

Predict Spam Email Using R

Sayan Maity

First we set up working directory to save our work

```
getwd() #to know which directory we are working in
```

```
## [1] "C:/Users/s9868852/Downloads"
```

```
setwd("C:/Users/s9868852/Downloads") #to setup new working directory
```

Download the data:

```
library(kernlab)
```

```
## Warning: package 'kernlab' was built under R version 3.6.1
```

```
data("spam")
```

```
df <- spam
```

```
str(df) # this to show us all the features of the dataset
```

```
## 'data.frame': 4601 obs. of 58 variables:
## $ make : num 0 0.21 0.06 0 0 0 0 0 0.15 0.06 ...
## $ address : num 0.64 0.28 0 0 0 0 0 0 0.12 ...
## $ all : num 0.64 0.5 0.71 0 0 0 0 0 0.46 0.77 ...
## $ num3d : num 0 0 0 0 0 0 0 0 0 0 ...
## $ our : num 0.32 0.14 1.23 0.63 0.63 1.85 1.92 1.88 0.61 0.19 ...
## $ over : num 0 0.28 0.19 0 0 0 0 0 0.32 ...
## $ remove : num 0 0.21 0.19 0.31 0.31 0 0 0 0.3 0.38 ...
## $ internet : num 0 0.07 0.12 0.63 0.63 1.85 0 1.88 0 0 ...
## $ order : num 0 0 0.64 0.31 0.31 0 0 0 0.92 0.06 ...
## $ mail : num 0 0.94 0.25 0.63 0.63 0 0.64 0 0.76 0 ...
## $ receive : num 0 0.21 0.38 0.31 0.31 0 0.96 0 0.76 0 ...
## $ will : num 0.64 0.79 0.45 0.31 0.31 0 1.28 0 0.92 0.64 ...
## $ people : num 0 0.65 0.12 0.31 0.31 0 0 0 0 0.25 ...
## $ report : num 0 0.21 0 0 0 0 0 0 0 0 ...
## $ addresses : num 0 0.14 1.75 0 0 0 0 0 0.12 ...
## $ free : num 0.32 0.14 0.06 0.31 0.31 0 0.96 0 0 0 ...
## $ business : num 0 0.07 0.06 0 0 0 0 0 0 0 ...
## $ email : num 1.29 0.28 1.03 0 0 0 0.32 0 0.15 0.12 ...
## $ you : num 1.93 3.47 1.36 3.18 3.18 0 3.85 0 1.23 1.67 ...
## $ credit : num 0 0 0.32 0 0 0 0 0 3.53 0.06 ...
## $ your : num 0.96 1.59 0.51 0.31 0.31 0 0.64 0 2 0.71 ...
## $ font : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num000 : num 0 0.43 1.16 0 0 0 0 0 0 0.19 ...
## $ money : num 0 0.43 0.06 0 0 0 0 0 0.15 0 ...
```

```
## $ hp : num 0 0 0 0 0 0 0 0 0 0 ...
## $ hpl : num 0 0 0 0 0 0 0 0 0 0 ...
## $ george : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num650 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ lab : num 0 0 0 0 0 0 0 0 0 0 ...
## $ labs : num 0 0 0 0 0 0 0 0 0 0 ...
## $ telnet : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num857 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ data : num 0 0 0 0 0 0 0 0 0.15 0 ...
## $ num415 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num85 : num 0 0 0 0 0 0 0 0 0 0 ...
## $ technology : num 0 0 0 0 0 0 0 0 0 0 ...
## $ num1999 : num 0 0.07 0 0 0 0 0 0 0 0 ...
## $ parts : num 0 0 0 0 0 0 0 0 0 0 ...
## $ pm : num 0 0 0 0 0 0 0 0 0 0 ...
## $ direct : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ cs : num 0 0 0 0 0 0 0 0 0 0 ...
## $ meeting : num 0 0 0 0 0 0 0 0 0 0 ...
## $ original : num 0 0 0.12 0 0 0 0 0 0.3 0 ...
## $ project : num 0 0 0 0 0 0 0 0 0 0.06 ...
## $ re : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ edu : num 0 0 0.06 0 0 0 0 0 0 0 ...
## $ table : num 0 0 0 0 0 0 0 0 0 0 ...
## $ conference : num 0 0 0 0 0 0 0 0 0 0 ...
## $ charSemicolon : num 0 0 0.01 0 0 0 0 0 0 0.04 ...
## $ charRoundbracket : num 0 0.132 0.143 0.137 0.135 0.223 0.054 0.206 0.271 0.03 ...
## $ charSquarebracket : num 0 0 0 0 0 0 0 0 0 0 ...
## $ charExclamation : num 0.778 0.372 0.276 0.137 0.135 0 0.164 0 0.181 0.244 ...
## $ charDollar : num 0 0.18 0.184 0 0 0 0.054 0 0.203 0.081 ...
## $ charHash : num 0 0.048 0.01 0 0 0 0 0 0.022 0 ...
## $ capitalAve : num 3.76 5.11 9.82 3.54 3.54 ...
## $ capitalLong : num 61 101 485 40 40 15 4 11 445 43 ...
## $ capitalTotal : num 278 1028 2259 191 191 ...
## $ type : Factor w/ 2 levels "nonspam","spam": 2 2 2 2 2 2 2 2 2 2 ...
```

We check for any missing data in the data set.

```
anyNA(df)
```

```
## [1] FALSE
```

```
any(is.na(df))
```

```
## [1] FALSE
```

```
# This is two different way to look for missing data in a data set
```

As we can see every variable other than the variable “type” is of numeric variable. So we need to convert the factor variable into a numeric variable.

```
df$type <- as.numeric(df$type) # this will make the factor variable into a numeric variable
```

```
df$type[df$type==1] <- 0 # for no-spam
```

```
df$type[df$type==2] <- 1 # for spam
```

```
# this is set a binary value for our spam and non spam emails
```

Now we will learn about the data. Some descriptives are useful. Here ,we are calculating the mean of each feature in nonspam and spam classes.

```
nonspam <-apply(df[df[,58]==0,],2, mean)
spam <-apply(df[df[,58]==1,],2, mean)
table <-cbind(nonspam, spam)
table
```

##	nonspam	spam
## make	7.347920e-02	1.523387e-01
## address	2.444656e-01	1.646498e-01
## all	2.005811e-01	4.037948e-01
## num3d	8.859397e-04	1.646718e-01
## our	1.810402e-01	5.139548e-01
## over	4.454448e-02	1.748759e-01
## remove	9.383070e-03	2.754054e-01
## internet	3.841463e-02	2.081412e-01
## order	3.804878e-02	1.700607e-01
## mail	1.671700e-01	3.505074e-01
## receive	2.171090e-02	1.184335e-01
## will	5.363235e-01	5.499724e-01
## people	6.166428e-02	1.435466e-01
## report	4.240316e-02	8.357419e-02
## addresses	8.317791e-03	1.120794e-01
## free	7.358680e-02	5.183618e-01
## business	4.834648e-02	2.875069e-01
## email	9.729197e-02	3.192278e-01
## you	1.270341e+00	2.264539e+00
## credit	7.578910e-03	2.055212e-01
## your	4.387016e-01	1.380370e+00
## font	4.522597e-02	2.380364e-01
## num000	7.087518e-03	2.470546e-01
## money	1.713773e-02	2.128792e-01
## hp	8.954735e-01	1.747932e-02
## hpl	4.319943e-01	9.172642e-03
## george	1.265265e+00	1.549917e-03
## num650	1.938056e-01	1.879757e-02
## lab	1.627941e-01	6.839493e-04
## labs	1.658537e-01	5.968009e-03
## telnet	1.060330e-01	1.274131e-03
## num857	7.730631e-02	5.184777e-04
## data	1.509864e-01	1.456150e-02
## num415	7.778694e-02	1.776062e-03
## num85	1.694548e-01	6.927744e-03
## technology	1.416714e-01	2.951462e-02
## num1999	1.977439e-01	4.346939e-02
## parts	1.872310e-02	4.710425e-03
## pm	1.216786e-01	1.242692e-02
## direct	8.311693e-02	3.671815e-02
## cs	7.202654e-02	5.515720e-05
## meeting	2.168077e-01	2.443464e-03

```
## original      7.058106e-02 8.450083e-03
## project      1.266356e-01 6.243795e-03
## re           4.157604e-01 1.250910e-01
## edu          2.871844e-01 1.472697e-02
## table        8.192253e-03 1.218974e-03
## conference   5.122669e-02 2.101489e-03
## charSemicolon 5.028085e-02 2.057308e-02
## charRoundbracket 1.585782e-01 1.089702e-01
## charSquarebracket 2.268364e-02 8.198566e-03
## charExclamation 1.099835e-01 5.137126e-01
## charDollar   1.164849e-02 1.744782e-01
## charHash     2.171306e-02 7.887700e-02
## capitalAve   2.377301e+00 9.519165e+00
## capitalLong  1.821449e+01 1.043933e+02
## capitalTotal 1.614709e+02 4.706194e+02
## type        0.000000e+00 1.000000e+00
```

We want to know what features have the largest mean difference between spam and nonspam.

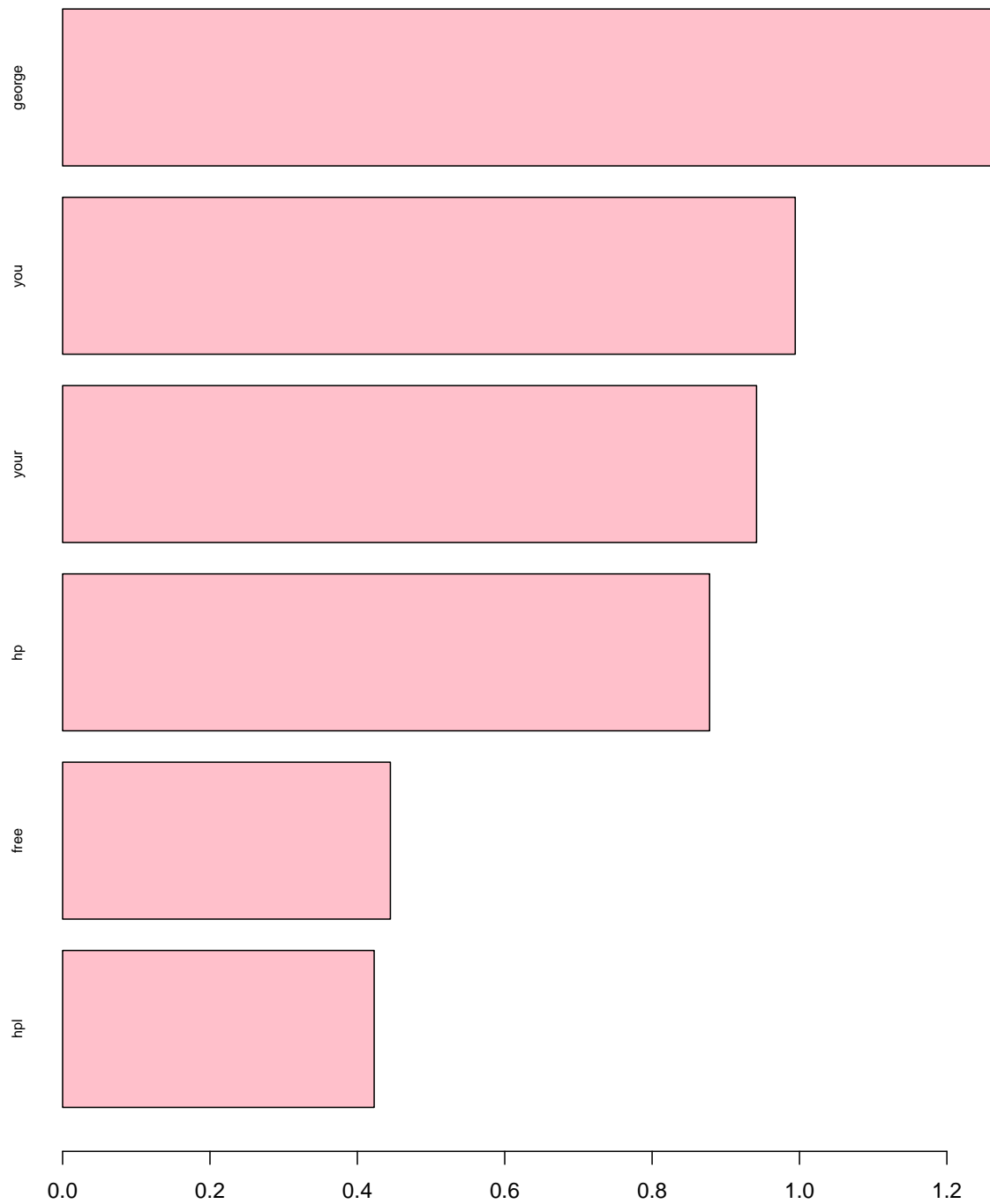
```
#check the str for tbl1: it's a list. Chnage it to data.frame
table <-as.data.frame(table)
# removing type from table
table <-table[-58, ]
# Difference (absolute value for more meaningful!)
table$diff <-abs(table[,1]-table[,2])
# order the tbl1 by the difference. See order()
table <-round(table[order(table[,3],decreasing =TRUE), ],4)
table
```

```
##          nonspam      spam      diff
## capitalTotal 161.4709 470.6194 309.1485
## capitalLong  18.2145 104.3933  86.1788
## capitalAve    2.3773   9.5192   7.1419
## george        1.2653   0.0015   1.2637
## you           1.2703   2.2645   0.9942
## your          0.4387   1.3804   0.9417
## hp            0.8955   0.0175   0.8780
## free          0.0736   0.5184   0.4448
## hpl           0.4320   0.0092   0.4228
## charExclamation 0.1100   0.5137   0.4037
## our           0.1810   0.5140   0.3329
## re            0.4158   0.1251   0.2907
## edu           0.2872   0.0147   0.2725
## remove        0.0094   0.2754   0.2660
## num000         0.0071   0.2471   0.2400
## business      0.0483   0.2875   0.2392
## email         0.0973   0.3192   0.2219
## meeting       0.2168   0.0024   0.2144
## all           0.2006   0.4038   0.2032
## credit        0.0076   0.2055   0.1979
## money         0.0171   0.2129   0.1957
## font          0.0452   0.2380   0.1928
## mail          0.1672   0.3505   0.1833
## num650        0.1938   0.0188   0.1750
```

