Bansilal Ramnath Agarwal Charitable Trust's
VISHWAKARMA INSTITUTE OF TECHNOLOGY – PUNE
Department of Multidisciplinary Engineering

MD2201 Data Science
Course Project

S.No	Div	Batch No	Group No	Roll No	Gr.No	Name of Student
1	CS A	3	4	63	12111190	Siddhi Deshmukh
2				64	12111289	Atharva Deshpande
3				65	12111410	Samarth Deshpande
4				66	12110163	Devyani Manmode
5				67	12110721	Sakshi Dhamne

1. **Project Title:** Campus Recruitment Analysis

2. **Data Set Name:** Campus Recruitment

3. **Data set Link:** https://www.kaggle.com/datasets/niki188/campus-recruitment

4. **Objective:** The objective of using Machine Learning is to discover patterns in the chosen Dataset " Campus Recruitment" and make the prediction based on these intricate patterns for answering the placed and Not placed applicants analysis. with the help of suitable machine learning algorithms analyzing the data as well as identifying trends and status of the applicant.

5. **Project Description:** Machine learning uses programmed algorithms that receive and analyze input data to predict output values within an acceptable range. As new data is fed to these algorithms, they learn and optimize their operations to improve performance, developing 'intelligence' over time. The dataset chosen for the analysis is " Campus Recruitment ". The algorithms like Logistic regression , Random Forest are used for Prediction of

'placed' and 'Not placed' Applicants. Logistic regression is commonly used for prediction and classification problems Logistic regression is a classification algorithm. It is used to predict a binary outcome based on a set of independent variables. A binary outcome is one where there are only two possible scenarios—either the event happens (1) or it does not happen (0). Here for this dataset Placed and Not placed is considered. Independent variables are those variables or factors which may influence the outcome (or dependent variable).

Second Algorithm implemented on the dataset is "Random Forest", It is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."

In Addition the next Algorithm used is SVM it is a support vector machines are a set of supervised learning methods used for classification, regression, and outliers detection, SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes. The decision boundary created by SVMs is called the maximum margin classifier or the maximum margin hyper plane.

6. Code:

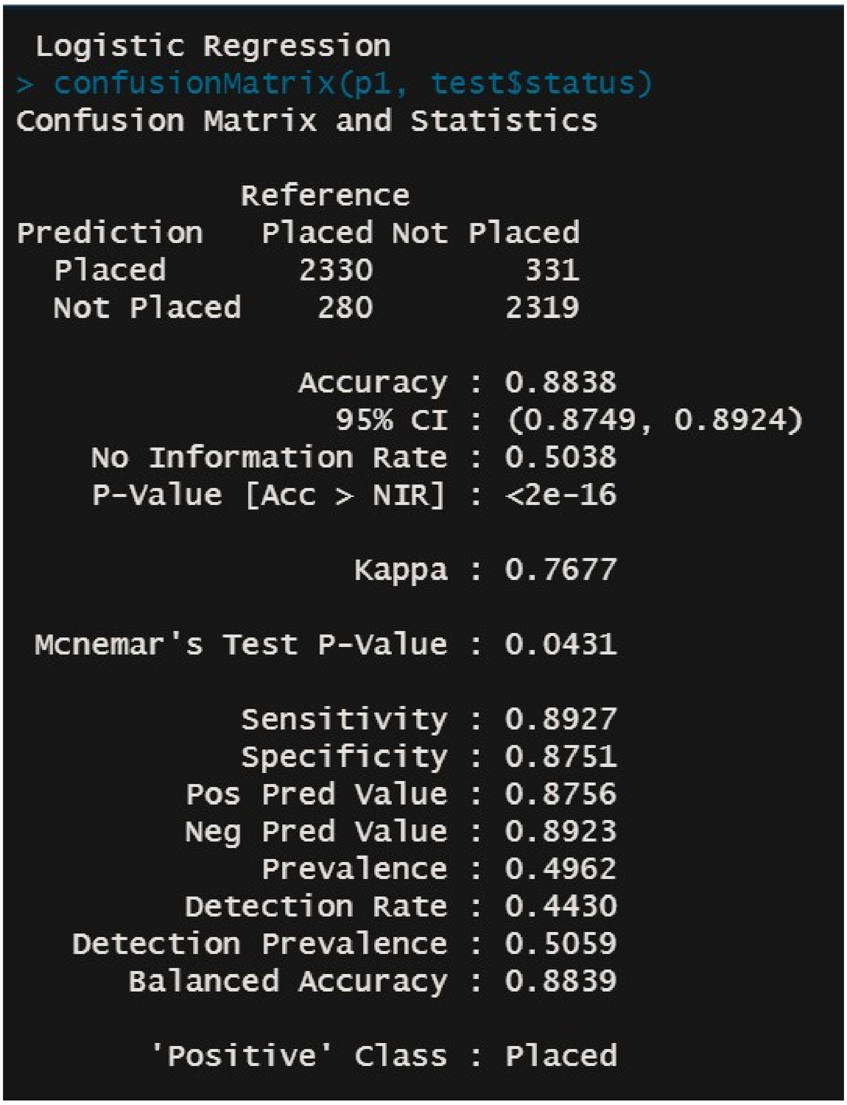
```
f<- read.csv("Placement_Data_Full_Class.csv")

f <- f[,-15]

f <- f[,-1]
```

```
f$gender = as.factor(f$gender)

f$ssc_b = as.factor(f$ssc_b)

f$hsc_s = as.factor(f$hsc_s)

f$hsc_b = as.factor(f$hsc_b)

f$degree_t = as.factor(f$degree_t)

f$workex = as.factor(f$workex)

f$specialisation = as.factor(f$specialisation)

f$status = as.factor(f$status)

library("caret")

library("ROSE")

f <- ovun.sample(status~., data = f, method = "both", p = 0.5, seed = 222)$data

set.seed(123)

a <- sample(2, nrow(f), replace = TRUE, prob = c(0.75, 0.25))

traning <- f[a==1,]

test <- f[a==2,]

#Logistic Regression----

cat("\n Logistic Regression \n")

library(nnet)

lr <- multinom(status~., data = traning)

p1 <- predict(lr, test)

confusionMatrix(p1, test$status)
```

