Final Project _Group7

12/15/2021

Analysis of Individual and Interactive Effects of select decision variables in Foreign Direct Investment (FDI)

1. Objectives

The overall objective is to assess the role of foreign investors' inherent risk choices in investment services that could be explained by individuals' specific decision factors and how their interactive effects influence and predictive capacity determine FDI inflow.

Problem Statement

Understanding the complexity of FDI select factors by ascertaining their individual and collective effects and using them to make predictions for FDI decisions.

2. Introduction

A foreign direct investment (FDI) is the purchase of a stake in a firm by a corporation or investor based outside of the country's boundaries. In general, the word refers to a commercial decision to buy a significant stake in or buy a foreign company completely in order to expand its activities to a new territory. It's not commonly used to refer to a stock purchase in a foreign company. A company's considerable investment in a foreign enterprise is known as foreign direct investment (FDI).

An investment promotion agency (IPA) is most often a government agency whose mission is attracting the investments from a country, state, region or a city.

Thirty overseas investors, government Ministries, Departments, and Agencies (MDAs), and Nigerian investors in the diaspora were among the study's participants. On the supply side, the article reveals that numerous major FDI perception determinants influence decision-making. There are four attractors and four repellents on the demand side that influence potential investors' judgments.

These findings, in general, demonstrate far more sophisticated and complex dynamics of FDI inflows, as well as how individual investor attitudes influence them. The findings, in particular, provide more information on an investor's risk assessments in a given country. This is determined by the investor's overall perceptions of the country's economic, political, and social environment, as well as their time horizon and risk preferences, indicating considerable individual and situational variances in how decision-makers approach FDI.

3. **Approach**

Overview of packages and dataset

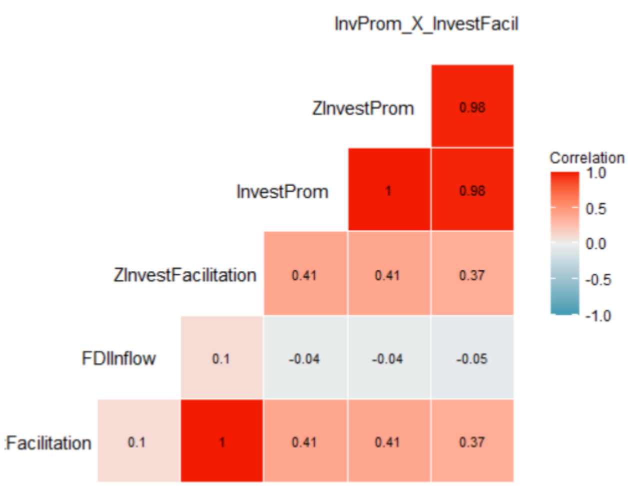
- We used several relevant R packages in analyzing the dataset, most notable amongst them are “DPLYR, Class, lattice, ISLR, “esquisse” and proc”. The packages are uploaded onto R from their various libraries. A total of …packages were required to run the codes.

```
FinalProjectx<-read.csv("GroupsevenQMM.csv")
str(FinalProjectx)
head(FinalProjectx)
```

- The original dataset was a dataset gathered from a real online survey on Qualtrics from 250 foreign investors. The selected variables data were extracted from SPSS onto R in the form of excel.csv.
- There are no missing values in the dataset, hence, there was no need for us to transform them into numeric or impute the data points at any points. At the end of it, we have five columns in the data frame which include the vector (Foreign Direct Investment inflow) and four predictors all mean to ascertain their effect on vectors for investment decisions.

Correlation

We plotted a correlation table among the various variables. There seems to be a considerable positive correlation between the FDI inflow decision and Z investment facilitation and between Investment Promotion and InvProm_X_InvestFail variables. Remarkably, the FDI Inflow is also negatively correlated with investment promotion services, Z investment promotion, and InvProm_X_InvestFail variables. Most pairwise correlations between predictors are generally low. (Santos, 2021)



Data normalization

The dataset was normalized using the method “range” for an efficient outcome

```
 ï..InvestFacilitation   FDIInflow      ZInvestFacilitation   InvestProm      ZInvestProm    
InvProm_X_InvestFacil
 Min.   :0.0000        Min.   :0.0000   Min.   :0.0000       Min.   :0.0000   Min.   
:0.0000   Min.   :0.0000  
 1st Qu.:0.6500        1st Qu.:0.2500   1st Qu.:0.6500       1st Qu.:0.5044   1st 
Qu.:0.5048   1st Qu.:0.4970  
 Median :0.7000        Median :0.5000   Median :0.7000       Median :0.7257   Median 
:0.7255   Median :0.6852
```

Approach: Analyses and Outcomes

1. Linear Regression of individual and interacted variables
2. Use of Naive Bayes for FDI prediction
3. Use of Data Envelopment Analysis (DEA) for FDI decision

Linear Regression

- We created a linear model for all the variables against the FDI Inflow and displayed a plot of the points using a Linear Model code “lm” and subsequently summarized the outcomes.
- *Investment Promotion:* We creates a linear model for Investment Promotion vs FDI Inflow and displays a plot of the points

