



Università degli Studi di Milano

Statistical Learning

Danial Forouzanfar - 12988A

Supervised:

- Decision Tree,
- Random Forest,
- eXtreme Gradient Boosting

Data on smoking habits

gender

Female:886
Male :677

age

Min. :16.00
1st Qu.:34.00
Median :47.00
Mean :49.59
3rd Qu.:65.00
Max. :97.00

marital_status

single :386
Married :760
Separated:417

highest_qualification

without :525
University:252
GCSE :381
Sub Degree:241
ONC : 73
A Levels : 91

nationality

British :512
English :773
Other : 87
Scottish:134
Welsh : 57

gross_income

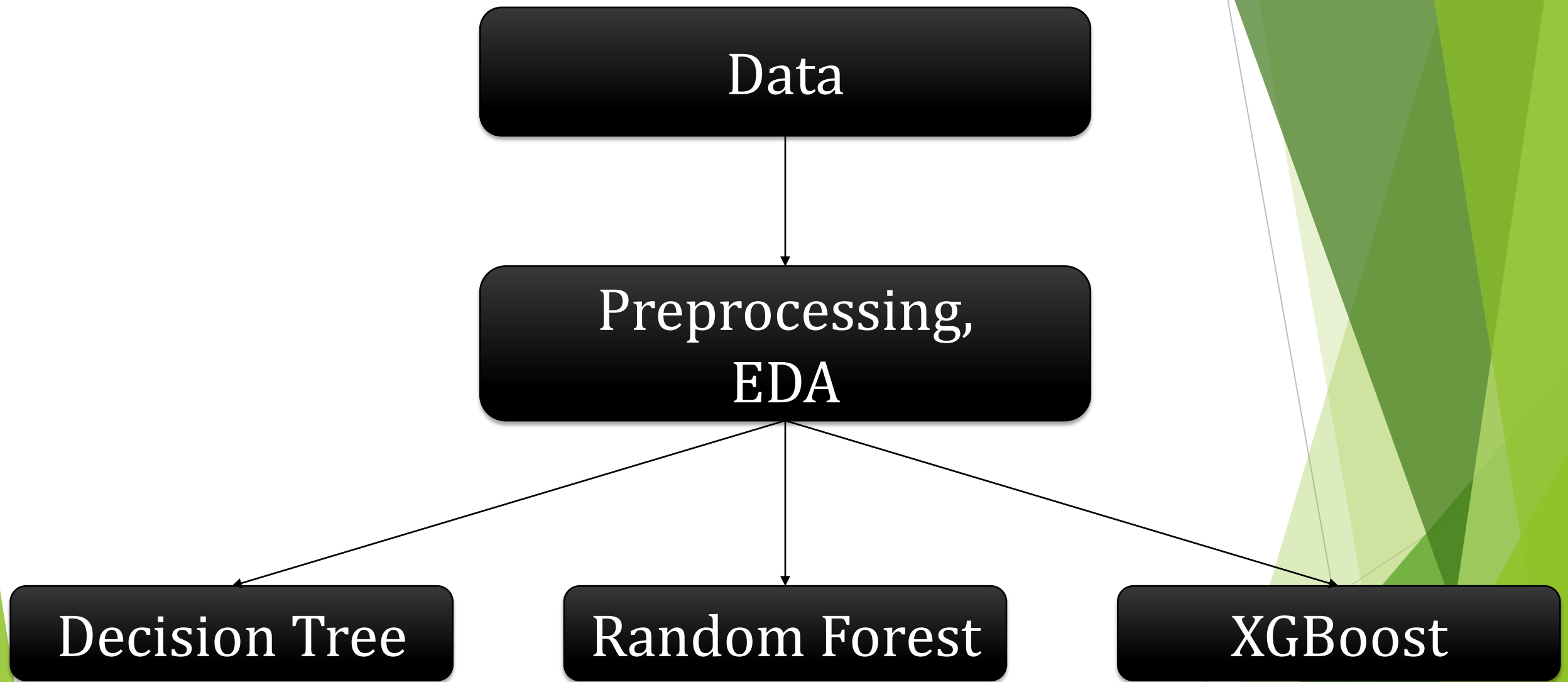
Low :784
High :168
Moderate:611

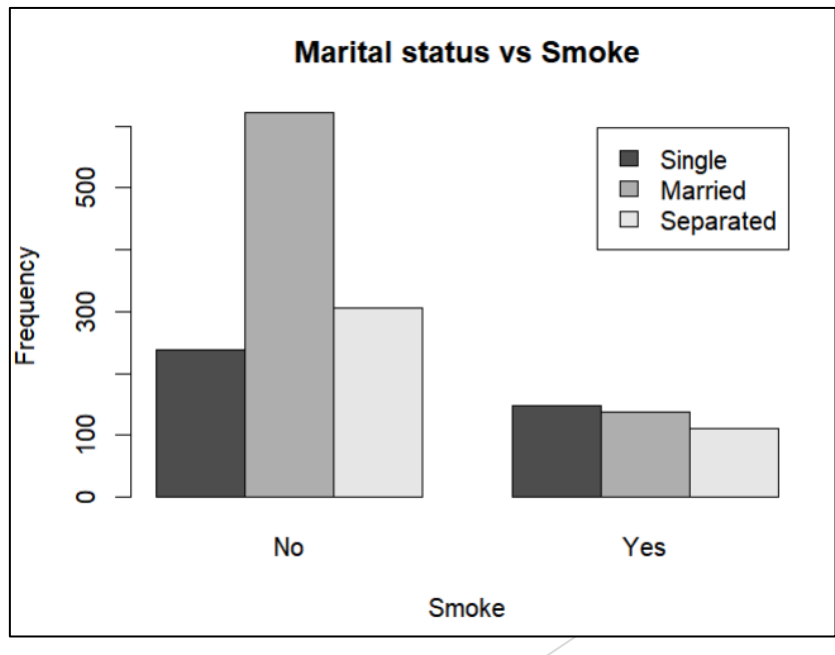
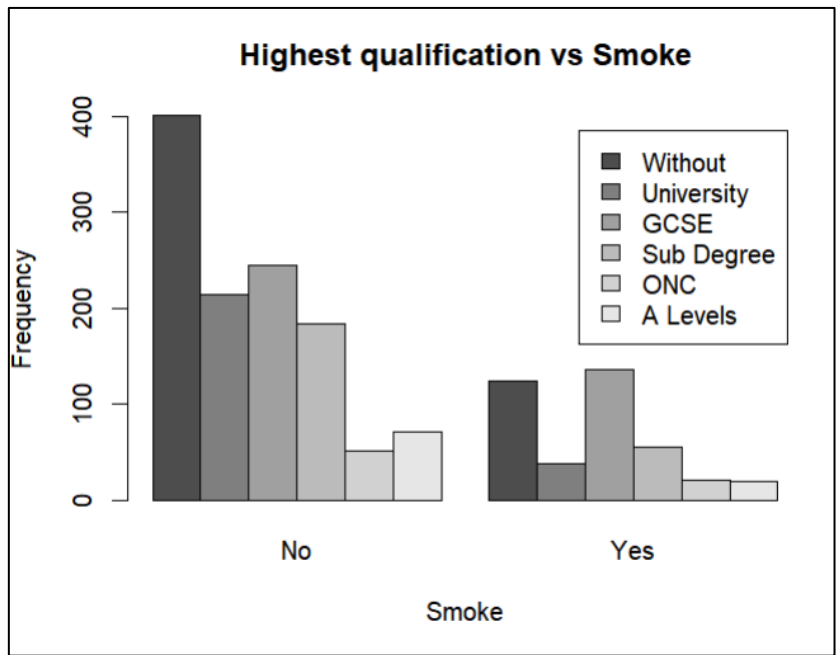
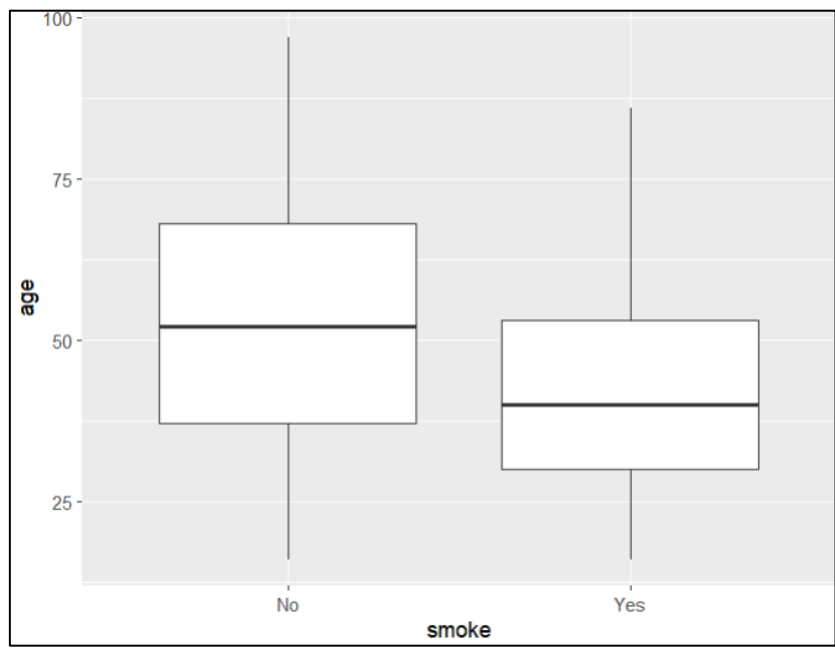
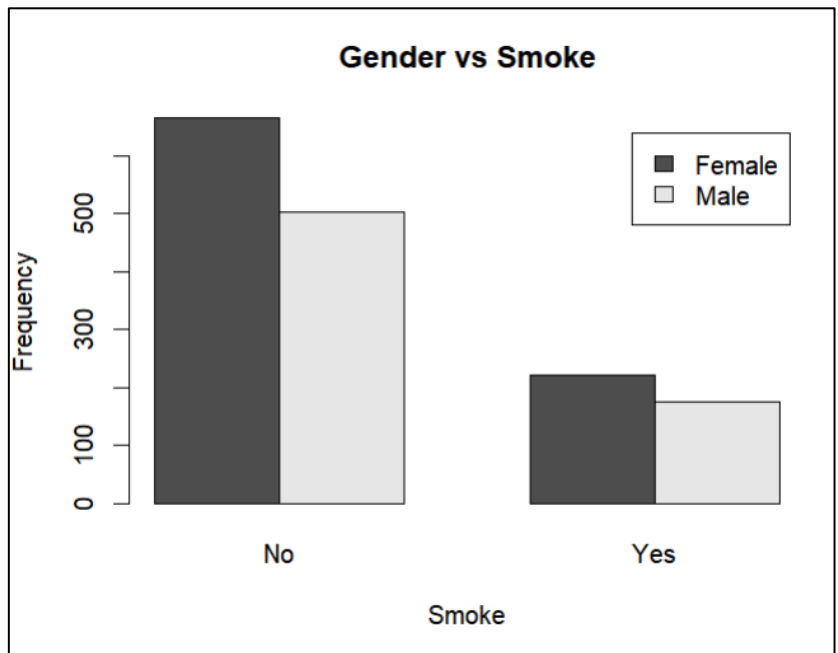
region