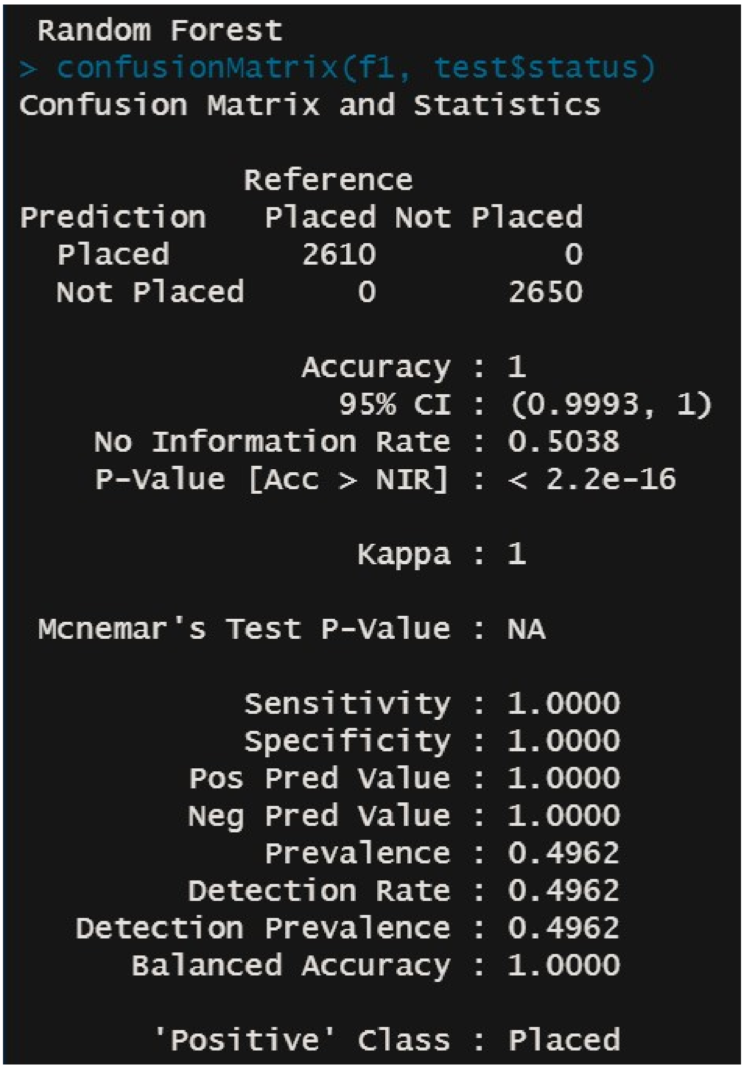
```
#Random Forest----
cat("\n Random Forest \n")
library(randomForest)
rf <- randomForest(status~., data = traning)
f1 <- predict(rf, test)
confusionMatrix(f1, test$status)
#Support Vector Machine----
cat("\n Support Vector Machine \n")
library("e1071")
svm <- svm(status~.,data = traning)
k1 <- predict(svm, test)
confusionMatrix(k1, test$status)
```

