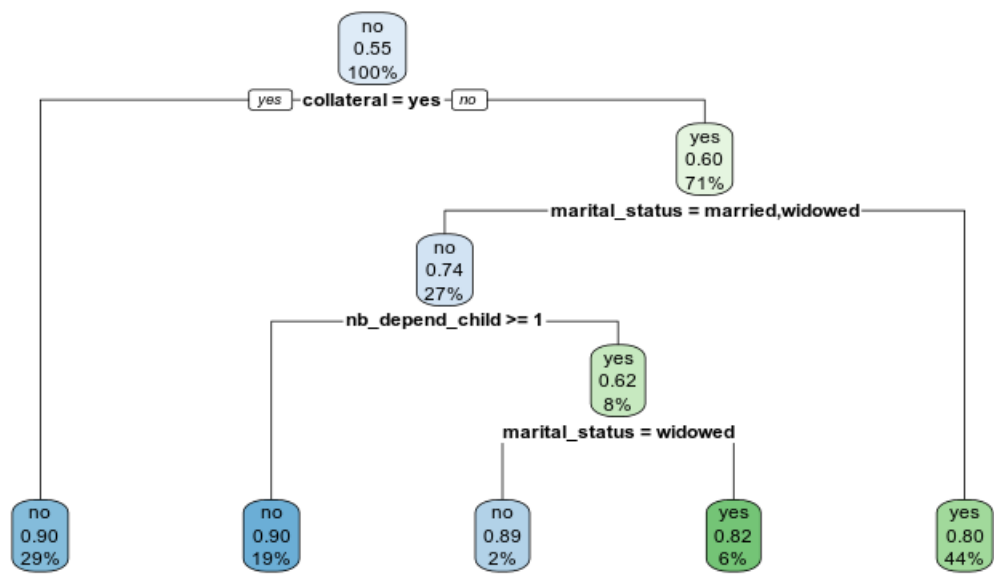


```
library(FSelector)
d <- read.csv("Loans.csv")
weights <- information.gain(writeoff~.,d,unit="log2")
print(weights)
```

```
library(rpart)
class_tree <- rpart(writeoff~.,method="class", data=d)
library(rpart.plot)
rpart.plot(class_tree,uniform=TRUE,main="Tree of loans", extra=108)
```

Tree of loans



```
library(e1071)
d2 <- read.csv("Loans_test.csv")
totalData <- rbind(d, d2)
for (f in 1:length(names(totalData))) {
  levels(d2[, f]) <- levels(totalData[, f])
}
m <- svm(writeoff~.,data=d, kernel='linear', scale=TRUE)
writeoff_pred <- predict(m,d2)
writeoff_pred
```

```
1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
28 29 30
no no yes yes yes no yes yes no no yes no no yes no yes yes yes no no no no no yes yes no no y
es yes no
Levels: no yes
```

```
mymodel <- glm(writeoff~.,data=d,family = binomial)
summary(mymodel)
```

```

Call:
glm(formula = writeoff ~ ., family = binomial, data = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2289  -0.6639  -0.1906   0.7136   3.5383

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.804e+00  1.120e-01  16.109 < 2e-16 ***
genderM         1.045e-02  2.980e-02   0.351  0.7258
age           -2.977e-03  3.306e-03  -0.901  0.3678
marital_statusmarried -2.074e+00  7.432e-02 -27.906 < 2e-16 ***
marital_statussingle -6.315e-01  4.517e-02 -13.979 < 2e-16 ***
marital_statuswidowed -3.146e+00  6.817e-02 -46.147 < 2e-16 ***
educationhighsch  -3.605e-02  4.492e-02  -0.802  0.4223
educationpostgrad -4.162e-02  1.099e-01  -0.379  0.7050
educationuniv     6.904e-02  5.887e-02   1.173  0.2409
nb_depend_child  -5.074e-01  1.909e-02 -26.577 < 2e-16 ***
employ_statuspart_time  3.797e-01  4.975e-02   7.633 2.29e-14 ***
employ_statusretired  2.094e+00  3.132e+00   0.669  0.5038
employ_statusself_employ  2.704e-01  4.401e-02   6.145 8.00e-10 ***
employ_statusunemployed  4.947e-01  7.853e-02   6.299 2.99e-10 ***
yrs_current_job    1.276e-02  6.492e-03   1.966  0.0493 *
yrs_employed      -1.172e-03  4.744e-03  -0.247  0.8048
net_income         3.178e-06  1.323e-06   2.402  0.0163 *
spouse_workyes     -5.669e-02  1.028e-01  -0.551  0.5814
spouse_income      1.830e-06  1.767e-06   1.035  0.3004
residential_statusowner_morg -2.456e-02  5.001e-02  -0.491  0.6234
residential_statustenant  2.564e-03  4.671e-02   0.055  0.9562
residential_statusw_parents  3.553e-01  2.315e-01   1.535  0.1248
yrs_current_address -2.884e-03  4.012e-03  -0.719  0.4722
loan_amount       -5.307e-07  9.328e-07  -0.569  0.5694
loan_purposeholiday  -4.352e-02  8.610e-02  -0.505  0.6132
loan_purposehome_improv -9.985e-02  6.417e-02  -1.556  0.1197
loan_purposeother    -3.956e-02  4.038e-02  -0.980  0.3272
loan_purposevehicle  -3.605e-01  4.744e-02  -7.599 2.98e-14 ***
loan_purposewedding  -3.724e-01  5.179e-02  -7.190 6.50e-13 ***
loan_length       -3.060e-04  1.004e-03  -0.305  0.7606
collateralyes     -2.975e+00  4.803e-02 -61.940 < 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: 41334  on 29999  degrees of freedom
Residual deviance: 27888  on 29969  degrees of freedom
AIC: 27950

Number of Fisher Scoring iterations: 5

