Configuration Manual

MSc Research Project
Fintech

Similoluwa Kenku
Student ID: X18113079

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Similoluwa Kenku
Student ID:	X18113079
Programme:	Fintech **Year:** 2019
Module:	MSc Research Project
Lecturer:	Noel Cosgrave
Submission Due Date:	12/9/2019
Project Title:	Financial and Digital Inclusion: Determining factors influencing use of Financial Products
Word Count:	969 **Page Count:** 9

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: .....

Date: 12/9/2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Similoluwa Kenku
Student ID: x18113079

1 Software Tool

R studio: This is a software used for data analysis and machine learning purposes with statistical and visualisation capabilities. The R studio version used for this analysis is 948 bytes in size.

1.1. PC used to run the R program

- ✓ HP Pavilion 15 Laptop
- ✓ Intel Core i5 8th Generation
- ✓ 64-bit processor @ 1.60GHz 1.80GHz
- ✓ 8gb Ram
- ✓ Windows 10 home
- ✓ 1TB memory

2 Codes and Plots

```
library(readxl)
library(caret)
library(glmnet)
library(mlbench)
library(psych)
library(dplyr)
library(mice)
library(VIM)
library(stargazer)
library(mctest)
library(ppcor)
library(GGally)
library(corpcor)

#READ DATA
Findex <- read_excel("~/Findex.xlsx")
View(Findex)
summary(Findex)
str(Findex)
fin<-Findex
```

```
# DECISION TO HAVE AN ACCOUNT WITH A FINANCIAL INSTITUTION
#FEMALES 15+
fe<-cbind(fin$`own financial institution account, female 15+`,fin$`Internet bill payment,
female 15+`,fin$`Internet purchase, female  15+`,fin$`Saved for business or farm, female
15+`,fin$`Saved for old age, female 15+`,fin$savedataFIfemale15,fin$`Informal Savings,
female 15+`,fin$`Saved for education, female  15+`)
colnames(fe)<c("ownaccount","Internet_BillPmt","Internet_Purch","Saved_BusinessFarm","
OldAge","Savings","Saved_Informal","Education")
as.data.frame(fe)->fe1
pairs.panels(fe1,cex=2)

# IMPUTATION FEMALES 15+
impfe<-mice(fe1[,1:8],m=3,seed = 123)
print(impfe)
complete(impfe,1)->fe1
View(fe1)

#Test for Multicollinearity
X1<-fe1[c(2:8)]
Y1<-fe1[c(1)]
ggpairs(X1)
cor2pcor(cov(X1))
omcdiag(X1,Y1)
imcdiag(X1,Y1)

#DATA PARTITIONING
set.seed(222)
ind1<-sample(2,nrow(fe1),replace = T,prob = c(0.8,0.2))
train1<-subset(fe1,ind1==1)
test1<-subset(fe1,ind1==2)
#custom control parameters
custom<-trainControl(method = "repeatedcv",number = 10,repeats = 5,verboseIter = T)

##MODEL FOR FEMALE 15+
set.seed(1234)
lm_fe<-train(ownaccount~.,train1,method='lm',trControl=custom)
lm_fe2=lm(ownaccount~.,train1)

# Forward Stepwise Regression
lm_fe3<-lm(ownaccount~.,train1)
step(lm_fe3,direction = "forward",scope = formula(lm_fe2))
lm_fe<train(ownaccount~Internet_BillPmt+Internet_Purch+Saved_BusinessFarm,train1,met
hod='lm',trControl=custom)

#Results
lm_fe$results
lm_fe
summary(lm_fe)
```

```
#PLOT
plot(varImp(lm_fe,scale=T))

## RESIDUAL PLOT
lm_fe.res=resid(lm_fe)
plot(train1$ownaccount,lm_fe.res,ylab="Residuals",xlab="Own Account",main="Female
Residual Plot")
abline(0,0)

#Prediction
p1<-predict(lm_fe,test1)
sqrt(mean((test1$ownaccount-p1)^2))

#MALES
