

Code ▾

Classification

Classification can be used to find linear relationship between input features and output class labels. The main focus for this predictive algorithm is to separate the data into different classes by making a boundary, so that it maximizes the margin between the classes. The strengths of classification include that it is very simple and easy to work with. The weaknesses would be that if the data is not linearly separable, it would not work.

Hide

```
set.seed(3)
data <- read.csv(file = 'desktop/adult.csv')
dim(data)
```

```
[1] 48842    15
```

Hide

```
str(data)
```

```
'data.frame':   48842 obs. of  15 variables:
 $ age          : int   25 38 28 44 18 34 29 63 24 55 ...
 $ workclass    : chr   "Private" "Private" "Local-gov" "Private" ...
 $ fnlwt       : int  226802 89814 336951 160323 103497 198693 227026 104626 369667 1
 04996 ...
 $ education    : chr   "11th" "HS-grad" "Assoc-acdm" "Some-college" ...
 $ educational.num: int    7  9 12 10 10  6  9 15 10  4 ...
 $ marital.status : chr   "Never-married" "Married-civ-spouse" "Married-civ-spouse" "Married-civ-spouse" ...
 $ occupation   : chr   "Machine-op-inspct" "Farming-fishing" "Protective-serv" "Machine-op-inspct" ...
 $ relationship : chr   "Own-child" "Husband" "Husband" "Husband" ...
 $ race         : chr   "Black" "White" "White" "Black" ...
 $ gender       : chr   "Male" "Male" "Male" "Male" ...
 $ capital.gain : int    0  0  0 7688  0  0  0 3103  0  0 ...
 $ capital.loss : int    0  0  0  0  0  0  0  0  0  0 ...
 $ hours.per.week : int   40 50 40 40 30 30 40 32 40 10 ...
 $ native.country : chr   "United-States" "United-States" "United-States" "United-States" ...
 $ income       : chr   "<=50K" "<=50K" ">50K" ">50K" ...
```

Hide

```
data$workclass<-as.factor(data$workclass)
data$education<-as.factor(data$education)
data$marital.status<-as.factor(data$marital.status)
data$occupation <-as.factor(data$occupation )
data$relationship <-as.factor(data$relationship )
data$race <-as.factor(data$race )
data$gender <-as.factor(data$gender )
data$native.country <-as.factor(data$native.country )
data$income <-as.factor(data$income )
```

Hide

```
str(data)
```

```
'data.frame': 48842 obs. of 15 variables:
 $ age          : int  25 38 28 44 18 34 29 63 24 55 ...
 $ workclass     : Factor w/ 9 levels "?","Federal-gov",...: 5 5 3 5 1 5 1 7 5 5 ...
 $ fnlwtgt      : int  226802 89814 336951 160323 103497 198693 227026 104626 369667 1
04996 ...
 $ education     : Factor w/ 16 levels "10th","11th",...: 2 12 8 16 16 1 12 15 16 6 ...
 $ educational.num: int   7 9 12 10 10 6 9 15 10 4 ...
 $ marital.status: Factor w/ 7 levels "Divorced","Married-AF-spouse",...: 5 3 3 3 5 5 5
3 5 3 ...
 $ occupation    : Factor w/ 15 levels "?","Adm-clerical",...: 8 6 12 8 1 9 1 11 9 4 ...
 $ relationship  : Factor w/ 6 levels "Husband","Not-in-family",...: 4 1 1 1 4 2 5 1 5 1
...
 $ race          : Factor w/ 5 levels "Amer-Indian-Eskimo",...: 3 5 5 3 5 5 3 5 5 5 ...
 $ gender        : Factor w/ 2 levels "Female","Male": 2 2 2 2 1 2 2 2 1 2 ...
 $ capital.gain  : int   0 0 0 7688 0 0 0 3103 0 0 ...
 $ capital.loss  : int   0 0 0 0 0 0 0 0 0 0 ...
 $ hours.per.week: int   40 50 40 40 30 30 40 32 40 10 ...
 $ native.country: Factor w/ 42 levels "?","Cambodia",...: 40 40 40 40 40 40 40 40 40 40
...
 $ income        : Factor w/ 2 levels "<=50K",">50K": 1 1 2 2 1 1 1 2 1 1 ...
```

Hide

```
set.seed(3)
i<-sample(1:nrow(data), nrow(data)*0.8, replace = FALSE)
train<-data[i,]
test<-data[-i,]
```

Hide

```
dim(train)
```

```
[1] 39073    15
```

Hide

```
head(train)
```

	a..	workclass	fnlwgt	education	educational.num	marital.status	
	<int>	<fctr>	<int>	<fctr>	<int>	<fctr>	▶
21479	63	Private	275034	Bachelors	13	Married-civ-spouse	
39610	31	Federal-gov	101345	HS-grad	9	Never-married	
19307	45	Self-emp-not-inc	94962	HS-grad	9	Married-civ-spouse	
41352	59	Private	205949	HS-grad	9	Separated	
33098	49	Local-gov	116163	Bachelors	13	Divorced	
36520	21	Private	40767	Some-college	10	Never-married	

6 rows | 1-7 of 15 columns

Hide

```
names(train)
```

```
[1] "age"           "workclass"      "fnlwgt"
[4] "education"     "educational.num" "marital.status"
[7] "occupation"    "relationship"    "race"
[10] "gender"        "capital.gain"    "capital.loss"
[13] "hours.per.week" "native.country"  "income"
```

Hide

```
str(train)
```

```
'data.frame':  39073 obs. of  15 variables:
 $ age           : int  63 31 45 59 49 21 26 53 48 24 ...
 $ workclass     : Factor w/  9 levels "?","Federal-gov",...: 5 2 7 5 3 5 1 5 5 5 ...
 $ fnlwgt        : int  275034 101345 94962 205949 116163 40767 182332 302868 226696 37
3628 ...
 $ education     : Factor w/ 16 levels "10th","11th",...: 10 12 12 12 10 16 9 10 12 10
...
 $ educational.num: int  13 9 9 9 13 10 11 13 9 13 ...
 $ marital.status : Factor w/  7 levels "Divorced","Married-AF-spouse",...: 3 5 3 6 1 5 3
3 3 5 ...
 $ occupation    : Factor w/ 15 levels "?","Adm-clerical",...: 13 7 4 4 11 7 1 13 4 13
...
 $ relationship  : Factor w/  6 levels "Husband","Not-in-family",...: 1 4 1 5 5 2 6 1 1 4
...
 $ race          : Factor w/  5 levels "Amer-Indian-Eskimo",...: 5 5 5 1 5 5 5 5 3 5 ...
 $ gender        : Factor w/  2 levels "Female","Male": 2 1 2 1 1 2 1 2 2 1 ...
 $ capital.gain   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ capital.loss   : int  1740 0 0 0 0 0 0 0 0 1504 ...
 $ hours.per.week : int  35 40 40 40 50 40 40 50 40 40 ...
 $ native.country : Factor w/ 42 levels "?","Cambodia",...: 40 40 10 40 11 40 40 40 40 40
...
 $ income        : Factor w/  2 levels "<=50K", ">50K": 1 1 1 1 1 1 1 2 1 1 ...
```