```

Hide

```
predict.glm(mymodel,d2,type="response")
```

```

      1      2      3      4      5      6      7      8      9
10      11      12      13      14
0.16307279 0.18390052 0.74803410 0.84014846 0.78579976 0.22342371 0.71956548 0.67242284 0.09639258 0.1895009
9 0.87138687 0.26591336 0.31189128 0.72566721
      15      16      17      18      19      20      21      22      23
24      25      26      27      28
0.15506935 0.75633247 0.78124483 0.80326330 0.21100289 0.22396596 0.28857575 0.17058339 0.11828716 0.8176637
2 0.60873945 0.18361894 0.15158788 0.80470211
      29      30
0.78223969 0.06163769

```

Hide

```
myprobs=predict(mymodel,d2,type="response")
mypred=rep("NO",nrow(d2))
mypred[myprobs>0.5] = "YES"
mypred
```

```
[1] "NO" "NO" "YES" "YES" "YES" "NO" "YES" "YES" "NO" "NO" "YES" "NO" "NO" "YES" "NO" "YES" "YES" "
YES" "NO" "NO" "NO" "NO" "NO" "YES" "YES"
[26] "NO" "NO" "YES" "YES" "NO"
```

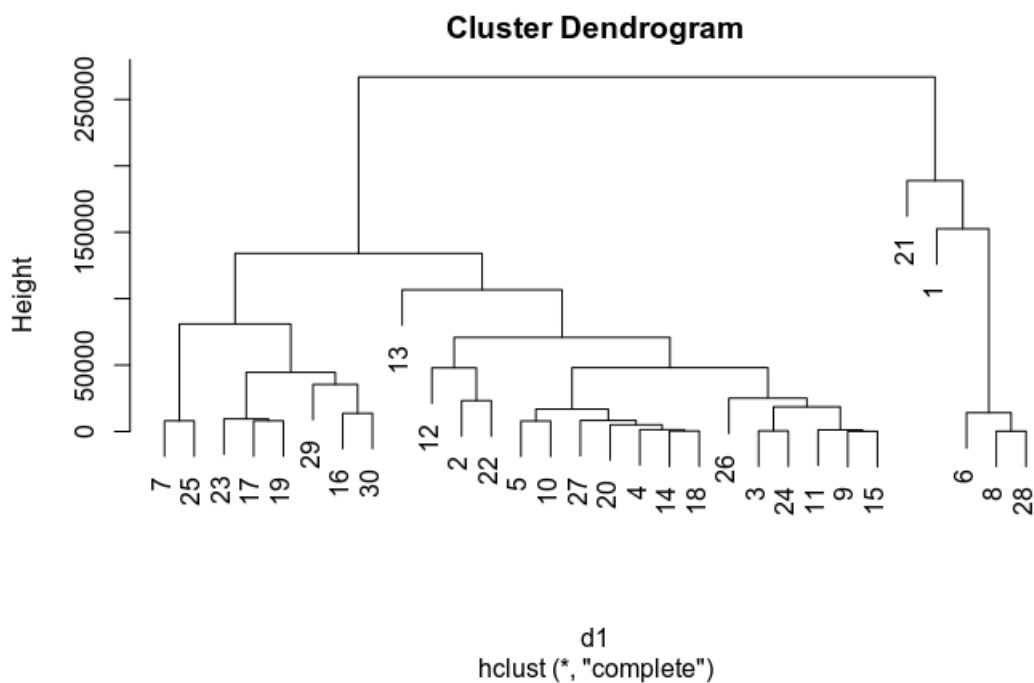
Hide

```
x<-read.csv("Loans_test.csv")
d1 <- dist(x,method="euclidean")
```