## internet	0.0384	0.2081	0.1697
## num3d	0.0009	0.1647	0.1638
## charDollar	0.0116	0.1745	0.1628
## num85	0.1695	0.0069	0.1625
## lab	0.1628	0.0007	0.1621
## labs	0.1659	0.0060	0.1599
## num1999	0.1977	0.0435	0.1543
## data	0.1510	0.0146	0.1364
## order	0.0380	0.1701	0.1320
## over	0.0445	0.1749	0.1303
## project	0.1266	0.0062	0.1204
## technology	0.1417	0.0295	0.1122
## pm	0.1217	0.0124	0.1093
## telnet	0.1060	0.0013	0.1048
## addresses	0.0083	0.1121	0.1038
## receive	0.0217	0.1184	0.0967
## people	0.0617	0.1435	0.0819
## address	0.2445	0.1646	0.0798
## make	0.0735	0.1523	0.0789
## num857	0.0773	0.0005	0.0768
## num415	0.0778	0.0018	0.0760
## cs	0.0720	0.0001	0.0720
## original	0.0706	0.0085	0.0621
## charHash	0.0217	0.0789	0.0572
## charRoundbracket	0.1586	0.1090	0.0496
## conference	0.0512	0.0021	0.0491
## direct	0.0831	0.0367	0.0464
## report	0.0424	0.0836	0.0412
## charSemicolon	0.0503	0.0206	0.0297
## charSquarebracket	0.0227	0.0082	0.0145
## parts	0.0187	0.0047	0.0140
## will	0.5363	0.5500	0.0136
## table	0.0082	0.0012	0.0070

It seems that the first 3 features have different scales, not relative frequencies as others. So let's see a barplot for a better visualization without the first three features

```
barplot(table[9:4,3],col ="pink",horiz =TRUE,names.arg =rownames(table)[9:4],cex.names =0.7)
```



Now we will develop a learning method that has to decide which features to use and how:

LPM

We will split the data into training and test sets.

```
set.seed(1)

ind <- sample(nrow(df), nrow(df)*.7, replace = FALSE)

train <- df[ind,]
test <- df[-ind,]
# This would split the data into training set and test set
```

We will train a Linear Probability Model on the training set by `lm()`.

```
model_lpm <- lm(type ~ . - type, data = train)
summary(model_lpm)
```

```
##
## Call:
## lm(formula = type ~ . - type, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91499 -0.22800 -0.05997  0.22926  0.99142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.999e-01  1.397e-02  14.309  < 2e-16 ***
## make          -9.193e-02  2.167e-02  -4.243  2.27e-05 ***
## address       -1.195e-02  4.589e-03  -2.604  0.009248 **
## all           4.058e-02  1.237e-02   3.279  0.001053 **
## num3d         1.011e-02  4.249e-03   2.380  0.017369 *
## our           9.178e-02  9.872e-03   9.297  < 2e-16 ***
## over          1.360e-01  2.297e-02   5.922  3.52e-09 ***
## remove        2.082e-01  1.548e-02  13.450  < 2e-16 ***
## internet      9.036e-02  1.514e-02   5.968  2.67e-09 ***
## order         5.556e-02  2.183e-02   2.545  0.010975 *
## mail          1.428e-02  9.980e-03   1.431  0.152654
## receive       7.455e-02  3.137e-02   2.377  0.017535 *
## will         -3.145e-02  7.357e-03  -4.275  1.97e-05 ***
## people        1.186e-02  2.032e-02   0.584  0.559495
## report       -5.854e-03  1.682e-02  -0.348  0.727761
## addresses    -1.375e-02  2.744e-02  -0.501  0.616280
## free          5.982e-02  6.861e-03   8.719  < 2e-16 ***
## business      6.068e-02  1.432e-02   4.239  2.31e-05 ***
## email         4.972e-02  1.150e-02   4.322  1.59e-05 ***
## you           1.680e-02  3.706e-03   4.534  6.02e-06 ***
## credit        6.441e-02  1.319e-02   4.883  1.10e-06 ***
## your          4.935e-02  5.726e-03   8.619  < 2e-16 ***
## font          4.252e-02  6.179e-03   6.882  7.09e-12 ***
## num000        1.739e-01  1.904e-02   9.135  < 2e-16 ***
## money          1.436e-01  1.728e-02   8.308  < 2e-16 ***
## hp           -2.104e-02  4.388e-03  -4.795  1.70e-06 ***
## hpl          -1.935e-02  7.881e-03  -2.455  0.014137 *
```

```
## george -1.295e-02 1.842e-03 -7.031 2.50e-12 ***
## num650 1.681e-02 1.565e-02 1.074 0.282822
## lab -2.016e-03 1.488e-02 -0.135 0.892238
## labs -7.565e-02 2.038e-02 -3.711 0.000210 ***
## telnet -1.355e-02 3.471e-02 -0.390 0.696300
## num857 -3.168e-02 1.721e-01 -0.184 0.853973
## data -4.469e-02 1.079e-02 -4.143 3.51e-05 ***
## num415 7.097e-02 1.681e-01 0.422 0.672867
## num85 -8.321e-02 2.307e-02 -3.606 0.000316 ***
## technology 5.597e-02 2.793e-02 2.004 0.045161 *
## num1999 -2.869e-02 1.459e-02 -1.966 0.049363 *
## parts -5.895e-02 2.412e-02 -2.444 0.014572 *
## pm -1.559e-02 1.299e-02 -1.200 0.230391
## direct 6.447e-02 3.528e-02 1.827 0.067724 .
## cs -1.804e-02 1.803e-02 -1.001 0.317031
## meeting -3.709e-02 9.710e-03 -3.820 0.000136 ***
## original -7.353e-02 2.792e-02 -2.634 0.008488 **
## project -4.547e-02 1.178e-02 -3.859 0.000116 ***
## re -3.334e-02 6.143e-03 -5.427 6.18e-08 ***
## edu -3.556e-02 6.576e-03 -5.407 6.87e-08 ***
## table -2.492e-01 8.533e-02 -2.921 0.003515 **
## conference -4.940e-02 1.936e-02 -2.552 0.010763 *
## charSemicolon -1.297e-01 2.579e-02 -5.028 5.22e-07 ***
## charRoundbracket -3.684e-02 2.886e-02 -1.276 0.201925
## charSquarebracket -3.328e-03 7.122e-02 -0.047 0.962735
## charExclamation 5.617e-02 6.854e-03 8.195 3.60e-16 ***
## charDollar 2.242e-01 2.524e-02 8.882 < 2e-16 ***
## charHash 2.550e-02 1.249e-02 2.041 0.041320 *
## capitalAve -1.859e-04 2.321e-04 -0.801 0.423363
## capitalLong 2.828e-04 7.422e-05 3.810 0.000142 ***
## capitalTotal 6.921e-05 1.169e-05 5.922 3.52e-09 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3284 on 3162 degrees of freedom
## Multiple R-squared: 0.559, Adjusted R-squared: 0.551
## F-statistic: 70.32 on 57 and 3162 DF, p-value: < 2.2e-16
```

#type is numeric 0 and 1. We previously change the class so we can use it for model

Now we will predict spam email in the test data.

```
phat_lpm <- predict(model_lpm, data = test)
summary(phat_lpm)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.6153  0.1598  0.3021  0.4006  0.6424  2.1555
```

Now, we will use 0.5 as a threshold value to assign the emails to spam and no spam category

```
spam_lpm <- ifelse(phat_lpm < 0.5,0,1)
table(spam_lpm)
```

```
## spam_lpm
```



```
##      0      1
## 2107 1113
```

Here we can see how many emails we predict spam and how many are not spam. Now we will build the confusion table.

```
ct <- table(train$type,spam_lpm)
ct
```

```
##      spam_lpm
##           0      1
## 0 1841      89
## 1  266    1024
```

Using this table we can see how many emails we mean predict correctly. To measure our prediction we will use two measurement

(i) Hit ratio,

```
sum(diag(ct))/sum(ct)
```

```
## [1] 0.8897516
```

(ii) TPR

```
ct[2,2]/(sum(ct[,2]))
```

```
## [1] 0.9200359
```

So we can see this model has almost 89 % hit ratio and almost 92 % True Probabilty Ratio

Now we want to change the threshold to see if we can improve TPR somehow

```
x <- .90
spam_lpm <- ifelse(phat_lpm < x,0,1)
ct <- table(train$type,spam_lpm)
ct[2,2]/(sum(ct[,2]))
```

```
## [1] 0.973262
```

Here we can improve TPR from 92% to 98% which is a great improvement also a sign of a really good prediction

Here we want to see that can we build a function that we can use to try different threshold value to check the improvement of TPR

```
f1 <- function(x){spam_lpm <- ifelse(phat_lpm < x,0,1)
return(spam_lpm)}
f2 <- function(x){ct <- table(train$type,f1(x))
ct}
f3 <- function(x){TPR <- f2(x)[2,2]/(sum(f2(x)[,2]))
return(TPR)}
f3(.87)
```

```
## [1] 0.973494
```

Logistic

Now we use the same training and test sets to train a logistic model on the training set by `glm()`.

```
model_logistic <- glm(type ~ ., family = gaussian, data = df)
summary(model_logistic)
```

```
##
## Call:
## glm(formula = type ~ . - type, family = gaussian, data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.27083  -0.21836  -0.05765   0.21419   0.92732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.003e-01  1.152e-02  17.390 < 2e-16 ***
## make          -4.982e-02  1.678e-02  -2.969 0.003007 **
## address       -1.205e-02  3.790e-03  -3.178 0.001492 **
## all           3.928e-02  1.006e-02   3.906 9.51e-05 ***
## num3d         1.192e-02  3.461e-03   3.444 0.000579 ***
## our           8.421e-02  7.570e-03  11.124 < 2e-16 ***
## over          1.188e-01  1.842e-02   6.451 1.22e-10 ***
## remove        2.129e-01  1.303e-02  16.342 < 2e-16 ***
## internet      9.399e-02  1.257e-02   7.477 9.05e-14 ***
## order         7.247e-02  1.896e-02   3.822 0.000134 ***
## mail          1.507e-02  7.780e-03   1.937 0.052842 .
## receive       5.686e-02  2.623e-02   2.167 0.030253 *
## will         -2.786e-02  5.901e-03  -4.721 2.41e-06 ***
## people        1.190e-02  1.668e-02   0.714 0.475423
## report        4.860e-03  1.474e-02   0.330 0.741597
## addresses     1.852e-02  2.196e-02   0.843 0.399050
## free          7.506e-02  6.024e-03  12.460 < 2e-16 ***
## business      5.172e-02  1.186e-02   4.359 1.34e-05 ***
## email         5.540e-02  9.763e-03   5.674 1.48e-08 ***
## you           1.413e-02  3.103e-03   4.554 5.39e-06 ***
## credit        6.172e-02  9.840e-03   6.272 3.89e-10 ***
## your          5.269e-02  4.663e-03  11.299 < 2e-16 ***
## font          4.477e-02  5.346e-03   8.374 < 2e-16 ***
## num000        1.748e-01  1.591e-02  10.987 < 2e-16 ***
## money         9.089e-02  1.143e-02   7.953 2.28e-15 ***
## hp           -2.317e-02  3.655e-03  -6.340 2.52e-10 ***
## hpl          -2.163e-02  6.736e-03  -3.211 0.001332 **
## george       -1.220e-02  1.508e-03  -8.091 7.52e-16 ***
## num650        3.987e-03  1.267e-02   0.315 0.753016
## lab          -7.450e-03  1.118e-02  -0.666 0.505356
## labs         -5.195e-02  1.610e-02  -3.227 0.001260 **
## telnet       -2.329e-02  1.939e-02  -1.201 0.229792
## num857        6.332e-03  1.687e-01   0.038 0.970057
## data         -4.198e-02  8.845e-03  -4.746 2.13e-06 ***
```

```
## num415          5.114e-02  1.660e-01  0.308 0.758033
## num85           -3.117e-02  1.221e-02 -2.552 0.010729 *
## technology      2.648e-02  1.963e-02  1.349 0.177312
## num1999         -3.321e-02  1.271e-02 -2.612 0.009027 **
## parts           -5.344e-02  2.247e-02 -2.378 0.017449 *
## pm              -1.975e-02  1.172e-02 -1.686 0.091853 .
## direct          4.076e-02  2.724e-02  1.497 0.134588
## cs              -8.364e-03  1.438e-02 -0.582 0.560912
## meeting         -3.693e-02  7.603e-03 -4.857 1.23e-06 ***
## original        -6.324e-02  2.356e-02 -2.684 0.007304 **
## project         -3.238e-02  7.863e-03 -4.118 3.89e-05 ***
## re              -3.525e-02  4.958e-03 -7.110 1.34e-12 ***
## edu             -3.781e-02  5.777e-03 -6.546 6.57e-11 ***
## table           -1.952e-01  6.345e-02 -3.076 0.002108 **
## conference      -5.822e-02  1.697e-02 -3.432 0.000605 ***
## charSemicolon   -1.401e-01  2.205e-02 -6.354 2.30e-10 ***
## charRoundbracket -5.996e-02  2.231e-02 -2.687 0.007232 **
## charSquarebracket -5.905e-02  4.488e-02 -1.316 0.188291
## charExclamation  6.805e-02  6.136e-03 11.090 < 2e-16 ***
## charDollar      2.332e-01  2.171e-02 10.741 < 2e-16 ***
## charHash        2.769e-02  1.162e-02  2.383 0.017200 *
## capitalAve      2.327e-04  1.831e-04  1.270 0.203973
## capitalLong     6.675e-05  3.713e-05  1.798 0.072240 .
## capitalTotal    7.986e-05  9.656e-06  8.270 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.106417)
##
## Null deviance: 1098.60 on 4600 degrees of freedom
## Residual deviance: 483.45 on 4543 degrees of freedom
## AIC: 2808.7
##
## Number of Fisher Scoring iterations: 2
```

we will predict spams in the test set.

```
phat_logistic <- predict(model_logistic, data = test)
summary(phat_logistic)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.6515  0.1505  0.2930  0.3940  0.6323  2.2708
```

Again we use 0.5 as a threshold

```
spam_logistic <- ifelse(phat_lpm < 0.5,0,1)
table(spam_logistic)
```

```
## spam_logistic
##      0      1
## 2107 1113
```

and we will build the confusion table

```
ct_logistic <- table(train$type,spam_logistic)
ct_logistic
```

```
##      spam_logistic
##           0       1
##    0 1841     89
##    1  266    1024
```

Hit ratio

```
sum(diag(ct_logistic))/sum(ct_logistic)
```

```
## [1] 0.8897516
```

TPR

```
ct_logistic[2,2]/(sum(ct_logistic[,2]))
```

```
## [1] 0.9200359
```

Again we will try different threshold to see if we can improve TPR.

```
y <- .85
spam_logistic <- ifelse(phat_lpm < y,0,1)
ct_logistic <- table(train$type,spam_logistic)
ct_logistic[2,2]/(sum(ct_logistic[,2]))
```

```
## [1] 0.9750567
```

Prediction

Now we have choose one of the following models

At threshold value of .85 the TPR is highest for both the models. Thus it doesn't matter which model to choose. But to keep the unit of measurement same we will choose LPM.