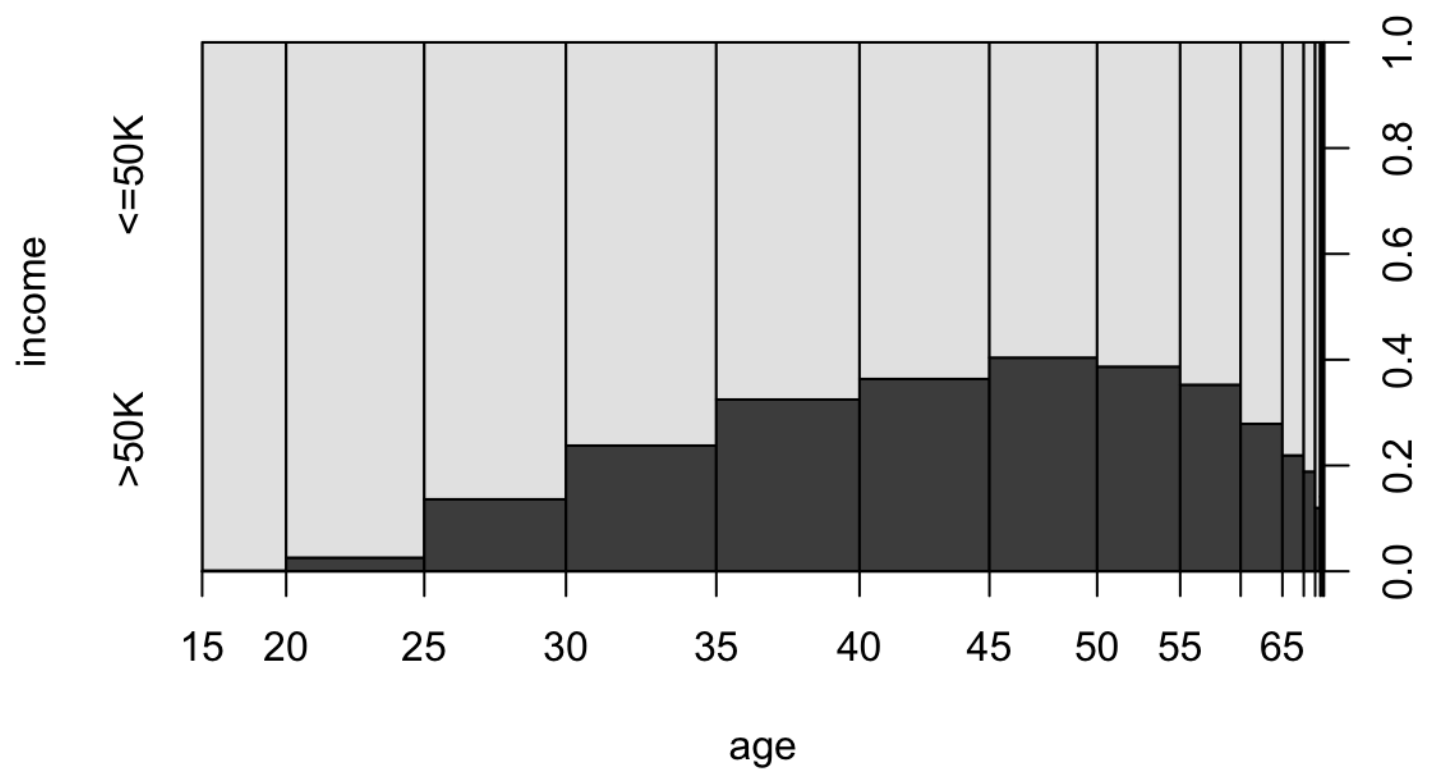
Hide

```
colSums(is.na(train))
```

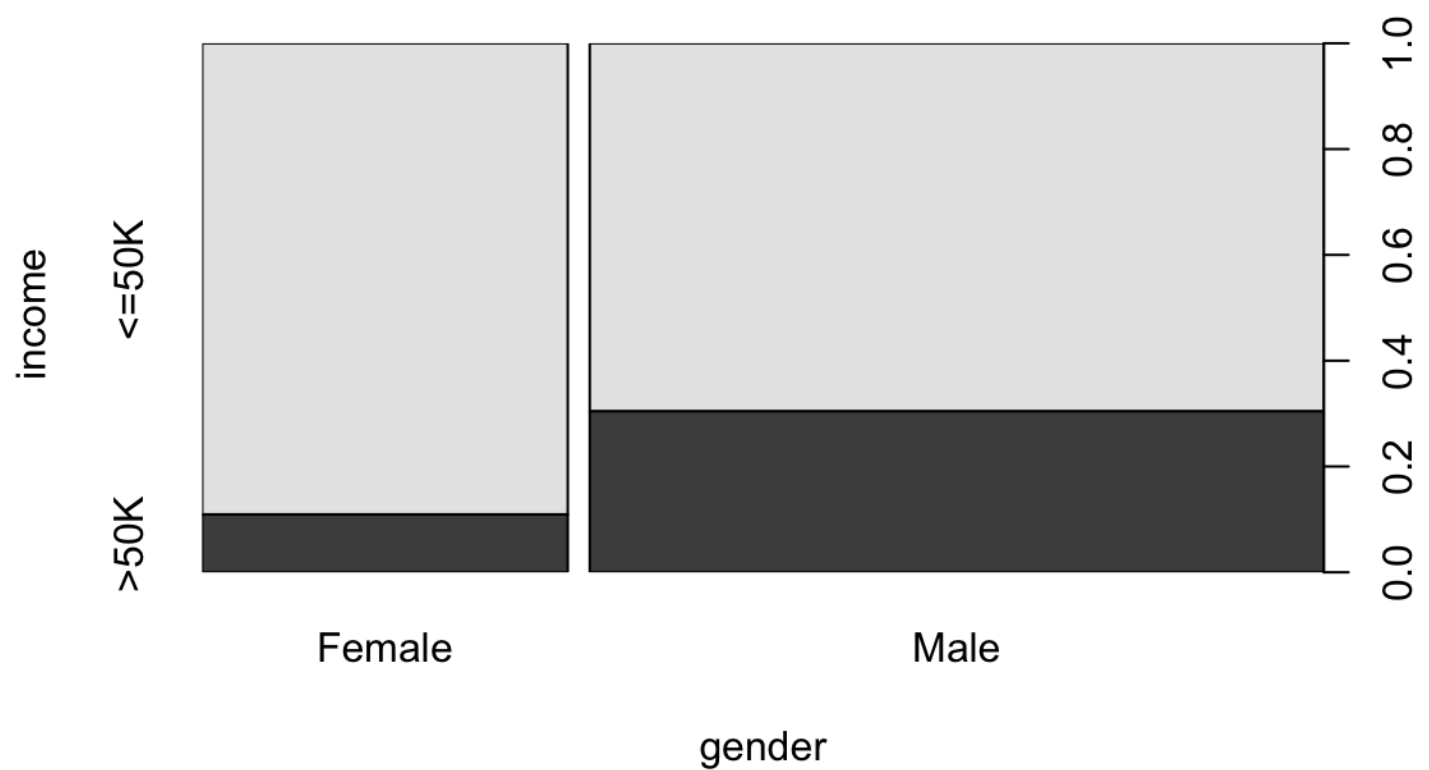
age	workclass	fnlwgt	education
0	0	0	0
educational.num	marital.status	occupation	relationship
0	0	0	0
race	gender	capital.gain	capital.loss
0	0	0	0
hours.per.week	native.country	income	
0	0	0	

Hide

```
plot(train$income~train$age, xlab="age", ylab = " income")
```

[Hide](#)

```
plot(train$income ~train$gender, xlab="gender", ylab = "income ")
```



[Hide](#)

```
glm1 <- glm(gender~., data = train, family = binomial)
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

[Hide](#)

```
summary(glm1)
```

Call:

```
glm(formula = gender ~ ., family = binomial, data = train)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7707	-0.4534	0.0000	0.3383	3.5843

Coefficients: (2 not defined because of singularities)