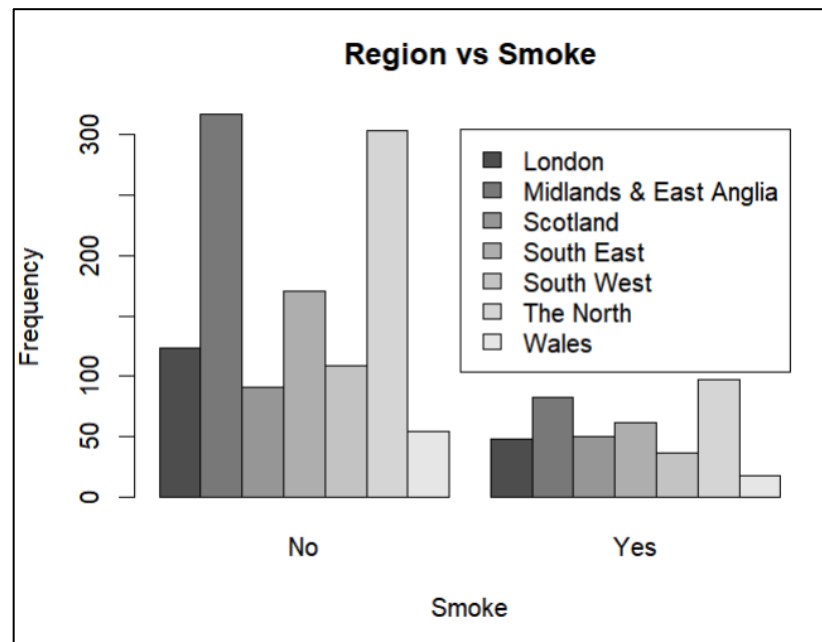
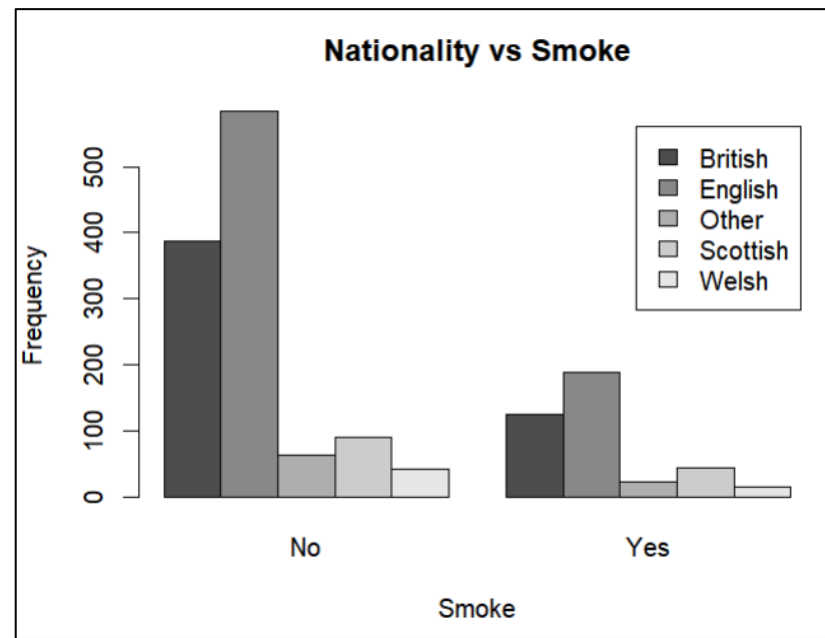
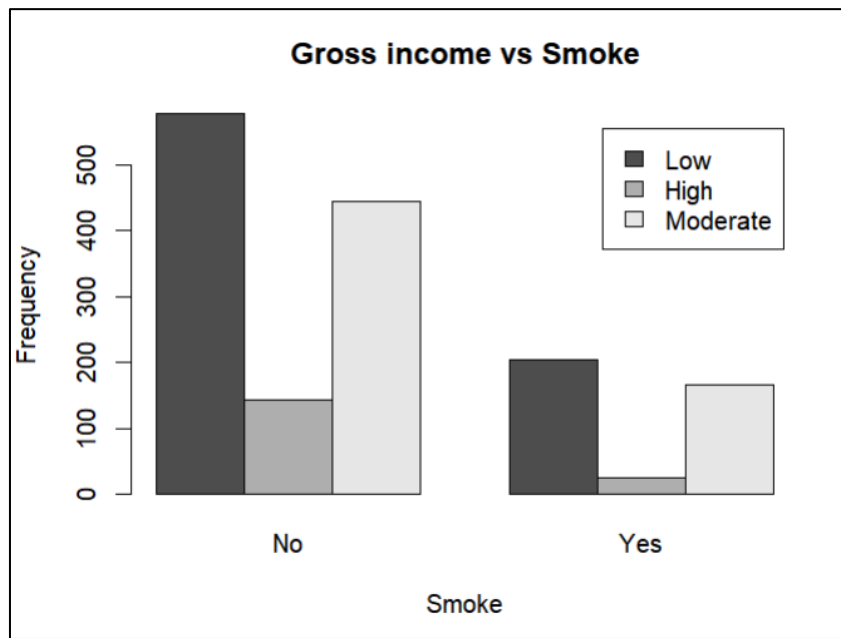
London :172
Midlands & East Anglia:400
Scotland :141
South East :232
South West :146
The North :400
Wales : 72