7. Results: Quantitative findings and Plots.

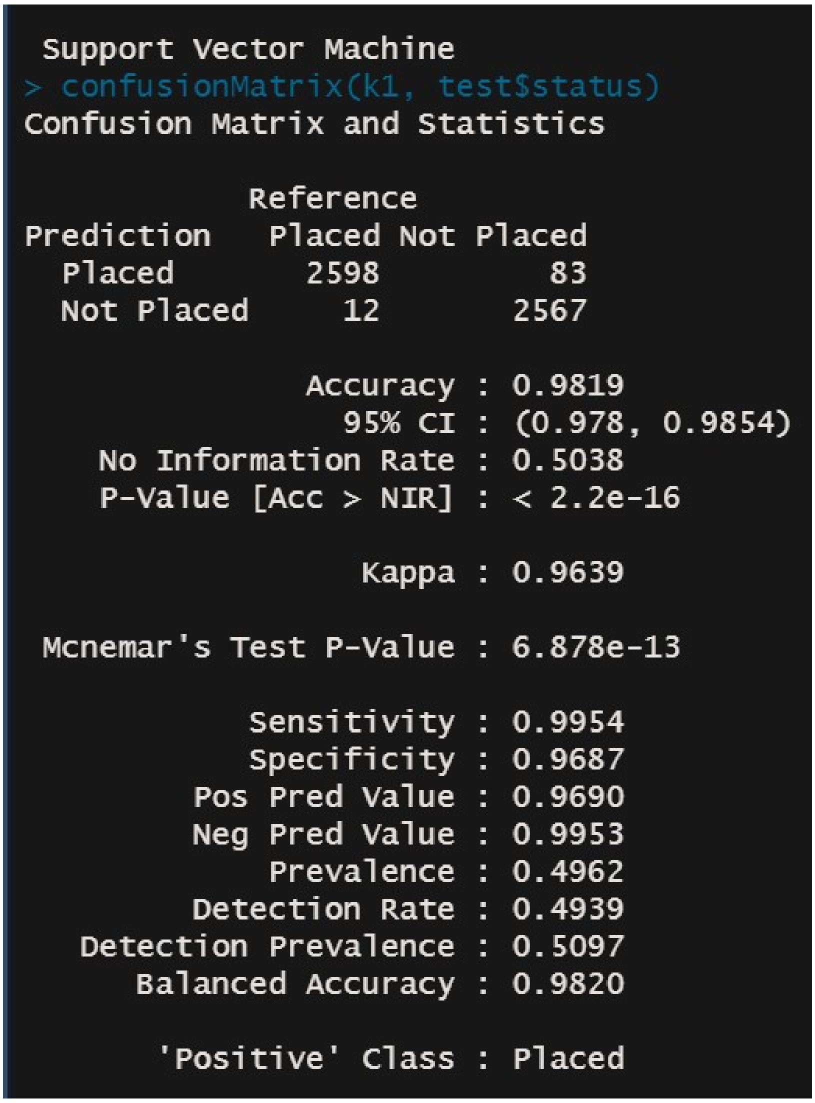
Confusion Matrices for Logistic Regression, Random Forest & SVM



```
 Logistic Regression
> confusionMatrix(p1, test$status)
Confusion Matrix and Statistics

            Reference
Prediction    Placed Not Placed
  Placed        2330        331
  Not Placed     280       2319

               Accuracy : 0.8838
                 95% CI : (0.8749, 0.8924)
    No Information Rate : 0.5038
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7677

 Mcnemar's Test P-Value : 0.0431

            Sensitivity : 0.8927
            Specificity : 0.8751
         Pos Pred Value : 0.8756
         Neg Pred Value : 0.8923
             Prevalence : 0.4962
         Detection Rate : 0.4430
   Detection Prevalence : 0.5059
      Balanced Accuracy : 0.8839

       'Positive' Class : Placed
```