	Estimate	Std. Error	z value
(Intercept)	2.028e+01	1.365e+02	0.149
age	-3.339e-03	1.731e-03	-1.929
workclassFederal-gov	2.801e+00	1.740e-01	16.103
workclassLocal-gov	2.201e+00	1.565e-01	14.059
workclassNever-worked	1.140e+00	8.272e-01	1.378
workclassPrivate	2.418e+00	1.408e-01	17.171
workclassSelf-emp-inc	3.355e+00	1.939e-01	17.302
workclassSelf-emp-not-inc	3.095e+00	1.623e-01	19.069
workclassState-gov	2.476e+00	1.614e-01	15.343
workclassWithout-pay	1.317e+00	8.234e-01	1.600
fnlwgt	1.254e-06	1.553e-07	8.074
education11th	-8.481e-04	1.162e-01	-0.007
education12th	1.943e-01	1.520e-01	1.278
education1st-4th	5.099e-01	2.814e-01	1.812
education5th-6th	9.726e-03	2.187e-01	0.044
education7th-8th	9.312e-02	1.632e-01	0.571
education9th	9.070e-02	1.620e-01	0.560
educationAssoc-acdm	-8.564e-02	1.268e-01	-0.675
educationAssoc-voc	-2.089e-01	1.256e-01	-1.664
educationBachelors	5.145e-02	1.034e-01	0.498
educationDoctorate	1.684e-01	2.120e-01	0.794
educationHS-grad	-1.047e-01	9.526e-02	-1.099
educationMasters	-1.779e-01	1.256e-01	-1.417
educationPreschool	4.221e-01	3.476e-01	1.215
educationProf-school	4.024e-01	2.010e-01	2.002
educationSome-college	-1.255e-01	9.630e-02	-1.303
educational.num	NA	NA	NA
marital.statusMarried-AF-spouse	-7.704e+00	1.639e+02	-0.047
marital.statusMarried-civ-spouse	-7.687e-02	1.520e-01	-0.506
marital.statusMarried-spouse-absent	3.064e-01	1.164e-01	2.633
marital.statusNever-married	5.154e-01	4.836e-02	10.658
marital.statusSeparated	4.839e-02	7.908e-02	0.612
marital.statusWidowed	-9.298e-01	9.614e-02	-9.671
occupationAdm-clerical	-3.478e+00	1.366e-01	-25.470
occupationArmed-Forces	1.682e+01	4.586e+03	0.004
occupationCraft-repair	-1.573e-01	1.485e-01	-1.059
occupationExec-managerial	-2.670e+00	1.395e-01	-19.148
occupationFarming-fishing	-3.417e-01	1.958e-01	-1.745
occupationHandlers-cleaners	-6.594e-01	1.534e-01	-4.300
occupationMachine-op-inspct	-1.862e+00	1.420e-01	-13.118
occupationOther-service	-2.657e+00	1.347e-01	-19.724
occupationPriv-house-serv	-5.108e+00	3.631e-01	-14.069

occupationProf-specialty	-2.897e+00	1.412e-01	-20.521
occupationProtective-serv	-1.077e+00	1.889e-01	-5.703
occupationSales	-2.710e+00	1.366e-01	-19.839
occupationTech-support	-2.626e+00	1.551e-01	-16.923
occupationTransport-moving	NA	NA	NA
relationshipNot-in-family	-2.080e+01	1.365e+02	-0.152
relationshipOther-relative	-2.082e+01	1.365e+02	-0.153
relationshipOwn-child	-2.073e+01	1.365e+02	-0.152
relationshipUnmarried	-2.207e+01	1.365e+02	-0.162
relationshipWife	-2.750e+01	1.365e+02	-0.201
raceAsian-Pac-Islander	-1.003e-01	2.054e-01	-0.488
raceBlack	-3.807e-01	1.692e-01	-2.250
raceOther	-1.701e-01	2.326e-01	-0.731
raceWhite	-1.191e-01	1.630e-01	-0.730
capital.gain	-4.626e-06	3.652e-06	-1.267
capital.loss	1.181e-04	4.666e-05	2.531
hours.per.week	2.095e-02	1.497e-03	13.994
native.countryCambodia	-4.009e-01	7.767e-01	-0.516
native.countryCanada	-5.619e-01	2.965e-01	-1.895
native.countryChina	-1.475e-01	3.737e-01	-0.395
native.countryColumbia	-6.095e-01	3.986e-01	-1.529
native.countryCuba	-1.482e+00	4.032e-01	-3.675
native.countryDominican-Republic	-1.102e+00	3.570e-01	-3.087
native.countryEcuador	-7.992e-01	5.175e-01	-1.544
native.countryEl-Salvador	-1.806e-01	3.057e-01	-0.591
native.countryEngland	-3.491e-01	3.183e-01	-1.097
native.countryFrance	-4.140e-01	6.196e-01	-0.668
native.countryGermany	-7.471e-01	2.905e-01	-2.572
native.countryGreece	-2.061e-01	5.915e-01	-0.348
native.countryGuatemala	-1.135e-02	3.968e-01	-0.029
native.countryHaiti	-2.624e-01	4.261e-01	-0.616
native.countryHoland-Netherlands	-2.188e+01	1.773e+04	-0.001
native.countryHonduras	-3.087e-01	7.069e-01	-0.437
native.countryHong	-3.161e-01	1.078e+00	-0.293
native.countryHungary	-5.995e-01	8.251e-01	-0.727
native.countryIndia	1.003e+00	3.954e-01	2.538
native.countryIran	3.731e-01	6.345e-01	0.588
native.countryIreland	-8.196e-01	5.212e-01	-1.572
native.countryItaly	-1.198e+00	4.600e-01	-2.603
native.countryJamaica	-1.062e+00	3.679e-01	-2.886
native.countryJapan	-5.170e-01	4.087e-01	-1.265
native.countryLaos	-9.630e-01	9.069e-01	-1.062
native.countryMexico	-4.038e-01	1.835e-01	-2.200
native.countryNicaragua	-6.803e-01	4.929e-01	-1.380
native.countryOutlying-US (Guam-USVI-etc)	-5.478e-03	6.341e-01	-0.009
native.countryPeru	-6.755e-01	4.614e-01	-1.464
native.countryPhilippines	-5.483e-01	2.639e-01	-2.077
native.countryPoland	-1.747e-01	4.194e-01	-0.416
native.countryPortugal	7.227e-01	6.230e-01	1.160
native.countryPuerto-Rico	-6.627e-01	2.909e-01	-2.278
native.countryScotland	5.249e-01	7.043e-01	0.745
native.countrySouth	-7.442e-01	3.430e-01	-2.170

native.countryTaiwan	-2.950e-01	4.508e-01	-0.654
native.countryThailand	-7.250e-01	6.334e-01	-1.145
native.countryTrinidad&Tobago	-1.375e+00	9.261e-01	-1.485
native.countryUnited-States	-5.364e-01	1.265e-01	-4.241
native.countryVietnam	-5.280e-01	3.754e-01	-1.407
native.countryYugoslavia	-5.939e-01	8.821e-01	-0.673
income>50K	7.562e-01	7.401e-02	10.218
	Pr(> z)		
(Intercept)	0.881895		
age	0.053713	.	
workclassFederal-gov	< 2e-16	***	
workclassLocal-gov	< 2e-16	***	
workclassNever-worked	0.168088		
workclassPrivate	< 2e-16	***	
workclassSelf-emp-inc	< 2e-16	***	
workclassSelf-emp-not-inc	< 2e-16	***	
workclassState-gov	< 2e-16	***	
workclassWithout-pay	0.109610		
fnlwgt	6.79e-16	***	
education11th	0.994177		
education12th	0.201298		
education1st-4th	0.070003	.	
education5th-6th	0.964520		
education7th-8th	0.568241		
education9th	0.575465		
educationAssoc-acdm	0.499582		
educationAssoc-voc	0.096121	.	
educationBachelors	0.618745		
educationDoctorate	0.426945		
educationHS-grad	0.271797		
educationMasters	0.156580		
educationPreschool	0.224542		
educationProf-school	0.045297	*	
educationSome-college	0.192475		
educational.num	NA		
marital.statusMarried-AF-spouse	0.962499		
marital.statusMarried-civ-spouse	0.613010		
marital.statusMarried-spouse-absent	0.008463	**	
marital.statusNever-married	< 2e-16	***	
marital.statusSeparated	0.540607		
marital.statusWidowed	< 2e-16	***	
occupationAdm-clerical	< 2e-16	***	
occupationArmed-Forces	0.997074		
occupationCraft-repair	0.289516		
occupationExec-managerial	< 2e-16	***	
occupationFarming-fishing	0.080980	.	
occupationHandlers-cleaners	1.71e-05	***	
occupationMachine-op-inspct	< 2e-16	***	
occupationOther-service	< 2e-16	***	
occupationPriv-house-serv	< 2e-16	***	
occupationProf-specialty	< 2e-16	***	
occupationProtective-serv	1.18e-08	***	

