) who carried out the project entitled "**Analytical Model For Job Placement Using Data Science Algorithm**" under my supervision from February 2022 to April 2022.

Internal Guide

Dr. Bevish Jinila. Y M.E., Ph.D

Submitted for Viva voce Examination held on _____

InternalExaminer

ExternalExaminer

DECLARATION

I, **D DIYA MARY** hereby declare that the project report entitled **Analytical Model For Job Placement Using Data Science Algorithm** done by me under the guidance of **Dr. Bevish Jinila. Y** is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology Degree in Information Technology.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean, School of Computing, Dr. R. Subhashini, M.E., Ph.D., Head of the Department of Information Technology** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. Bevish Jinila. Y, M.E., Ph.D.,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE

CERTIFICATE OF COMPLETION

This certifies that

D DIYA MARY

of Sathyabama Institute of Science and Technology

Has successfully completed Training Program on

Data Science Using R

From 23rd Feb 2022 to 6th Apr 2022



Certificate ID: VA1707312

A handwritten signature in black ink, appearing to read 'Dharani'.

Dr. Dharani Senthilkumar
CEO - Vani Analytics
<http://www.vanianalytics.com>

ABSTRACT

One of the biggest challenges that higher learning institutions face today is to improve the placement performance of students. Admission and reputation of institutions mainly depends on placements. Hence all institutions strive to strengthen placement department. All student's dream is to obtain a job offer in their hands before they leave their college. Educational institutes look for more efficient technology that assist better management and support decision making procedures or assist them to set new strategies. Placement predictor system could predict the type of company a student has chances to be placed. It helps the students to have an idea about where they stand and what to be done to obtain a good placement.

In this Project, student placement prediction model has been proposed that uses Naive Bayes algorithm to identify the placement chance of students. This model helps the placement cell within an organization to identify the prospective students and pay attention to and improve their technical as well as interpersonal skills. With this model, students can put in more hard work for getting placed in to the companies that belong to higher hierarchies.

TABLE OF CONTENTS		
CHAPTER	TITLE	PAGE NO
	ABSTRACT	06
1	INTRODUCTION	09
2	AIM AND SCOPE	10
3	MATERIALS AND METHODS	11
	3.1 MATERIALS	11
	3.1.1 HARDWARE SPECIFICATION	11
	3.1.2 SOFTWARE SPECIFICATION	11
	3.2 METHODS AND MODULES	13
	3.2.1 SYSTEM ARCHITECTURE	13
	3.2.2 MODULES	14
4	SYSTEM DESIGN	17
	4.1 DATA FLOW DIAGRAM	17
5	RESULT AND DISCUSSION	20
	4.1 RESULT	20
	4.2 DISCUSSION	21
6	CONCLUSION	22
	REFERENCES	23
	APPENDIX A: SNAP SHOTS	24
	APPENDIX B: SOURCE CODE	35

Fig No	Figure Number	Page No.
1	3.1	13
2	4.1.1	17
3	4.1.2	18
4	4.1.3	18
5	4.1.4	19
6	5.1	20
7	5.2	22
8	5.3	21

CHAPTER 1

INTRODUCTION

Majority of students in higher education join a course for securing a good job. Therefore, taking a wise career decision regarding the placement after completing a particular course is crucial in a student's life. It is a well-known fact all around the world that admission of students in an educational institution depends on the placements. Placement is one of the factors considered for determining the quality of the institution. Hence every institution strives hard to provide better placements to their students.

According to statistics 1.5 million engineers are graduating each year in India. The demand and need for qualified graduates in the field of IT industry is rising day by day. But most of the students are unaware about the need of the IT industry. The number of student graduates who satisfies the requirements and quality of the company is very low. The placement cell and professors at an institute should take proper steps to produce a set of students suitable for each company requirements.

Campus placement holds an essential position in each academic group in supporting college students to achieve their goals. Managing placement and education information in any large corporation is tough because the scholar quantity is high; in such a situation differentiation and classification on different categories will become tedious. All college students dream to get employment to provide by their hands earlier than they depart their college.

In this project, Naive Bayes algorithms is used to predict the student placement using factors, such as gender, ssc_p, ssc_b, hsc_p, hsc_b, hsc_s, degree_p, degree_t, workex, etest_p, specialisation, mba_p, status etc. A placement probability predictor allows college students to possess a thought regarding where they stand and what they can do to obtain a decent placement. The result acquired can assist the students to better understand their weak regions to work upon and get their desired placement.

CHAPTER 2

AIM AND SCOPE

Job placement for students play a significant role and helping students realize if they are eligible or require more skills to get placed will enhance their process of placement. We create a placement prediction model, which will consider certain factors of students and conclude if the student will be placed or not. Naive Bayes Algorithm is applied to the placement data, efficient prediction accuracy is achieved by delivering the result of the placement chance of students.

First, the dataset is imported, and necessary packages are installed. Then transformation of data is done in a format to understand the data in a more efficient way. Creating a naive bayes model, target variable can be predicted by giving predictor variables. This model will predict the probability of a student getting placed in a company.

The purpose of this project is to collect data, analyze and predict whether students get placed or not, making the process of placement efficient and simple.

CHAPTER 3

MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 HARDWARE SPECIFICATION

Processor : Processor Intel CORE i3 and above
Internet Connection : Existing telephone lines, Data card, Fiber net
RAM : 4GB

3.1.2 SOFTWARE SPECIFICATION

Operating System : Windows, Linux, Mac
Language : R Programming – R-4.1.3
GUI : RStudio

R is a Programming language and environment for Data Manipulation, statistical computing and graphics. Developed at Bell Laboratories by John Chambers and colleagues. Similar to the S language and available across all platforms -Linux, Mac, Windows. It provides variety of Graphical and Statistical techniques – Regression, Clustering, Classification, Association Rule Mining etc.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and Mac OS. The source code for the R software environment is written primarily in C, FORTRAN, and R.

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes

- An effective data handling and storage facility,
- A suite of operators for calculations on arrays, in particular matrices,
- A large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

RStudio is an Integrated Development Environment (IDE) for R. RStudio makes R easier to use. It includes a code editor, compiler or interpreter, debugger and visualization tools as well as tools for plotting, history, syntax-highlighting editor that supports direct code execution and workspace management. RStudio was founded by JJ Allaire, creator of the programming language ColdFusion.

RStudio is available in open source and commercial editions available in two editions: RStudio Desktop, where the program is run locally as a regular desktop application (Windows, macOS, and Linux) and RStudio Server, which allows accessing RStudio using a web browser while it is running on a remote Linux server. RStudio is written in the C++ programming language. RStudio helps to keep R more organized and it adds more functionality to it.

3.2 METHOD AND MODULES

3.2.1 SYSTEM ARCHITECTURE

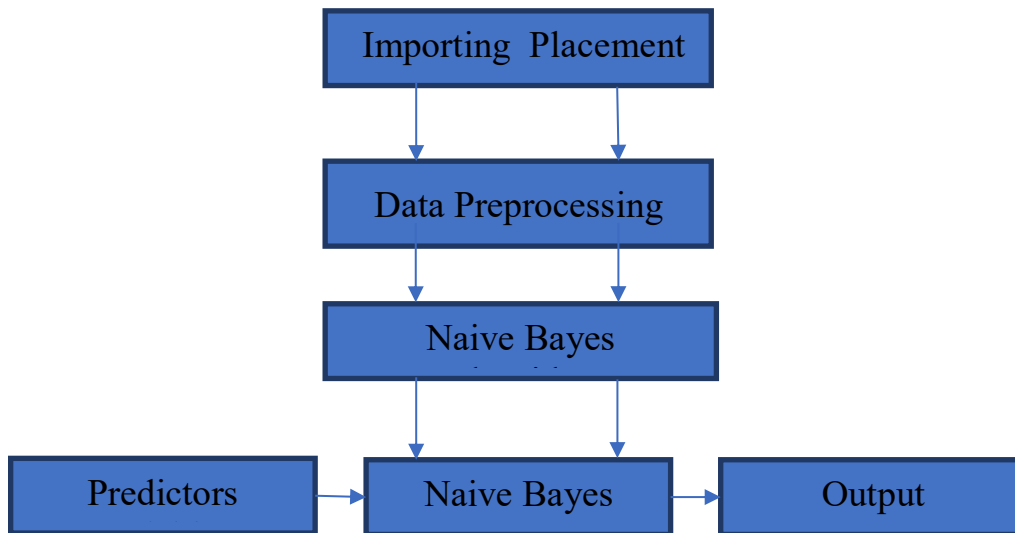


Fig.3.1 Architecture Diagram

The fig 3.1 shows the architectural diagram of this project defines the flow of data for predicting the student placement. The first step is data importing, Data have to be loaded in to the R environment for analysis. The second step is, preprocessing the collected raw data into understandable format. Standardization and Normalization is the technique which is used to transform the various format of data into the common format and min-max technique is used for normalization of data values.

It is not necessary to hold all the attributes for doing the analysis, we can hold only the attributes which is affecting the analysis. The missing values problem have to be solved by simple statistical techniques. Naive Bayes algorithm is applied on the preprocessed data to create a predictive Naive Bayes model. The algorithm is tested with some predictor variables, the Naive Bayes Model will generate the output of whether a student will get the placement in the company or not.