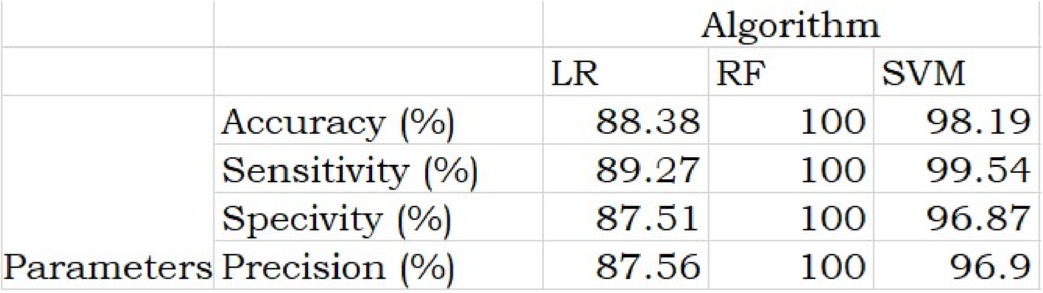


```
 Random Forest
> confusionMatrix(f1, test$status)
Confusion Matrix and Statistics

            Reference
Prediction   Placed Not Placed
  Placed       2610          0
  Not Placed      0       2650
                                          
               Accuracy : 1               
                 95% CI : (0.9993, 1)     
    No Information Rate : 0.5038          
    P-Value [Acc > NIR] : < 2.2e-16       
                                          
                  Kappa : 1               
                                          
 Mcnemar's Test P-Value : NA              
                                          
            Sensitivity : 1.0000          
            Specificity : 1.0000          
         Pos Pred Value : 1.0000          
         Neg Pred Value : 1.0000          
             Prevalence : 0.4962          
         Detection Rate : 0.4962          
   Detection Prevalence : 0.4962          
      Balanced Accuracy : 1.0000          
                                          
       'Positive' Class : Placed          
```

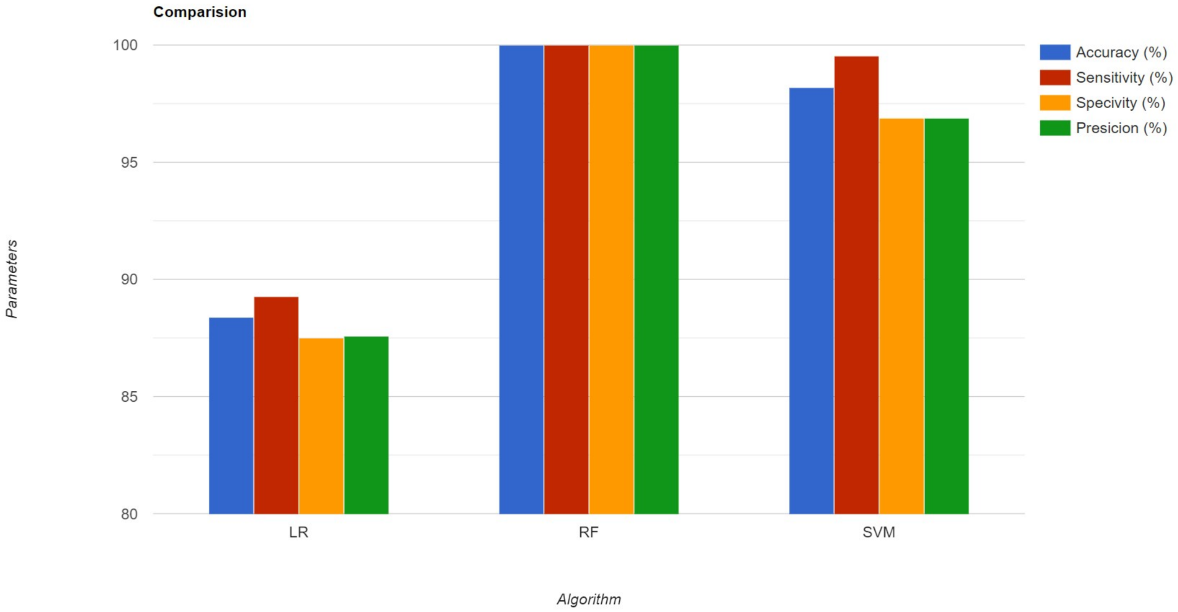


```
 Support Vector Machine
> confusionMatrix(k1, test$status)
Confusion Matrix and Statistics

            Reference
Prediction   Placed Not Placed
  Placed       2598         83
  Not Placed     12       2567

               Accuracy : 0.9819
                 95% CI : (0.978, 0.9854)
    No Information Rate : 0.5038
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9639

 Mcnemar's Test P-Value : 6.878e-13

            Sensitivity : 0.9954
            Specificity : 0.9687
         Pos Pred Value : 0.9690
         Neg Pred Value : 0.9953
             Prevalence : 0.4962
         Detection Rate : 0.4939
   Detection Prevalence : 0.5097
      Balanced Accuracy : 0.9820

       'Positive' Class : Placed
```

Bar Graph



		Algorithm		
		LR	RF	SVM
	Accuracy (%)	88.38	100	98.19
	Sensitivity (%)	89.27	100	99.54
	Specivity (%)	87.51	100	96.87
Parameters	Precision (%)	87.56	100	96.9



8. Conclusions:

Machine learning is a subset of AI that leverages algorithms to analyze vast amounts of data. The algorithms like Logistic Regression, Random Forest, and SVM are used to make a prediction from “placement Analysis” data. The most suitable algorithm found is Random forest then SVM and lastly Logistic regression on the basis of parameters like Accuracy, sensitivity, precision, specivity.