NAs introduced by coercion

Hide

```
hc <- hclust(d1)
plot(hc)
```



Hide

```
library(caret)
```

```
Loading required package: lattice
Loading required package: ggplot2
```

Hide

```
indxTrain <- createDataPartition(d$writeoff,p=0.75,list=F)
training <- d[indxTrain]
testing <- d[-indxTrain]
ctrl <- trainControl(method="cv",number=10)
library(e1071)
impute(d)
```


| [31,] | M | 21 | divorced | basic | 1 | self_employ | 0 | 2 | |
|---------|--------------------|---------------------|-------------|---------------|-------------|---------------|----------|------|---|
| 25500" | "no" | "0" | | | | | | | |
| [32,] | "M" | "47" | "single" | "univ" | "0" | "part_time" | "15" | "15" | " |
| 47813" | "no" | "0" | | | | | | | |
| [33,] | "M" | "23" | "single" | "basic" | "0" | "part_time" | "1" | "2" | " |
| 12750" | "no" | "0" | | | | | | | |
| [34,] | "F" | "26" | "single" | "univ" | "0" | "full_time" | "0" | "0" | " |
| 53125" | "no" | "0" | | | | | | | |
| [35,] | "F" | "42" | "widowed" | "highsch" | "2" | "self_employ" | "11" | "11" | " |
| 57375" | "no" | "0" | | | | | | | |
| [36,] | "F" | "30" | "divorced" | "highsch" | "2" | "part_time" | "5" | "6" | " |
| 28688" | "no" | "0" | | | | | | | |
| [37,] | "M" | "22" | "divorced" | "univ" | "0" | "full_time" | "0" | "3" | " |
| 53125" | "no" | "0" | | | | | | | |
| [38,] | "M" | "32" | "single" | "basic" | "0" | "full_time" | "2" | "8" | " |
| 31875" | "no" | "0" | | | | | | | |
| [39,] | "M" | "41" | "divorced" | "univ" | "1" | "full_time" | "12" | "18" | " |
| 79688" | "no" | "0" | | | | | | | |
| [40,] | "M" | "25" | "married" | "highsch" | "1" | "full_time" | "1" | "2" | " |
| 31875" | "yes" | "37025" | | | | | | | |
| [41,] | "M" | "32" | "single" | "highsch" | "0" | "part_time" | "3" | "9" | " |
| 28688" | "no" | "0" | | | | | | | |
| [42,] | "F" | "31" | "single" | "univ" | "0" | "full_time" | "3" | "3" | " |
| 79688" | "no" | "0" | | | | | | | |
| [43,] | "F" | "23" | "divorced" | "univ" | "0" | "full_time" | "1" | "1" | " |
| 53125" | "no" | "0" | | | | | | | |
| [44,] | "M" | "36" | "widowed" | "postgrad" | "0" | "full_time" | "3" | "14" | " |
| 127500" | "no" | "0" | | | | | | | |
| [45,] | "F" | "39" | "married" | "highsch" | "2" | "full_time" | "4" | "17" | " |
| 47813" | "yes" | "53618" | | | | | | | |
| [46,] | "M" | "31" | "divorced" | "basic" | "0" | "unemployed" | "0" | "5" | " |
| 0" | "no" | "0" | | | | | | | |
| [47,] | "F" | "29" | "married" | "univ" | "2" | "part_time" | "4" | "4" | " |
| 31875" | "no" | "0" | | | | | | | |
| [48,] | "M" | "30" | "divorced" | "univ" | "0" | "full_time" | "5" | "6" | " |
| 79688" | "no" | "0" | | | | | | | |
| [49,] | "M" | "36" | "married" | "highsch" | "2" | "self_employ" | "3" | "10" | " |
| 57375" | "yes" | "31290" | | | | | | | |
| [50,] | "F" | "30" | "married" | "postgrad" | "2" | "self_employ" | "0" | "0" | " |
| 153000" | "no" | "0" | | | | | | | |
| [51,] | "M" | "30" | "single" | "univ" | "0" | "self_employ" | "6" | "9" | " |
| 95625" | "no" | "0" | | | | | | | |
| [52,] | "F" | "31" | "married" | "highsch" | "1" | "full_time" | "6" | "7" | " |
| 47813" | "yes" | "53944" | | | | | | | |
| [53,] | "F" | "33" | "married" | "postgrad" | "1" | "full_time" | "3" | "10" | " |
| 127500" | "yes" | "54576" | | | | | | | |
| [54,] | "M" | "43" | "married" | "univ" | "2" | "unemployed" | "0" | "5" | " |
| 0" | "yes" | "50432" | | | | | | | |
| [55,] | "M" | "34" | "divorced" | "highsch" | "1" | "part_time" | "3" | "6" | " |
| 28688" | "no" | "0" | | | | | | | |
| | | | | | | | | | |
| | residential_status | yrs_current_address | loan_amount | loan_purpose | loan_length | collateral | writeoff | | |
| [1,] | "tenant" | "7" | "33617" | "debt_consol" | "36" | "no" | "yes" | | |
| [2,] | "tenant" | "1" | "45562" | "debt_consol" | "27" | "no" | "yes" | | |
| [3,] | "tenant" | "6" | "18977" | "debt_consol" | "34" | "no" | "yes" | | |
| [4,] | "tenant" | "13" | "23406" | "wedding" | "50" | "no" | "no" | | |