3.2.2 MODULES

There are four modules:

- Importing Placement dataset
- Data Preprocessing
- Model Generation Using Naive Bayes Algorithm
- Prediction using Naive Bayes Model

3.2.2.1 IMPORTING PLACEMENT DATASET

Data is available in any file format like .txt, .csv, .xlsx, .spss etc. Data have to be loaded in to R environment for analysis. Once data have been extracted from the file it should be stored in a data frame. Packages necessary for classification algorithm – Naive Bayes have to be installed into the R environment. For Naive Bayes, naivebayes package have to be installed and loaded in to R environment.

3.2.2.2 DATA PREPROCESSING

Data preprocessing is the data mining technique that involves transforming raw data into understandable format. The raw data is highly susceptible to noise, missing values, and inconsistency.

Real world data is often incomplete, inconsistent, and is likely to contain errors. The missing Values problem have to be solved by simple statistical techniques. Data preprocessing is the proven method for resolving such issues. In order to improve the quality of the data consequently, the mining results of raw data is preprocessed so the efficiency process improved. It is not necessary to hold all the attributes for doing the analysis, we can hold only the attributes which is affecting the analysis.

Data standardization is the process by which similar data is collected in various formats is transformed to a common format that enhances the comparison process, allows for collaborative research and large scale analytics. Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between

0 and 1. It is also known as Min-Max scaling.

Formula used for Normalization:

$$\frac{X - X_{min}}{X_{max} - X_{min}}$$

3.2.2.3 MODEL GENERATION USING NAIVE BAYES ALGORITHM

Classification can be performed on structured or unstructured data. Classification is a technique where we categorize data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under. Classification can be performed on structured or unstructured data.

Classification is a technique where we categorize data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under. Naive Bayes is a classification and prediction algorithm based on Bayes Theorem.

Bayes Theorem

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

$P(A|B)$ - Probability of occurrence of event A given the event B is true - Posterior Probability

$P(A)$ - Probabilities of the occurrence of event A - Class prior Probability

$P(B)$ - Probabilities of the occurrence of event B – Predictor Prior Probability

$P(B|A)$ - Probability of the occurrence of event B given the event A is true - Likelihood

Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features. Naive Bayes classifiers work well in many real-world situations such as document classification and spam filtering. This algorithm requires a small amount of training data to estimate the necessary parameters. Naive Bayes classifiers are extremely fast compared to more sophisticated methods.

Algorithmic steps for Naive bayes classification

Step 1: Convert the data set into a frequency table

Step 2: Create Likelihood table by finding the probabilities values of each attribute

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Once Naive Bayes model have been created, by using testing data, model can be evaluated and by giving only predictor variables, target variable can be predicted.

3.2.2.4 PREDICTION USING NAIVE BAYES MODEL

By using testing data model can be evaluated. The algorithm is used for prediction by passing predictors variable, the Naive Bayes Model will predict the probability of a student getting placed in a company.

CHAPTER 4 SYSTEM DESIGN

4.1 DATA FLOW DIAGRAM

Data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. It differs from the flowchart as it shows the data flow instead of the control flow of the program. A data flow diagram can also be used for the visualization of data processing. The DFD is designed to show how a system is divided into smaller portions and to highlight the flow of data between those parts.

4.1.1 LEVEL 0

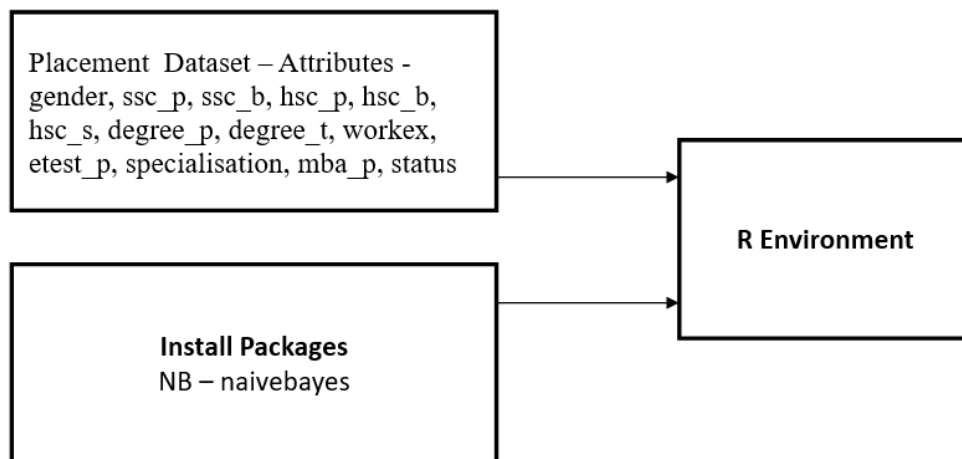


Fig 4.1.1 LEVEL 0

In the fig 4.1.1, Data is imported in to R environment for analysis. The input dataset is placement dataset, is in csv format have to be loaded into the program to start the analysis. Package necessary for Naive Bayes algorithm have to be installed and loaded into the program. Naive Bayes - naivebayes.

4.1.2 LEVEL 1

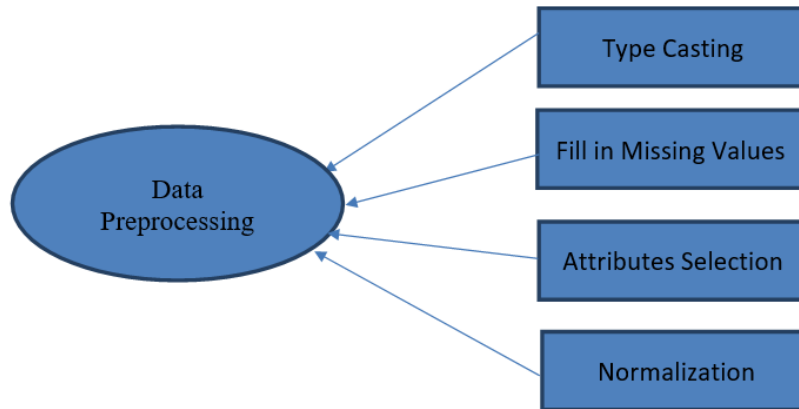


Fig 4.1.2 LEVEL 1

The fig 4.1.2 shows the data preprocessing, which includes attribute selection, standardization and normalization functions. In standardization, raw data is transformed into common, understandable format. In attribute selection, hold only the attributes which is affecting the analysis and it is not necessary to hold all the attributes for doing the analysis. In Normalization, mean of the attribute will be 0 and standard deviation will be 1.

4.1.3 LEVEL 2

In fig 4.1.3, preprocessed data is given as an input to Naive Bayes algorithm and Naive Bayes model has been created, which is used for classification and prediction.

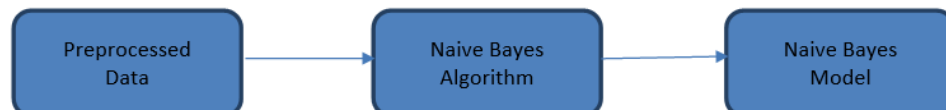


Fig 4.1.3 LEVEL 2

4.1.4 LEVEL 3

In fig 4.1.4, Predictors variables is given as an input and the Naive Bayes Model will generate the output of whether a student will get the placement in the company or not. Thus, prediction process is implemented successfully. Using this model classification is also achieved by passing more number of placement datasets, it will classify the students as two groups. One group is the set of students who will get the placement and the other group is the set of students who will not get the placement in the company.



Fig 4.1.4 LEVEL 3

CHAPTER 5

RESULT AND DISCUSSION

5.1 RESULT

After we create the naive bayes model, the result we get is if the student gets placed or not. The following are some of the plots which are generated while the naive bayes model is created.

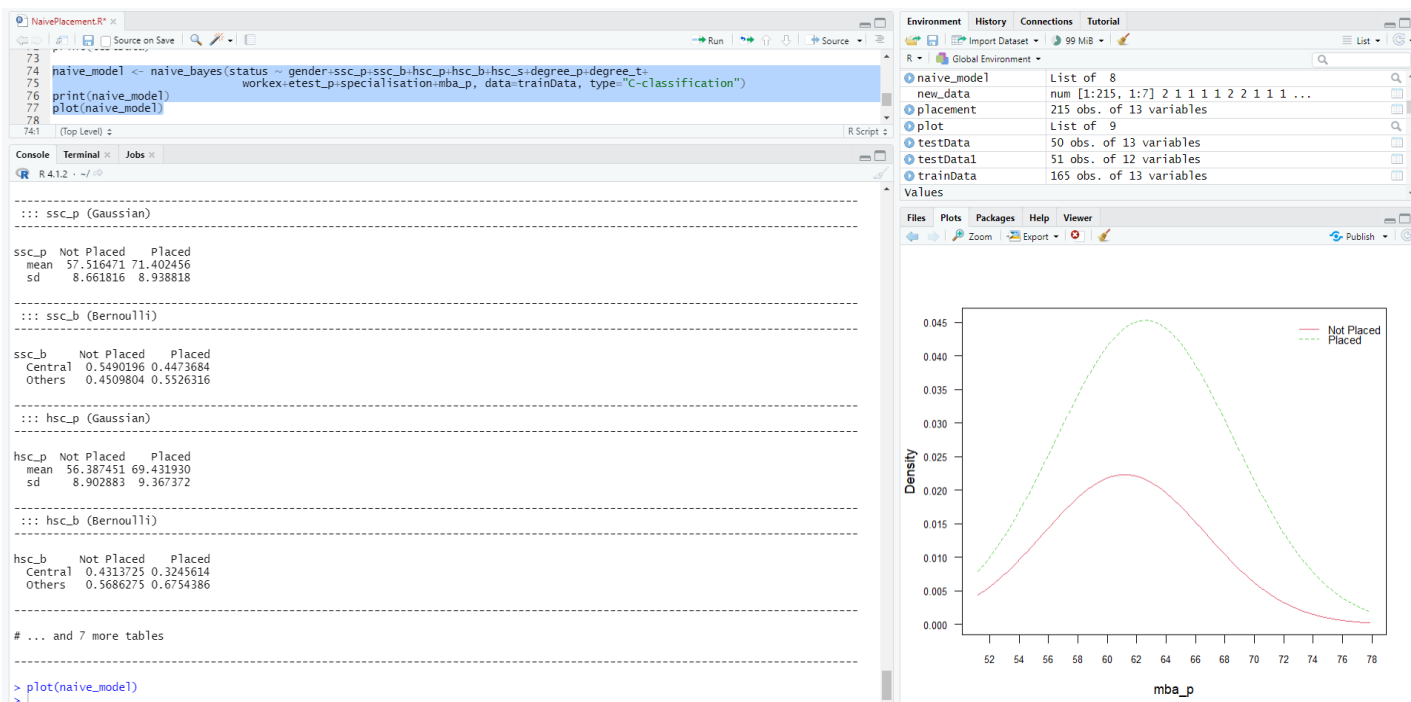


Fig 5.1 Output 1

col	chr [1:4] "Open Source" "Open Source" "Paid" "Paid"
count	5
	1 2 3 4 TRUE TRUE FALSE FALSE