Now how do we know that which features are the most important predictors. Just by looking into the coefficients of LPM model we can see which words are significant

```
summary(model_lpm)
```

```
##
## Call:
## lm(formula = type ~ . - type, data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.91499 -0.22800 -0.05997  0.22926  0.99142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

## (Intercept)	1.999e-01	1.397e-02	14.309	< 2e-16	***
## make	-9.193e-02	2.167e-02	-4.243	2.27e-05	***
## address	-1.195e-02	4.589e-03	-2.604	0.009248	**
## all	4.058e-02	1.237e-02	3.279	0.001053	**
## num3d	1.011e-02	4.249e-03	2.380	0.017369	*
## our	9.178e-02	9.872e-03	9.297	< 2e-16	***
## over	1.360e-01	2.297e-02	5.922	3.52e-09	***
## remove	2.082e-01	1.548e-02	13.450	< 2e-16	***
## internet	9.036e-02	1.514e-02	5.968	2.67e-09	***
## order	5.556e-02	2.183e-02	2.545	0.010975	*
## mail	1.428e-02	9.980e-03	1.431	0.152654	
## receive	7.455e-02	3.137e-02	2.377	0.017535	*
## will	-3.145e-02	7.357e-03	-4.275	1.97e-05	***
## people	1.186e-02	2.032e-02	0.584	0.559495	
## report	-5.854e-03	1.682e-02	-0.348	0.727761	
## addresses	-1.375e-02	2.744e-02	-0.501	0.616280	
## free	5.982e-02	6.861e-03	8.719	< 2e-16	***
## business	6.068e-02	1.432e-02	4.239	2.31e-05	***
## email	4.972e-02	1.150e-02	4.322	1.59e-05	***
## you	1.680e-02	3.706e-03	4.534	6.02e-06	***
## credit	6.441e-02	1.319e-02	4.883	1.10e-06	***
## your	4.935e-02	5.726e-03	8.619	< 2e-16	***
## font	4.252e-02	6.179e-03	6.882	7.09e-12	***
## num000	1.739e-01	1.904e-02	9.135	< 2e-16	***
## money	1.436e-01	1.728e-02	8.308	< 2e-16	***
## hp	-2.104e-02	4.388e-03	-4.795	1.70e-06	***
## hpl	-1.935e-02	7.881e-03	-2.455	0.014137	*
## george	-1.295e-02	1.842e-03	-7.031	2.50e-12	***
## num650	1.681e-02	1.565e-02	1.074	0.282822	
## lab	-2.016e-03	1.488e-02	-0.135	0.892238	
## labs	-7.565e-02	2.038e-02	-3.711	0.000210	***
## telnet	-1.355e-02	3.471e-02	-0.390	0.696300	
## num857	-3.168e-02	1.721e-01	-0.184	0.853973	
## data	-4.469e-02	1.079e-02	-4.143	3.51e-05	***
## num415	7.097e-02	1.681e-01	0.422	0.672867	
## num85	-8.321e-02	2.307e-02	-3.606	0.000316	***
## technology	5.597e-02	2.793e-02	2.004	0.045161	*
## num1999	-2.869e-02	1.459e-02	-1.966	0.049363	*
## parts	-5.895e-02	2.412e-02	-2.444	0.014572	*
## pm	-1.559e-02	1.299e-02	-1.200	0.230391	
## direct	6.447e-02	3.528e-02	1.827	0.067724	.
## cs	-1.804e-02	1.803e-02	-1.001	0.317031	
## meeting	-3.709e-02	9.710e-03	-3.820	0.000136	***
## original	-7.353e-02	2.792e-02	-2.634	0.008488	**
## project	-4.547e-02	1.178e-02	-3.859	0.000116	***
## re	-3.334e-02	6.143e-03	-5.427	6.18e-08	***
## edu	-3.556e-02	6.576e-03	-5.407	6.87e-08	***
## table	-2.492e-01	8.533e-02	-2.921	0.003515	**
## conference	-4.940e-02	1.936e-02	-2.552	0.010763	*
## charSemicolon	-1.297e-01	2.579e-02	-5.028	5.22e-07	***
## charRoundbracket	-3.684e-02	2.886e-02	-1.276	0.201925	
## charSquarebracket	-3.328e-03	7.122e-02	-0.047	0.962735	
## charExclamation	5.617e-02	6.854e-03	8.195	3.60e-16	***
## charDollar	2.242e-01	2.524e-02	8.882	< 2e-16	***

```
## charHash          2.550e-02  1.249e-02   2.041 0.041320 *
## capitalAve       -1.859e-04  2.321e-04  -0.801 0.423363
## capitalLong       2.828e-04  7.422e-05   3.810 0.000142 ***
## capitalTotal      6.921e-05  1.169e-05   5.922 3.52e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3284 on 3162 degrees of freedom
## Multiple R-squared:  0.559, Adjusted R-squared:  0.551
## F-statistic: 70.32 on 57 and 3162 DF, p-value: < 2.2e-16
```

We can clearly see the words that are significant that are most commonly used words in Spam emails.

To be more precise we can use principal component analysis to find the features that are explaining the most.