occupationSales	< 2e-16 ***
occupationTech-support	< 2e-16 ***
occupationTransport-moving	NA
relationshipNot-in-family	0.878893
relationshipOther-relative	0.878786
relationshipOwn-child	0.879318
relationshipUnmarried	0.871578
relationshipWife	0.840367
raceAsian-Pac-Islander	0.625402
raceBlack	0.024428 *
raceOther	0.464601
raceWhite	0.465201
capital.gain	0.205291
capital.loss	0.011386 *
hours.per.week	< 2e-16 ***
native.countryCambodia	0.605762
native.countryCanada	0.058082 .
native.countryChina	0.693018
native.countryColumbia	0.126227
native.countryCuba	0.000238 ***
native.countryDominican-Republic	0.002020 **
native.countryEcuador	0.122476
native.countryEl-Salvador	0.554645
native.countryEngland	0.272830
native.countryFrance	0.504000
native.countryGermany	0.010105 *
native.countryGreece	0.727567
native.countryGuatemala	0.977181
native.countryHaiti	0.538110
native.countryHoland-Netherlands	0.999015
native.countryHonduras	0.662365
native.countryHong	0.769326
native.countryHungary	0.467470
native.countryIndia	0.011160 *
native.countryIran	0.556557
native.countryIreland	0.115842
native.countryItaly	0.009229 **
native.countryJamaica	0.003896 **
native.countryJapan	0.205814
native.countryLaos	0.288313
native.countryMexico	0.027819 *
native.countryNicaragua	0.167508
native.countryOutlying-US (Guam-USVI-etc)	0.993107
native.countryPeru	0.143202
native.countryPhilippines	0.037757 *
native.countryPoland	0.677046
native.countryPortugal	0.246067
native.countryPuerto-Rico	0.022712 *
native.countryScotland	0.456120
native.countrySouth	0.030036 *
native.countryTaiwan	0.512854
native.countryThailand	0.252392

```

native.countryTrinidad&Tobago      0.137659
native.countryUnited-States        2.23e-05 ***
native.countryVietnam              0.159523
native.countryYugoslavia           0.500778
income>50K                         < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 49692  on 39072  degrees of freedom
Residual deviance: 22946  on 38974  degrees of freedom
AIC: 23144

Number of Fisher Scoring iterations: 19

```

From the summary we notice that it is very significant. It means that there is a strong connection. the null deviance explains the deviance of the model that includes the intercept term. the residual deviance represent the eviance of the fitted model. Number of fisher Scoring iterations are used to estimate model parameters.

Hide

```
install.packages("e1071", dep = TRUE)
```

```

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/e1071_1.7-13.tgz'
Content type 'application/x-gzip' length 660266 bytes (644 KB)
=====
downloaded 644 KB

```

```

The downloaded binary packages are in
  /var/folders/fx/fht6tvb95xldy2488r10ctj00000gn/T//RtmpUjg3Ck/downloaded_packages

```

Hide

```

library(e1071)
nb1<-naiveBayes(gender~. data = train)

```

```
Error: unexpected symbol in "nb1<-naiveBayes(gender~. data"
```

Hide

```

library(e1071)
nb1<-naiveBayes(gender~., data = train)
nb1

```

Naive Bayes Classifier for Discrete Predictors

Call:

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

A-priori probabilities:

Y

	Female	Male
Y	0.332429	0.667571

Conditional probabilities:

age

Y	[,1]	[,2]
Female	36.87967	14.12651
Male	39.52247	13.46567

workclass

Y	?	Federal-gov	Local-gov	Never-worked	Private
Female	0.0775271383	0.0281776888	0.0765262915	0.0001539764	0.7151435830
Male	0.0471936820	0.0296733630	0.0578132188	0.0002300261	0.6826406993

workclass

Y	Self-emp-inc	Self-emp-not-inc	State-gov	Without-pay
Female	0.0131649858	0.0398798984	0.0488875202	0.0005389175
Male	0.0460818893	0.0985661708	0.0374559117	0.0003450391

fnlwgt

Y	[,1]	[,2]
Female	185075.7	103256.0
Male	191409.9	106696.3

education

Y	10th	11th	12th	1st-4th	5th-6th
Female	0.028639618	0.040418816	0.012857033	0.003695435	0.007621834
Male	0.028638246	0.035730716	0.013188161	0.005597301	0.011539641

education

Y	7th-8th	9th	Assoc-acdm	Assoc-voc	Bachelors
Female	0.014858727	0.013703903	0.039648934	0.044268227	0.152436677
Male	0.022274191	0.016293513	0.030018402	0.040292900	0.169222512

education

Y	Doctorate	HS-grad	Masters	Preschool	Prof-school
Female	0.006928940	0.313419047	0.052044037	0.001693741	0.007698822
Male	0.014951695	0.326905383	0.055934673	0.002146910	0.021277411

education

Y	Some-college
Female	0.260066210
Male	0.205988345

educational.num

Y	[,1]	[,2]
Female	10.03834	2.379056
Male	10.09282	2.665784

```

marital.status
Y      Divorced Married-AF-spouse Married-civ-spouse
Female 0.2443606128      0.0014627762      0.1530525830
Male   0.0798573838      0.0004217145      0.6094540715

marital.status
Y      Married-spouse-absent Never-married      Separated      Widowed
Female      0.0184771730      0.4487643391 0.0580491185 0.0758333975
Male      0.0104661862      0.2725042171 0.0184020856 0.0088943414

occupation
Y      ? Adm-clerical Armed-Forces Craft-repair
Female 0.0776811148 0.2340441912 0.0000000000 0.0197859727
Male   0.0474237080 0.0567781015 0.0004217145 0.1755482288

occupation
Y      Exec-managerial Farming-fishing Handlers-cleaners
Female      0.1073985680      0.0057741166      0.0163984910
Male      0.1329167306      0.0428615243      0.0550912437

occupation
Y      Machine-op-inspct Other-service Priv-house-serv Prof-specialty
Female      0.0498113789      0.1654476865      0.0138578797      0.1397336208
Male      0.0678576905      0.0688928079      0.0004600521      0.1205336605

occupation
Y      Protective-serv      Sales Tech-support Transport-moving
Female      0.0069289399 0.1217183771 0.0340287936      0.0073908692
Male      0.0262229719 0.1098374482 0.0272964269      0.0678576905

relationship
Y      Husband Not-in-family Other-relative      Own-child      Unmarried
Female 0.0000000000      0.361459697      0.041881592 0.210254831 0.242204943
Male   0.604048459      0.205336605      0.024957829 0.128967950 0.036574145

relationship
Y      Wife
Female 0.144198938
Male   0.000115013

race
Y      Amer-Indian-Eskimo Asian-Pac-Islander      Black      Other
Female      0.011240280      0.031950112 0.139964585 0.010008469
Male      0.008280939      0.030938506 0.071998160 0.007897562

race
Y      White
Female 0.806836554
Male   0.880884834

capital.gain
Y      [,1]      [,2]
Female 595.1062 5255.098
Male   1314.4604 8252.817

capital.loss
Y      [,1]      [,2]