| [5,] | "owner_morg" | "15" | "7063" | "other" | "7" | "yes" | "no" | | |
| [6,] | "owner" | "3" | "19704" | "vehicle" | "40" | "yes" | "no" | | |
| [7,] | "tenant" | "4" | "30660" | "other" | "26" | "no" | "yes" | | |
| [8,] | "owner" | "5" | "5031" | "vehicle" | "6" | "yes" | "no" | | |
| [9,] | "tenant" | "2" | "12058" | "debt_consol" | "53" | "no" | "yes" | | |
| [10,] | "tenant" | "11" | "119208" | "debt_consol" | "75" | "no" | "yes" | | |
| [11,] | "tenant" | "10" | "11716" | "wedding" | "57" | "no" | "yes" | | |
| [12,] | "owner_morg" | "11" | "6028" | "wedding" | "10" | "yes" | "no" | | |
| [13,] | "tenant" | "5" | "113314" | "debt_consol" | "71" | "no" | "yes" | | |
| [14,] | "tenant" | "10" | "11345" | "vehicle" | "58" | "no" | "no" | | |
| [15,] | "owner_morg" | "13" | "29226" | "wedding" | "38" | "no" | "yes" | | |
| [16,] | "owner_morg" | "6" | "9770" | "wedding" | "6" | "no" | "yes" | | |
| [17,] | "tenant" | "4" | "30771" | "debt_consol" | "51" | "no" | "yes" | | |
| [18,] | "tenant" | "15" | "76001" | "debt_consol" | "42" | "no" | "yes" | | |
| [19,] | "tenant" | "14" | "8703" | "debt_consol" | "9" | "no" | "yes" | | |
| [20,] | "tenant" | "7" | "23234" | "wedding" | "50" | "no" | "no" | | |
| [21,] | "owner" | "9" | "12569" | "other" | "30" | "no" | "no" | | |
| [22,] | "owner" | "9" | "42439" | "debt_consol" | "25" | "yes" | "no" | | |

```

[23,] "tenant"      " 3"      " 27970"  "wedding"  "59"      "no"      "yes"
[24,] "tenant"      "13"      " 15694"  "other"    "50"      "no"      "no"
[25,] "tenant"      " 7"      " 41986"  "other"    "54"      "no"      "no"
[26,] "tenant"      "12"      " 48428"  "debt_consol" "60"      "no"      "yes"
[27,] "tenant"      " 9"      " 49504"  "debt_consol" "38"      "no"      "no"
[28,] "tenant"      " 6"      " 13814"  "debt_consol" "26"      "no"      "yes"
[29,] "tenant"      "11"      " 29577"  "other"    "30"      "no"      "no"
[30,] "tenant"      "13"      " 24278"  "vehicle"  "34"      "yes"     "no"
[31,] "tenant"      " 5"      " 5090"   "other"    " 9"      "no"      "yes"
[32,] "owner_morg"   "13"      " 47080"  "home_improv" "44"      "yes"     "no"
[33,] "tenant"      " 8"      " 14079"  "wedding"  "54"      "no"      "yes"
[34,] "tenant"      "10"      " 32007"  "debt_consol" "44"      "no"      "yes"
[35,] "owner_morg"   " 9"      " 10925"  "debt_consol" "45"      "no"      "yes"
[36,] "owner_morg"   " 3"      " 22385"  "wedding"  "40"      "no"      "yes"
[37,] "owner"        " 5"      " 8456"   "debt_consol" " 6"      "yes"     "no"
[38,] "owner"        "14"      " 27167"  "wedding"  "26"      "no"      "yes"
[39,] "tenant"      "13"      " 32154"  "other"    "43"      "no"      "yes"
[40,] "owner"        " 6"      " 27523"  "vehicle"  "53"      "yes"     "no"
[41,] "tenant"      " 9"      " 7442"   "wedding"  "11"      "no"      "no"
[42,] "owner_morg"   "15"      " 38375"  "debt_consol" "40"      "no"      "yes"
[43,] "tenant"      " 8"      " 29247"  "debt_consol" "29"      "no"      "yes"
[44,] "owner"        " 6"      " 28778"  "vehicle"  "55"      "yes"     "no"
[45,] "owner_morg"   " 8"      " 16695"  "holidays"  "44"      "no"      "no"
[46,] "owner_morg"   " 8"      " 18464"  "other"    "48"      "yes"     "yes"
[47,] "owner"        " 8"      " 12326"  "debt_consol" "55"      "yes"     "no"
[48,] "owner"        "11"      " 77983"  "home_improv" "43"      "no"      "yes"
[49,] "owner_morg"   "14"      " 62251"  "debt_consol" "58"      "yes"     "no"
[50,] "owner"        "14"      "150946"  "debt_consol" "87"      "yes"     "no"
[51,] "owner_morg"   "12"      " 51441"  "home_improv" "49"      "yes"     "no"
[52,] "owner"        "15"      " 67429"  "home_improv" "50"      "yes"     "no"
[53,] "owner_morg"   " 2"      " 44249"  "debt_consol" "38"      "no"      "no"
[54,] "owner_morg"   "11"      " 33217"  "debt_consol" "29"      "no"      "no"
[55,] "tenant"      " 8"      " 20015"  "debt_consol" "35"      "no"      "yes"