```
myPCA <- prcomp(df[,1:57], scale. = F, center = F)
summary(myPCA)
```

```
## Importance of components:
##
##          PC1          PC2          PC3          PC4          PC5          PC6
## Standard deviation  678.2331 170.04362 27.44152 3.46523 2.61376 1.93446
## Proportion of Variance  0.9393  0.05905  0.00154 0.00002 0.00001 0.00001
## Cumulative Proportion  0.9393  0.99838  0.99992 0.99994 0.99996 0.99996
##
##          PC7  PC8  PC9  PC10  PC11  PC12  PC13  PC14
## Standard deviation  1.397 1.305 1.162 1.032 1.006 0.9557 0.9241 0.8891
## Proportion of Variance 0.000 0.000 0.000 0.000 0.000 0.0000 0.0000 0.0000
## Cumulative Proportion 1.000 1.000 1.000 1.000 1.000 1.0000 1.0000 1.0000
##
##          PC15  PC16  PC17  PC18  PC19  PC20  PC21  PC22
## Standard deviation  0.8533 0.7834 0.7571 0.69 0.663 0.6354 0.6157 0.5644
## Proportion of Variance 0.0000 0.0000 0.0000 0.00 0.000 0.0000 0.0000 0.0000
## Cumulative Proportion 1.0000 1.0000 1.0000 1.00 1.000 1.0000 1.0000 1.0000
##
##          PC23  PC24  PC25  PC26  PC27  PC28  PC29  PC30
## Standard deviation  0.5446 0.5213 0.4828 0.4705 0.4405 0.4339 0.4244 0.4188
## Proportion of Variance 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## Cumulative Proportion 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
##
##          PC31  PC32  PC33  PC34  PC35  PC36  PC37  PC38
## Standard deviation  0.3818 0.3737 0.3644 0.3581 0.3531 0.3442 0.3333 0.3294
## Proportion of Variance 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## Cumulative Proportion 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
##
##          PC39  PC40  PC41  PC42  PC43  PC44  PC45  PC46
## Standard deviation  0.2932 0.2872 0.285 0.2787 0.2663 0.2555 0.2374 0.2296
## Proportion of Variance 0.0000 0.0000 0.000 0.0000 0.0000 0.0000 0.0000 0.0000
## Cumulative Proportion 1.0000 1.0000 1.000 1.0000 1.0000 1.0000 1.0000 1.0000
##
##          PC47  PC48  PC49  PC50  PC51  PC52  PC53  PC54
## Standard deviation  0.2246 0.2145 0.2129 0.208 0.2019 0.1954 0.181 0.1496
## Proportion of Variance 0.0000 0.0000 0.0000 0.000 0.0000 0.0000 0.000 0.0000
## Cumulative Proportion 1.0000 1.0000 1.0000 1.000 1.0000 1.0000 1.000 1.0000
##
##          PC55  PC56  PC57
## Standard deviation  0.1073 0.07584 0.02043
## Proportion of Variance 0.0000 0.00000 0.00000
## Cumulative Proportion 1.0000 1.00000 1.00000
```

myPCA\$rotation

##	PC1	PC2	PC3	PC4
## make	-1.022087e-04	-4.541890e-05	-4.490153e-04	-2.541895e-03
## address	-9.541413e-05	-1.369553e-04	-4.711984e-04	-4.886593e-03
## all	-2.256004e-04	-2.811083e-04	-1.523393e-03	-6.612773e-03
## num3d	-8.190910e-05	-1.097435e-04	1.707555e-04	-2.486603e-04
## our	-2.002601e-04	-2.767713e-04	-1.350685e-03	-8.341562e-03
## over	-9.102534e-05	-9.854589e-05	4.118651e-04	-2.068632e-03
## remove	-6.895689e-05	-1.846407e-04	-3.763784e-04	-1.881229e-03
## internet	-8.800294e-05	-5.920978e-05	-1.193335e-04	-1.672902e-03
## order	-1.502266e-04	-1.222832e-04	-4.949379e-04	-1.297854e-03
## mail	-2.278354e-04	-2.836937e-04	-1.145964e-03	-6.615848e-03
## receive	-6.884914e-05	-4.389702e-05	-4.554198e-05	-1.084189e-03
## will	-3.631081e-04	-1.432366e-04	-4.313481e-04	-1.496633e-02
## people	-1.008612e-04	1.386519e-05	1.576911e-04	-2.652412e-03
## report	-1.118957e-04	5.490192e-05	1.221291e-04	-1.435975e-03
## addresses	-8.574622e-05	-2.366520e-04	8.655873e-04	-5.941727e-04
## free	-1.606572e-04	-1.745152e-04	-5.592351e-04	-2.004176e-02
## business	-1.283768e-04	-1.069504e-04	-4.648741e-04	-2.529575e-03
## email	-1.505395e-04	-2.047377e-04	6.259147e-04	-4.831351e-03
## you	-1.025011e-03	-3.642642e-04	-1.011489e-03	-6.900163e-02
## credit	-1.073990e-04	-2.161722e-04	-5.679290e-04	-9.603014e-04
## your	-5.953271e-04	-5.851912e-04	-1.644766e-03	-2.482667e-02
## font	-2.165240e-04	1.653378e-04	-8.715102e-04	2.446864e-04
## num000	-1.422742e-04	-9.667666e-05	4.706605e-04	-1.172000e-03
## money	-1.070447e-04	-2.003734e-05	-1.336411e-06	-1.888137e-03
## hp	-2.441375e-04	2.338247e-04	-1.738598e-03	-5.139059e-02
## hpl	-9.426071e-05	8.128361e-05	-5.715296e-04	-2.839795e-02
## george	-4.509405e-05	-3.393614e-05	-1.278913e-03	-9.936528e-01
## num650	-2.940450e-05	-7.360141e-06	-3.277292e-04	-1.719048e-02
## lab	-1.682889e-05	4.670571e-06	-2.133059e-04	-1.500575e-02
## labs	-2.510551e-05	6.022455e-07	-2.247933e-04	-1.615368e-02
## telnet	-1.569858e-05	-9.058063e-07	-1.538770e-04	-1.308837e-02
## num857	-8.841273e-06	-2.972809e-07	-1.198721e-04	-1.294312e-02
## data	-6.487350e-05	9.523114e-05	-1.958966e-04	-3.636561e-03
## num415	-1.032588e-05	-5.012473e-06	-2.114300e-04	-1.294430e-02
## num85	-3.308966e-05	5.720804e-06	-2.177749e-04	-1.572512e-02
## technology	-3.600541e-05	2.310168e-05	-1.845072e-04	-1.400990e-02
## num1999	-8.286281e-05	6.809384e-05	-3.623697e-04	-8.711666e-03
## parts	-4.134900e-06	1.261986e-06	-1.167289e-05	-4.091669e-04
## pm	-2.052283e-05	-2.125582e-06	-1.403687e-04	-6.930584e-03
## direct	-2.739014e-05	-3.367192e-05	-8.978947e-05	-1.278484e-02
## cs	-1.436454e-05	1.695464e-05	-1.274491e-04	-1.524706e-03
## meeting	-2.469060e-05	5.150582e-06	-1.868299e-04	-9.595333e-03
## original	-1.797348e-05	-9.555836e-06	2.740897e-06	-4.578363e-03
## project	-1.559293e-05	6.987723e-06	-1.207375e-04	-3.710042e-03
## re	-5.946026e-05	-3.367319e-05	-4.145843e-04	-2.553255e-02
## edu	-5.556685e-05	2.777276e-05	-2.932207e-04	-3.800793e-03
## table	-3.866324e-06	5.743669e-06	-1.037681e-05	-2.097950e-04
## conference	-1.589284e-05	1.737557e-05	-7.459937e-05	-2.936348e-03
## charSemicolon	-4.236248e-05	-2.437599e-05	6.880506e-05	-1.205128e-03
## charRoundbracket	-1.336097e-04	-5.788937e-04	1.543163e-03	-8.651152e-03
## charSquarebracket	-1.137933e-05	1.000212e-05	-3.691105e-05	-7.395488e-04

## charExclamation	-2.115385e-04	-3.582195e-04	-1.070533e-03	-6.314027e-03
## charDollar	-1.152162e-04	-1.370329e-04	-1.033410e-04	-7.976478e-04
## charHash	-5.333751e-05	-1.182263e-04	2.172996e-04	-9.757131e-04
## capitalAve	-1.107468e-02	-8.925898e-02	-9.959353e-01	1.575716e-03
## capitalLong	-1.669671e-01	-9.818586e-01	8.985426e-02	-3.273784e-05
## capitalTotal	-9.858992e-01	1.672867e-01	-4.025787e-03	1.615520e-04
##	PC5	PC6	PC7	PC8
## make	0.0368499365	-0.0054429984	3.826598e-03	-3.520221e-03
## address	0.0461623514	-0.0044697098	-1.249949e-02	-9.898887e-01
## all	0.0905691185	-0.0005433867	-5.813714e-03	-5.798771e-03
## num3d	0.0118497501	-0.0098892552	9.976329e-01	-6.550119e-03
## our	0.1061804693	0.0007964598	7.268806e-03	-1.430041e-02
## over	0.0293076318	-0.0055443183	-1.349187e-03	-9.153518e-04
## remove	0.0433090472	-0.0114521737	7.829537e-03	-1.039355e-02
## internet	0.0316880697	-0.0029814026	5.285222e-03	-6.723486e-03
## order	0.0197036816	-0.0040929561	-6.644974e-04	-6.645605e-03
## mail	0.0764618273	0.0097045109	-1.136102e-03	-3.935423e-02
## receive	0.0233779427	-0.0056244897	-1.201112e-03	-4.469995e-03
## will	0.1763922241	0.0511682423	-8.251613e-03	-1.998872e-02
## people	0.0291695940	-0.0040548672	-3.124332e-03	2.618392e-04
## report	0.0103051726	-0.0023745635	2.305036e-03	-1.113279e-03
## addresses	0.0116266365	-0.0024233402	-6.640472e-05	-4.994906e-03
## free	0.0886778024	-0.0291324369	1.203663e-02	-1.968883e-02
## business	0.0499963728	-0.0051844295	3.762369e-03	-7.999906e-03
## email	0.0632231042	0.0050650796	1.062632e-02	-3.151273e-02
## you	0.8571994546	-0.1314398816	-2.640281e-02	9.903321e-02
## credit	0.0259750077	-0.0088594715	-7.536011e-04	-1.758001e-03
## your	0.3578586467	-0.0626150061	2.879896e-02	-7.591842e-02
## font	0.0090825049	-0.0136350331	4.381481e-02	-2.260821e-03
## num000	0.0315224436	-0.0106693851	2.911529e-03	-5.341205e-04
## money	0.0440501993	-0.0148134563	1.219427e-02	-4.513000e-03
## hp	0.0975891492	0.8616067764	9.398594e-03	6.311695e-03
## hpl	0.0471008922	0.3190812013	2.704381e-03	1.701190e-05
## george	-0.0896774692	-0.0608692952	3.390130e-04	8.109332e-04
## num650	0.0236479191	0.1406095917	1.754846e-03	-8.805553e-04
## lab	0.0209577013	0.1098014352	9.803814e-04	-1.374779e-03
## labs	0.0216747714	0.1400673058	1.815770e-03	-8.107567e-04
## telnet	0.0126683620	0.1024588331	1.671080e-03	-1.428121e-03
## num857	0.0103950463	0.0874800898	1.511478e-03	-4.079430e-03
## data	0.0076405453	0.0162107407	-1.870672e-03	1.456918e-03
## num415	0.0104470840	0.0874721722	1.490161e-03	-3.832984e-03
## num85	0.0206776007	0.1286574421	2.671516e-03	-1.771626e-04
## technology	0.0164290226	0.1143536279	2.717494e-03	-1.868186e-03
## num1999	0.0199318402	0.0492153813	-6.912456e-05	-4.101360e-03
## parts	0.0023025256	0.0029904898	7.931959e-05	-2.778530e-04
## pm	0.0178532479	0.0247597160	-3.401347e-04	8.257553e-04
## direct	0.0157369215	0.0855441183	1.397652e-03	-2.820955e-03
## cs	0.0068995377	0.0077771055	-1.288240e-03	1.968304e-03
## meeting	0.0231393725	0.0418846791	-1.798471e-03	7.826495e-04
## original	0.0097784876	0.0208311287	-6.391964e-04	-3.233645e-03
## project	0.0111543680	0.0144646427	-8.933052e-04	1.768141e-03
## re	0.1209869950	0.0592080772	-1.374339e-02	1.276824e-02
## edu	0.0445129045	0.0025528400	-8.346264e-03	1.632163e-02
## table	0.0013298924	0.0008190753	-9.204222e-05	1.241415e-04

## conference	0.0044160248	0.0046317364	1.389478e-05	8.572202e-04
## charSemicolon	0.0045241350	0.0069036191	2.454408e-03	-1.229621e-03
## charRoundbracket	0.0208782694	0.0436149036	-1.721595e-03	5.790107e-04
## charSquarebracket	0.0022453941	0.0047517255	-2.078217e-04	6.973703e-06
## charExclamation	0.1071467974	-0.0244764554	-2.344476e-03	-2.610767e-03
## charDollar	0.0207121882	-0.0068807169	1.611192e-03	-2.853049e-03
## charHash	0.0092139044	0.0140017182	2.334858e-03	-3.961010e-03
## capitalAve	-0.0023281226	-0.0014955178	9.246218e-05	5.737146e-04
## capitalLong	-0.0007284704	0.0004149993	-1.310101e-04	1.411869e-04
## capitalTotal	-0.0012258601	-0.0001602570	-6.486806e-05	4.182433e-05
##	PC9	PC10	PC11	PC12
## make	0.0404285808	-0.0084791318	3.328711e-03	-8.187519e-03
## address	-0.1175962141	0.0001394602	-3.444668e-02	1.304081e-02
## all	0.0654572633	-0.0142664941	-4.020442e-03	-4.381343e-02
## num3d	-0.0402198206	-0.0453376644	-1.274042e-02	-7.345534e-03
## our	0.1175486080	-0.0046822098	-9.503425e-03	-1.359073e-01
## over	0.0204927026	0.0049460655	-2.494653e-04	2.256000e-03
## remove	0.0417520472	0.0086036430	-3.441971e-03	1.403979e-02
## internet	0.0591453078	-0.0079907291	3.688794e-02	7.470107e-03
## order	0.0375711527	-0.0090954777	1.196949e-03	4.343674e-03
## mail	0.0521258379	0.0145552872	-8.793587e-03	-4.056902e-02
## receive	0.0427732874	-0.0015470150	4.280010e-03	2.927146e-03
## will	0.2054908825	-0.0459228527	-1.447083e-01	-5.527981e-01
## people	0.0028894721	-0.0078623677	-1.257150e-02	3.897794e-03
## report	0.0077912514	-0.0138285197	6.243478e-03	-5.401735e-03
## addresses	0.0151568112	-0.0023011730	-2.977698e-03	-1.855576e-03
## free	0.1200245000	0.0206186405	1.418392e-02	7.044786e-02
## business	0.0925212998	-0.0062094366	1.702274e-02	9.233864e-03
## email	0.0623936904	-0.0072949658	-7.419599e-03	7.191131e-03
## you	-0.3622254964	0.0379840167	-2.457036e-01	1.312789e-01
## credit	0.0635405295	0.0212752147	1.853973e-02	1.429427e-02
## your	0.7675480729	-0.0248430923	3.467689e-01	1.221366e-01
## font	-0.0398868761	0.9637790975	1.848404e-01	-1.031705e-01
## num000	