```

	Female	64.58634	350.6348		
	Male	102.01725	433.5859		
		hours.per.week			
Y		[,1]	[,2]		
	Female	36.35984	11.89401		
	Male	42.37226	12.11403		
		native.country			
Y		?	Cambodia	Canada	China Columbia
	Female	1.470475e-02	3.079529e-04	4.080376e-03	2.232658e-03 1.847717e-03
	Male	1.855544e-02	7.667536e-04	3.450391e-03	2.760313e-03 1.763533e-03
		native.country			
Y		Cuba	Dominican-Republic	Ecuador	El-Salvador
	Female	3.002541e-03	2.771576e-03	1.077835e-03	3.079529e-03
	Male	2.300261e-03	1.840209e-03	9.201043e-04	2.952001e-03
		native.country			
Y		England	France	Germany	Greece Guatemala
	Female	3.002541e-03	6.928940e-04	5.081223e-03	5.389175e-04 1.616753e-03
	Male	2.415274e-03	7.284159e-04	3.642079e-03	1.341819e-03 1.840209e-03
		native.country			
Y		Haiti	Holand-Netherlands	Honduras	Hong
	Female	1.847717e-03	7.698822e-05	7.698822e-04	6.159058e-04
	Male	1.341819e-03	0.000000e+00	3.450391e-04	6.900782e-04
		native.country			
Y		Hungary	India	Iran	Ireland Italy
	Female	3.849411e-04	1.000847e-03	6.928940e-04	6.928940e-04 2.001694e-03
	Male	4.600521e-04	4.255482e-03	1.418494e-03	9.201043e-04 2.223585e-03
		native.country			
Y		Jamaica	Japan	Laos	Mexico Nicaragua
	Female	3.849411e-03	1.847717e-03	5.389175e-04	1.385788e-02 1.385788e-03
	Male	1.303481e-03	1.725196e-03	3.833768e-04	2.235087e-02 8.434289e-04
		native.country			
Y		Outlying-US(Guam-USVI-etc)	Peru	Philippines	Poland
	Female		4.619293e-04	1.231812e-03	6.851952e-03 1.539764e-03
	Male		4.600521e-04	9.584420e-04	5.252262e-03 1.993559e-03
		native.country			
Y		Portugal	Puerto-Rico	Scotland	South Taiwan
	Female	7.698822e-04	4.927246e-03	5.389175e-04	2.694588e-03 1.154823e-03
	Male	1.380156e-03	3.335378e-03	4.600521e-04	2.223585e-03 1.495169e-03
		native.country			
Y		Thailand	Trinidad&Tobago	United-States	Vietnam
	Female	7.698822e-04	6.928940e-04	9.023019e-01	2.078682e-03
	Male	4.983898e-04	2.683637e-04	8.959899e-01	1.648520e-03
		native.country			
Y		Yugoslavia			
	Female	3.849411e-04			
	Male	4.983898e-04			
		income			
Y		<=50K	>50K		

```
Female 0.8904458 0.1095542
Male   0.6951771 0.3048229
```

Hide

```
probs <- predict(glm1, newdata = test, type = "response")
```

Warning: prediction from a rank-deficient fit may be misleading

Hide

```
pred <- ifelse(probs>0.5,"male", "female")
acc <- mean (pred == test$gender)
print(paste("Accuracy = ", acc))
```

```
[1] "Accuracy = 0"
```

Hide

```
table (pred, test$gender)
```

```
pred      Female Male
female    2588  865
male      615 5701
```

Hide

```
install.packages('caret', dependencies = T)
```

```
also installing the dependencies 'bit', 'prettyunits', 'bit64', 'progress', 'listenv',
'parallelly', 'clipr', 'vroom', 'sys', 'future', 'globals', 'R.methodsS3', 'R.oo', 'R.ut
ils', 'forcats', 'hms', 'readr', 'bitops', 'colorspace', 'askpass', 'future.apply', 'num
Deriv', 'progressr', 'SQUAREM', 'httpuv', 'fontawesome', 'sourcetools', 'later', 'promis
es', 'R.cache', 'haven', 'caTools', 'TH.data', 'farver', 'labeling', 'munsell', 'RColorB
rewer', 'viridisLite', 'fansi', 'pkgconfig', 'tzdb', 'cpp11', 'timechange', 'profileMode
l', 'minqa', 'nloptr', 'RcppEigen', 'lazyeval', 'curl', 'openssl', 'utf8', 'plotrix', 'l
ava', 'shiny', 'miniUI', 'rstudioapi', 'styler', 'classInt', 'labelled', 'gplots', 'libc
oin', 'matrixStats', 'multcomp', 'rprojroot', 'diffobj', 'rematch2', 'gtable', 'isoban
d', 'scales', 'tibble', 'iterators', 'Rcpp', 'data.table', 'clock', 'generics', 'gower',
'hardhat', 'lubridate', 'purrr', 'tidyr', 'tidyselect', 'timeDate', 'brglm', 'gtools',
'lme4', 'qvcalc', 'rex', 'httr', 'crayon', 'pillar', 'Formula', 'plotmo', 'TeachingDemo
s', 'prodlim', 'combinat', 'questionr', 'ROCR', 'mvtnorm', 'modeltools', 'strucchange',
'coin', 'sandwich', 'ISwR', 'corpcor', 'brio', 'callr', 'desc', 'pkgload', 'praise', 'pr
ocessx', 'ps', 'waldo', 'ROSE', 'ggplot2', 'foreach', 'ModelMetrics', 'plyr', 'pROC', 'r
ecipes', 'reshape2', 'withr', 'BradleyTerry2', 'covr', 'Cubist', 'dplyr', 'earth', 'elli
pse', 'fastICA', 'gam', 'ipred', 'kernlab', 'klaR', 'mda', 'MLmetrics', 'pamr', 'party',
'pls', 'RANN', 'spls', 'subselect', 'superpc', 'testthat', 'themis'
```

There is a binary version available but the source version is later:

	binary <chr>	source <chr>	needs_compilation <lgl>
recipes	1.0.4	1.0.5	FALSE
1 row			


```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/bit_4.0.5.tgz'
Content type 'application/x-gzip' length 1241800 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/prettyunits_1.1.1.tgz'
Content type 'application/x-gzip' length 35099 bytes (34 KB)
=====
downloaded 34 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/bit64_4.0.5.tgz'
Content type 'application/x-gzip' length 577944 bytes (564 KB)
=====
downloaded 564 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/progress_1.2.2.tgz'
Content type 'application/x-gzip' length 83891 bytes (81 KB)
=====
downloaded 81 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/listenv_0.9.0.tgz'
Content type 'application/x-gzip' length 104936 bytes (102 KB)
=====
downloaded 102 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/parallelly_1.34.0.tgz'
Content type 'application/x-gzip' length 325182 bytes (317 KB)
=====
downloaded 317 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/clipr_0.8.0.tgz'
Content type 'application/x-gzip' length 51002 bytes (49 KB)
=====
downloaded 49 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/vroom_1.6.1.tgz'
Content type 'application/x-gzip' length 2741452 bytes (2.6 MB)
=====
downloaded 2.6 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/sys_3.4.1.tgz'
Content type 'application/x-gzip' length 46958 bytes (45 KB)
=====
downloaded 45 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/future_1.31.0.tgz'
Content type 'application/x-gzip' length 605943 bytes (591 KB)
=====
downloaded 591 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/globals_0.16.2.tgz'
Content type 'application/x-gzip' length 105158 bytes (102 KB)
```

```
=====
downloaded 102 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/R.methodsS3_1.8.2.tgz'
Content type 'application/x-gzip' length 80414 bytes (78 KB)
=====
downloaded 78 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/R.oo_1.25.0.tgz'
Content type 'application/x-gzip' length 964424 bytes (941 KB)
=====
downloaded 941 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/R.utils_2.12.2.tgz'
Content type 'application/x-gzip' length 1405515 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/forcats_1.0.0.tgz'
Content type 'application/x-gzip' length 422160 bytes (412 KB)
=====
downloaded 412 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/hms_1.1.2.tgz'
Content type 'application/x-gzip' length 97496 bytes (95 KB)
=====
downloaded 95 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/readr_2.1.4.tgz'
Content type 'application/x-gzip' length 1856819 bytes (1.8 MB)
=====
downloaded 1.8 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/bitops_1.0-7.tgz'
Content type 'application/x-gzip' length 29377 bytes (28 KB)
=====
downloaded 28 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/colourspace_2.1-0.tgz'
Content type 'application/x-gzip' length 2621598 bytes (2.5 MB)
=====
downloaded 2.5 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/askpass_1.1.tgz'
Content type 'application/x-gzip' length 21686 bytes (21 KB)
=====
downloaded 21 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/future.apply_1.10.0.tgz'
Content type 'application/x-gzip' length 152629 bytes (149 KB)
=====
downloaded 149 KB
```