smoke

No :1167
Yes: 396







Libraries

General

tidymodels

tidyverse

caret

Decision Tree

visNetwork

rpart

Random Forest

randomForest

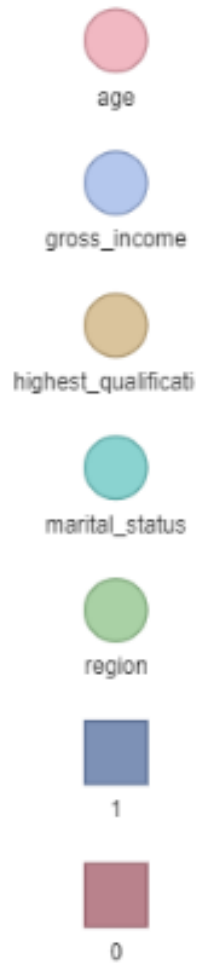
e1071

XG Boosting

xgboost

rattle

Decision Tree



Decision Tree

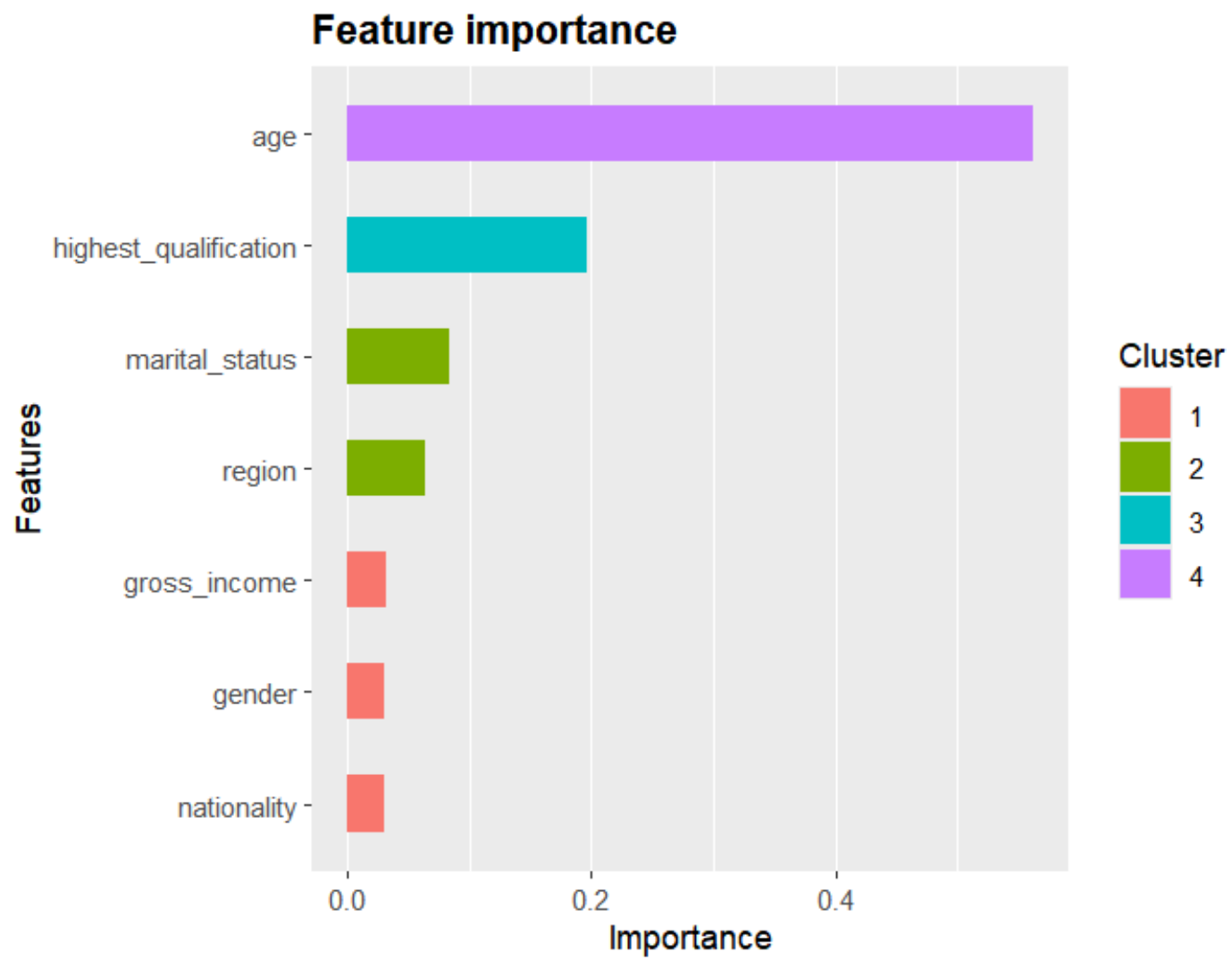
	Metric	Value
1	Accuracy	0.7468085
2	Precision	0.9259259
3	Recall	0.7775120
4	F1 score	0.8452536

Random Forest

	Metric	Value
1	Accuracy	0.7510638
2	Precision	0.9572650
3	Recall	0.7671233
4	F1 score	0.8817110

XG Boosting

	Metric	Value
1	Accuracy	0.7595745
2	Precision	0.9743590
3	Recall	0.7668161
4	F1 score	0.8582183



Unsupervised:

- PCA

Fraud Detection

type
Min. :1.000
1st Qu.:2.000
Median :2.000
Mean :2.724
3rd Qu.:4.000
Max. :5.000