0.0265804252	0.0078222837	-4.772787e-03	9.799809e-03
## money	0.0312932710	-0.0020837141	-1.615134e-02	3.367672e-02
## hp	-0.0044522847	0.0237203346	4.644396e-02	2.875428e-01
## hpl	0.0199093144	0.0090327900	-9.477095e-02	-1.433028e-01
## george	0.0056156269	0.0018010885	-4.259758e-03	1.135411e-02
## num650	0.0106530049	0.0089481628	-6.527286e-02	-1.138802e-01
## lab	0.0356051005	-0.0006630903	-8.939642e-02	-3.143918e-01
## labs	0.0117093522	0.0121278720	-5.640240e-02	-7.188088e-02
## telnet	0.0113662538	0.0103059140	-5.373481e-02	-7.251286e-02
## num857	0.0081861007	0.0087236424	-4.071468e-02	-6.312666e-02
## data	-0.0067982513	-0.0156415435	1.482740e-02	-4.745143e-02
## num415	0.0079874887	0.0085766692	-4.082591e-02	-6.302302e-02
## num85	0.0084242214	0.0103474656	-7.115745e-02	-1.025155e-01
## technology	0.0130883209	0.0075515485	-5.206756e-02	-8.049362e-02
## num1999	-0.0201694796	-0.0212982253	4.225048e-02	-6.626444e-02
## parts	0.0058827324	-0.0012646045	1.954638e-03	-1.861487e-02
## pm	-0.0218253178	-0.0151685394	4.211914e-02	-5.641765e-02
## direct	0.0196572306	0.0078654203	-4.403631e-02	-6.304448e-02
## cs	-0.0262509702	-0.0140065452	3.126340e-02	-5.738630e-02
## meeting	0.0312151686	-0.0293428334	-4.699214e-02	-4.823615e-01
## original	-0.0079065229	-0.0061192595	1.759669e-02	-2.008921e-02

## project	0.0012849673	-0.0052451645	-2.726906e-03	-7.277591e-02
## re	-0.3438923708	-0.1857597779	8.143178e-01	-1.504320e-01
## edu	-0.1782564345	-0.0763801369	1.778558e-01	-2.994117e-01
## table	0.0012738800	-0.0004145812	-5.673342e-04	-1.065783e-03
## conference	-0.0030758827	-0.0037947757	-1.243264e-03	-1.753234e-02
## charSemicolon	-0.0101564854	0.1001616494	2.066972e-02	-1.571471e-02
## charRoundbracket	-0.0014118355	-0.0080426581	-6.744008e-05	-5.274984e-02
## charSquarebracket	-0.0006671359	-0.0001292852	2.987299e-03	-4.087915e-03
## charExclamation	-0.0019025807	-0.0137742263	1.213926e-01	-1.734862e-02
## charDollar	0.0213341233	-0.0048507486	2.512550e-03	6.239255e-03
## charHash	-0.0006545930	0.0937028832	1.517944e-02	1.061003e-02
## capitalAve	-0.0012276327	-0.0007290979	-8.786405e-04	2.402822e-04
## capitalLong	-0.0004048641	0.0001988410	-4.059816e-05	4.224787e-05
## capitalTotal	-0.0001606212	-0.0002336501	-3.543528e-06	5.459762e-05
##	PC13	PC14	PC15	PC16
## make	9.574352e-04	-0.0133897692	8.219577e-03	-1.555325e-02
## address	-4.054911e-03	-0.0019933915	-1.540793e-02	8.400719e-03
## all	2.275604e-02	-0.0340319554	6.937287e-02	-3.782318e-02
## num3d	3.059608e-03	-0.0062628734	-7.472936e-03	-8.462474e-03
## our	1.020430e-01	-0.0654244997	1.103436e-01	2.464520e-01
## over	-1.204705e-03	-0.0062485877	1.467619e-02	-1.180359e-02
## remove	1.094338e-02	-0.0059372053	6.266807e-02	2.603790e-02
## internet	-1.685018e-03	0.0021458698	2.181849e-02	-7.257538e-03
## order	-1.079711e-03	-0.0004326567	1.300372e-03	-1.709012e-02
## mail	-9.356363e-04	-0.0251146196	3.438669e-02	-1.133081e-01
## receive	-2.144878e-03	-0.0064890549	2.555921e-03	-2.992540e-04
## will	2.134602e-01	-0.4483521694	-9.767638e-02	-3.618693e-01
## people	-2.879979e-04	-0.0078266962	8.856566e-03	-6.246508e-03
## report	4.701885e-03	0.0026402390	-6.204830e-03	1.092879e-02
## addresses	-5.226369e-04	-0.0012586265	4.312660e-03	-9.624836e-03
## free	5.828011e-02	-0.0279301315	7.682314e-01	3.576199e-01
## business	6.832869e-03	-0.0201969081	3.341294e-02	-1.113386e-02
## email	-1.032209e-02	0.0216778177	6.361875e-02	-8.851104e-03
## you	-1.165653e-03	0.0356025396	-9.018396e-02	6.237026e-02
## credit	-5.068159e-03	-0.0018096423	1.543259e-02	-1.628675e-02
## your	-1.706823e-01	0.1201479431	-1.872759e-01	4.046123e-02
## font	-7.039388e-04	-0.0232882665	-2.171685e-02	-1.066754e-02
## num000	8.509624e-03	-0.0026344696	2.798963e-02	-1.054577e-02
## money	-7.002473e-03	0.0066010502	2.780802e-02	1.117110e-02
## hp	8.173738e-03	-0.3270452259	-6.414391e-03	9.033706e-02
## hpl	-3.016049e-02	0.1903490869	1.338508e-02	-1.380065e-01
## george	-2.274573e-03	-0.0200637989	-1.048619e-02	-6.699425e-03
## num650	-5.937880e-02	0.3103408221	2.601802e-02	-7.718536e-02
## lab	4.953604e-02	0.2329437538	1.290067e-02	1.617312e-01
## labs	-4.727544e-02	0.2258273387	2.342928e-02	-5.923565e-02
## telnet	-3.638902e-02	0.2427979716	2.243526e-02	-4.750484e-02
## num857	-3.312775e-02	0.2258068236	1.841587e-02	-3.071079e-02
## data	1.790376e-02	-0.0251990497	-1.683716e-02	6.425748e-03
## num415	-3.301954e-02	0.2260496768	1.908421e-02	-3.130007e-02
## num85	-5.384147e-02	0.3082843260	2.628116e-02	-5.317337e-02
## technology	-3.028284e-02	0.2247835206	2.302618e-02	-5.908350e-02
## num1999	-6.189213e-02	-0.0425693274	8.654010e-03	-1.538905e-02
## parts	1.015032e-02	-0.0059521168	8.504914e-03	3.986685e-02
## pm	1.744925e-02	-0.0041925326	-9.576861e-03	-1.322210e-02

## direct	-3.406586e-02	0.2245206622	2.098664e-02	-3.565566e-02
## cs	-1.393503e-01	-0.0364465753	2.188617e-02	-6.248858e-03
## meeting	1.890822e-01	-0.0616867279	-9.159825e-04	5.072160e-01
## original	-5.078719e-03	0.0084227450	-6.665857e-03	-2.418715e-03
## project	2.501065e-02	-0.0228254845	-1.258759e-02	-1.713662e-02
## re	2.763449e-01	0.1314212305	-6.759961e-02	1.620013e-02
## edu	-8.592532e-01	-0.2066883216	1.279477e-01	2.238254e-02
## table	-2.081349e-04	0.0006507410	-1.897682e-03	-7.592230e-05
## conference	7.164404e-03	-0.0086199291	-6.528491e-03	-1.239190e-02
## charSemicolon	-6.266940e-03	-0.0177208609	3.483632e-03	-9.969016e-03
## charRoundbracket	-2.106645e-02	0.0569039722	1.271959e-02	-1.787338e-02
## charSquarebracket	-1.174548e-03	0.0007364374	-4.124730e-04	9.606961e-04
## charExclamation	1.310642e-01	-0.0079639399	5.457674e-01	-5.671607e-01
## charDollar	5.343284e-03	0.0001867174	2.539815e-02	-2.081056e-02
## charHash	1.759305e-03	-0.0212512099	2.976337e-02	1.657856e-02
## capitalAve	-4.577898e-07	0.0003625541	-8.621961e-04	1.604684e-04
## capitalLong	-4.451712e-05	-0.0001014009	-2.159394e-04	1.524568e-04
## capitalTotal	1.303320e-05	0.0001274608	-3.759411e-05	6.557268e-05
##	PC17	PC18	PC19	PC20
## make	1.092034e-02	-4.251620e-03	1.696686e-02	-2.570172e-03
## address	-1.507221e-02	2.131998e-03	2.064308e-02	2.899741e-02
## all	-9.910850e-03	2.017005e-02	-5.977920e-02	1.456538e-03
## num3d	-1.123227e-03	-5.846268e-04	3.125879e-03	-1.431627e-03
## our	-8.219515e-02	-4.985148e-02	-7.648418e-01	1.888839e-01
## over	-9.084272e-03	-9.415384e-03	-3.169823e-02	7.662545e-04
## remove	1.824859e-02	-1.085209e-02	-1.026429e-01	-1.436128e-02
## internet	1.316327e-02	9.249068e-03	-5.378283e-02	-6.713689e-02
## order	-5.041856e-05	6.574543e-03	-2.983480e-02	-4.557865e-02
## mail	1.238417e-01	1.750239e-02	-3.688464e-01	-7.883272e-01
## receive	1.578573e-02	-7.459124e-03	-7.734434e-03	-2.298167e-02
## will	3.488578e-01	-1.816167e-01	1.500113e-01	5.932427e-02
## people	-8.391215e-03	-7.342110e-03	-2.792768e-02	-1.689459e-02
## report	8.659498e-04	-8.571114e-03	4.054453e-03	-6.175546e-03
## addresses	8.260371e-03	2.285655e-03	-5.783403e-02	-5.456106e-02
## free	4.375542e-01	-1.820223e-03	2.048471e-01	1.179920e-02
## business	-1.760127e-02	-3.031683e-02	-1.306641e-01	-3.271948e-02
## email	3.863866e-02	-9.679088e-03	-1.178500e-01	-1.790484e-02
## you	-2.549224e-02	1.268808e-02	5.343518e-02	1.915064e-02
## credit	-2.402454e-02	-1.212134e-02	-3.810799e-02	-1.921470e-02
## your	-8.672990e-02	4.760007e-02	1.263164e-01	4.195535e-02
## font	1.491160e-02	2.486895e-03	6.710797e-03	1.522784e-02
## num000	-3.293362e-03	-9.044139e-03	-5.876440e-02	-4.690376e-02
## money	7.504313e-03	2.165560e-05	4.823812e-02	-2.987689e-02
## hp	-1.149328e-01	-1.606622e-01	2.523106e-02	-1.121523e-02
## hpl	1.625828e-01	8.553456e-01	-2.181641e-02	3.991989e-02
## george	-9.084061e-03	-1.203466e-04	-3.587618e-03	-1.344556e-03
## num650	7.467359e-02	-1.087385e-01	-3.353988e-02	-1.773824e-02
## lab	-1.636658e-01	-1.568750e-01	1.298493e-01	-6.165129e-02
## labs	5.145557e-02	-1.389236e-01	-1.773272e-02	-3.897541e-03
## telnet	3.973467e-02	-1.341038e-01	-1.514030e-02	5.179802e-02
## num857	2.327441e-02	-1.230554e-01	9.136250e-04	1.729025e-02
## data	3.419906e-02	2.831046e-02	-5.519398e-02	9.012078e-02
## num415	2.274499e-02	-1.228571e-01	1.137025e-03	1.721915e-02
## num85	3.878172e-02	-1.341494e-01	-9.792205e-03	1.910985e-02

## technology	5.231978e-02	-1.418612e-01	-1.820481e-02	1.957208e-02
## num1999	3.732623e-02	1.107370e-01	-3.157646e-02	-8.668504e-03
## parts	-2.356201e-02	1.984461e-02	-2.846725e-02	5.282139e-03
## pm	4.021149e-02	3.655982e-02	-4.130657e-03	1.024576e-02
## direct	2.569893e-02	-1.301247e-01	-6.741536e-03	8.959518e-03
## cs	7.286764e-03	9.739317e-03	-5.095033e-03	4.927028e-03
## meeting	-4.577455e-01	1.476148e-01	1.726590e-01	-1.339581e-01
## original	1.623332e-02	2.493424e-02	-8.229177e-03	-1.158236e-02
## project	6.897401e-02	2.854677e-02	-2.775803e-01	5.291020e-01
## re	1.728975e-01	-2.322570e-02	-6.074111e-03	-1.804775e-02
## edu	-4.220360e-02	-2.931740e-02	-2.109365e-02	9.778285e-03
## table	2.264625e-04	4.557133e-03	2.297644e-03	8.706034e-04
## conference	1.767087e-02	1.801065e-02	3.544568e-03	4.696820e-03
## charSemicolon	-5.561826e-03	1.200718e-02	-1.609856e-03	-1.570842e-03
## charRoundbracket	2.585618e-02	-2.163868e-02	-1.549605e-02	9.971546e-03
## charSquarebracket	2.058075e-03	1.062927e-02	-4.758575e-03	-2.668438e-03
## charExclamation	-5.636073e-01	3.072397e-02	1.764031e-02	4.659308e-02
## charDollar	-1.936401e-02	-3.037230e-03	-1.545801e-02	-1.778890e-02
## charHash	-1.236427e-02	-4.022988e-02	-1.895698e-02	-3.375229e-02
## capitalAve	5.058281e-04	-1.055678e-04	1.035606e-03	3.980422e-04
## capitalLong	4.682420e-05	6.775934e-05	1.976138e-04	1.438786e-04
## capitalTotal	-2.736466e-05	-4.426204e-06	5.922834e-05	6.234807e-05
##	PC21	PC22	PC23	PC24
## make	-4.648537e-03	-4.072467e-03	2.771000e-02	3.253482e-04
## address	7.032208e-03	-3.798620e-03	-1.988736e-02	-4.708230e-03
## all	9.189702e-02	9.611010e-02	4.241157e-01	2.380367e-02
## num3d	-2.212420e-03	-4.311332e-04	-5.556742e-03	-3.678850e-04
## our	3.918687e-01	-4.594797e-02	-1.816903e-01	1.271986e-01
## over	1.667481e-02	9.088054e-03	5.256917e-02	-4.110854e-02
## remove	4.868365e-02	1.867053e-03	4.516620e-02	-8.952683e-02
## internet	-1.091001e-02	-2.503986e-02	-8.726029e-03	-2.223615e-01
## order	-1.139927e-02	8.805360e-03	4.365272e-02	-8.169140e-02
## mail	-3.963253e-01	-1.742889e-02	-6.716440e-02	1.056631e-01
## receive	-3.896099e-03	-8.941146e-03	-3.510865e-03	-5.310140e-02
## will	1.024746e-01	-5.373836e-02	-3.750856e-02	6.524999e-03
## people	3.203909e-03	-3.059152e-03	5.210527e-02	5.485262e-03
## report	-7.542457e-03	-1.239974e-02	-2.434887e-02	-1.302122e-02
## addresses	-2.178698e-03	2.091063e-02	1.321851e-01	-1.094565e-02
## free	-5.728137e-02	-8.680334e-03	-6.421706e-02	2.346774e-02
## business	2.880134e-02	-4.497428e-02	-6.638459e-02	-3.786108e-01
## email	8.685585e-02	1.553779e-01	7.824294e-01	-1.058462e-01
## you	-1.564321e-02	-4.042559e-03	-3.986660e-02	-1.578202e-03
## credit	-3.592766e-03	-7.558448e-02	-1.095233e-01	-8.457483e-01
## your	-4.882838e-02	7.066309e-03	-5.890735e-02	1.137275e-01
## font	1.528000e-02	8.985969e-03	1.493761e-02	5.563247e-03
## num000	1.016822e-02	-6.498654e-03	9.546470e-02	-3.551802e-02
## money	-3.742798e-02	-1.120135e-02	7.805023e-02	-7.366533e-02
## hp	