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/numDeriv_2016.8-1.1.tgz'
Content type 'application/x-gzip' length 113250 bytes (110 KB)
=====
downloaded 110 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/progressr_0.13.0.tgz'
Content type 'application/x-gzip' length 376161 bytes (367 KB)
=====
downloaded 367 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/SQUAREM_2021.1.tgz'
Content type 'application/x-gzip' length 176568 bytes (172 KB)
=====
downloaded 172 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/httpuv_1.6.9.tgz'
Content type 'application/x-gzip' length 2756435 bytes (2.6 MB)
=====
downloaded 2.6 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/fontawesome_0.5.0.tgz'
Content type 'application/x-gzip' length 1375186 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/sourcetools_0.1.7-1.tgz'
Content type 'application/x-gzip' length 139039 bytes (135 KB)
=====
downloaded 135 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/later_1.3.0.tgz'
Content type 'application/x-gzip' length 623657 bytes (609 KB)
=====
downloaded 609 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/promises_1.2.0.1.tgz'
Content type 'application/x-gzip' length 1788376 bytes (1.7 MB)
=====
downloaded 1.7 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/R.cache_0.16.0.tgz'
Content type 'application/x-gzip' length 108810 bytes (106 KB)
=====
downloaded 106 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/haven_2.5.1.tgz'
Content type 'application/x-gzip' length 1046146 bytes (1021 KB)
=====
downloaded 1021 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/caTools_1.18.2.tgz'
```

```
Content type 'application/x-gzip' length 243206 bytes (237 KB)
=====
downloaded 237 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/TH.data_1.1-1.tgz'
Content type 'application/x-gzip' length 8792206 bytes (8.4 MB)
=====
downloaded 8.4 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/farver_2.1.1.tgz'
Content type 'application/x-gzip' length 1943238 bytes (1.9 MB)
=====
downloaded 1.9 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/labeling_0.4.2.tgz'
Content type 'application/x-gzip' length 60389 bytes (58 KB)
=====
downloaded 58 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/munsell_0.5.0.tgz'
Content type 'application/x-gzip' length 241895 bytes (236 KB)
=====
downloaded 236 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/RColorBrewer_1.1-3.tgz'
Content type 'application/x-gzip' length 53152 bytes (51 KB)
=====
downloaded 51 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/viridisLite_0.4.1.tgz'
Content type 'application/x-gzip' length 1297022 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/fansi_1.0.4.tgz'
Content type 'application/x-gzip' length 364195 bytes (355 KB)
=====
downloaded 355 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pkgconfig_2.0.3.tgz'
Content type 'application/x-gzip' length 17870 bytes (17 KB)
=====
downloaded 17 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/tzdb_0.3.0.tgz'
Content type 'application/x-gzip' length 1243874 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/cpp11_0.4.3.tgz'
Content type 'application/x-gzip' length 296464 bytes (289 KB)
=====
```

downloaded 289 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/timechange_0.2.0.tgz'
Content type 'application/x-gzip' length 842573 bytes (822 KB)

downloaded 822 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/profileModel_0.6.1.tgz'
Content type 'application/x-gzip' length 104325 bytes (101 KB)

downloaded 101 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/minqa_1.2.5.tgz'
Content type 'application/x-gzip' length 349364 bytes (341 KB)

downloaded 341 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/nloptr_2.0.3.tgz'
Content type 'application/x-gzip' length 1949528 bytes (1.9 MB)

downloaded 1.9 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/RcppEigen_0.3.3.9.3.tgz'
Content type 'application/x-gzip' length 4692990 bytes (4.5 MB)

downloaded 4.5 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/lazyeval_0.2.2.tgz'
Content type 'application/x-gzip' length 157457 bytes (153 KB)

downloaded 153 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/curl_5.0.0.tgz'
Content type 'application/x-gzip' length 758776 bytes (740 KB)

downloaded 740 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/openssl_2.0.5.tgz'
Content type 'application/x-gzip' length 2880572 bytes (2.7 MB)

downloaded 2.7 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/utf8_1.2.3.tgz'
Content type 'application/x-gzip' length 196823 bytes (192 KB)

downloaded 192 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/plotrix_3.8-2.tgz'
Content type 'application/x-gzip' length 1137125 bytes (1.1 MB)

downloaded 1.1 MB

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/lava_1.7.1.tgz'
Content type 'application/x-gzip' length 2652541 bytes (2.5 MB)
=====
downloaded 2.5 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/shiny_1.7.4.tgz'
Content type 'application/x-gzip' length 4262326 bytes (4.1 MB)
=====
downloaded 4.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/miniUI_0.1.1.1.tgz'
Content type 'application/x-gzip' length 34281 bytes (33 KB)
=====
downloaded 33 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/rstudioapi_0.14.tgz'
Content type 'application/x-gzip' length 291350 bytes (284 KB)
=====
downloaded 284 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/styler_1.9.0.tgz'
Content type 'application/x-gzip' length 719018 bytes (702 KB)
=====
downloaded 702 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/classInt_0.4-8.tgz'
Content type 'application/x-gzip' length 496565 bytes (484 KB)
=====
downloaded 484 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/labelled_2.10.0.tgz'
Content type 'application/x-gzip' length 289216 bytes (282 KB)
=====
downloaded 282 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/gplots_3.1.3.tgz'
Content type 'application/x-gzip' length 593253 bytes (579 KB)
=====
downloaded 579 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/libcoin_1.0-9.tgz'
Content type 'application/x-gzip' length 1042970 bytes (1018 KB)
=====
downloaded 1018 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/matrixStats_0.63.0.tgz'
Content type 'application/x-gzip' length 592779 bytes (578 KB)
=====
downloaded 578 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/multcomp_1.4-22.tgz'
Content type 'application/x-gzip' length 737201 bytes (719 KB)
```

=====
downloaded 719 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/rprojroot_2.0.3.tgz'
Content type 'application/x-gzip' length 99592 bytes (97 KB)

=====
downloaded 97 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/diffobj_0.3.5.tgz'
Content type 'application/x-gzip' length 1004587 bytes (981 KB)

=====
downloaded 981 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/rematch2_2.1.2.tgz'
Content type 'application/x-gzip' length 45040 bytes (43 KB)

=====
downloaded 43 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/gtable_0.3.1.tgz'
Content type 'application/x-gzip' length 158928 bytes (155 KB)

=====
downloaded 155 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/isoband_0.2.7.tgz'
Content type 'application/x-gzip' length 1869775 bytes (1.8 MB)

=====
downloaded 1.8 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/scales_1.2.1.tgz'
Content type 'application/x-gzip' length 608312 bytes (594 KB)

=====
downloaded 594 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/tibble_3.1.8.tgz'
Content type 'application/x-gzip' length 720020 bytes (703 KB)

=====
downloaded 703 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/iterators_1.0.14.tgz'
Content type 'application/x-gzip' length 345398 bytes (337 KB)

=====
downloaded 337 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/Rcpp_1.0.10.tgz'
Content type 'application/x-gzip' length 3240791 bytes (3.1 MB)

=====
downloaded 3.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/data.table_1.14.8.tgz'
Content type 'application/x-gzip' length 2375057 bytes (2.3 MB)