ma<-cbind(fin$`own financial institution account,male  15+`,fin$`Internet bill payment, male
15+`,fin$`Internet purchase, male  15+`,fin$`Saved for farm or business, male
15+`,fin$`Saved for old age, male 15+`,fin$savedataFImale15,fin$`Informal Savings, male
15+`,fin$`Saved for education, male 15+`)
colnames(ma)<c("ownaccount","Internet_BillPmt","Internet_Purch","Saved_BusinessFarm",
"OldAge","Savings","Saved_Informal","Education")
as.data.frame(ma)->ma1

#IMPUTATIONS FOR MALES
impma<-mice(ma1[,1:8],m=3,seed = 123)
print(impma)
complete(impma,1)->ma1
View(ma1)

#Test for Multicollinearity
X2<-ma1[c(2:8)]
Y2<-ma1[c(1)]
ggpairs(X2)
cor2pcor(cov(X2))
omcdiag(X2,Y2)
imcdiag(X2,Y2)

#DATA PARTITIONING
set.seed(222)
ind2<-sample(2,nrow(ma1),replace = T,prob = c(0.8,0.2))
train2<-subset(ma1,ind2==1)
test2<-subset(ma1,ind2==2)

#Custom control parameters
custom<-trainControl(method = "repeatedcv",number = 10,repeats = 5,verboseIter = T)

##MODEL FOR MALE 15+
set.seed(1234)
lm_ma<-train(ownaccount~.,train2,method='lm',trControl=custom)
lm_ma2<-lm(ownaccount~.,train2)
```

```
# Forward Stepwise Regression
lm_ma3<-lm(ownaccount~.,train2)
step(lm_ma3,direction = "forward",scope = formula(lm_ma2))
lm_ma<train(ownaccount~Internet_BillPmt+Internet_Purch+Saved_BusinessFarm,train2,me
thod='lm',trControl=custom)

#Results
lm_ma$results
lm_ma
summary(lm_ma)

#PLOT
plot(varImp(lm_ma,scale=T))

##RESIDUAL PLOT
lm_ma.res=resid(lm_ma)
plot(train2$ownaccount,lm_ma.res,ylab="Residuals",xlab="Own Account",main="Male
Residual Plot")
abline(0,0)

#Prediction
p2<-predict(lm_ma,test2)
sqrt(mean((test2$ownaccount-p2)^2))

### GENDER COMPARISON
stargazer(lm_fe3,lm_ma3,type = "text",out = "FIaccount.txt",no.space = T)

#DETERMINANTS OF MOBILE BANKING
#MOBILE MONEY (FEMALE 15+)
mb_fe<-cbind(fin$`Mobile money account, female 15+`,fin$`Received digital payments,
female  15+`,fin$`Made digital payments, female  15+`,fin$`Made or received digital
payments, female  15+`,fin$`Own credit card, female  15+`,fin$`Debit card, female
15+`,fin$`Internet bill payment, female 15+`,fin$`Internet purchase, female  15+`)
colnames(mb_fe)<c("mobileacct","received_digital","made_digital","madeReceived_dig","o
wncreditcard","owndebitcard","intern_bill","intern_purch")
as.data.frame(mb_fe)->mb_fe1

#IMPUTATION FOR FEMALES 15+
impfe<-mice(mb_fe1[,1:8],m=3,seed = 123)
print(impfe)
complete(impfe,1)->mb_fe2
View(mb_fe2)
summary(mb_fe2)

#Test for Multicollinearity
X3<-mb_fe2[c(2:8)]
Y3<-mb_fe2[c(1)]
ggpairs(X3)
cor2pcor(cov(X3))
omcdiag(X3,Y3)
```

```
imcdiag(X3,Y3)


#DATA PARTITIONING
set.seed(222)
ind3<-sample(2,nrow(mb_fe2),replace = T,prob = c(0.8,0.2))
train3<-subset(mb_fe2,ind3==1)
test3<-subset(mb_fe2,ind3==2)

#Custom control parameters
custom<-trainControl(method = "repeatedcv",number = 10,repeats = 5,verboseIter = T)

##MODEL FOR FEMALE 15+
set.seed(1234)
lm_mbfe<-train(mobileacct~.,train3,method='lm',trControl=custom)
lm_mbfe2<-lm(mobileacct~.,train3)

# Forward Stepwise Regression
lm_mbfe3<-lm(mobileacct~.,train3)
step(lm_mbfe3,direction = "forward",scope = formula(lm_mbfe2))
lm_mbfe<train(mobileacct~received_digital+made_digital+madeReceived_dig+owncreditcar