-3.150745e-02	-5.488826e-03	5.234007e-03	6.557107e-03
## hpl	1.131395e-01	-7.582031e-02	-4.723801e-02	-2.045709e-02
## george	2.724386e-03	-2.289108e-03	2.277135e-03	-6.772618e-04
## num650	1.171852e-02	5.120896e-02	6.703364e-02	1.451064e-02
## lab	-5.612451e-02	-4.090606e-02	-9.797934e-02	-1.040855e-02
## labs	1.875967e-02	9.802523e-03	2.346374e-02	1.588332e-03
## telnet	-2.311407e-02	-1.158759e-02	5.361636e-03	-4.707818e-03

## num857	1.010257e-02	-1.718157e-02	-6.073566e-02	3.522237e-03
## data	-1.160367e-01	9.224865e-01	-2.047226e-01	-7.685374e-02
## num415	9.897127e-03	-1.732380e-02	-6.139723e-02	4.412971e-03
## num85	1.774332e-02	3.352120e-02	5.142841e-02	2.003701e-03
## technology	1.737975e-02	1.460500e-02	6.711610e-03	-8.170487e-03
## num1999	-3.394471e-02	1.462805e-01	-1.829608e-03	-1.040254e-02
## parts	1.175327e-02	1.210853e-02	8.147975e-03	1.795590e-03
## pm	-3.765676e-02	1.769865e-01	1.155667e-02	3.960386e-03
## direct	9.992424e-03	-1.494497e-02	-3.903392e-02	-1.863879e-02
## cs	8.385604e-03	1.399232e-02	3.799202e-03	-3.567336e-03
## meeting	-1.485940e-01	1.032782e-02	1.132079e-01	-3.632448e-02
## original	-1.688300e-02	3.422248e-02	2.260542e-03	3.151394e-03
## project	-7.676349e-01	-1.610417e-01	9.557612e-02	-3.645002e-03
## re	4.165449e-02	-4.946445e-02	2.413932e-02	-2.324074e-02
## edu	2.225328e-02	-3.143303e-02	-5.318222e-03	-8.712329e-03
## table	1.821322e-04	1.744055e-03	5.342150e-03	2.709738e-04
## conference	1.946332e-03	1.096466e-02	1.659333e-03	4.099687e-03
## charSemicolon	-5.429306e-03	1.329180e-02	-6.189969e-03	1.832585e-02
## charRoundbracket	-9.227141e-03	5.594600e-02	-1.578074e-02	1.157408e-02
## charSquarebracket	-3.756291e-03	2.640247e-02	-3.122622e-03	-1.422585e-03
## charExclamation	-2.645665e-02	1.558713e-02	-5.844960e-02	5.429523e-02
## charDollar	4.347946e-04	-6.060932e-03	2.989916e-02	-8.082826e-03
## charHash	-3.871715e-02	1.115581e-02	3.564256e-02	1.591784e-02
## capitalAve	1.747716e-05	3.336491e-05	6.205440e-04	1.308667e-04
## capitalLong	-1.613015e-05	2.032367e-05	-2.358740e-04	1.226435e-04
## capitalTotal	2.057591e-05	-7.496999e-05	-4.329633e-05	6.221252e-05
##	PC25	PC26	PC27	PC28
## make	-7.925472e-03	-1.459352e-03	-9.250548e-03	-1.254998e-01
## address	1.020547e-02	6.425083e-03	-2.111743e-03	4.060619e-03
## all	7.432352e-01	4.075734e-01	1.181297e-01	-4.810394e-02
## num3d	1.016879e-02	6.258612e-03	-9.241014e-04	6.543789e-03
## our	-2.300461e-02	-1.892982e-02	-9.255444e-02	-1.140496e-01
## over	2.140100e-02	1.925684e-02	6.750300e-02	-1.545322e-02
## remove	-9.210863e-02	-6.032342e-02	2.047268e-01	1.787633e-01
## internet	-3.283294e-03	-5.215593e-02	2.497642e-01	2.012658e-01
## order	2.477217e-03	1.256149e-02	5.240585e-02	-2.038430e-02
## mail	-1.197213e-03	5.038042e-02	-5.480366e-02	-3.470892e-02
## receive	-1.526958e-02	-1.205174e-02	1.484153e-02	2.273618e-02
## will	-7.079482e-02	1.022043e-02	-3.301744e-02	1.207955e-02
## people	-4.043179e-04	-1.130288e-02	7.206549e-02	-5.203409e-02
## report	2.214166e-02	2.889925e-02	-2.322119e-02	-2.080585e-02
## addresses	-2.083748e-02	1.449427e-02	7.404448e-02	-2.830475e-02
## free	4.316496e-03	2.035669e-03	-2.635653e-02	2.412496e-02
## business	-3.288833e-02	-7.278329e-02	3.462535e-01	3.327564e-01
## email	-4.316405e-01	-1.634347e-01	1.599130e-02	-1.715301e-02
## you	-1.371137e-02	-1.395565e-02	-3.641870e-03	3.009843e-02
## credit	1.150741e-01	5.829103e-02	-2.933281e-01	-1.662353e-01
## your	-1.290481e-02	-1.154607e-02	-4.507565e-02	1.387658e-02
## font	1.531014e-03	2.860233e-03	2.216804e-03	-1.178021e-02
## num000	4.082764e-02	5.029894e-02	1.554372e-01	-1.396713e-02
## money	-1.313603e-01	-2.716699e-03	-1.036780e-01	-7.328729e-01
## hp	-4.479154e-03	1.275281e-02	-9.147641e-03	-5.353081e-03
## hpl	-7.450112e-02	8.917559e-02	7.053998e-02	-2.750836e-02
## george	-2.361368e-03	4.075797e-03	2.393253e-04	1.410392e-04

## num650	5.857804e-03	1.045025e-02	-3.757268e-01	1.852158e-01
## lab	-6.861761e-02	2.717487e-02	4.366033e-01	-2.410659e-01
## labs	2.329304e-02	3.969063e-03	6.008086e-03	2.752244e-02
## telnet	1.724090e-02	-1.393816e-02	8.188366e-02	-5.220501e-02
## num857	2.446654e-02	-8.795276e-03	1.080730e-01	-7.569964e-02
## data	-1.224295e-01	2.221176e-01	4.853866e-02	-1.509013e-02
## num415	2.453263e-02	-9.412166e-03	1.085040e-01	-7.560228e-02
## num85	5.041651e-03	1.911697e-02	-4.045622e-01	2.435665e-01
## technology	3.177801e-02	-3.881347e-02	6.561162e-02	-2.981560e-02
## num1999	2.407057e-01	-5.520159e-01	6.305364e-02	-4.403613e-02
## parts	1.558720e-02	1.044293e-02	-4.614898e-02	2.502251e-03
## pm	3.246811e-01	-6.098393e-01	-4.837928e-02	-8.250667e-02
## direct	9.387565e-03	-9.289509e-03	1.358025e-01	-6.290711e-02
## cs	3.442955e-02	-5.275839e-02	-3.453396e-02	2.889227e-02
## meeting	2.438040e-03	-1.204804e-02	-1.713958e-01	1.326305e-01
## original	7.296291e-02	-1.566045e-01	1.190681e-02	-2.094277e-02
## project	5.621518e-03	3.572536e-02	1.947198e-02	6.241396e-03
## re	-3.702513e-02	6.678175e-02	9.416572e-03	-1.340766e-02
## edu	-4.065095e-02	7.225492e-02	1.586885e-02	-1.126090e-02
## table	2.680764e-03	2.792588e-03	-1.906429e-03	2.947337e-03
## conference	8.042413e-03	-2.618015e-02	-1.460516e-02	6.674408e-03
## charSemicolon	2.270587e-02	-3.807999e-02	-8.074243e-04	-7.132138e-03
## charRoundbracket	7.740811e-02	-7.048218e-02	-2.279671e-02	7.526662e-03
## charSquarebracket	4.455280e-03	-7.626401e-03	-4.405297e-03	3.212637e-04
## charExclamation	-5.050674e-02	-3.315541e-02	-3.320582e-02	-2.349247e-03
## charDollar	4.728280e-03	1.272216e-02	5.185165e-02	-2.532057e-02
## charHash	-3.035360e-02	9.226416e-03	7.856641e-02	1.042526e-01
## capitalAve	-1.194594e-03	-6.176842e-04	2.045767e-04	1.093977e-04
## capitalLong	-2.257848e-05	1.732021e-05	-3.406141e-05	-3.715693e-06
## capitalTotal	-4.077946e-05	-1.233740e-05	-3.481761e-05	4.998830e-05
##	PC29	PC30	PC31	PC32
## make	1.243093e-01	6.605541e-02	-8.653970e-03	-8.850110e-02
## address	5.211065e-03	5.025717e-04	4.094527e-03	-3.828659e-03
## all	2.437434e-02	1.915031e-02	-8.896599e-02	8.530495e-02
## num3d	-8.068066e-03	-2.191441e-03	1.500201e-03	-1.058753e-03
## our	-3.720648e-02	-8.529343e-03	6.775583e-02	8.392007e-03
## over	3.793928e-02	1.467203e-02	1.043509e-01	-4.364936e-02
## remove	3.291223e-02	1.255764e-01	-5.653340e-01	1.093943e-01
## internet	3.071553e-01	1.379243e-02	6.540147e-01	1.166831e-01
## order	5.857881e-02	7.196203e-03	3.475922e-02	-2.487163e-02
## mail	-5.695479e-02	-6.337083e-02	-4.965945e-02	-2.503399e-03
## receive	7.556045e-03	-1.917240e-03	-1.631825e-02	2.413147e-03
## will	-2.371388e-03	1.399473e-02	9.245958e-03	-2.616879e-03
## people	7.255189e-02	-2.779749e-04	6.985304e-02	-7.805094e-02
## report	2.087822e-02	3.835661e-02	5.938450e-02	-5.306422e-02
## addresses	4.845957e-03	-5.900875e-02	1.276151e-01	-9.487901e-02
## free	-2.022953e-02	-4.335008e-02	1.966002e-02	-1.156040e-02
## business	4.089565e-01	1.282240e-01	-2.975895e-01	-9.194465e-02
## email	-1.799032e-01	-1.394491e-01	1.641297e-02	2.270976e-03
## you	-2.052385e-02	-1.796286e-02	1.109115e-02	5.038583e-03
## credit	-3.271916e-01	-8.428191e-02	9.443039e-03	6.806118e-03
## your	-5.694797e-02	-1.699969e-02	-7.509671e-03	-8.384459e-04
## font	5.363317e-02	-7.193846e-02	-1.083834e-02	-2.613552e-04
## num000	1.014738e-01	-5.882053e-03	2.529591e-01	-2.518287e-01

## money	4.883304e-01	3.347611e-01	-8.007449e-02	9.609367e-02
## hp	1.363650e-02	-3.917349e-02	-1.769865e-03	-9.076601e-03
## hpl	-3.539056e-02	3.178531e-02	-5.586669e-03	-5.950446e-02
## george	9.362458e-04	6.894024e-04	-2.776264e-04	-1.835879e-04
## num650	1.452346e-01	4.407355e-02	-1.888358e-02	-1.079970e-02
## lab	-2.215637e-01	-1.435068e-01	2.146395e-02	1.284959e-01
## labs	-5.030100e-02	1.350258e-01	5.349288e-03	1.748219e-02
## telnet	-2.923187e-02	-1.957699e-02	-1.955688e-02	-5.624253e-02
## num857	-2.964164e-02	-2.905321e-02	-1.547893e-02	-4.797707e-02
## data	1.496758e-02	-4.268903e-03	8.078831e-03	-1.066092e-02
## num415	-2.949790e-02	-2.907132e-02	-1.558401e-02	-4.781329e-02
## num85	2.248517e-01	7.513192e-02	6.311519e-02	1.464775e-01
## technology	-1.282766e-02	2.500157e-03	-2.282900e-02	-6.737513e-02
## num1999	-2.669783e-02	-1.470038e-02	2.703462e-02	6.270333e-01
## parts	6.094631e-03	6.125689e-03	4.078727e-03	-1.402702e-02
## pm	-1.984001e-02	3.720175e-02	-3.126268e-02	-5.925390e-01
## direct	-2.172801e-02	-3.887237e-02	-7.527632e-03	-5.830319e-02
## cs	3.709953e-02	2.308889e-02	5.970691e-02	1.614192e-01
## meeting	1.100183e-01	5.432497e-02	-2.193524e-02	-5.924813e-02
## original	-1.314541e-02	-1.344014e-02	2.654352e-02	7.697877e-02
## project	7.152097e-03	-1.447448e-02	-3.613143e-03	6.459714e-03
## re	7.241878e-03	1.671871e-02	-1.697378e-02	8.122184e-03
## edu	9.765145e-04	-8.505804e-04	-1.637637e-02	-8.480557e-02
## table	-1.026643e-03	7.008502e-04	-2.266079e-05	-2.391981e-03
## conference	-7.037425e-03	2.521893e-03	4.371144e-03	6.769479e-02
## charSemicolon	1.796599e-02	-2.694853e-02	-6.332919e-03	2.525735e-02
## charRoundbracket	2.121196e-02	2.487949e-02	-1.258164e-02	2.121850e-02
## charSquarebracket	1.070762e-03	-1.458433e-03	4.425649e-03	1.036836e-02
## charExclamation	-1.672677e-02	-1.840736e-02	3.935896e-03	4.395313e-03
## charDollar	7.200557e-02	2.257686e-02	5.680005e-02	-8.106386e-02
## charHash	-4.017514e-01	8.639898e-01	1.534869e-01	-7.325953e-04
## capitalAve	-9.248729e-05	2.068862e-04	6.612515e-04	-3.927191e-04
## capitalLong	1.074340e-04	-1.025371e-04	-4.294046e-05	8.061874e-05
## capitalTotal	-7.069685e-05	-1.174135e-05	-5.745968e-05	-1.108371e-06
##	PC33	PC34	PC35	PC36
## make	-1.201218e-01	-2.431280e-02	-1.625232e-01	1.094497e-03
## address	-1.509873e-03	2.977176e-03	-2.811294e-03	1.548955e-04
## all	1.148133e-01	4.040933e-02	4.064323e-02	-4.781765e-02
## num3d	1.025508e-03	-1.306109e-03	-3.523047e-03	-1.390527e-03
## our	3.874616e-02	1.120558e-02	-7.079376e-03	-2.356076e-02
## over	-1.989858e-01	1.754515e-02	-2.654806e-02	5.571840e-02
## remove	-3.629391e-01	2.131843e-01	5.362591e-01	2.286640e-01
## internet	2.186931e-01	1.172436e-01	4.406698e-01	1.359597e-01
## order	-7.342329e-02	-1.960957e-02	-4.269109e-02	-1.504332e-02
## mail	4.556979e-02	-3.388202e-02	1.993350e-02	-4.015938e-02
## receive	-3.953732e-02	-6.364758e-03	1.802268e-02	1.940315e-02
## will	-1.163031e-02	-2.799092e-02	2.395041e-02	8.425755e-03
## people	-1.466090e-01	5.619839e-03	-1.124416e-01	3.295868e-02
## report	-5.130714e-02	-4.322626e-02	-1.501348e-01	2.697912e-01
## addresses	-2.306405e-01	-7.535135e-03	-3.477080e-02	-8.579043e-03
## free	1.003448e-02	-1.328792e-02	-3.314248e-02	-1.451405e-02
## business	2.293017e-01	-8.197791e-02	-4.252581e-01	-2.002644e-01
## email	1.391831e-01	-3.610762e-02	-3.813681e-02	-5.209843e-02
## you	1.909647e-02	-5.680250e-03	-2.559151e-03	-1.333611e-03

## credit	-7.181178e-02	6.724013e-03	2.785441e-02	1.846854e-02
## your	6.772359e-04	-3.809766e-03	1.986820e-03	-2.974270e-03
## font	2.341369e-03	-4.434312e-03	-1.849662e-03	4.901532e-03
## num000	-6.324802e-01	2.412691e-02	-2.102980e-01	7.839928e-02
## money	6.716035e-02	3.749160e-02	8.618481e-02	-5.480926e-03
## hp	-6.040099e-03	2.322838e-02	8.723048e-03	-3.655086e-03
## hpl	4.305921e-03	4.833463e-03	8.637343e-03	-3.676979e-02
## george	-2.654484e-05	2.645999e-03	1.044692e-03	