=====
downloaded 2.3 MB

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/clock_0.6.1.tgz'
Content type 'application/x-gzip' length 8540466 bytes (8.1 MB)
=====
downloaded 8.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/generics_0.1.3.tgz'
Content type 'application/x-gzip' length 77771 bytes (75 KB)
=====
downloaded 75 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/gower_1.0.1.tgz'
Content type 'application/x-gzip' length 214341 bytes (209 KB)
=====
downloaded 209 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/hardhat_1.2.0.tgz'
Content type 'application/x-gzip' length 795388 bytes (776 KB)
=====
downloaded 776 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/lubridate_1.9.2.tgz'
Content type 'application/x-gzip' length 979863 bytes (956 KB)
=====
downloaded 956 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/purrr_1.0.1.tgz'
Content type 'application/x-gzip' length 510124 bytes (498 KB)
=====
downloaded 498 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/tidyr_1.3.0.tgz'
Content type 'application/x-gzip' length 1324272 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/tidysselect_1.2.0.tgz'
Content type 'application/x-gzip' length 218996 bytes (213 KB)
=====
downloaded 213 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/timeDate_4022.108.tgz'
Content type 'application/x-gzip' length 1377548 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/brglm_0.7.2.tgz'
Content type 'application/x-gzip' length 124880 bytes (121 KB)
=====
downloaded 121 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/gtools_3.9.4.tgz'
```



```
Content type 'application/x-gzip' length 356179 bytes (347 KB)
=====
downloaded 347 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/lme4_1.1-31.tgz'
Content type 'application/x-gzip' length 7082646 bytes (6.8 MB)
=====
downloaded 6.8 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/qvcalc_1.0.2.tgz'
Content type 'application/x-gzip' length 44342 bytes (43 KB)
=====
downloaded 43 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/rex_1.2.1.tgz'
Content type 'application/x-gzip' length 122412 bytes (119 KB)
=====
downloaded 119 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/htrr_1.4.4.tgz'
Content type 'application/x-gzip' length 502987 bytes (491 KB)
=====
downloaded 491 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/crayon_1.5.2.tgz'
Content type 'application/x-gzip' length 161352 bytes (157 KB)
=====
downloaded 157 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pillar_1.8.1.tgz'
Content type 'application/x-gzip' length 673948 bytes (658 KB)
=====
downloaded 658 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/Formula_1.2-4.tgz'
Content type 'application/x-gzip' length 175041 bytes (170 KB)
=====
downloaded 170 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/plotmo_3.6.2.tgz'
Content type 'application/x-gzip' length 1555460 bytes (1.5 MB)
=====
downloaded 1.5 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/TeachingDemos_2.12.tgz'
Content type 'application/x-gzip' length 1232382 bytes (1.2 MB)
=====
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/prodlim_2019.11.13.tgz'
Content type 'application/x-gzip' length 422363 bytes (412 KB)
=====
```

downloaded 412 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/combinat_0.0-8.tgz'
Content type 'application/x-gzip' length 40732 bytes (39 KB)

downloaded 39 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/questionr_0.7.8.tgz'
Content type 'application/x-gzip' length 3888081 bytes (3.7 MB)

downloaded 3.7 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ROCR_1.0-11.tgz'
Content type 'application/x-gzip' length 451112 bytes (440 KB)

downloaded 440 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/mvtnorm_1.1-3.tgz'
Content type 'application/x-gzip' length 260575 bytes (254 KB)

downloaded 254 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/modeltools_0.2-23.tgz'
Content type 'application/x-gzip' length 214665 bytes (209 KB)

downloaded 209 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/strucchange_1.5-3.tgz'
Content type 'application/x-gzip' length 944267 bytes (922 KB)

downloaded 922 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/coin_1.4-2.tgz'
Content type 'application/x-gzip' length 1455790 bytes (1.4 MB)

downloaded 1.4 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/sandwich_3.0-2.tgz'
Content type 'application/x-gzip' length 1454087 bytes (1.4 MB)

downloaded 1.4 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ISwR_2.0-8.tgz'
Content type 'application/x-gzip' length 208941 bytes (204 KB)

downloaded 204 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/corpcor_1.6.10.tgz'
Content type 'application/x-gzip' length 113556 bytes (110 KB)

downloaded 110 KB

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/brio_1.1.3.tgz'
Content type 'application/x-gzip' length 40561 bytes (39 KB)
=====
downloaded 39 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/callr_3.7.3.tgz'
Content type 'application/x-gzip' length 428266 bytes (418 KB)
=====
downloaded 418 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/desc_1.4.2.tgz'
Content type 'application/x-gzip' length 326015 bytes (318 KB)
=====
downloaded 318 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pkgload_1.3.2.tgz'
Content type 'application/x-gzip' length 176099 bytes (171 KB)
=====
downloaded 171 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/praise_1.0.0.tgz'
Content type 'application/x-gzip' length 16253 bytes (15 KB)
=====
downloaded 15 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/processx_3.8.0.tgz'
Content type 'application/x-gzip' length 309385 bytes (302 KB)
=====
downloaded 302 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ps_1.7.2.tgz'
Content type 'application/x-gzip' length 309375 bytes (302 KB)
=====
downloaded 302 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/waldo_0.4.0.tgz'
Content type 'application/x-gzip' length 100194 bytes (97 KB)
=====
downloaded 97 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ROSE_0.0-4.tgz'
Content type 'application/x-gzip' length 111585 bytes (108 KB)
=====
downloaded 108 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ggplot2_3.4.1.tgz'
Content type 'application/x-gzip' length 4222110 bytes (4.0 MB)
=====
downloaded 4.0 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/foreach_1.5.2.tgz'
Content type 'application/x-gzip' length 136295 bytes (133 KB)
```

```
=====
downloaded 133 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ModelMetrics_1.2.2.2.tgz'
Content type 'application/x-gzip' length 557661 bytes (544 KB)
=====
downloaded 544 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/plyr_1.8.8.tgz'
Content type 'application/x-gzip' length 1015309 bytes (991 KB)
=====
downloaded 991 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pROC_1.18.0.tgz'
Content type 'application/x-gzip' length 1121281 bytes (1.1 MB)
=====
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/reshape2_1.4.4.tgz'
Content type 'application/x-gzip' length 333392 bytes (325 KB)
=====
downloaded 325 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/withr_2.5.0.tgz'
Content type 'application/x-gzip' length 229317 bytes (223 KB)
=====
downloaded 223 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/BradleyTerry2_1.1-2.tgz'
Content type 'application/x-gzip' length 593387 bytes (579 KB)
=====
downloaded 579 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/covr_3.6.1.tgz'
Content type 'application/x-gzip' length 325440 bytes (317 KB)
=====
downloaded 317 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/Cubist_0.4.2.tgz'
Content type 'application/x-gzip' length 986973 bytes (963 KB)
=====
downloaded 963 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/dplyr_1.1.0.tgz'
Content type 'application/x-gzip' length 1554483 bytes (1.5 MB)
=====
downloaded 1.5 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/earth_5.3.2.tgz'
Content type 'application/x-gzip' length 2146244 bytes (2.0 MB)
=====
downloaded 2.0 MB
```

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ellipse_0.4.3.tgz'
Content type 'application/x-gzip' length 73594 bytes (71 KB)
=====
downloaded 71 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/fastICA_1.2-3.tgz'
Content type 'application/x-gzip' length 68416 bytes (66 KB)
=====
downloaded 66 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/gam_1.22-1.tgz'
Content type 'application/x-gzip' length 400719 bytes (391 KB)
=====
downloaded 391 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ipred_0.9-13.tgz'
Content type 'application/x-gzip' length 384988 bytes (375 KB)
=====
downloaded 375 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/kernlab_0.9-32.tgz'
Content type 'application/x-gzip' length 2504920 bytes (2.4 MB)
=====
downloaded 2.4 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/klaR_1.7-1.tgz'
Content type 'application/x-gzip' length 566873 bytes (553 KB)
=====
downloaded 553 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/mda_0.5-3.tgz'
Content type 'application/x-gzip' length 925130 bytes (903 KB)
=====
downloaded 903 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/MLmetrics_1.1.1.tgz'
Content type 'application/x-gzip' length 73834 bytes (72 KB)
=====
downloaded 72 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pamr_1.56.1.tgz'
Content type 'application/x-gzip' length 820865 bytes (801 KB)
=====
downloaded 801 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/party_1.3-11.tgz'
Content type 'application/x-gzip' length 964542 bytes (941 KB)
=====
downloaded 941 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/pls_2.8-1.tgz'
```