```

```
[ reached getOption("max.print") -- omitted 29945 rows ]
```

Hide

```

knnFit <- train(writeoff~age+net_income, data=d,method="knn",trControl=ctrl,tuneLength=20)
knnFit

```

k-Nearest Neighbors

```
30000 samples
  2 predictor
  2 classes: 'no', 'yes'
```

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 27001, 27000, 27000, 27000, 27000, 27000, ...

Resampling results across tuning parameters:

| k | Accuracy | Kappa |
|----|-----------|------------|
| 5 | 0.5357331 | 0.01574643 |
| 7 | 0.5354663 | 0.01508062 |
| 9 | 0.5355331 | 0.01506079 |
| 11 | 0.5360996 | 0.01610141 |
| 13 | 0.5358663 | 0.01529654 |
| 15 | 0.5371330 | 0.01790338 |
| 17 | 0.5369664 | 0.01744767 |
| 19 | 0.5361330 | 0.01541445 |
| 21 | 0.5373664 | 0.01798866 |
| 23 | 0.5375664 | 0.01803220 |
| 25 | 0.5384000 | 0.01933665 |
| 27 | 0.5371998 | 0.01681426 |
| 29 | 0.5375664 | 0.01743903 |
| 31 | 0.5375998 | 0.01758464 |
| 33 | 0.5380998 | 0.01844866 |
| 35 | 0.5371998 | 0.01647183 |
| 37 | 0.5373999 | 0.01690796 |
| 39 | 0.5370997 | 0.01624497 |
| 41 | 0.5389664 | 0.01973648 |
| 43 | 0.5385000 | 0.01879040 |

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was k = 41.