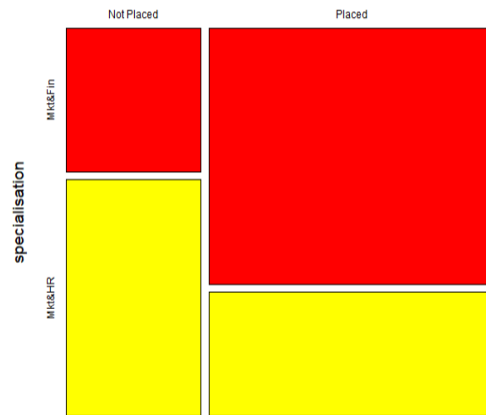


Fig 5.2 Output 2

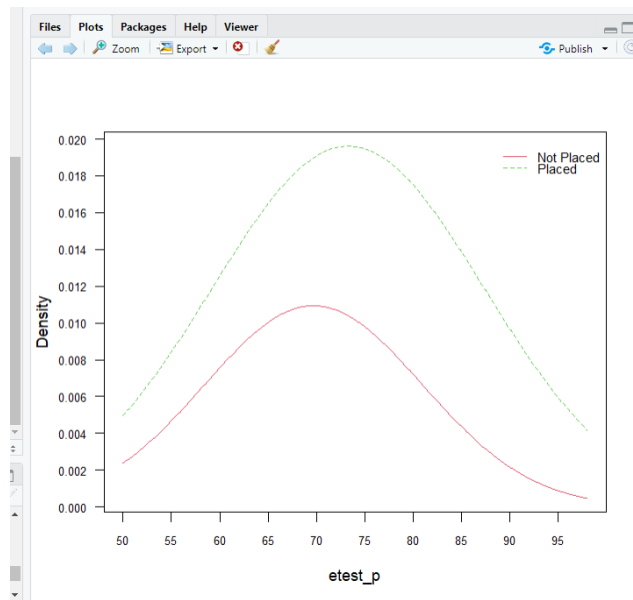


Fig 5.3 Output 3

5.2 DISCUSSION

From the result, students can know if they will be placed or not and can use this information to upskill themselves or work on certain areas.

This way, institutions can make the process of placement simple for students by helping them predict their probability of getting placed.

CHAPTER 5

CONCLUSION

This project presents a recommendation system that uses Data Science algorithm for classifying and predicting, the student placement. The Naive Bayes Data Science algorithm is used for classifying and predicting the placement of students in a company. The students in pre-final and final years of their course can also use this system to know their individual placement status that they are most likely to achieve. This application endows the targeted group of students to boost their placement probability.

Thus, this project concludes, Naive Bayes Algorithm performs better and faster when compared to other statistical techniques. This project can be further extended using some more data science algorithm and deep learning algorithms to improve the performance of prediction system.

REFERENCES

1. Tadi Aravind, "A Comparative Study on Machine Learning Algorithms for Predicting the Placement Information of Under Graduate Students", IEEE International conference on I-SMAC, 2019
2. Kanika Thakur, "Ensemble method to predict impact of student intelligent quotient and academic achievement on placement", IEEE International Conference on Intelligent Engineering and Management, 2021
3. Jinny Rhee; Sheri Sheppard, "Supporting student career development of undergraduate engineering", IEEE Frontiers in Education Conference (FIE), 2017.
4. Jaime Ramírez Castillo; Aldabbagh Ghadah, "Towards improved student placement and preparation methods on information technologies post-secondary education", IEEE Federated Conference on Computer Science and Information Systems, 2013.
5. Liya Claire Joy, "A Review on Student Placement Chance Prediction" IEEE International Conference on Advanced Computing & Communication Systems, 2019.

APPENDIX A

SNAP SHOTS

SAMPLE DATASET

gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisat	mba_p	status	salary
M	67	Others	91	Others	Commerce	58	Sci&Tech	No	55	Mkt&HR	58.8	Placed	270000
M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	200000
M	65	Central	68	Central	Arts	64	Comm&M	No	75	Mkt&Fin	57.8	Placed	250000
M	56	Central	52	Central	Science	52	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed	
M	85.8	Central	73.6	Central	Commerce	73.3	Comm&M	No	96.8	Mkt&Fin	55.5	Placed	425000
M	55	Others	49.8	Others	Science	67.25	Sci&Tech	Yes	55	Mkt&Fin	51.58	Not Placed	
F	46	Others	49.2	Others	Commerce	79	Comm&M	No	74.28	Mkt&Fin	53.29	Not Placed	
M	82	Central	64	Central	Science	66	Sci&Tech	Yes	67	Mkt&Fin	62.14	Placed	252000
M	73	Central	79	Central	Commerce	72	Comm&M	No	91.34	Mkt&Fin	61.29	Placed	231000
M	58	Central	70	Central	Commerce	61	Comm&M	No	54	Mkt&Fin	52.21	Not Placed	
M	58	Central	61	Central	Commerce	60	Comm&M	Yes	62	Mkt&HR	60.85	Placed	260000
M	69.6	Central	68.4	Central	Commerce	78.3	Comm&M	Yes	60	Mkt&Fin	63.7	Placed	250000
F	47	Central	55	Others	Science	65	Comm&M	No	62	Mkt&HR	65.04	Not Placed	
F	77	Central	87	Central	Commerce	59	Comm&M	No	68	Mkt&Fin	68.63	Placed	218000
M	62	Central	47	Central	Commerce	50	Comm&M	No	76	Mkt&HR	54.96	Not Placed	
F	65	Central	75	Central	Commerce	69	Comm&M	Yes	72	Mkt&Fin	64.66	Placed	200000
M	63	Central	66.2	Central	Commerce	65.6	Comm&M	Yes	60	Mkt&Fin	62.54	Placed	300000
F	55	Central	67	Central	Commerce	64	Comm&M	No	60	Mkt&Fin	67.28	Not Placed	
F	63	Central	66	Central	Commerce	64	Comm&M	No	68	Mkt&HR	64.08	Not Placed	
M	60	Others	67	Others	Arts	70	Comm&M	Yes	50.48	Mkt&Fin	77.89	Placed	236000
M	62	Others	65	Others	Commerce	66	Comm&M	No	50	Mkt&HR	56.7	Placed	265000
F	79	Others	76	Others	Commerce	85	Comm&M	No	95	Mkt&Fin	69.06	Placed	393000
F	69.8	Others	60.8	Others	Science	72.23	Sci&Tech	No	55.53	Mkt&HR	68.81	Placed	360000
F	77.4	Others	60	Others	Science	64.74	Sci&Tech	Yes	92	Mkt&Fin	63.62	Placed	300000
M	76.5	Others	97.7	Others	Science	78.86	Sci&Tech	No	97.4	Mkt&Fin	74.01	Placed	360000
F	52.58	Others	54.6	Central	Commerce	50.2	Comm&M	Yes	76	Mkt&Fin	65.33	Not Placed	
M	71	Others	79	Others	Commerce	66	Comm&M	Yes	94	Mkt&Fin	57.55	Placed	240000
M	63	Others	67	Others	Commerce	66	Comm&M	No	68	Mkt&HR	57.69	Placed	265000
M	76.76	Others	76.5	Others	Commerce	67.5	Comm&M	Yes	73.35	Mkt&Fin	64.15	Placed	350000
M	62	Central	67	Central	Commerce	58	Comm&M	No	77	Mkt&Fin	51.29	Not Placed	

DATA IMPORTING

```
NaivePlacement.R
1 library(naivebayes)
2
3 placement<- read.csv("C:\\Program Files\\R\\R-4.1.2\\datasets\\Placement.csv", header = TRUE)
4 print(placement)
5
6
7 dim(placement)
8 length(placement)
9 placement[6,]
10 placement[,c(10,11,12,13)]
11 placement[1:10,]
12 placement[,9]
13 str(placement)
14
```

78/22 (Top Level) R Script

Console Terminal Jobs

```
R 4.1.2 ~ /
> placement<- read.csv("C:\\Program Files\\R\\R-4.1.2\\datasets\\Placement.csv", header = TRUE)
> print(placement)
```