d+owndebitcard+intern_bill+intern_purch,train3,method='lm',trControl=custom)

#Results
lm_mbfe$results
lm_mbfe
summary(lm_mbfe)

#PLOT
pairs.panels(mb_fe1)
plot(varImp(lm_mbfe,scale=T))

## RESIDUAL PLOT
lm_mbfe.res=resid(lm_mbfe)
plot(train3$mobileacct,lm_mbfe.res,ylab="Residuals",xlab="Own Mobile
Account",main="Female Residual Plot")
abline(0,0)

#Prediction
p3<-predict(lm_mbfe,test3)
sqrt(mean((test3$mobileacct-p3)^2))

#MOBILE MONEY (MALES 15+)
mbma<-cbind(fin$`Mobile money account, male  15+`,fin$`Received digital payments, male
15+`,fin$`Made digital payments, male  15+`,fin$`Made or received digital payments, male
15+`,fin$`Own credit card, male  15+`,fin$`Debit card, male  15+`,fin$`Internet bill payment,
male  15+`,fin$`Internet purchase, male  15+`)
colnames(mbma)<c("mobileacct","received_digital","made_digital","madeReceived_dig","o
wncreditcard","owndebitcard","intern_bill","intern_purch")
as.data.frame(mbma)->mbma1
```

```
#IMPUTATION FOR MALES 15+
imp_ma<-mice(mbma1[,1:8],m=3,seed = 123)
print(imp_ma)
complete(imp_ma,1)->mb_ma2
View(mb_ma2)
summary(mb_ma2)

#Test for Multicollinearity
X4<-mb_ma2[c(2:8)]
Y4<-mb_ma2[c(1)]
ggpairs(X4)
cor2pcor(cov(X4))
omcdiag(X4,Y4)
imcdiag(X4,Y4)

#DATA PARTITIONING
set.seed(222)
ind4<-sample(2,nrow(mb_ma2),replace = T,prob = c(0.8,0.2))
train4<-subset(mb_ma2,ind4==1)
test4<-subset(mb_ma2,ind4==2)

#Custom control parameters
custom<-trainControl(method = "repeatedcv",number = 10,repeats = 5,verboseIter = T)

##MODEL FOR MALE 15+
set.seed(1234)
lm_mbma<-train(mobileacct~.,train4,method='lm',trControl=custom)
lm_mbma2<-lm(mobileacct~.,train4)

# Forward Stepwise Regression
lm_mbma3<-lm(mobileacct~.,train4)
step(lm_mbma3,direction = "forward",scope = formula(lm_mbma2))
lm_mbfe<-
train(mobileacct~received_digital+made_digital+madeReceived_dig+owncreditcard+owndeb
itcard+intern_bill+intern_purch,train4,method='lm',trControl=custom)

#Results
lm_mbma$results
lm_mbma
summary(lm_mbma)

#PLOT
pairs.panels(mbma1)
plot(varImp(lm_mbma,scale=T))

## RESIDUAL PLOT
lm_mbma.res=resid(lm_mbma)
```

```
plot(train4$mobileacct,lm_mbma.res,ylab="Residuals",xlab="Own Mobile
Account",main="Male Residual Plot")
abline(0,0)

#Prediction
p4<-predict(lm_mbma,test4)
sqrt(mean((test4$mobileacct-p4)^2))

### ACROSS GENDER COMPARISON
stargazer(lm_mbfe3,lm_mbma3,type = "text",out = "mobile.txt",no.space =
T)mobile.txt",no.space = T)
```

Joint Output

	Dependent variable:	
	ownaccount	
	females	males
	(1)	(2)
Internet_BillPmt	0.038	-0.235**
	(0.098)	(0.110)
Internet_Purch	0.291**	0.255*
	(0.136)	(0.150)
Saved_BusinessFarm	-0.631***	-0.533***
	(0.195)	(0.128)
OldAge	0.093	0.049
	(0.134)	(0.126)
Savings	0.952***	1.280***
	(0.142)	(0.117)
Saved_Informal	-0.359***	-0.279**
	(0.105)	(0.125)
Education	0.258*	-0.180
	(0.132)	(0.113)
Constant	0.278***	0.396***
	(0.020)	(0.023)
Observations	408	408
R2	0.771	0.749
Adjusted R2	0.767	0.744
Residual Std. Error (df = 400)	0.154	0.151
F Statistic (df = 7; 400)	192.851***	170.094***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Fig.1: Own account before stepwise Regression