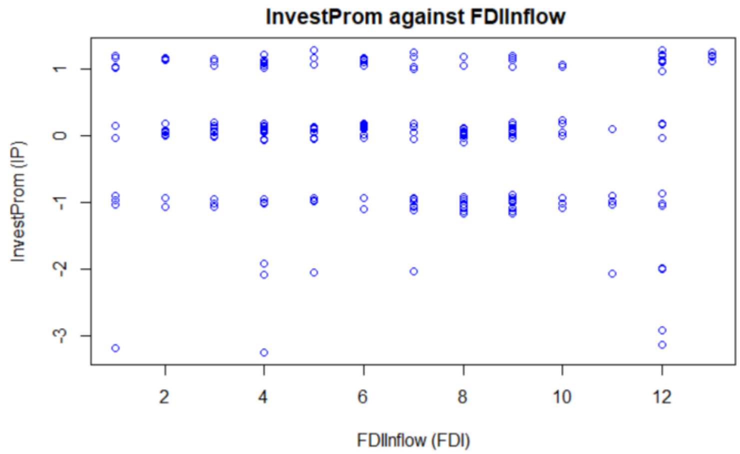
```
Modela = lm(FDIInflow ~  FinalProjectx$InvestProm, data = FinalProjectx)
summary(Modela)
plot(FinalProjectx$FDIInflow, FinalProjectx$InvestProm, xlab = "FDIInflow (FDI)", ylab = 
"InvestProm (IP)", main = " InvestProm against FDIInflow", col = "blue")
```

Result

```
Call:
lm(formula = FDIInflow ~ FinalProjectx$InvestProm, data = FinalProjectx)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.0378 -2.6358  0.2294  2.2412  6.5002 

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                6.6508     0.1781  37.338   <2e-16 ***
FinalProjectx$InvestProm  -0.1217     0.1894  -0.643    0.521    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.102 on 310 degrees of freedom
Multiple R-squared:  0.00133,	Adjusted R-squared:  -0.001891 
F-statistic: 0.413 on 1 and 310 DF,  p-value: 0.5209
```

- From this linear model, we can see that Investment Promotion Services results in a model where the P-Value is insignificant (0.52), and that it accounts for 0.13% of the variation in Foreign Direct Investment Inflow.
- The Multiple R-squared is 0.13% on 310 degrees of freedom because we used very few variables from the original dataset

We plotted only investment promotion services being one of the select variables.



❖ Investment Facilitation: We creates a linear model for Investment Facilitation vs FDI Inflow and displays a plot of the points

```
Modelb = lm(FDIInflow ~ FinalProjectx$ï..InvestFacilitation, data = FinalProjectx)
summary(Modelb)
plot(FinalProjectx$FDIInflow, FinalProjectx$InvestFacilitationx, xlab = "FDIInflow (FDI)",
ylab = "InvestFacilitation (IF)", main = " InvestFacilitation against FDIInflow", col =
"blue")
```

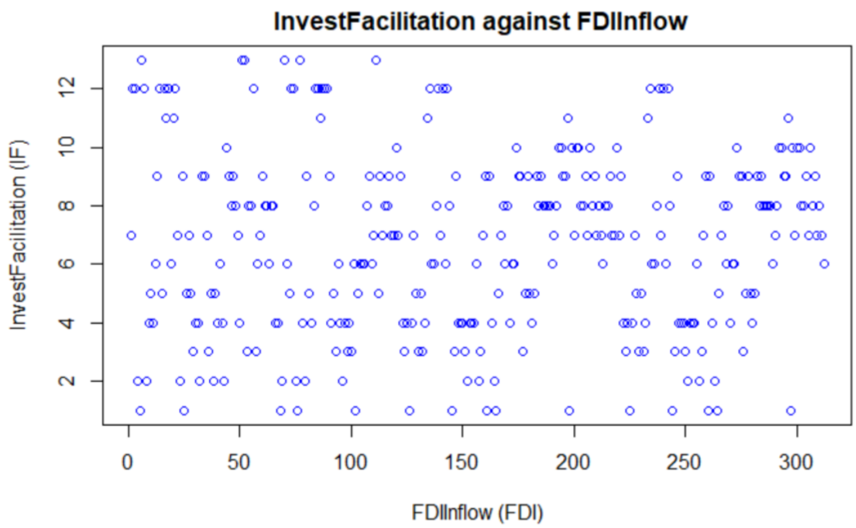
Result

```
Call:
lm(formula = FDIInflow ~ FinalProjectx$ï..InvestFacilitation,
    data = FinalProjectx)

Residuals:
    Min      1Q  Median      3Q     Max
-6.3836 -2.4969  0.3764  2.2497  6.8831

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                           4.2170     1.3448   3.136  0.00188 **
FinalProjectx$ï..InvestFacilitation   0.6333     0.3443   1.840  0.06680 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.087 on 310 degrees of freedom
Multiple R-squared:  0.0108,	Adjusted R-squared:  0.007607
F-statistic: 3.384 on 1 and 310 DF,  p-value: 0.0668
```

❖ From this linear model, we can see that Investment Facilitation Services results in a model where the P-Value is almost significant (0.067) and that it accounts for 1.08% of the variation in Foreign Direct Investment Inflow.

❖ The Multiple R-squared is 0.108% on 310 degrees of freedom because we used very few variables from the original dataset

We plotted only investment facilitation services being one of the select variables.