	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.00	Mkt&HR	58.80	Placed	270000
2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.50	Mkt&Fin	66.28	Placed	200000
3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75.00	Mkt&Fin	57.80	Placed	250000
4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66.00	Mkt&HR	59.43	Not Placed	NA
5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.80	Mkt&Fin	55.50	Placed	425000
6	M	55.00	Others	49.80	Others	Science	67.25	Sci&Tech	Yes	55.00	Mkt&Fin	51.58	Not Placed	NA
7	F	46.00	Others	49.20	Others	Commerce	79.00	Comm&Mgmt	No	74.28	Mkt&Fin	53.29	Not Placed	NA
8	M	82.00	Central	64.00	Central	Science	66.00	Sci&Tech	Yes	67.00	Mkt&Fin	62.14	Placed	252000
9	M	73.00	Central	79.00	Central	Commerce	72.00	Comm&Mgmt	No	91.34	Mkt&Fin	61.29	Placed	231000
10	M	58.00	Central	70.00	Central	Commerce	61.00	Comm&Mgmt	No	54.00	Mkt&Fin	52.21	Not Placed	NA
11	M	58.00	Central	61.00	Central	Commerce	60.00	Comm&Mgmt	Yes	62.00	Mkt&HR	60.85	Placed	260000
12	M	69.60	Central	68.40	Central	Commerce	78.30	Comm&Mgmt	Yes	60.00	Mkt&Fin	63.70	Placed	250000
13	F	47.00	Central	55.00	Others	Science	65.00	Comm&Mgmt	No	62.00	Mkt&HR	65.04	Not Placed	NA
14	F	77.00	Central	87.00	Central	Commerce	59.00	Comm&Mgmt	No	68.00	Mkt&Fin	68.63	Placed	218000
15	M	62.00	Central	47.00	Central	Commerce	50.00	Comm&Mgmt	No	76.00	Mkt&HR	54.96	Not Placed	NA
16	F	65.00	Central	75.00	Central	Commerce	69.00	Comm&Mgmt	Yes	72.00	Mkt&Fin	64.66	Placed	200000
17	M	63.00	Central	66.20	Central	Commerce	65.60	Comm&Mgmt	Yes	60.00	Mkt&Fin	62.54	Placed	300000
18	F	55.00	Central	67.00	Central	Commerce	64.00	Comm&Mgmt	No	60.00	Mkt&Fin	67.28	Not Placed	NA
19	F	63.00	Central	66.00	Central	Commerce	64.00	Comm&Mgmt	No	68.00	Mkt&HR	64.08	Not Placed	NA
20	M	60.00	Others	67.00	Others	Arts	70.00	Comm&Mgmt	Yes	50.48	Mkt&Fin	77.89	Placed	236000
21	M	62.00	Others	65.00	Others	Commerce	66.00	Comm&Mgmt	No	50.00	Mkt&HR	56.70	Placed	265000
22	F	79.00	Others	76.00	Others	Commerce	85.00	Comm&Mgmt	No	95.00	Mkt&Fin	69.06	Placed	393000
23	F	69.80	Others	60.80	Others	Science	72.23	Sci&Tech	No	55.53	Mkt&HR	68.81	Placed	360000
24	F	77.40	Others	60.00	Others	Science	64.74	Sci&Tech	Yes	92.00	Mkt&Fin	63.62	Placed	300000
25	M	76.50	Others	97.70	Others	Science	78.86	Sci&Tech	No	97.40	Mkt&Fin	74.01	Placed	360000

```
NaivePlacement.R* x
1 library(naivebayes)
2
3 placement<- read.csv("C:\\Program Files\\R\\R-4.1.2\\datasets\\Placement.csv", header = TRUE)
4 print(placement)
5
6
7 dim(placement)
8 length(placement)
9 placement[6,]
10 placement[,c(10,11,12,13)]
11 placement[1:10,]
12 placement[,9]
13 str(placement)
14
```

14:1 (Top Level) R Script

Console Terminal Jobs

R 4.1.2 · ~ /

```
>
>
> dim(placement)
[1] 215 14
> length(placement)
[1] 14
> placement[6,]
gender ssc_p ssc_b hsc_p hsc_b hsc_s degree_p degree_t workex etest_p specialisation mba_p status salary
6 M 55 Others 49.8 Others Science 67.25 Sci&Tech Yes 55 Mkt&Fin 51.58 Not Placed NA
> placement[,c(10,11,12,13)]
etest_p specialisation mba_p status
1 55.00 Mkt&HR 58.80 Placed
2 86.50 Mkt&Fin 66.28 Placed
3 75.00 Mkt&Fin 57.80 Placed
4 66.00 Mkt&HR 59.43 Not Placed
5 96.80 Mkt&Fin 55.50 Placed
6 55.00 Mkt&Fin 51.58 Not Placed
7 74.28 Mkt&Fin 53.29 Not Placed
8 67.00 Mkt&Fin 62.14 Placed
9 91.34 Mkt&Fin 61.29 Placed
10 54.00 Mkt&Fin 52.21 Not Placed
11 62.00 Mkt&HR 60.85 Placed
12 60.00 Mkt&Fin 63.70 Placed
13 62.00 Mkt&HR 65.04 Not Placed
14 68.00 Mkt&Fin 68.63 Placed
15 76.00 Mkt&HR 54.96 Not Placed
16 72.00 Mkt&Fin 64.66 Placed
17 60.00 Mkt&Fin 62.54 Placed
18 60.00 Mkt&Fin 67.28 Not Placed
19 68.00 Mkt&HR 64.08 Not Placed
20 50.48 Mkt&Fin 77.89 Placed
21 50.00 Mkt&HR 56.70 Placed
22 95.00 Mkt&Fin 69.06 Placed
23 55.53 Mkt&HR 68.81 Placed
```

DATA PREPROCESSING

```

NaivePlacement.R*
14
15 # DATA PREPROCESSING
16 summary(placement)
17 placement$gender=as.factor(placement$gender)
18 placement$ssc_b=as.factor(placement$ssc_b)
19 placement$hsc_b=as.factor(placement$hsc_b)
20 placement$hsc_s=as.factor(placement$hsc_s)
21 placement$degree_t=as.factor(placement$degree_t)
22 placement$workex=as.factor(placement$workex)
23 placement$specialisation=as.factor(placement$specialisation)
24 placement$status =as.factor(placement$status)
25 summary(placement)
26
27
14:1 (Top Level)
R Script

```

```

> # DATA PREPROCESSING
> summary(placement)
gender          ssc_p          ssc_b          hsc_p          hsc_b          hsc_s          degree_p
Length:215    Min.   :40.89    Length:215    Min.   :37.00    Length:215    Min.   :50.00
Class :character 1st Qu.:60.60    Class :character 1st Qu.:60.90    Class :character 1st Qu.:61.00
Mode  :character Median :67.00    Mode  :character Median :65.00    Mode  :character Median :66.00
                Mean  :67.30                Mean  :66.33                Mean  :66.37
                3rd Qu.:75.70                3rd Qu.:73.00                3rd Qu.:72.00
                Max.   :89.40                Max.   :97.70                Max.   :91.00

degree_t      workex      etest_p      specialisation      mba_p      status      salary
Length:215    Length:215    Min.   :50.0    Length:215    Min.   :51.21    Length:215    Min.   :200000
Class :character Class :character 1st Qu.:60.0    Class :character 1st Qu.:57.95    Class :character 1st Qu.:240000
Mode  :character Mode  :character Median :71.0    Mode  :character Median :62.00    Mode  :character Median :265000
                Mean  :72.1                Mean  :62.28                Mean  :288655
                3rd Qu.:83.5                3rd Qu.:66.25                3rd Qu.:300000
                Max.   :98.0                Max.   :77.89                Max.   :940000
                NA's   :67

> placement$gender=as.factor(placement$gender)
> placement$ssc_b=as.factor(placement$ssc_b)
> placement$hsc_b=as.factor(placement$hsc_b)
> placement$hsc_s=as.factor(placement$hsc_s)
> placement$degree_t=as.factor(placement$degree_t)
> placement$workex=as.factor(placement$workex)
> placement$specialisation=as.factor(placement$specialisation)
> placement$status =as.factor(placement$status)
> summary(placement)
gender          ssc_p          ssc_b          hsc_p          hsc_b          hsc_s          degree_p          degree_t      workex
F: 76    Min.   :40.89    Central:116    Min.   :37.00    Central: 84    Arts   : 11    Min.   :50.00    Comm&Mgmt:145    No :141
M:139    1st Qu.:60.60    Others  : 99    1st Qu.:60.90    Others :131    Commerce:113 1st Qu.:61.00    Others  : 11    Yes: 74
                Median :67.00                Median :65.00                Median :66.00
                Mean  :67.30                Mean  :66.33                Mean  :66.37
                3rd Qu.:75.70                3rd Qu.:73.00                3rd Qu.:72.00
                Max.   :89.40                Max.   :97.70                Max.   :91.00

```

```

24 placement$status =as.factor(placement$status)
25 summary(placement)
26
27
28 barplot(table(placement$status),
29         main="Student Placement", col="orange")
30
31 y<-placement$status
32 print(y)
33 x1<- placement$gender
34 barplot(table(placement$gender),
35         main="Students Gender", col="blue")
36
37 x2<- placement$ssc_p
38 hist(x2)
39 x3<- placement$ssc_b
40 barplot(table(x3))
41
42 w4<- placement$ssc_p
43 hist(w4, main = "SSC Pass Percentage", xlab="Pass Percentage", ylab="No. of Students", col = "red")
44
45 x5<- placement$hsc_b
46 barplot(table(x5))
47
48 w6<- placement$hsc_s
49 barplot(table(w6))
50
51
28:1 (Top Level)
R Script