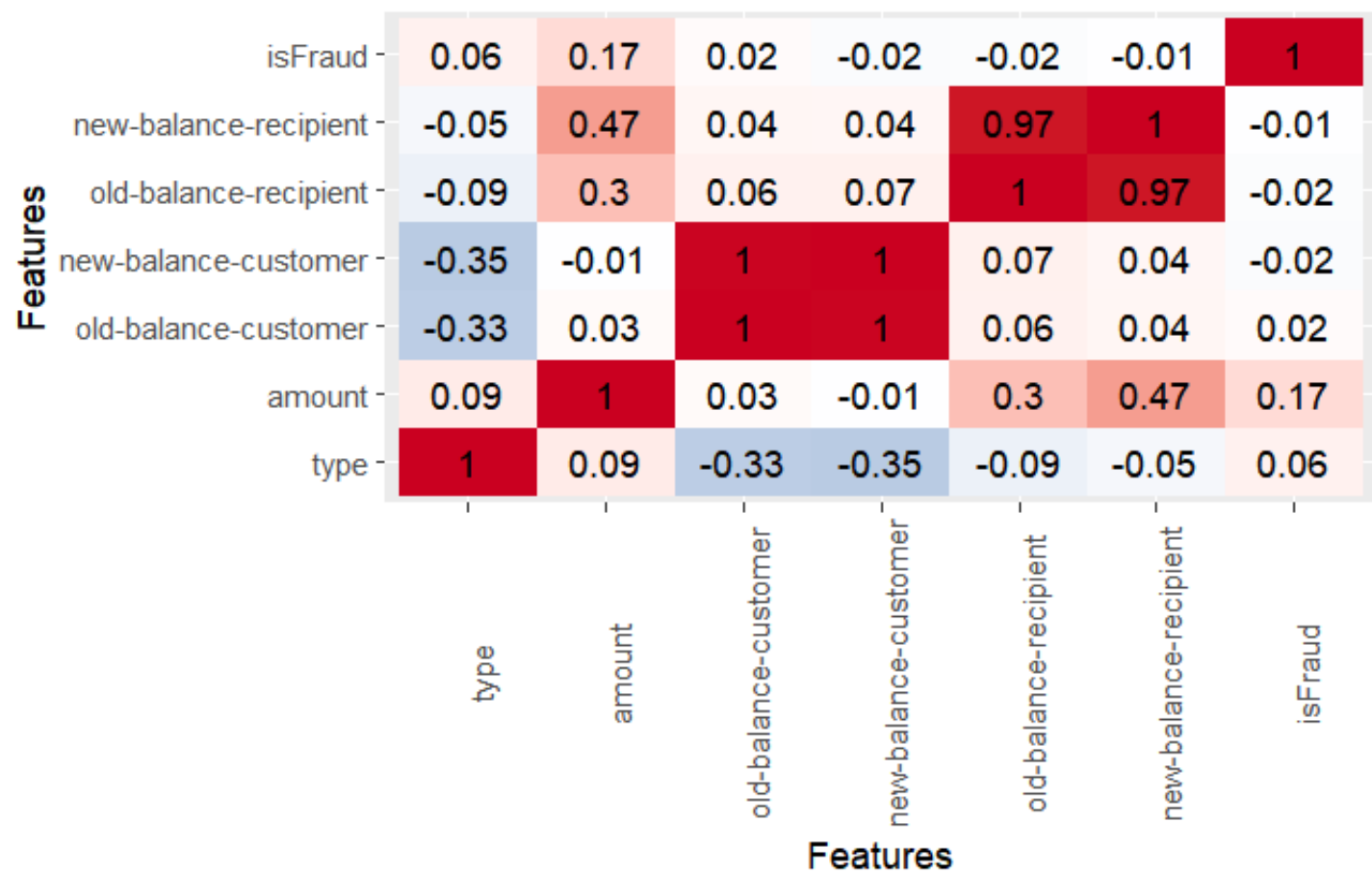
old-balance-customer
Min. :0.0000000
1st Qu.:0.0000000
Median :0.0002518
Mean :0.0140007
3rd Qu.:0.0018875
Max. :0.5721143

new-balance-recipient
Min. :0.0000000
1st Qu.:0.0000000
Median :0.0006044
Mean :0.0034529
3rd Qu.:0.0031185
Max. :0.9204010

amount
Min. :0.0000000
1st Qu.:0.0001467
Median :0.0008187
Mean :0.0020730
3rd Qu.:0.0022747
Max. :0.5551823

new-balance-customer
Min. :0.000000
1st Qu.:0.000000
Median :0.000000
Mean :0.017020
3rd Qu.:0.002801
Max. :0.688303

old-balance-recipient
Min. :0.0000000
1st Qu.:0.0000000
Median :0.0003593
Mean :0.0030826
3rd Qu.:0.0026181
Max. :0.8749410



PCA

PCA	
Method	Randomized
Transform	Standardize
K	2
Threshold Level	15

PCA

Eigen Values

Var	PC1	PC2
Type	-0.261	-0.274
Amount	0.270	-0.315
Old-Balance-Customer	0.440	0.480
New-Balance-Customer	0.438	0.487
Old-Balance-Recipient	0.484	-0.402
New-Balance-Recipient	0.489	-0.444

PCA