- *Investment Facilitation and Investment Promotion:* We creates a linear model for Investment Facilitation and Promotion vs FDI Inflow and displays a plot of the points

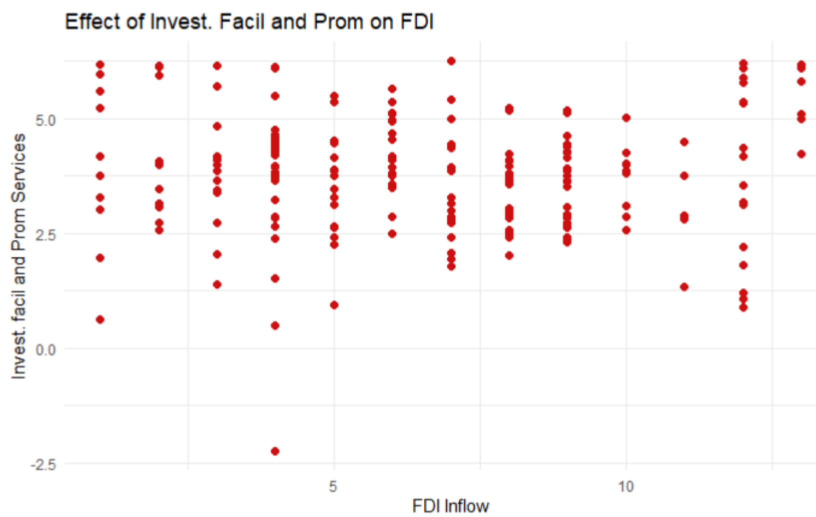
```
Modelc = lm(FDIInflow ~ ï..InvestFacilitation+InvestProm, data = FinalProjectx)
summary(Modelc)
ggplot(FinalProjectx) +
  aes(x = FDIInflow, y = ï..InvestFacilitation+InvestProm,) + geom_point(shape = "circle",
size = 2.25, colour = "#CA1010") + labs( x = "FDI Inflow",y = "Invest. facil and Prom
Services", title = "Effect of Invest. Facil and Prom on FDI") + theme_minimal()
```

Result

```
Call:
lm(formula = FDIInflow ~ ï..InvestFacilitation + InvestProm,
    data = FinalProjectx)

Residuals:
   Min     1Q Median     3Q    Max
-6.564 -2.408  0.122  2.116  7.532

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             3.2500     1.4822   2.193   0.0291 *
ï..InvestFacilitation   0.8701     0.3765   2.311   0.0215 *
InvestProm             -0.3168     0.2062  -1.536   0.1254
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.081 on 309 degrees of freedom
Multiple R-squared:  0.0183,	Adjusted R-squared:  0.01194
F-statistic:  2.88 on 2 and 309 DF,  p-value: 0.05766
```

- Upon creation of a model that contains both Investment Facilitation and Investment Promotion, this time we have a highly significant P value of IF (0.022), and we see that the total model accounts for 1.83% of the variability in FDI Inflow.
- The Multiple R-squared is 0.183% on 309 degrees of freedom because we used very few variables from the original dataset

We plotted only investment facilitation and Promotion services being one of the select variables.



- ❖ We now Built a model that uses the ZInvestProm, ZInvestFacilitation, and InvProm_X_InvestFacil to predict theFDI Inflow
- ❖ combined moderated effects of Standardized (Z-Score)IP and IF on FDI Inflow

```
Modeld = lm(FDIInflow ~ ZInvestProm + ZInvestFacilitation + InvProm_X_InvestFacil, data =
FinalProjectx)
summary(Modeld)
ggplot(FinalProjectx) +
  aes(x = FDIInflow, y = InvProm_X_InvestFacil) + geom_point(shape = "circle", size = 2.25,
colour = "#CA1010") + labs( x = "FDI Inflow",y = "InvProm_X_InvestFacil Services", title =
"Effect of InvProm_X_InvestFacil on FDI") + theme_minimal()
```

Result

```
Call:
lm(formula = FDIInflow ~ ZInvestProm + ZInvestFacilitation +
    InvProm_X_InvestFacil, data = FinalProjectx)

Residuals:
    Min      1Q  Median      3Q     Max
-6.5898 -2.4615  0.0806  2.0880  7.0462

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             6.6270     0.1824  36.341   <2e-16 ***
ZInvestProm             0.9537     0.9564   0.997   0.3195
ZInvestFacilitation     0.4200     0.2032   2.067   0.0395 *
InvProm_X_InvestFacil  -0.3376     0.2502  -1.349   0.1782
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.076 on 308 degrees of freedom
Multiple R-squared:  0.02406,   Adjusted R-squared:  0.01455
F-statistic: 2.531 on 3 and 308 DF,  p-value: 0.05725
```

- ❖ We have a highly significant P value of ZIF (0.039). Although the result indicates there was a non-significant interaction effect (p-value = 0.178) of FacilitationXInvestment, we may still consider meaningful moderation to be present (Hayes 2013, Matthes and Jörg 2020)Johnson-Neyman Plot. It could also be seen that the total model accounts for 2.4 % of the variability in FDI Inflow.