```

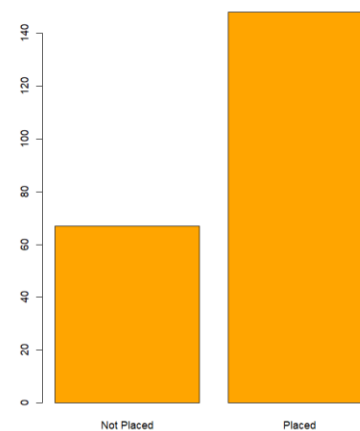
```

> summary(placement)
gender          ssc_p          ssc_b          hsc_p          hsc_b          hsc_s          degree_p          degree_t      workex
F: 76    Min.   :40.89    Central:116    Min.   :37.00    Central: 84    Arts   : 11    Min.   :50.00    Comm&Mgmt:145    No :141
M:139    1st Qu.:60.60    Others  : 99    1st Qu.:60.90    Others :131    Commerce:113 1st Qu.:61.00    Others  : 11    Yes: 74
                Median :67.00                Median :65.00                Median :66.00
                Mean  :67.30                Mean  :66.33                Mean  :66.37
                3rd Qu.:75.70                3rd Qu.:73.00                3rd Qu.:72.00
                Max.   :89.40                Max.   :97.70                Max.   :91.00

etest_p      specialisation      mba_p      status      salary
Min.   :50.0    Mkt&Fin:120    Min.   :51.21    Not Placed: 67    Min.   :200000
1st Qu.:60.0    Mkt&Fin: 95    1st Qu.:57.95    Placed  :148    1st Qu.:240000
Median :71.0                Median :62.00                Median :265000
Mean  :72.1                Mean  :62.28                Mean  :288655
3rd Qu.:83.5                3rd Qu.:66.25                3rd Qu.:300000
Max.   :98.0                Max.   :77.89                Max.   :940000
                NA's   :67

```

Student Placement



```

31 y<-placement$status
32 print(y)
33 x1<- placement$gender
34 barplot(table(placement$gender),
35         main="Students Gender", col="blue")
36
37 x2<- placement$ssc_p
38 hist(x2)
39 x3<- placement$ssc_b
40 barplot(table(x3))
41 x4<- placement$hsc_p
42 hist(x4, main = "SSC Pass Percentage",xlab="Pass Percentage", ylab="No.of Students",col = "red")
43
44
45
46
47
31:1 (Top Level)

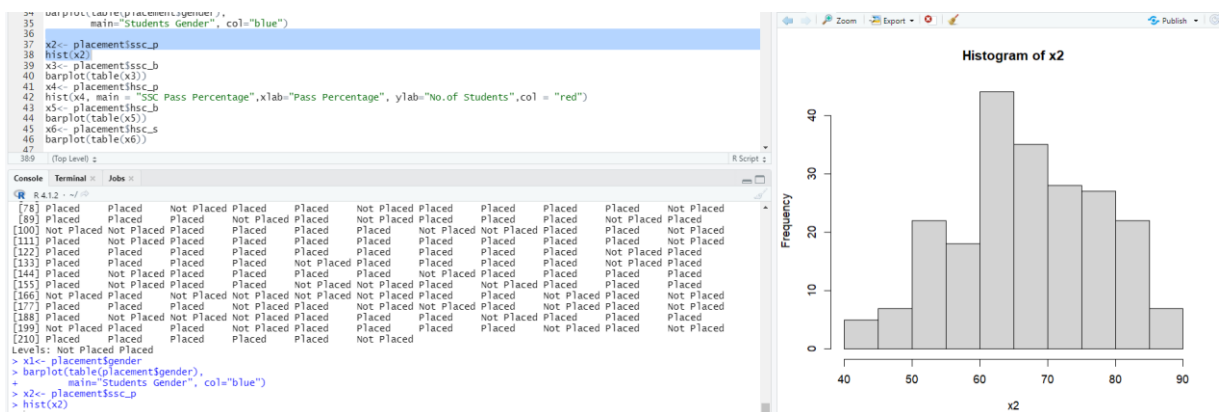
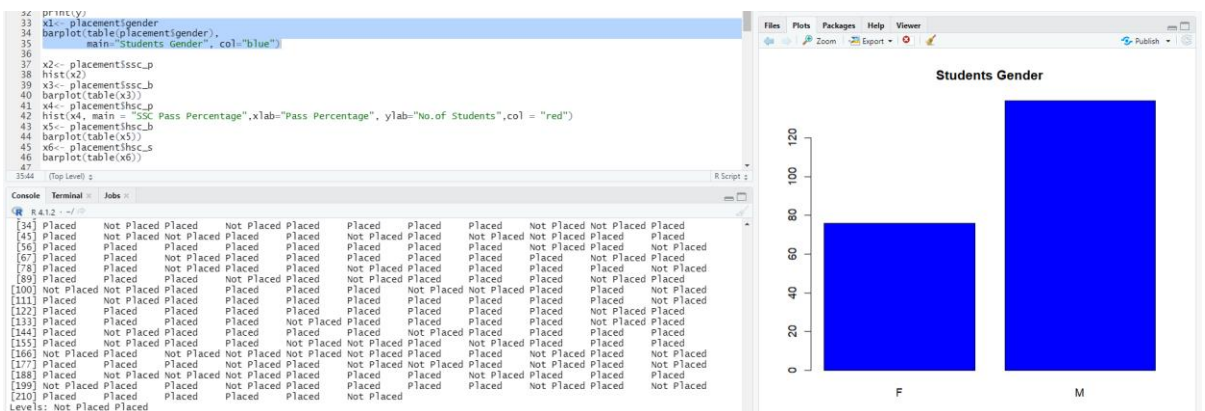
```

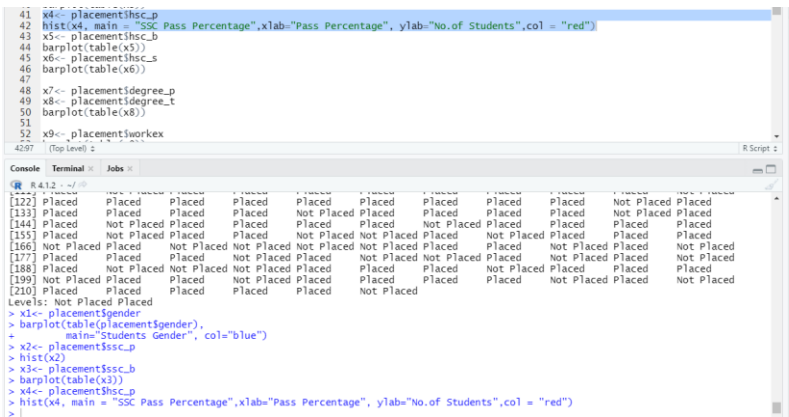
Console Terminal Jobs

```

> y<-placement$status
> print(y)
[1] Placed Placed Placed Not Placed Placed Not Placed Not Placed Placed Placed Not Placed Placed
[12] Placed Not Placed Placed Not Placed Placed Placed Not Placed Placed Placed Not Placed Placed
[23] Placed Placed Placed Not Placed Placed Placed Placed Not Placed Placed Not Placed Placed
[34] Placed Not Placed Placed Not Placed Placed Placed Placed Not Placed Placed Not Placed Placed
[45] Placed Not Placed Not Placed Placed Placed Not Placed Placed Not Placed Placed Placed Placed
[56] Placed Placed Placed Placed Placed Placed Placed Placed Placed Not Placed Placed
[67] Placed Placed Not Placed Placed Placed Placed Placed Placed Placed Not Placed Placed
[78] Placed Placed Not Placed Placed Placed Placed Not Placed Placed Placed Placed Not Placed
[89] Placed Placed Placed Not Placed Placed Placed Placed Placed Placed Not Placed Placed
[100] Not Placed Not Placed Placed Placed Placed Placed Not Placed Not Placed Placed Placed Not Placed
[111] Placed Not Placed Placed Placed Placed Placed Placed Placed Placed Placed Not Placed
[122] Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed Not Placed
[133] Placed Placed Placed Placed Not Placed Placed Placed Placed Placed Placed Not Placed
[144] Placed Not Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed
[155] Placed Not Placed Placed Placed Not Placed Not Placed Placed Not Placed Placed Placed Placed
[166] Not Placed Placed Not Placed Not Placed Not Placed Placed Placed Placed Placed Placed Not Placed
[177] Placed Placed Placed Not Placed Placed Not Placed Not Placed Placed Not Placed Placed Not Placed
[188] Placed Not Placed Not Placed Not Placed Placed Placed Placed Not Placed Placed Placed Placed
[199] Not Placed Placed Placed Not Placed Placed Placed Placed Placed Placed Placed Placed
[210] Placed Placed Placed Placed Placed Placed Not Placed Placed Placed Placed Not Placed
Levels: Not Placed Placed
>

```





```

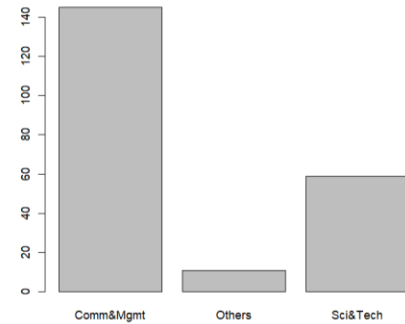
47
48 x7<- placement$degree_p
49 x8<- placement$degree_t
50 barplot(table(x8))
51
52 x9<- placement$workex
53 barplot(table(x9))
54
481 (Top Level) z
R Script z