-1.215170e-03
## num650	1.140969e-01	5.072425e-01	-2.595384e-01	4.655075e-01
## lab	2.240692e-02	5.055656e-01	-1.089509e-01	-2.648575e-01
## labs	5.133370e-02	-2.362624e-01	1.027741e-02	3.182085e-02
## telnet	2.605048e-02	-2.910434e-01	8.309102e-02	9.096813e-02
## num857	-4.149224e-03	-1.812078e-01	5.816135e-02	6.713146e-02
## data	-1.252498e-02	-2.280037e-02	7.434988e-03	-1.074777e-03
## num415	-4.722202e-03	-1.824228e-01	5.680712e-02	6.745586e-02
## num85	-2.715471e-01	-6.114530e-03	1.421874e-01	-6.087013e-01
## technology	7.013958e-02	-2.021005e-01	3.466274e-02	1.654732e-01
## num1999	-1.626166e-01	-1.346250e-01	-2.134415e-01	1.036967e-01
## parts	1.901889e-02	-7.976123e-02	1.637822e-02	3.929318e-02
## pm	9.709783e-02	1.779894e-01	1.775698e-01	-1.611186e-01
## direct	-1.678031e-02	-2.082081e-01	5.070952e-02	6.247333e-02
## cs	-6.604657e-02	6.717403e-02	-1.080448e-02	-1.022883e-01
## meeting	-1.602250e-02	-2.254429e-01	4.821149e-02	1.340249e-01
## original	-3.698292e-02	-8.860458e-03	-2.900052e-02	1.762708e-02
## project	-5.773622e-03	3.289133e-02	2.656005e-03	-7.414347e-03
## re	-1.601649e-02	-4.624059e-03	-8.531192e-03	-9.951046e-03
## edu	1.479318e-02	-4.500447e-03	2.501314e-02	4.563803e-03
## table	4.248530e-03	-7.259656e-04	-1.499229e-04	-1.255299e-03
## conference	9.744001e-03	7.606117e-04	-2.734952e-02	3.167219e-03
## charSemicolon	1.572650e-02	-5.619400e-03	-7.196786e-03	-3.018522e-03
## charRoundbracket	4.769026e-02	7.072378e-02	-5.630610e-02	1.326314e-01
## charSquarebracket	-1.210808e-02	4.823221e-03	-3.923873e-03	-1.870743e-03
## charExclamation	1.617085e-02	-1.996380e-03	1.159920e-02	6.224340e-03
## charDollar	-2.027634e-01	9.451721e-03	-8.507801e-02	5.359670e-02
## charHash	1.478930e-02	2.854000e-02	-6.553211e-02	-3.750560e-02
## capitalAve	-5.336833e-04	7.846357e-06	-2.977562e-04	3.396163e-04
## capitalLong	1.415701e-04	-7.511645e-05	1.856254e-05	-1.042770e-04
## capitalTotal	8.107849e-05	2.238206e-05	7.323901e-05	-3.049759e-05
##	PC37	PC38	PC39	PC40
## make	6.160053e-02	3.189431e-03	4.308397e-02	9.189486e-01
## address	7.150449e-04	-2.509458e-03	-3.545022e-04	-1.166524e-03
## all	-2.295284e-02	1.772609e-02	2.747791e-03	1.361638e-02
## num3d	9.234587e-04	-2.634694e-03	3.482559e-03	-1.451322e-03
## our	-3.327142e-03	-1.023788e-03	-1.066317e-02	9.720974e-03
## over	2.105676e-03	-6.523354e-02	-2.047358e-02	-6.475872e-02
## remove	-3.293029e-04	6.655589e-02	1.786263e-02	4.792019e-02
## internet	-4.257106e-02	1.215740e-02	2.095638e-02	1.005431e-01
## order	-8.553774e-03	4.290077e-03	-1.593338e-01	5.899948e-02
## mail	4.347754e-03	1.549469e-02	7.383951e-03	9.097146e-03
## receive	-4.121750e-03	1.347750e-02	1.561612e-03	1.325179e-01
## will	-8.712015e-03	-1.222961e-02	-1.318536e-03	-3.521392e-02
## people	-2.670403e-02	1.594197e-01	9.030894e-01	-1.223282e-01
## report	-2.117482e-01	8.838242e-01	-2.143825e-01	-3.508031e-02
## addresses	2.788336e-02	-1.048323e-01	-1.339019e-01	-1.212467e-01

## free	-1.892024e-04	-5.416318e-03	2.598986e-03	-1.622750e-02
## business	6.031012e-03	-1.034242e-02	-3.688381e-02	-8.006882e-02
## email	-4.623598e-03	6.107859e-02	-9.804661e-03	2.688434e-02
## you	6.144704e-03	-6.239998e-03	-1.848335e-02	-2.877769e-03
## credit	1.075900e-02	-1.341671e-02	4.919380e-02	9.322674e-03
## your	3.628021e-03	-1.119930e-02	5.250068e-03	-3.999964e-02
## font	-6.730132e-03	1.405260e-02	1.270632e-03	1.904630e-03
## num000	3.121468e-02	-1.675567e-01	-2.115756e-01	-1.297081e-01
## money	-1.042617e-02	-3.502801e-02	-5.673090e-02	-1.570367e-01
## hp	-4.401809e-03	2.159949e-03	-2.788747e-03	5.987912e-03
## hpl	-4.611728e-03	1.821306e-02	3.565830e-03	-1.327929e-03
## george	-3.403093e-04	-2.179237e-05	-7.903900e-04	8.178939e-05
## num650	2.966919e-02	-1.613998e-01	4.275060e-03	-2.610430e-02
## lab	-5.782563e-02	8.610390e-02	-3.260911e-02	1.902317e-02
## labs	1.626423e-01	-4.865876e-03	-5.956014e-02	-1.370717e-01
## telnet	2.274016e-02	-1.899362e-02	1.732085e-02	3.766084e-02
## num857	2.948071e-02	-2.713036e-02	1.888108e-02	3.537203e-02
## data	4.786586e-03	4.649829e-03	1.107895e-02	1.489726e-03
## num415	2.958762e-02	-2.768938e-02	1.977365e-02	4.029982e-02
## num85	-2.020409e-01	1.416159e-01	1.204208e-02	2.375221e-02
## technology	-3.297425e-03	2.508238e-02	1.114212e-03	3.818791e-03
## num1999	-1.583167e-01	-7.374752e-02	-1.606627e-02	-3.351904e-03
## parts	1.497605e-02	-1.491006e-02	6.463046e-03	1.047806e-02
## pm	1.294618e-02	6.215131e-02	-2.780747e-02	-2.419331e-02
## direct	3.264227e-02	-4.587035e-02	2.231888e-02	-1.465498e-03
## cs	9.056370e-01	2.500711e-01	-8.417485e-03	-3.108206e-02
## meeting	4.211572e-02	-4.925996e-02	1.493092e-02	-1.014044e-02
## original	2.928600e-02	1.423912e-02	-3.174811e-04	1.228998e-03
## project	2.615216e-03	2.620867e-03	-4.449844e-03	-3.630521e-03
## re	1.673051e-03	-8.364481e-03	6.375039e-03	-3.231506e-03
## edu	-1.364359e-01	-3.239855e-02	-2.749185e-03	3.614547e-03
## table	-3.767964e-03	1.290802e-02	-4.498146e-03	1.779842e-03
## conference	-2.463189e-02	-2.870772e-02	6.552200e-02	5.447733e-02
## charSemicolon	4.886726e-02	1.903020e-02	1.917306e-02	1.226696e-02
## charRoundbracket	6.590541e-03	-4.078714e-02	4.511160e-02	5.970031e-02
## charSquarebracket	6.938872e-03	-1.349361e-03	-4.914379e-03	-2.759098e-03
## charExclamation	5.432535e-04	6.414213e-03	-1.449309e-02	-5.209443e-03
## charDollar	-9.715354e-03	2.534643e-02	1.555279e-01	-3.294682e-02
## charHash	-3.161252e-02	-3.104624e-02	2.551507e-02	2.108101e-02
## capitalAve	-6.140626e-06	-1.222763e-04	2.253984e-05	-4.839149e-04
## capitalLong	-1.919106e-05	1.048675e-04	1.774325e-05	1.409042e-05
## capitalTotal	3.083992e-05	-9.108975e-05	-1.114156e-05	-1.723059e-05
##	PC41	PC42	PC43	PC44
## make	-5.121299e-02	-1.408853e-01	-3.369049e-02	-0.0352110023
## address	1.024749e-03	-3.152343e-03	-1.229356e-03	-0.0005878375
## all	-1.717548e-03	2.623745e-02	-3.279052e-02	0.0124313956
## num3d	-1.217356e-03	5.552354e-04	4.099389e-03	-0.0003349695
## our	-1.175485e-03	-4.618335e-03	-1.193797e-03	0.0165100548
## over	-2.116521e-02	-7.873716e-02	5.381849e-01	-0.7707581716
## remove	1.179802e-02	-4.637465e-02	-1.238348e-02	0.0250341810
## internet	-2.433359e-03	-3.186091e-02	-6.358032e-02	0.0199527089
## order	-3.610388e-02	-3.762241e-02	7.043191e-01	0.5205353233
## mail	4.995446e-03	1.327735e-02	-2.648281e-02	-0.0475099098
## receive	-1.245665e-02	-1.879934e-02	-1.227396e-03	0.0482986218

## will	-2.201540e-02	2.497398e-02	-1.848839e-03	0.0094651363
## people	-5.107102e-02	-6.875057e-02	8.352378e-02	0.1021964794
## report	3.194658e-02	2.680324e-03	-7.394954e-03	-0.0267170725
## addresses	6.462765e-03	5.403334e-02	1.956457e-01	0.1982706107
## free	4.206658e-04	-2.191857e-03	1.635558e-02	0.0047655509
## business	1.132960e-02	4.689185e-02	-3.170929e-02	-0.0432453940
## email	-1.304713e-03	3.247694e-02	-5.861434e-02	-0.0421944736
## you	-1.137632e-03	2.954608e-03	-1.806879e-03	0.0098497553
## credit	7.813843e-04	-6.027886e-03	-4.021149e-02	-0.0072745288
## your	7.115255e-03	-6.541833e-04	-1.301832e-02	-0.0030158683
## font	-8.467797e-05	-1.621152e-02	2.709669e-03	0.0101377230
## num000	6.657248e-02	3.658790e-03	-3.567749e-01	0.0348668593
## money	1.795629e-02	2.256971e-02	-3.938666e-02	-0.0074841563
## hp	5.306501e-03	2.549408e-02	3.973956e-03	0.0036112488
## hpl	-1.227380e-02	5.585541e-03	-4.706702e-03	-0.0115904304
## george	-1.790744e-03	-3.288276e-04	-1.021950e-03	0.0002254667
## num650	-2.485009e-02	3.801992e-02	4.190902e-02	0.0782476559
## lab	5.654724e-03	-7.022501e-02	-1.686722e-02	-0.0069920444
## labs	-2.926818e-02	-8.576238e-01	-5.679847e-02	0.0349843561
## telnet	-1.027862e-02	2.052383e-01	-2.867643e-02	0.0176507526
## num857	2.307064e-06	1.452244e-01	3.863773e-02	0.0042678528
## data	-4.148262e-03	-2.052123e-03	-1.325301e-02	0.0072963317
## num415	-7.626503e-04	1.458796e-01	3.828734e-02	0.0042554231
## num85	-2.677350e-03	7.919171e-02	-1.934970e-04	-0.0335271264
## technology	7.929313e-02	1.188893e-01	-3.551217e-02	-0.0797043876
## num1999	-6.562122e-02	-1.244534e-02	-2.703233e-02	0.0303231368
## parts	3.743712e-02	-4.462081e-04	1.201148e-02	0.0333247895
## pm	2.397610e-02	-4.901394e-02	1.086230e-02	0.0055797126
## direct	2.932850e-03	2.060844e-01	3.863222e-02	0.0495750252
## cs	2.887588e-02	1.433934e-01	9.024251e-03	-0.0090604996
## meeting	-1.264710e-04	1.482626e-02	8.625167e-03	-0.0024859315
## original	-4.916160e-02	8.152523e-02	4.040763e-02	0.0272177341
## project	-4.160298e-04	-2.174515e-02	7.862281e-04	0.0045257406
## re	-5.962271e-03	-3.594931e-03	4.354630e-03	0.0016375061
## edu	4.051597e-03	-1.678616e-02	-6.738426e-04	0.0082864156
## table	-1.057391e-03	5.555290e-03	1.681737e-03	-0.0003748138
## conference	9.840160e-01	-4.472086e-02	6.476549e-02	0.0141374533
## charSemicolon	3.618001e-03	4.288774e-02	2.365954e-02	-0.0271619705
## charRoundbracket	-1.151207e-02	9.891764e-02	-3.101701e-02	-0.2094812982
## charSquarebracket	-5.010012e-03	1.516719e-02	1.079155e-02	0.0025091896
## charExclamation	3.779991e-03	-3.826031e-03	-2.090871e-03	0.0091964701
## charDollar	-2.966264e-03	-5.355201e-03	-7.433360e-02	0.0752376686
## charHash	-4.341956e-03	1.453326e-01	1.795914e-02	0.0222295502
## capitalAve	-2.290457e-05	2.632427e-04	-3.302663e-05	-0.0007451282
## capitalLong	2.055221e-05	-1.096442e-04	-8.016251e-05	0.0001402248
## capitalTotal	-7.914826e-06	-1.697537e-06	-5.198014e-05	-0.0000403476
##	PC45	PC46	PC47	PC48
## make	-7.292309e-02	5.536729e-02	3.428110e-02	3.435746e-02
## address	-2.174965e-03	-2.163909e-03	-2.820642e-04	7.761720e-05
## all	-1.583371e-02	2.013222e-02	-4.390416e-03	-8.487621e-03
## num3d	8.313348e-04	4.804666e-04	-5.257935e-04	1.789046e-03
## our	-8.598562e-03	-8.522127e-03	-9.685969e-03	-1.117982e-02
## over	-9.121953e-02	-1.364605e-02	-6.856397e-02	-1.680280e-02
## remove	-3.807539e-03	1.356950e-02	2.946155e-03	2.617994e-02

## internet	3.236346e-03	9.061932e-03	-7.341597e-03	2.054959e-03
## order	1.475933e-01	-2.398976e-01	-2.157947e-02	-8.825015e-02
## mail	-2.651871e-03	-1.182631e-02	-1.117442e-02	-1.732653e-02
## receive	-7.930158e-02	5.474949e-03	-1.974038e-02	3.140185e-02
## will	4.133178e-03	1.971856e-02	-9.112542e-03	-1.888991e-03
## people	-1.021607e-01	6.477046e-02	3.154230e-02	7.385144e-02
## report	5.828017e-03	9.611894e-02	-3.107489e-02	2.011360e-02
## addresses	-1.494288e-01	3.147975e-01	1.578781e-01	2.212548e-01
## free	-4.775238e-03	1.199873e-04	2.130209e-03	3.009923e-03
## business	-2.402308e-02	4.534948e-02	-8.716150e-03	1.882965e-02
## email	6.853266e-02	7.021295e-03	-9.642465e-03	-3.625233e-02
## you	-2.710349e-03	-3.427940e-03	3.202120e-03	-1.235353e-03
## credit	1.101828e-02	-1.820289e-03	8.098743e-03	1.362796e-02
## your	-7.246156e-03	-4.505358e-03	5.982882e-03	8.557878e-03
## font	-3.238984e-02	1.387667e-02	1.773409e-03	-9.349622e-02
## num000	-6.578442e-04	-5.277229e-02	-1.696789e-02	3.825590e-02
## money	-1.267186e-02	1.655290e-02	6.594878e-04	1.299579e-02
## hp	-2.752952e-03	6.268485e-03	-6.533185e-03	-1.058483e-02
## hpl	-2.517322e-03	-6.098172e-03	3.197048e-03	-3.502621e-03
## george	-1.588371e-03	-1.466606e-03	-3.680499e-04	5.309021e-05
## num650	-1.116833e-01	5.998048e-02	-6.084786e-02	2.802698e-02
## lab	-2.511229e-02	-6.341916e-02	-2.313213e-02	3.304377e-02
## labs	7.456515e-02	