Naive Bayes classifier model to predict decision for FDI Inflow

- ❖ We now use Naive Bayes on Moderated (Interactive) variables to predict Foreign Direct Investment Inflows
- ❖ Now, run the Naive Bayes classifier model, and predict FDI status on the test set

```
FinalProjectx_nb_model <-naiveBayes(FDIInflow~ZInvestProm + ZInvestFacilitation +
InvProm_X_InvestFacil,data = Train)
FinalProjectx_nb_model
```

Result

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
         1          2          3          4          5          6          7          8 
9         10         11         12         13 
0.04382470 0.05179283 0.06374502 0.13944223 0.06374502 0.09163347 0.11553785 0.13545817 
0.12350598 0.05179283 0.02390438 0.07171315 0.02390438 

Conditional probabilities:
    ZInvestProm
Y           [,1]       [,2]
  1  -0.22971545 1.33112858
```

- ❖ The first part of the output above shows the ratios of default (yes) and default (no) in the training set (called a priori probabilities), followed by a table giving for each target class, mean, and standard deviation of the (sub-)variable. Also, note that the Naive Bayes algorithm assumes a Normal distribution for the independent variables. By the rule of the use of categorical predictors (the independent variables have been converted to categorical), we now have the conditional probabilities p(X|Y) for each attribute level given the default status.

- ❖ We now Predict the default status of the test dataset

```
FinalProjectx_Predicted_Test_labels <-predict(FinalProjectx_nb_model,Test)
library(gmodels)
# Show the confusion matrix of the classifier
CrossTable(x=Test$FDIInflow,y=FinalProjectx_Predicted_Test_labels, prop.chisq = FALSE)
```

Result

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  61


               | FinalProjectx_Predicted_Test_labels
Test$FDIInflow |         4 |         6 |         7 |         8 |         9 |        11 |
12 |        13 | Row Total |
---------------|-----------|-----------|-----------|-----------|-----------|-----------|--
---------|-----------|-----------|
             1 |         0 |         2 |         0 |         0 |         1 |         0 |
1 |         0 |         4 |
               |     0.000 |     0.500 |     0.000 |     0.000 |     0.250 |     0.000 |
```

- It is sometimes useful to output the raw prediction probabilities rather than the predicted class. To do that, we use the raw option in the model.

```
FinalProjectx_nb_model <- naiveBayes(FDIInflow~ZInvestProm + ZInvestFacilitation +
InvProm_X_InvestFacil,data = Train)
#Make predictions and return probability of each class
FinalProjectx_Predicted_Test_labels <-predict(FinalProjectx_nb_model,Test, type = "raw")
#show the first few values
head(FinalProjectx_Predicted_Test_labels)
```

```
set.seed(123)
data <- data.frame(FinalProjectx = sample(c("True","False"), 250, replace = TRUE),
                   FinalProjectx_Predicted_Test_labels = sample(c("True","False"), 250,
replace = TRUE)
                   )
table(data$FinalProjectx_Predicted_Test_labels, data$FinalProjectx)
```

Result

```
               1           2          3          4          5            6          7
8           9          10           11
[1,] 0.033051216 0.104068612 0.06750112 0.08764532 0.03036577 3.882281e-01 0.06860737
0.004872683 0.03713679 0.0112266301 1.730718e-05
[2,] 0.037097052 0.146122111 0.13600135 0.15631300 0.05305502 2.369012e-01 0.09698860
0.002014897 0.02370126 0.0007867073 5.006152e-03
[3,] 0.157266460 0.001611739 0.01853804 0.17006282 0.11964241 3.356725e-05 0.30822430
0.004621455 0.02525077 0.0045743840 1.822529e-01
[4,] 0.047669022 0.011992422 0.03866089 0.08824561 0.07673208 5.709298e-03 0.17117611
0.074146754 0.16553991 0.0288948400 1.938306e-01
[5,] 0.009963116 0.026401851 0.03477716 0.06333487 0.04183165 1.112929e-01 0.06212594
0.338555770 0.19546468 0.1064941029 2.729768e-03
[6,] 0.029175641 0.101276439 0.10361038 0.13342727 0.06022652 2.750415e-01 0.10500850
0.021525663 0.08092289 0.0081030418 5.330803e-03
              12            13
[1,] 0.092140556  7.513854e-02
[2,] 0.106012629  5.939444e-93
[3,] 0.007921201  4.949262e-43
[4,] 0.097402434  1.682487e-31
[5,] 0.007028221 2.876708e-112
[6,] 0.076351343  3.600026e-98
```

```
        False True
  False    58   54
  True     69   69
```

Confusion Matrix

- The confusion matrix function is very helpful as not only does it display a confusion matrix, it calculates many relevant statistics alongside

- Next, we created a confusion matrix, it is useful to create a confusion matrix to determine the performance of the classification algorithm. A confusion matrix is a simple table displaying the number of true positives/negatives and false-positive/negatives, or in other words how often the algorithm correctly or incorrectly predicts the outcome. The confusion Matrix function is very helpful as not only does it display a confusion matrix, it calculates many relevant statistics alongside (Tricks, 2021)

```
set.seed(123)
data <- data.frame(FinalProjectx = sample(c("True","False"), 250, replace = TRUE),
FinalProjectx_Predicted_Test_labels = sample(c("True","False"), 250, replace = TRUE)
)
library(caret)
confusionMatrix(as.factor(data$FinalProjectx_Predicted_Test_labels),
as.factor(data$FinalProjectx), positive = "True")
```