```
Content type 'application/x-gzip' length 1176618 bytes (1.1 MB)
=====
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/RANN_2.6.1.tgz'
Content type 'application/x-gzip' length 207881 bytes (203 KB)
=====
downloaded 203 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/spls_2.2-3.tgz'
Content type 'application/x-gzip' length 2502985 bytes (2.4 MB)
=====
downloaded 2.4 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/subselect_0.15.4.tgz'
Content type 'application/x-gzip' length 1368848 bytes (1.3 MB)
=====
downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/superpc_1.12.tgz'
Content type 'application/x-gzip' length 290918 bytes (284 KB)
=====
downloaded 284 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/testthat_3.1.6.tgz'
Content type 'application/x-gzip' length 2988406 bytes (2.8 MB)
=====
downloaded 2.8 MB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/themis_1.0.0.tgz'
Content type 'application/x-gzip' length 233686 bytes (228 KB)
=====
downloaded 228 KB

trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/caret_6.0-93.tgz'
Content type 'application/x-gzip' length 3578405 bytes (3.4 MB)
=====
downloaded 3.4 MB
```

The downloaded binary packages are in
/var/folders/fx/fht6tvb95xldy2488rl0ctj00000gn/T//RtmpUjg3Ck/downloaded_packages

```
installing the source package 'recipes'
```

```
trying URL 'https://cran.rstudio.com/src/contrib/recipes_1.0.5.tar.gz'
```

```
Content type 'application/x-gzip' length 799374 bytes (780 KB)
```

```
=====
```

```
downloaded 780 KB
```

```
* installing *source* package 'recipes' ...
```

```
** package 'recipes' successfully unpacked and MD5 sums checked
```

```
** using staged installation
```

```
** R
```

```
** inst
```

```
** byte-compile and prepare package for lazy loading
```

```
** help
```

```
*** installing help indices
```

```
*** copying figures
```

```
** building package indices
```

```
** installing vignettes
```

```
** testing if installed package can be loaded from temporary location
```

```
** testing if installed package can be loaded from final location
```

```
** testing if installed package keeps a record of temporary installation path
```

```
* DONE (recipes)
```

```
The downloaded source packages are in
```

```
  '/private/var/folders/fx/fht6tvb95xldy2488rl0ctj00000gn/T/RtmpUjg3Ck/downloaded_packages'
```

[Hide](#)

```
install.packages("ROCR")
```

```
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/ROCR_1.0-11.tgz'
```

```
Content type 'application/x-gzip' length 451112 bytes (440 KB)
```

```
=====
```

```
downloaded 440 KB
```

```
The downloaded binary packages are in
```

```
  /var/folders/fx/fht6tvb95xldy2488rl0ctj00000gn/T//RtmpUjg3Ck/downloaded_packages
```

[Hide](#)

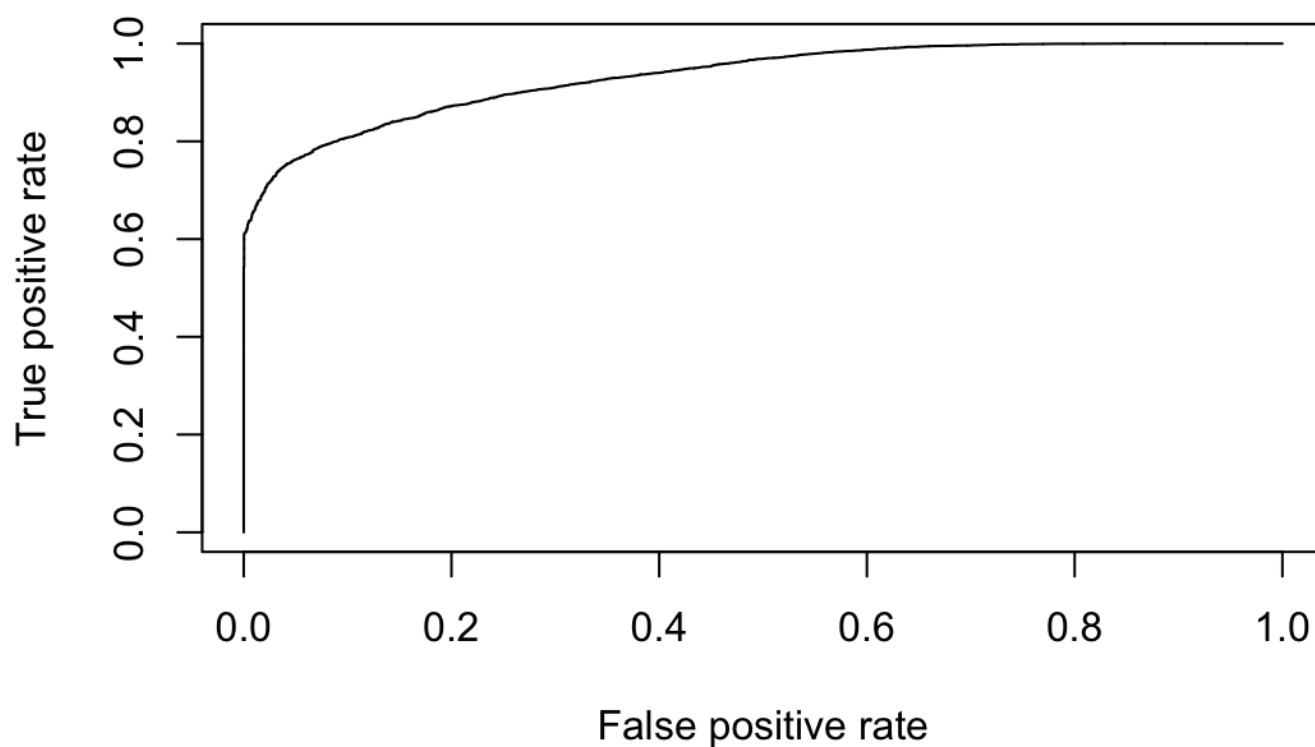
```
library(ROCR)
```

```
p <- predict(glm1, newdata = test, type = "response")
```

```
Warning: prediction from a rank-deficient fit may be misleading
```

[Hide](#)

```
pr <- prediction(p, test$gender)
prf <- performance(pr, measure = "tpr", x.measure = "fpr")
plot(prf)
```



Hide

```
auc <- performance(pr, measure = "auc")
auc <- auc@y.values[[1]]
auc
```

```
[1] 0.9331228
```

Hide

```
p1 <- predict(nbl, newdata = test, type="class")
table(p1, test$gender)
```

```
p1      Female Male
Female  2955 1756
Male    248 4810
```

Hide

```
mean(p1==test$gender)
```



```
[1] 0.7948613
```

Output: There are many strengths of using Naive Bayes which includes the simplicity and efficiency when using high dimensional data. It does not require high training data and handles categorical and numerical values. However one disadvantage is that it assumes the input features are independent which is not true all the time. There are also a lot of strengths and advantages while using logistic regression. One that stands out the most is that it provides a clear understanding of each input feature of the output. However one disadvantage is that it assumes a linear relationship between the input features and the output which is not always the same. I used accuracy in this project. It is very easy to understand, but it may not be suitable for imbalanced datasets. I also used precision which measures the number of right predictions out of all the predictions. However it is only useful when the cost of false negatives are high.