6.358467e-02	-2.665040e-02	2.864976e-02
## telnet	-2.553698e-01	-2.449459e-01	-7.314005e-01	1.083027e-01
## num857	1.421760e-01	2.388228e-01	8.050383e-02	-7.531384e-02
## data	-2.075110e-02	5.779917e-03	-4.861147e-03	-1.550067e-03
## num415	1.446256e-01	2.408226e-01	8.113014e-02	-7.834201e-02
## num85	3.105398e-02	-1.033830e-02	3.129486e-02	3.820248e-03
## technology	-2.560687e-01	-5.449057e-01	6.226245e-01	5.029581e-02
## num1999	-5.741573e-02	-3.409736e-02	-7.011705e-03	-3.084300e-02
## parts	7.295274e-02	2.876724e-02	2.692702e-02	2.056373e-01
## pm	-2.375297e-02	-3.811401e-03	-3.098629e-02	-1.973503e-02
## direct	1.328850e-01	3.138328e-01	6.058171e-02	-6.821975e-02
## cs	-3.231667e-02	-4.017301e-02	6.914405e-03	-6.049144e-02
## meeting	-7.940788e-04	1.369585e-02	1.228295e-02	-2.153653e-02
## original	-1.084382e-01	2.756416e-01	3.142143e-02	4.982579e-02
## project	9.992985e-03	1.298095e-02	3.555361e-02	-6.422937e-03
## re	-4.460904e-03	-9.226211e-03	-3.399608e-03	5.180481e-03
## edu	-6.626434e-03	4.499894e-03	-1.204586e-03	3.095739e-03
## table	5.316107e-03	5.331342e-03	2.328599e-03	1.694768e-03
## conference	9.152259e-03	5.045986e-02	-5.767980e-02	-1.488358e-02
## charSemicolon	3.551899e-01	-1.472729e-01	-1.659397e-02	8.519326e-01
## charRoundbracket	6.487294e-01	-9.949718e-03	1.211009e-02	-1.046529e-01
## charSquarebracket	7.450012e-03	1.335116e-02	-5.512128e-03	-4.991778e-04
## charExclamation	-1.895455e-02	1.809966e-02	5.477915e-03	9.555974e-04
## charDollar	3.728559e-01	-3.640359e-01	-1.097663e-01	-3.164590e-01
## charHash	-7.942970e-03	1.381498e-02	4.012157e-03	2.141312e-02
## capitalAve	9.631964e-04	3.328065e-04	1.402547e-04	2.696365e-04
## capitalLong	-5.140352e-04	-4.206283e-05	-1.988596e-05	-2.095647e-06
## capitalTotal	-1.377765e-05	3.332177e-05	2.140800e-06	1.138819e-05
##	PC49	PC50	PC51	PC52
## make	2.039693e-02	2.121404e-02	4.569107e-02	6.836421e-02
## address	-1.284943e-04	-2.734607e-03	1.123163e-03	1.992603e-03
## all	5.378613e-03	2.692449e-02	-2.784453e-02	-6.443118e-03

## num3d	3.207411e-04	-8.432291e-04	1.417837e-05	-2.478625e-04
## our	4.089169e-02	1.041829e-02	-8.189645e-03	-5.797110e-03
## over	-5.787610e-02	8.876840e-02	-5.967793e-02	7.062990e-04
## remove	-3.120065e-03	-1.327384e-02	2.443616e-02	-5.390228e-04
## internet	-1.487922e-03	1.159760e-02	-5.358373e-03	6.827657e-03
## order	-1.607965e-03	-1.432067e-01	-1.039418e-01	-2.190114e-01
## mail	-1.008899e-02	1.387754e-03	-2.049333e-02	-8.921835e-03
## receive	-2.805813e-02	-4.164575e-02	-1.312215e-01	-2.895453e-02
## will	-1.444677e-02	9.412249e-03	-9.079847e-03	-2.223552e-03
## people	2.198398e-02	-1.673685e-01	-4.022611e-02	-4.526457e-02
## report	4.844175e-03	-1.395693e-02	-4.435190e-03	4.164578e-02
## addresses	1.547556e-01	-2.660157e-02	4.460446e-01	5.685968e-01
## free	6.826679e-03	2.606918e-03	5.171121e-03	3.161209e-03
## business	-1.001349e-02	9.902755e-03	2.739940e-02	1.989890e-02
## email	-1.579354e-02	1.062474e-02	-6.108732e-02	-5.653817e-02
## you	-2.637591e-03	1.898772e-03	2.705696e-03	6.366752e-03
## credit	1.761251e-02	5.456974e-03	1.734406e-02	1.185584e-02
## your	5.288168e-03	-7.736292e-03	6.389092e-03	3.194267e-03
## font	-1.503967e-02	-2.626525e-02	1.981693e-02	5.499854e-03
## num000	-2.640546e-02	-1.825107e-01	-1.766704e-01	-2.598836e-01
## money	3.575137e-03	-1.507326e-02	-1.523107e-03	-9.236751e-03
## hp	-3.298311e-03	1.114591e-03	-6.274599e-03	3.810773e-03
## hpl	3.786473e-03	-1.494306e-03	-5.328495e-03	3.561602e-03
## george	-6.853774e-04	2.944277e-04	-4.097611e-04	6.871587e-04
## num650	-3.102706e-02	9.238617e-02	-8.482763e-02	2.636678e-02
## lab	-6.500902e-02	-1.955933e-02	4.383917e-02	7.829554e-03
## labs	4.136489e-02	2.064585e-03	6.735664e-02	-1.539159e-02
## telnet	4.784242e-02	-1.344867e-01	1.961661e-01	4.113752e-02
## num857	-2.848775e-02	8.669271e-02	-1.887863e-01	-4.461033e-02
## data	5.222535e-03	2.191903e-02	-1.224307e-02	-3.854972e-03
## num415	-2.946712e-02	9.118702e-02	-1.904919e-01	-4.200953e-02
## num85	-8.956554e-03	-1.971536e-02	1.849662e-02	-3.981780e-03
## technology	3.906234e-02	-5.181217e-02	1.472616e-01	-3.428083e-02
## num1999	-3.190474e-02	-1.478649e-02	-1.332734e-01	1.413538e-01
## parts	-9.453479e-01	-1.094155e-01	1.570271e-01	3.834535e-02
## pm	-2.134811e-02	8.045704e-03	-6.562381e-02	5.090882e-02
## direct	-1.815496e-02	1.113097e-01	-1.808480e-01	9.817761e-02
## cs	-1.810128e-02	-2.965653e-02	-3.697498e-02	4.035685e-02
## meeting	6.893915e-02	1.103794e-02	-2.444906e-02	-3.069931e-03
## original	3.627458e-02	3.154621e-01	5.447742e-01	-6.668932e-01
## project	-6.529688e-03	1.158791e-02	-1.620855e-02	-1.442830e-03
## re	-1.345792e-03	7.309715e-04	-1.101793e-02	6.799540e-03
## edu	-3.250722e-06	1.293935e-02	4.111894e-03	-1.339395e-02
## table	3.869027e-03	-2.344476e-04	2.780359e-03	-1.177365e-02
## conference	3.676821e-02	2.173437e-02	1.244347e-02	-1.909121e-02
## charSemicolon	1.539998e-01	2.313220e-01	-1.562964e-01	-4.178445e-02
## charRoundbracket	1.425747e-01	-5.355712e-01	3.649609e-01	-1.623783e-02
## charSquarebracket	4.273723e-03	5.104474e-03	2.761966e-02	-6.139128e-02
## charExclamation	5.428179e-04	-2.046566e-02	3.636977e-03	-5.516218e-03
## charDollar	-8.601909e-02	6.256266e-01	2.089128e-01	2.310121e-01
## charHash	-2.173702e-03	1.403833e-02	-1.421772e-02	1.137453e-02
## capitalAve	2.955601e-04	-9.239213e-04	8.873756e-04	2.831532e-04
## capitalLong	-1.439351e-04	3.196003e-04	-3.752081e-04	-1.090362e-04
## capitalTotal	4.555897e-06	-4.179184e-06	1.627650e-05	6.548468e-06

##	PC53	PC54	PC55	PC56
## make	-1.279765e-01	-1.550281e-02	6.106093e-03	7.503928e-04
## address	5.356029e-04	-2.602850e-03	3.163613e-04	-2.569748e-04
## all	1.179770e-02	-1.944627e-04	-1.726439e-04	5.885703e-03
## num3d	3.073267e-03	-5.365140e-04	1.338742e-04	-1.002669e-04
## our	3.434393e-03	-5.178788e-03	-1.667393e-03	-2.611772e-03
## over	2.066002e-02	-2.281028e-02	-4.134152e-03	-2.077842e-03
## remove	-6.456165e-02	1.816028e-03	-4.225902e-04	-1.247059e-03
## internet	-3.581359e-02	9.830560e-03	2.325392e-03	1.075054e-04
## order	-5.601802e-02	-2.008386e-02	-1.878092e-02	4.015904e-03
## mail	-2.206478e-02	2.388740e-03	-4.011076e-03	-9.428903e-04
## receive	9.683013e-01	-5.759296e-02	1.930417e-03	-1.034461e-02
## will	-1.543215e-02	9.643188e-03	2.181036e-03	5.197785e-04
## people	-2.218517e-02	7.490610e-03	2.202634e-03	-2.297314e-03
## report	-1.153874e-02	-1.720229e-03	3.183923e-03	1.280818e-02
## addresses	6.941775e-02	1.119002e-01	1.576952e-02	-5.387084e-03
## free	-1.003126e-02	1.771181e-03	8.088829e-05	-9.295596e-04
## business	-1.252108e-02	3.300364e-02	4.632503e-03	3.457803e-04
## email	-1.209730e-02	1.565421e-02	-1.471277e-03	5.093872e-03
## you	-2.099731e-03	1.199555e-03	-2.013180e-07	7.434799e-04
## credit	-3.833872e-02	6.999001e-03	-4.956354e-04	-6.739464e-04
## your	-2.803500e-02	4.844947e-03	-6.957214e-04	1.872894e-03
## font	2.959114e-03	-8.031699e-04	-6.992298e-04	-6.227636e-04
## num000	-4.822360e-02	-9.239216e-03	-1.349815e-02	-1.402763e-03
## money	2.660695e-02	-3.283042e-03	4.929656e-05	-2.590729e-03
## hp	-8.168255e-04	1.332222e-03	-1.148834e-03	-3.047203e-04
## hpl	5.377201e-03	-3.415295e-03	-6.687095e-03	4.533044e-03
## george	1.490274e-04	-3.108307e-04	2.485455e-05	1.787216e-05
## num650	-1.177772e-02	2.251531e-03	-2.725342e-03	-1.737707e-03
## lab	6.774505e-03	-1.361724e-02	1.074904e-03	-1.853551e-03
## labs	1.301037e-02	-3.016456e-02	8.223860e-03	-3.918510e-03
## telnet	-1.852919e-02	2.311311e-02	4.019112e-04	-3.231373e-03
## num857	1.126067e-02	4.247944e-01	1.015627e-02	7.449018e-05
## data	1.391680e-03	3.358354e-03	-2.355463e-02	8.842031e-04
## num415	1.140790e-02	4.376403e-01	1.142713e-02	-9.769501e-05
## num85	1.411200e-03	-3.087732e-03	-7.928637e-03	4.176369e-03
## technology	1.195897e-02	-3.758957e-02	7.902714e-03	-1.030443e-03
## num1999	-2.898281e-02	1.430596e-02	-5.028756e-03	-5.332044e-03
## parts	-1.170100e-02	-6.101305e-04	1.855805e-03	-2.502636e-03
## pm	5.812158e-04	-3.280326e-03	5.239992e-04	9.869070e-04
## direct	-5.451669e-02	-7.756622e-01	-1.693397e-02	3.251601e-03
## cs	3.214656e-03	5.331125e-03	-8.557312e-03	-9.222556e-04
## meeting	3.053766e-04	4.194139e-03	-2.359340e-03	2.178394e-03
## original	5.264379e-02	-4.751857e-02	-6.851698e-02	1.098765e-02
## project	4.407782e-03	-1.572700e-03	1.498965e-03	1.954126e-04
## re	3.961406e-03	-3.215614e-03	-1.130546e-03	-7.417769e-04
## edu	1.951848e-03	-1.065380e-03	5.621249e-04	5.776598e-05
## table	-1.032697e-02	-2.983675e-03	-1.058697e-02	-9.995918e-01
## conference	7.906619e-03	6.891626e-04	1.096212e-03	-1.156620e-03
## charSemicolon	-3.826239e-03	-6.142026e-03	-2.732090e-03	3.378556e-03
## charRoundbracket	9.161702e-02	-2.791456e-02	-1.760336e-02	4.446684e-03
## charSquarebracket	-7.051209e-04	-2.809096e-02	9.962311e-01	-9.472116e-03
## charExclamation	6.714891e-03	-9.725769e-04	3.633126e-04	-1.121739e-03
## charDollar	9.182197e-02	1.745055e-02	7.257438e-03	-5.863367e-03

## charHash	2.783090e-03	5.467708e-03	1.445672e-04	1.377737e-03
## capitalAve	3.151370e-04	-2.285978e-05	-3.529679e-05	9.962559e-06
## capitalLong	-9.605859e-05	1.010862e-05	2.098798e-05	-8.750258e-06
## capitalTotal	-1.385645e-05	2.340386e-06	-5.678822e-06	2.295920e-06
##	PC57			
## make	3.278814e-03			
## address	-1.921544e-04			
## all	-2.148517e-04			
## num3d	-1.302054e-05			
## our	-1.573266e-04			
## over	-2.906221e-04			
## remove	-4.160159e-04			
## internet	-9.554592e-05			
## order	-8.914295e-04			
## mail	1.658225e-04			
## receive	-1.994017e-04			
## will	-6.598284e-05			
## people	-5.185394e-04			
## report	-2.965425e-04			
## addresses	6.941122e-04			
## free	9.140235e-05			
## business	3.790603e-04			
## email	1.751209e-05			
## you	1.110827e-04			
## credit	-4.879996e-04			
## your	-2.285463e-04			
## font	-6.380144e-05			
## num000	-1.171600e-03			
## money	-9.061593e-04			
## hp	-2.417913e-06			
## hpl	5.062669e-05			
## george	-1.541938e-05			
## num650	7.746974e-05			
## lab	-6.456284e-04			
## labs	-9.296934e-04			
## telnet	-1.271472e-03			
## num857	7.126483e-01			
## data	-6.591920e-05			
## num415	-7.014767e-01			
## num85	-2.455208e-04			
## technology	-1.595672e-03			
## num1999	1.188372e-03			
## parts	-3.289629e-05			
## pm	-2.322763e-04			
## direct	-5.124129e-03			
## cs	-2.150024e-05			
## meeting	2.010088e-04			
## original	-1.102983e-03			
## project	3.086172e-05			
## re	-1.381265e-04			
## edu	-1.127825e-04			
## table	6.481517e-05			
## conference	-1.986450e-04			
## charSemicolon	-6.451959e-04			

```
## charRoundbracket -8.259567e-04
## charSquarebracket 5.389826e-04
## charExclamation 6.355381e-04
## charDollar 3.321584e-03
## charHash 2.348677e-04
## capitalAve 6.145951e-05
## capitalLong -1.851025e-06
## capitalTotal 6.529919e-07
```

from this analysis we can see that a single variable is explaining 93 percent of the variation.

This is a very basic analysis to detect which emails are spam and which are not. Also this can be used to find the keywords to detect which emails are spam. The result of these two model can be used in a more advanced model. But this analysis can show that without any advanced techniques like text analysis we still can detect spam emails by using simple binary choice models like LPM or Logistic models.