Result

```
Confusion Matrix and Statistics

          Reference
Prediction False True
     False    58   54
     True     69   69

               Accuracy : 0.508
                 95% CI : (0.4443, 0.5716)
    No Information Rate : 0.508
    P-Value [Acc > NIR] : 0.5253

                  Kappa : 0.0176

 Mcnemar's Test P-Value : 0.2068

            Sensitivity : 0.5610
            Specificity : 0.4567
         Pos Pred Value : 0.5000
         Neg Pred Value : 0.5179
             Prevalence : 0.4920
         Detection Rate : 0.2760
   Detection Prevalence : 0.5520
      Balanced Accuracy : 0.5088

       'Positive' Class : True
```

- Our results indicate that we misclassified a total of 127 cases out of 250. 69 as False Positives, and 58 as False Negatives. Interestingly, we classified a total of 123 cases of which 69 is True Positive and 54 is True Negative giving us an accuracy of 0.5253
- We have the confidence Interval of 95% that there is a 0.572 probability that the select predictors predicted the Foreign Direct Investment inflow based on the investment decision taken by the investor.
- Rows

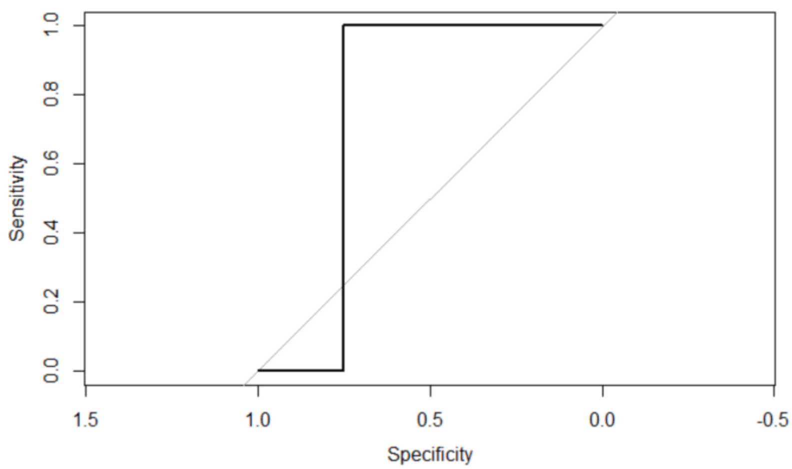
ROC Curve

❖ We could now output the ROC curves. we should remember that ROC curves plot sensitivity (true positive rate) versus (1 - specificity), which is (1 - TNR) or false positive rate

```
#Passing the second column of the predicted probabilities
#That column contains the probability associate to 'yes'
roc(Test$FDIInflow, FinalProjectx_Predicted_Test_labels[, 2])
plot.roc(Test$FDIInflow,FinalProjectx_Predicted_Test_labels[,2])
```

Result

```
Call:
roc.default(response = Test$FDIInflow, predictor =
FinalProjectx_Predicted_Test_labels[,     2])

Data: FinalProjectx_Predicted_Test_labels[, 2] in 4 controls (Test$FDIInflow 1) > 3
cases (Test$FDIInflow 2).
Area under the curve: 0.75
```



❖ The AUC is 0.75. The ROC curve is also plotted, though note that the X-Axis is Specificity (True Negative Rate), rather than 1-Specificity (False Positive Rate). This function can also be thought of as a plot of the FDI as a function of the Type I Error of the decision rule.

Box-Cox Transformation

❖ We first illustrate the transformation of data using the Box-Cox transformation approach

```
#Create a Box-Cox Transformation Model
FinalProjectx_Box_Cox_Transform<-preProcess(FinalProjectx,method = "BoxCox")
FinalProjectx_Box_Cox_Transform
```

Result

Data transformation, and particularly the Box-Cox power transformation, is one of these remedial actions that may help to make data normal. The Lambda value indicates the power to which all data should be raised. To do this, the Box-Cox power transformation searches from

Lambda = -5 to Lamba = +5 until the best value is found. The Box-Cox transformation tries to improve the normality of the residuals. Since that is the assumption of ANOVA as well. (StackStats, 2021)

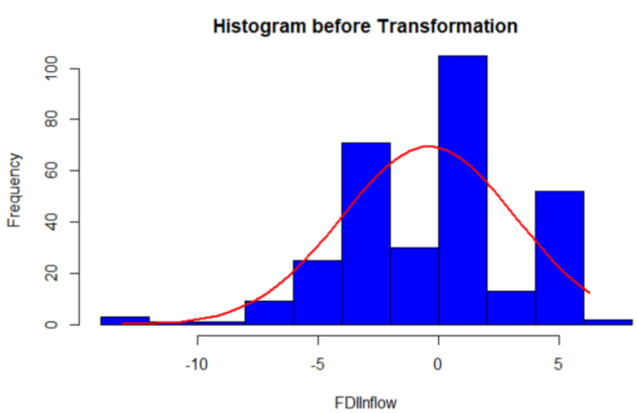
The lower and upper confidence levels (CLs) show that the best results for normality were reached with Lambda values between 0.8 and 1.6, the best value is 1.6.

```
Created from 312 samples and 2 variables

Pre-processing:
  - Box-Cox transformation (2)
  - ignored (0)

Lambda estimates for Box-Cox transformation:
1.6, 0.8
```

Now, we apply the transformation

```
FinalProjectx_Transformed=predict(FinalProjectx_Box_Cox_Transform, FinalProjectx)
y <- FinalProjectx_Transformed$InvProm_X_InvestFacil
h<-hist(y, breaks=10, col="blue", xlab="FDIInflow",
   main="Histogram before Transformation")
xfit<-seq(min(y),max(y),length=40)
yfit<-dnorm(xfit,mean=mean(y),sd=sd(y))
yfit <- yfit*diff(h$mids[1:2])*length(y)
lines(xfit, yfit, col="red", lwd=2)
```

Result



Result

The data before transformation seemingly does not assume a relative normal frequency distribution of FDI inflow. It is skewed to the right.