```

```

R 4.1.2 ~ /
[199] Not Placed Placed Not Placed Placed Placed Placed Placed Placed Not Placed Placed Not Placed
[210] Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed
Levels: Not Placed Placed
> x1<- placement$gender
> barplot(table(x1),
+ main="Students Gender", col="blue")
> x2<- placement$ssc_p
> hist(x2)
> x3<- placement$ssc_b
> barplot(table(x3))
> x4<- placement$ssc_p
> hist(x4, main = "SSC Pass Percentage", xlab="Pass Percentage", ylab="No.of Students", col = "red")
> x5<- placement$ssc_b
> barplot(table(x5))
> x6<- placement$ssc_s
> barplot(table(x6))
> x7<- placement$degree_p
> x8<- placement$degree_t
> barplot(table(x8))
>

```



```

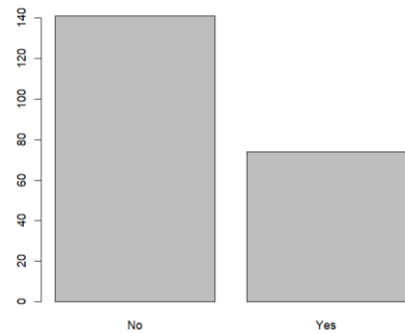
51
52 x9<- placement$workex
53 barplot(table(x9))
54
55 x10<- placement$test_p
56 x12<- placement$specialisation
57 x13<- placement$mba_p
58 x14<- placement$salary
59
60 hist(x14, main = "Students Salary", xlab="Salary Assigned", ylab="No.of Students", col = "brown")
61
551 (Top Level) z
R Script z

```

```

R 4.1.2 ~ /
Levels: Not Placed Placed
> x1<- placement$gender
> barplot(table(x1),
+ main="Students Gender", col="blue")
> x2<- placement$ssc_p
> hist(x2)
> x3<- placement$ssc_b
> barplot(table(x3))
> x4<- placement$ssc_p
> hist(x4, main = "SSC Pass Percentage", xlab="Pass Percentage", ylab="No.of Students", col = "red")
> x5<- placement$ssc_b
> barplot(table(x5))
> x6<- placement$ssc_s
> barplot(table(x6))
> x7<- placement$degree_p
> x8<- placement$degree_t
> barplot(table(x8))
> x9<- placement$workex
> barplot(table(x9))
>

```



```

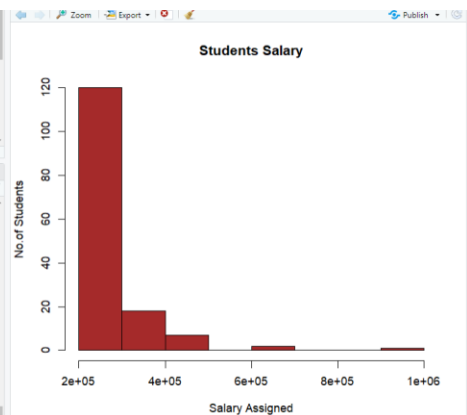
54
55 x10<- placement$test_p
56 x12<- placement$specialisation
57 x13<- placement$mba_p
58 x14<- placement$salary
59
60 hist(x14, main = "Students Salary", xlab="Salary Assigned", ylab="No.of Students", col = "brown")
61
62 new_data <- cbind(x3,x4,x5,x6,x7,x8,y)
63 print(new_data)
64 placement = placement[, -c(14)]
65
66 # Sampling of Datasets
67
551 (Top Level) z
R Script z

```

```

R 4.1.2 ~ /
> hist(x2)
> x3<- placement$ssc_b
> barplot(table(x3))
> x4<- placement$ssc_p
> hist(x4, main = "SSC Pass Percentage", xlab="Pass Percentage", ylab="No.of Students", col = "red")
> x5<- placement$ssc_b
> barplot(table(x5))
> x6<- placement$ssc_s
> barplot(table(x6))
> x7<- placement$degree_p
> x8<- placement$degree_t
> barplot(table(x8))
> x9<- placement$workex
> barplot(table(x9))
> x10<- placement$test_p
> x12<- placement$specialisation
> x13<- placement$mba_p
> x14<- placement$salary
> hist(x14, main = "Students Salary", xlab="Salary Assigned", ylab="No.of Students", col = "brown")
>

```



```

61 new_data <- cbind(x3,x4,x5,x6,x7,x8,y)
62 print(placement)
63 placement = placement[,-c(14)]
64
65 # Sampling of Datasets
66 (Top Level)

```

Console Terminal Jobs

```

R 4.1.2 ~ /
55 F 74.00 Central 60.00 Others Science 69.00 Comm&Mgmt No 78.00 Mkt&HR 65.56 Placed 216000
56 M 60.40 Central 66.60 Others Science 65.00 Comm&Mgmt No 71.00 Mkt&HR 52.71 Placed 220000
57 M 63.00 Others 71.40 Others Commerce 61.40 Comm&Mgmt No 68.00 Mkt&Fin 66.88 Placed 240000
58 M 68.00 Central 76.00 Central Commerce 74.00 Comm&Mgmt No 80.00 Mkt&Fin 63.59 Placed 360000
59 M 74.00 Central 62.00 Others Science 68.00 Comm&Mgmt No 74.00 Mkt&Fin 57.99 Placed 268000
60 M 52.60 Central 65.58 Others Science 72.11 Sci&Tech No 57.60 Mkt&Fin 56.66 Placed 265000
61 M 74.00 Central 70.00 Central Science 72.00 Comm&Mgmt Yes 60.00 Mkt&Fin 57.24 Placed 260000
62 M 84.20 Central 73.40 Central Commerce 66.89 Comm&Mgmt No 61.60 Mkt&Fin 62.48 Placed 300000
63 F 86.50 Others 64.20 Others Science 67.40 Sci&Tech No 59.00 Mkt&Fin 59.69 Placed 240000
64 M 61.00 Others 70.00 Others Commerce 64.00 Comm&Mgmt No 68.50 Mkt&HR 59.50 Not Placed NA
65 M 80.00 Others 73.00 Others Commerce 75.00 Comm&Mgmt No 61.00 Mkt&Fin 58.78 Placed 240000
66 M 54.00 Others 47.00 Others Science 57.00 Comm&Mgmt No 89.69 Mkt&HR 57.10 Not Placed NA
67 M 83.00 Others 74.00 Others Science 66.00 Comm&Mgmt No 68.92 Mkt&HR 58.46 Placed 275000
68 M 80.92 Others 78.50 Others Commerce 67.00 Comm&Mgmt No 68.71 Mkt&Fin 60.99 Placed 275000
69 F 69.70 Central 47.00 Central Commerce 72.70 Sci&Tech No 79.00 Mkt&HR 59.24 Not Placed NA
70 M 73.00 Central 73.00 Central Science 66.00 Sci&Tech Yes 70.00 Mkt&Fin 68.07 Placed 275000
71 M 82.00 Others 61.00 Others Science 62.00 Sci&Tech No 89.00 Mkt&Fin 65.45 Placed 360000
[ reached 'max' / getOption("max.print") -- omitted 144 rows ]
> placement = placement[,-c(14)]
\

```

```

65
66 # Sampling of Datasets
67 ind <- sample(2, nrow(placement), replace=TRUE, prob=c(0.8, 0.2))
68 trainData <- placement[ind==1,]
69 testData <- placement[ind==2,]
70
71 print(trainData)
72 print(testData)
73
66:1 (Top Level)