Alternatively, we applied Hypertuning

- ❖ Hyper tuning: Hyperparameter tuning in the ML model can largely affect its predictive performance, thus it is important to set a suitable hyperparameter for the model. Traditionally, hyperparameter tuning in the ML model is usually performed by a trial-and-error process. Depending on how many hyperparameters exist in the ML model, this process can be very exhausting (dhikaaurel, 2021)s
- ❖ We used hyper tuning to analyze the dataset by dividing the data into 80% tests and the remaining for training before plotting.

- ❖ Divide data into test and train

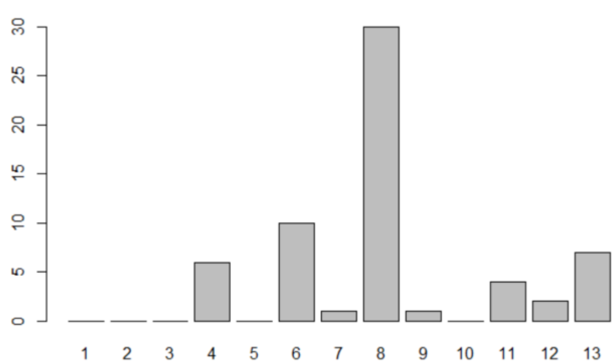
```
Index_Train<-createDataPartition(FinalProjectx$FDIInflow, p=0.8, list=FALSE)
Train <-FinalProjectx[Index_Train,]
Test  <-FinalProjectx[-Index_Train,]
```

```{r}
nb_model <-train(FDIInflow~ZInvestProm+ZInvestFacilitation+InvProm_X_InvestFacil, data
Train, preProc = c("BoxCox", "center", "scale"))
# Predict the default status of test dataset
Predicted_Test_labels <-predict(FinalProjectx_nb_model,Test)
summary(Predicted_Test_labels)
plot(Predicted_Test_labels)
```

Result

The result predicted that 80% of the used tested dataset in 13 categories, categories 1,2,3 4, and 10 are zero while category 8 has the highest number of 30.

1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	6	0	10	1	30	1	0	4	2	7



- ❖ We now generated a confusion matrix of the classifier

```
CrossTable(x=Test$FDIInflow,y=Predicted_Test_labels, prop.chisq = FALSE)
```

```
set.seed(123)
data <- data.frame(FinalProjectx = sample(c("True","False"), 250, replace = TRUE),
Predicted_Test_labels = sample(c("True","False"), 250, replace = TRUE)
)
library(caret)
confusionMatrix(as.factor(data$Predicted_Test_labels), as.factor(data$FinalProjectx),
positive = "True")
```

Result

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  61


              | Predicted_Test_labels
Test$FDIInflow |          4 |          6 |          7 |          8 |          9 |         11 |
12 |         13 | Row Total |
---------------|-----------|-----------|-----------|-----------|-----------|-----------|-
---------|-----------|-----------|
             1 |          0 |          2 |          0 |          1 |          0 |          0 |
0 |          0 |          3 |
               |      0.000 |      0.667 |      0.000 |      0.333 |      0.000 |      0.000 |
```

We have the same result as in the initial confusion matrix

```
Confusion Matrix and Statistics

          Reference
Prediction False True
     False    58   54
     True     69   69

               Accuracy : 0.508
                 95% CI : (0.4443, 0.5716)
    No Information Rate : 0.508
    P-Value [Acc > NIR] : 0.5253

                  Kappa : 0.0176

 Mcnemar's Test P-Value : 0.2068

            Sensitivity : 0.5610
            Specificity : 0.4567
         Pos Pred Value : 0.5000
         Neg Pred Value : 0.5179
             Prevalence : 0.4920
         Detection Rate : 0.2760
   Detection Prevalence : 0.5520
      Balanced Accuracy : 0.5088