```

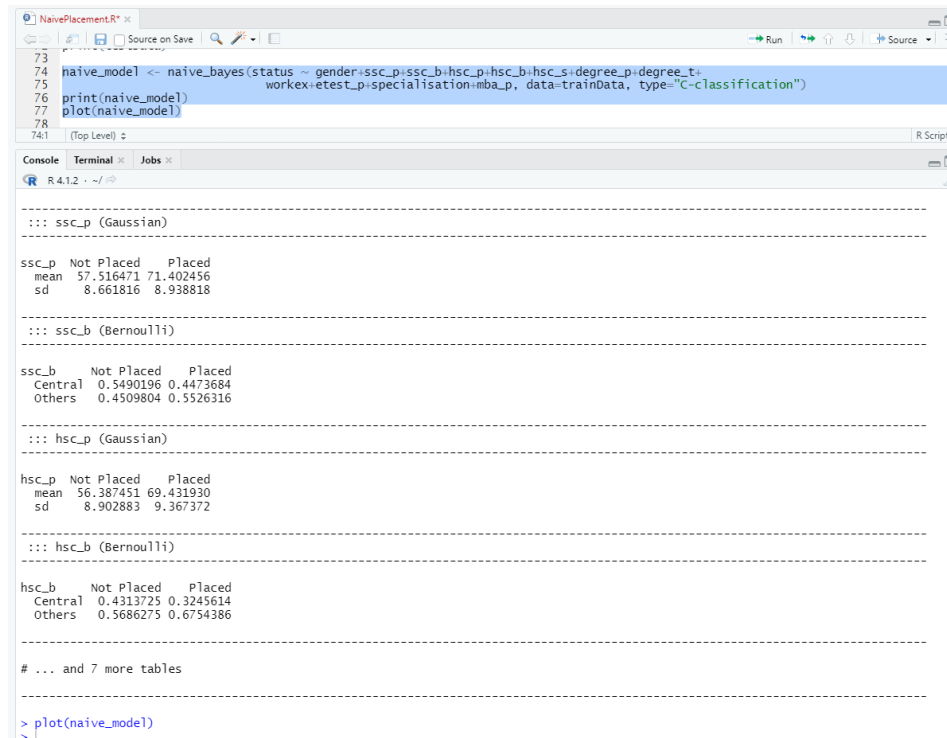
Console Terminal Jobs

```

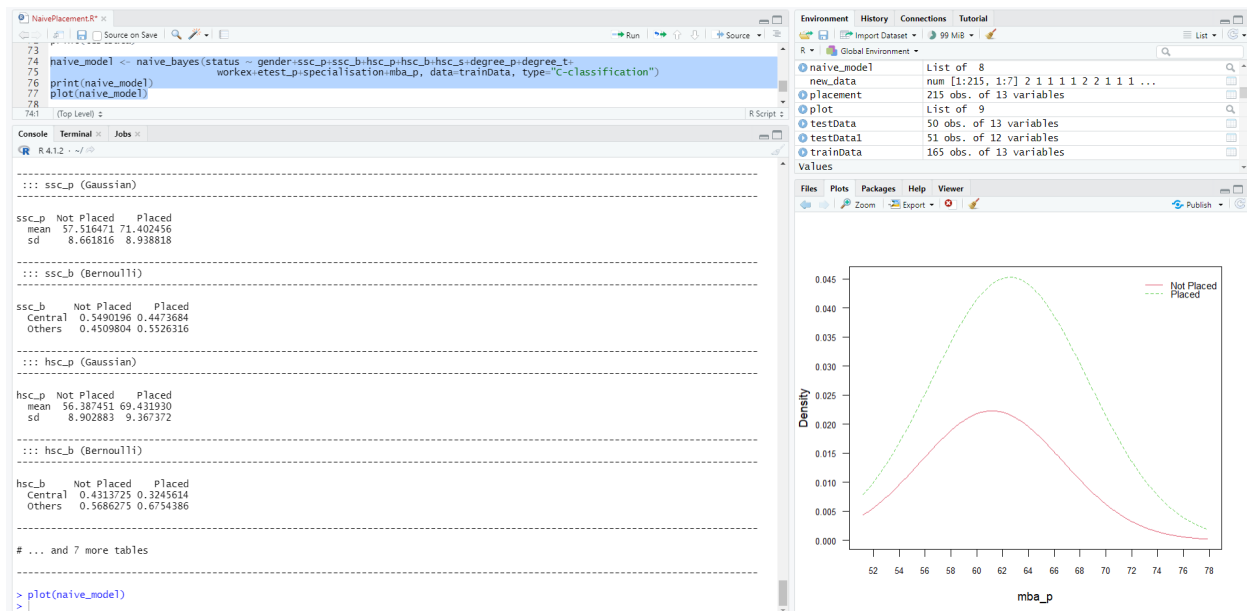
R 4.1.2 ~ /
> placement = placement[,-c(14)]
> # Sampling of Datasets
> ind <- sample(2, nrow(placement), replace=TRUE, prob=c(0.8, 0.2))
> trainData <- placement[ind==1,]
> testData <- placement[ind==2,]
> print(trainData)
  gender ssc_p ssc_b hsc_p hsc_b hsc_s degree_p degree_t workex etest_p specialisation mba_p status
1 M 67.00 Others 91.00 Others Commerce 58.00 Sci&Tech No 55.00 Mkt&HR 58.80 Placed
2 M 79.33 Central 78.33 Others Science 77.48 Sci&Tech Yes 86.50 Mkt&Fin 66.28 Placed
3 M 65.00 Central 68.00 Central Arts 64.00 Comm&Mgmt No 75.00 Mkt&Fin 57.80 Placed
4 M 56.00 Central 52.00 Central Science 52.00 Sci&Tech No 66.00 Mkt&HR 59.43 Not Placed
6 M 55.00 Others 49.80 Others Science 67.25 Sci&Tech Yes 55.00 Mkt&Fin 51.58 Not Placed
7 F 46.00 Others 49.20 Others Commerce 79.00 Comm&Mgmt No 74.28 Mkt&Fin 53.29 Not Placed
8 M 82.00 Central 64.00 Central Science 66.00 Sci&Tech Yes 67.00 Mkt&Fin 62.14 Placed
9 M 73.00 Central 79.00 Central Commerce 72.00 Comm&Mgmt No 91.34 Mkt&Fin 61.29 Placed
10 M 58.00 Central 70.00 Central Commerce 61.00 Comm&Mgmt No 54.00 Mkt&Fin 52.21 Not Placed
11 M 58.00 Central 61.00 Central Commerce 60.00 Comm&Mgmt Yes 62.00 Mkt&HR 60.85 Placed
15 M 62.00 Central 47.00 Central Commerce 50.00 Comm&Mgmt No 76.00 Mkt&HR 54.96 Not Placed
17 M 63.00 Central 66.20 Central Commerce 65.60 Comm&Mgmt Yes 60.00 Mkt&Fin 62.54 Placed
18 F 55.00 Central 67.00 Central Commerce 64.00 Comm&Mgmt No 60.00 Mkt&Fin 67.28 Not Placed
20 M 60.00 Others 67.00 Others Arts 70.00 Comm&Mgmt Yes 50.48 Mkt&Fin 77.89 Placed
21 M 62.00 Others 65.00 Others Commerce 66.00 Comm&Mgmt No 50.00 Mkt&HR 56.70 Placed
22 F 79.00 Others 76.00 Others Commerce 85.00 Comm&Mgmt No 95.00 Mkt&Fin 69.06 Placed
23 F 69.80 Others 60.80 Others Science 72.23 Sci&Tech No 55.53 Mkt&HR 68.81 Placed
24 F 77.40 Others 60.00 Others Science 64.74 Sci&Tech Yes 92.00 Mkt&Fin 63.62 Placed
25 M 76.50 Others 97.70 Others Science 78.86 Sci&Tech No 97.40 Mkt&Fin 74.01 Placed
26 F 52.58 Others 54.60 Central Commerce 50.20 Comm&Mgmt Yes 76.00 Mkt&Fin 65.33 Not Placed
27 M 71.00 Others 79.00 Others Commerce 66.00 Comm&Mgmt Yes 94.00 Mkt&Fin 57.55 Placed

```

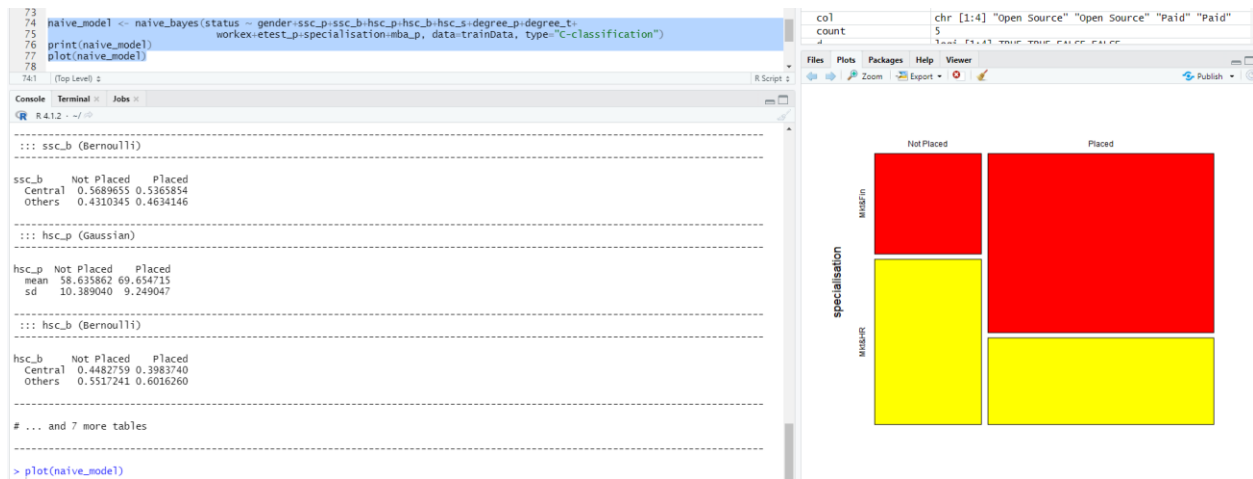
Naive Bayes Classification Model



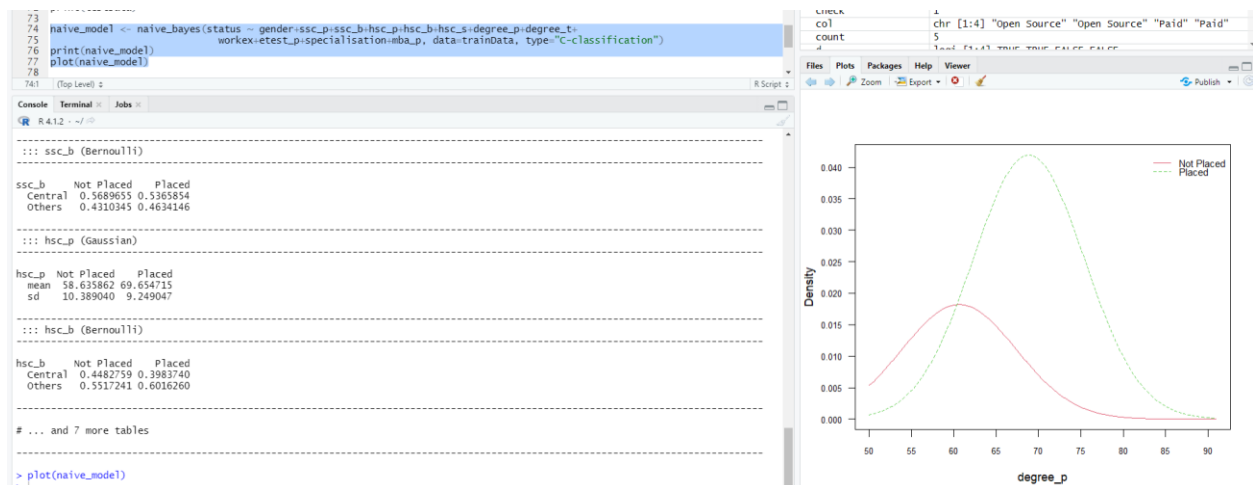
Naive Bayes Model Inference – MBA_P Vs PLACEMENT



Naive Bayes Model Inference – Placement Vs Specialization



Naive Bayes Model Inference – Placement Vs degree_p



Naive Bayes Prediction Model

```

NaivePlacement.R* x
Source on Save
Run
79 #predict on test data
80 testData1 = testData[, -c(13)]
81 testData1
82 testPred <- predict(naive_model, newdata = testData1)
83 print(testPred)
84 print(testData1$status)
85
86 a<-data.frame(gender="M",ssc_p =76.50, ssc_b = "Others", hsc_p =97.70, hsc_b = "Others", hsc_s ="Science",
87 degree_p = 78.86, degree_t ="Sci&Tech", workex= "No",etest_p= 97.40, specialisation ="Mkt&Fin",
88 mba_p =74.01)
89
90 result <- predict(naive_model,a)
91 print(result)
79:1 (Top Level)
R Script

Console Terminal Jobs
R 4.1.2 ~ /
151 M 71.00 Central 58.66 Central Science 58.00 Sci&Tech Yes 56.00 Mkt&Fin 61.30
152 M 65.00 Central 65.00 Central Commerce 75.00 Comm&Mgmt No 83.00 Mkt&Fin 58.87
155 M 53.00 Central 63.00 Others Science 60.00 Comm&Mgmt Yes 70.00 Mkt&Fin 53.20
156 M 51.57 Others 74.66 Others Commerce 59.90 Comm&Mgmt Yes 56.15 Mkt&HR 65.99
157 M 84.20 Central 69.40 Central Science 65.00 Sci&Tech Yes 80.00 Mkt&HR 52.72
161 M 87.00 Central 74.00 Central Science 65.00 Sci&Tech Yes 75.00 Mkt&HR 72.29
166 F 63.30 Central 78.33 Others Commerce 74.00 Comm&Mgmt No 80.00 Mkt&Fin 74.56
170 M 59.96 Others 42.16 Others Science 61.26 Sci&Tech No 54.48 Mkt&HR 65.48
175 M 73.24 Others 50.83 Others Science 64.27 Sci&Tech Yes 64.00 Mkt&Fin 66.23
176 M 63.00 Others 62.00 Others Science 65.00 Sci&Tech No 87.50 Mkt&HR 60.69
184 M 65.00 Central 77.00 Central Commerce 69.00 Comm&Mgmt No 60.00 Mkt&HR 61.82
185 F 56.28 Others 62.83 Others Commerce 59.79 Comm&Mgmt No 60.00 Mkt&HR 57.29
187 F 52.00 Central 64.00 Central Commerce 61.00 Comm&Mgmt No 55.00 Mkt&Fin 62.93
190 F 54.00 Central 77.60 Others Commerce 69.20 Comm&Mgmt No 95.65 Mkt&Fin 66.94
194 F 60.00 Central 63.00 Central Arts 56.00 Others Yes 80.00 Mkt&HR 56.63
196 M 66.00 Central 76.00 Central Commerce 72.00 Comm&Mgmt Yes 84.00 Mkt&HR 58.95
199 F 67.00 Central 70.00 Central Commerce 65.00 Others No 88.00 Mkt&HR 71.96
203 M 70.00 Central 63.00 Central Science 66.00 Sci&Tech No 61.28 Mkt&HR 60.11
> testPred <- predict(naive_model, newdata = testData1)
> print(testPred)
[1] Placed Placed Not Placed Placed Placed Not Placed Placed Not Placed Placed Placed Placed
[12] Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed Placed
[23] Not Placed Placed Placed Placed Placed Placed Not Placed Placed Placed Not Placed Placed
[34] Placed Not Placed Placed Placed Placed Placed Not Placed Placed Placed Placed Placed
[45] Not Placed Placed Not Placed Placed Not Placed Placed
Levels: Not Placed Placed
> print(testData1$status)
[1] Placed Placed Not Placed Placed Placed Not Placed Placed Not Placed Placed Placed Not Placed
[12] Placed Placed Placed Placed Placed Placed Not Placed Not Placed Placed Placed Placed
[23] Placed Placed Placed Placed Placed Placed Not Placed Placed Placed Not Placed Placed
[34] Placed Placed Not Placed Placed Placed Placed Not Placed Not Placed Placed Placed Placed
[45] Not Placed Not Placed Placed Placed Not Placed Placed
Levels: Not Placed Placed
>

85
86 a<-data.frame(gender="M",ssc_p =76.50, ssc_b = "Others", hsc_p =97.70, hsc_b = "Others", hsc_s ="Science",
87 degree_p = 78.86, degree_t ="Sci&Tech", workex= "No",etest_p= 97.40, specialisation ="Mkt&Fin",
88 mba_p =74.01)
89
90 result <- predict(naive_model,a)
91 print(result)
86:1 (Top Level)
R Script

Console Terminal Jobs
R 4.1.2 ~ /
> a<-data.frame(gender="M",ssc_p =76.50, ssc_b = "Others", hsc_p =97.70, hsc_b = "Others", hsc_s ="Science",
+ degree_p = 78.86, degree_t ="Sci&Tech", workex= "No",etest_p= 97.40, specialisation ="Mkt&Fin",
+ mba_p =74.01)
>
> result <- predict(naive_model,a)
> print(result)
[1] Placed
Levels: Not Placed Placed
>

```

APPENDIX B

SOURCE CODE

```
library(naivebayes)
```

```
placement<- read.csv("C:\\Program Files\\R\\R-4.1.2\\datasets\\Placement.csv",  
header = TRUE)  
print(placement)
```

```
dim(placement)  
length(placement)  
placement[6,]  
placement[,c(10,11,12,13)]  
placement[1:10,]  
placement[,9]  
str(placement)
```

```
# DATA PREPROCESSING
```

```
summary(placement)  
placement$gender=as.factor(placement$gender)  
placement$ssc_b=as.factor(placement$ssc_b)  
placement$hsc_b=as.factor(placement$hsc_b)  
placement$hsc_s=as.factor(placement$hsc_s)  
placement$degree_t=as.factor(placement$degree_t)  
placement$workex=as.factor(placement$workex)  
placement$specialisation=as.factor(placement$specialisation)  
placement$status =as.factor(placement$status)  
summary(placement)
```

```
barplot(table(placement$status),  
        main="Student Placement", col="orange")
```

```
y<-placement$status  
print(y)  
x1<- placement$gender  
barplot(table(placement$gender),  
        main="Students Gender", col="blue")
```

```
x2<- placement$ssc_p  
hist(x2)  
x3<- placement$ssc_b  
barplot(table(x3))  
x4<- placement$hsc_p  
hist(x4, main = "SSC Pass Percentage",xlab="Pass Percentage", ylab="No.of  
Students",col = "red")  
x5<- placement$hsc_b  
barplot(table(x5))  
x6<- placement$hsc_s  
barplot(table(x6))
```

```
x7<- placement$degree_p  
x8<- placement$degree_t  
barplot(table(x8))
```

```
x9<- placement$workex  
barplot(table(x9))
```

```
x10<- placement$etest_p  
x12<- placement$specialisation  
x13<- placement$mba_p
```

```
x14<- placement$salary
```

```
hist(x14, main = "Students Salary",xlab="Salary Assigned", ylab="No.of  
Students",col = "brown")
```

```
new_data <- cbind(x3,x4,x5,x6,x7,x8,y)  
print(placement)  
placement = placement[,-c(14)]
```

```
# Sampling of Datasets
```

```
ind <- sample(2, nrow(placement), replace=TRUE, prob=c(0.8, 0.2))  
trainData <- placement[ind==1,]  
testData <- placement[ind==2,]
```

```
print(trainData)  
print(testData)
```

```
naive_model <- naive_bayes(status ~  
gender+ssc_p+ssc_b+hsc_p+hsc_b+hsc_s+degree_p+degree_t+  
workex+etest_p+specialisation+mba_p, data=trainData,  
type="C-classification")  
print(naive_model)  
plot(naive_model)
```

```
#predict on test data
```

```
testData1 = testData[,-c(13)]  
testData1  
testPred <- predict(naive_model, newdata = testData1)  
print(testPred)  
print(testData$status)
```

```
a<-data.frame(gender= "M",ssc_p =76.50, ssc_b = "Others", hsc_p =97.70, hsc_b
= "Others", hsc_s ="Science",
              degree_p = 78.86, degree_t ="Sci&Tech", workex= "No",etest_p= 97.40,
specialisation ="Mkt&Fin",
              mba_p =74.01)
```

```
result <- predict(naive_model,a)
print(result)
```