       'Positive' Class : True
```

Data Envelopment Analysis (DEA) for Final Project

Effective government service provision benefits from the support of rigorous measurement techniques. Data Envelopment Analysis (DEA) is an analytical tool that can assist in the identification of best practices in the use of resources among a group of organizations. Such identification can highlight possible efficiency improvements that may help agencies to achieve their potential.

DEA is particularly appropriate where the researcher is interested in investigating the efficiency of converting multiple inputs into multiple outputs, DEA is a linear programming technique that enables management to benchmark the best-practice decision-making unit (DMU) (Necmi, 2001)

Country (Descriptor)	FDI Inflow in $B (Invest. Facilitation Services)	FDI Inflow in $B (Invest. Promotion Services))	Total FDI Inflow in $B (2015 – 2016)	Efficiency
1. UNITED KINGDOM	10.95	6.09	17.05	1.00
2. UNITED STATES	7.01	2.70	9.71	0.57
3. NETHERLANDS	3.29	1.48	4.77	0.28
4. SINGAPORE	0.46	0.79	1.25	0.073
5. SWITZERLAND	0.34	0.78	1.11	0.065
6. REPUBLIC OF SOUTH AFRICA	0.75	0.51	1.25	0.053

Source of Data: Nigerian Investment Promotion Commission

An Algebraic Representation

x_{ik}=i^{th} input quantity for DMU(k)
y_{jk}=j^{th} output quantity for DMU(k)
v_i=Weight for i^{th} input
u_j=Weight for the j^{th} output
X_K=Sum(x_{ik}*v_i)= Weighted input
Y_K=Sum(y_{jk}*u_j)= Weighted output
Efficiency = $E_k = Y_k / X_k$

Model Formulation

Maximize Y_0
subject to $X_0 = 1$ scaling of the input value (one constraint)
$-X_k + Y_k \leq 0$ efficiency no greater than 1, for each DMU (k)

Country 1

/* Objective function */
max: 10.95 U1 + 6.09 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1 <= 0;
7.01 U1 + 2.70 U2 – 9.71 V1 <= 0;

```
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 17.05 V1  = 1;
```

Country 2

```
/* Objective function */
max: 7.01 U1 + 2.70 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1  <= 0;
7.01 U1 + 2.70 U2   – 9.71 V1  <= 0;
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 9.71 V1  = 1;
```

Country 3

```
/* Objective function */
max: 3.29 U1 + 1.48 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1  <= 0;
7.01 U1 + 2.70 U2   – 9.71 V1  <= 0;
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 4.77 V1  = 1;
```

Country 4

```
/* Objective function */
max: 0.46 U1 + 0.79 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1  <= 0;
7.01 U1 + 2.70 U2   – 9.71 V1  <= 0;
```

```
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 1.25 V1  = 1;
```

<u>Country 5</u>

```
/* Objective function */
max: 0.34 U1 + 0.78 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1  <= 0;
7.01 U1 + 2.70 U2   – 9.71 V1  <= 0;
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 1.11 V1  = 1;
```

<u>Country 6</u>

```
/* Objective function */
max: 0.75 U1 + 0.51 U2;
/* Constraints */
10.95 U1 + 6.09 U2 – 17.05 V1  <= 0;
7.01 U1 + 2.70 U2   – 9.71 V1  <= 0;
3.29 U1 + 1.48 U2   – 4.77 V1  <= 0;
0.46 U1 + 0.79 U2   – 1.25 V1  <= 0;
0.34 U1 + 0.78 U2   – 1.11 V1  <= 0;
0.75 U1 + 0.51 U2   – 1.25 V1  <= 0;
+ 1.25 V1  = 1;
```

Result Summary and Comparisons.

We used R and solved the objective function, obtaining the objective and variable functions of all the six DMUs. The tabulated summaries are depicted below;

DMU			
Country 1			
Objective Function	max: 10.95 U1 + 6.09 U2;		
Max. efficiency(100%)	1		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.0586510263929619
Lower	0	0	0.0586510263929619
[1] 0			
[1] 0.9944471			
[1] 0.05967262 0.05599866 0.05865103			

DMU			
Country 2			
Objective Function	max: 7.01 U1 + 2.70 U2;		
Max. efficiency(100%)	0.57		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.102986611740474
Lower	0	0	0.102986611740474
[1] 0			
[1] 1			
[1] 0.1426534 0.0000000 0.1029866			

DMU			
Country 3			
Objective Function	max: 3.29 U1 + 1.48 U2;		
Max. efficiency(100%)	0.28		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.209643605870021
Lower	0	0	0.209643605870021
[1] 0			
[1] 0.9979824			
[1] 0.2132952 0.2001629 0.2096436			

DMU			
Country 4			
Objective Function	max: 0.46 U1 + 0.79 U2;		
Max. efficiency(100%)	0.073		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.8
Lower	0	0	0.8
[1] 0			
[1] 0.9912906			
[1] 0.7947522 0.7920311 0.8000000			

DMU			
Country 5			
Objective Function	max: 0.34 U1 + 0.78 U2;		
Max. efficiency(100%)	0.065		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.900900900900901
Lower	0	0	0.900900900900901
[1] 0			
[1] 1			
[1] 0.0000000 1.2820513 0.9009009			

DMU			
Country 6			
Objective Function	max: 0.75 U1 + 0.51 U2;		
Max. efficiency(100%)	0.053		
Kind	Std	Std	Std
Type	Real	Real	Real
Upper	Inf	Inf	0.8
Lower	0	0	0.8
[1] 0			
[1] 1			
[1] 0.8139346 0.7638217 0.8000000			

For All the DMUs (Countries)

CRS	DRS	IRS	VRS	FDH	ADD
[1] 1.0000 1.0000 0.9998 1.0000 1.0000 0.9916	[1] 1.0000 1.0000 0.9998 1.0000 1.0000 0.9916	[1] 1 1 1 1 1 1	[1] 1 1 1 1 1 1	[1] 1 1 1 1 1 1	[1] 1 1 1 1 1 1
peer1 peer2	peer1 peer2	peer1	peer1	peer1	peer1
[1,] 1 NA	[1,] 1 NA	[1,] 1	[1,] 1	[1,] 1	[1,] 1
[2,] 2 NA	[2,] 2 NA	[2,] 2	[2,] 2	[2,] 2	[2,] 2
[3,] 1 2	[3,] 1 2	[3,] 3	[3,] 3	[3,] 3	[3,] 3
[4,] 4 NA	[4,] 4 NA	[4,] 4	[4,] 4	[4,] 4	[4,] 4
[5,] 5 NA	[5,] 5 NA	[5,] 5	[5,] 5	[5,] 5	[5,] 5
[6,] 1 4	[6,] 1 4	[6,] 6	[6,] 6	[6,] 6	[6,] 6
L1 L2 L4 L5	L1 L2 L4 L5	L1 L2 L3 L4 L5 L6	L1 L2 L3 L4 L5 L6	L1 L2 L3 L4 L5 L6	L1 L2 L3 L4 L5 L6
[1,] 1.00000000 0.0000000 0.0000000 0	[1,] 1.00000000 0.0000000 0.0000000 0	[1,] 1 0 0 0 0 0	[1,] 1 0 0 0 0 0	[1,] 1 0 0 0 0 0	[1,] 1 0 0 0 0 0
[2,] 0.00000000 1.0000000 0.0000000 0	[2,] 0.00000000 1.0000000 0.0000000 0	[2,] 0 1 0 0 0 0	[2,] 0 1 0 0 0 0	[2,] 0 1 0 0 0 0	[2,] 0 1 0 0 0 0
[3,] 0.11362609 0.2917276 0.0000000 0	[3,] 0.11362609 0.2917276 0.0000000 0	[3,] 0 0 1 0 0 0	[3,] 0 0 1 0 0 0	[3,] 0 0 1 0 0 0	[3,] 0 0 1 0 0 0
[4,] 0.00000000 0.0000000 1.0000000 0	[4,] 0.00000000 0.0000000 1.0000000 0	[4,] 0 0 0 1 0 0	[4,] 0 0 0 1 0 0	[4,] 0 0 0 1 0 0	[4,] 0 0 0 1 0 0
[5,] 0.00000000 0.0000000 0.0000000 1	[5,] 0.00000000 0.0000000 0.0000000 1	[5,] 0 0 0 0 1 0	[5,] 0 0 0 0 1 0	[5,] 0 0 0 0 1 0	[5,] 0 0 0 0 1 0
[6,] 0.06067381 0.0000000 0.1724092 0	[6,] 0.06067381 0.0000000 0.1724092 0	[6,] 0 0 0 0 0 1	[6,] 0 0 0 0 0 1	[6,] 0 0 0 0 0 1	[6,] 0 0 0 0 0 1
The results indicate that DMUs 1, 2, 4, and 5 are efficient. DMUs 3 and 6 are not efficient	The results indicate that DMUs 1, 2, 4, and 5 are efficient. DMUs 3 and 6 are not efficient	The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.	The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.	The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.	The results indicate that DMUs 1, 2, 3, 4,and 5 are efficient.

- ❖ CRS: The results indicate that DMUs 1, 2, 4, and 5 are efficient. DMUs 3 and 6 are not efficient
- ❖ IRS: The results indicate that DMUs 1, 2, 4, and 5 are efficient. DMUs 3 and 6 are not efficient
- ❖ VRS: The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.
- ❖ FDH: The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.
- ❖ ADD: The results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient.

4. Conclusions

- Our findings indicate that the perception of combined investment facilitation and promotion services has a profound direct positive effect on individual investors' FDI decisions. And that the same factor could be used to effectively predict an investor's decision using Naïve Bayes.
- We have the confidence Interval of 95% that there is a 0.572 probability that the select predictors predicted the Foreign Direct Investment inflow based on the investment decision taken by the investor.

- Upon assessment of the DMUs in DEA, the results indicate that DMUs 1, 2, 3, 4, 5, and 6 are efficient when the organization uses either VRS, FDH, or ADD

Given the above inferences, I would suggest the following:

- With sufficient capacity building of the Investment Promotion Agency's (IPA) employees, it is envisaged that the employees could perform much better in other areas of services that proved to have weak effects on investor's decisions so that they could attain their full potentials as promoters and facilitators of investment service providers.
- Finally, the inferences established from the analyses should be used to create knowledge for IP agencies and to utilize the experience for a corporate policy framework that would enhance performance in preparation against the dynamic and evolving FDI competition for economic development.

REFERENCES:

dhikaaurel. (2021). *Genetic Algorithm in R: Hyperparameter Tuning*. Retrieved from https://towardsdatascience.com/genetic-algorithm-in-r-hyperparameter-tuning-5fc6439d2962

Necmi, K. (2001). Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis - ScienceDirect. *Socio-Economic Planning Sciences, Volume 35, Issue 1*(Issue 1), Pages 57-80. Retrieved from https://doi.org/10.1016/S0038-0121(00)00010-0

Santos, J. D. (2021). RPubs - Search Methods for Hyperparameter Tuning in R. In.

StackStats. (2021). Estimating Lambda for Box Cox transformation for ANOVA. Retrieved from https://stats.stackexchange.com/questions/10095/estimating-lambda-for-box-cox-transformation-for-anova

Tricks, D. (2021, 2021-04-13). Confusion matrix in R: two simple methods - Data Tricks. Retrieved from https://datatricks.co.uk/confusion-matrix-in-r-two-simple-methods