**Computed Logistic Regression Model on R**

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : int 0 0 0 0 0 0 0 0 0 0 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : Factor w/ 5272 levels "","?","0","1",..: 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : Factor w/ 13 levels "","?","0","1",..: 1 3 6 3 3 8 3 4 3 3 ...

> credittrain$seriousd = as.factor(credittrain$seriousd)

> credittrain$monthlyincome = as.integer(credittrain$monthlyincome)

> credittrain$dependents = as.integer(credittrain$dependents)

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : int 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : int 1 3 6 3 3 8 3 4 3 3 ...

> table(is.na(credittrain))

FALSE TRUE

230992 8

> newcredit = na.omit(credittrain)

> table(is.na(newcredit))

FALSE

230989

> creditlog = glm(seriousd ~ age + no\_30\_59 + debtratio + no\_nightydayslate + no\_realestatell + no\_60\_89 + no\_opencreditll + dependents,data=newcredit,family='binomial')

> summary(creditlog)

Call:

glm(formula = seriousd ~ age + no\_30\_59 + debtratio + no\_nightydayslate +

no\_realestatell + no\_60\_89 + no\_opencreditll + dependents,

family = "binomial", data = newcredit)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.1567 -0.3968 -0.3186 -0.2532 3.6650

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.669e+00 1.357e-01 -12.297 < 2e-16 \*\*\*

age -2.927e-02 2.229e-03 -13.135 < 2e-16 \*\*\*

no\_30\_59 4.517e-01 2.939e-02 15.370 < 2e-16 \*\*\*

debtratio -6.794e-05 3.178e-05 -2.138 0.032525 \*

no\_nightydayslate 5.366e-01 4.069e-02 13.188 < 2e-16 \*\*\*

no\_realestatell 9.737e-02 2.693e-02 3.616 0.000299 \*\*\*

no\_60\_89 -9.597e-01 4.685e-02 -20.486 < 2e-16 \*\*\*

no\_opencreditll -1.127e-02 6.712e-03 -1.678 0.093288 .

dependents 6.867e-02 1.765e-02 3.891 9.97e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10308.0 on 20998 degrees of freedom

Residual deviance: 9488.2 on 20990 degrees of freedom

AIC: 9506.2

Number of Fisher Scoring iterations: 6

> table(newcredit$seriousd,predicttrain>0.5)

Error in table(newcredit$seriousd, predicttrain > 0.5) :

object 'predicttrain' not found

> predicttrain = predict(creditlog,type="response")

> summary(predicttrain)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00007 0.03525 0.05404 0.06686 0.07997 0.99310

> table(newcredit$seriousd,predicttrain>0.5)

FALSE TRUE

0 19551 44

1 1354 50

> credittest <- read.csv("C:/Users/anup/Desktop/credittest.csv")

> View(credittest)

> str(credittest)

'data.frame': 9000 obs. of 11 variables:

$ seriousd : int 0 0 0 0 0 0 0 0 0 0 ...

$ revolvingul : num 0.866 0.145 0.422 0.301 0.835 ...

$ age : int 58 29 33 38 54 41 56 67 22 66 ...

$ no\_30\_59 : int 0 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num 0.475 0.235 0.13 0.472 4642 ...

$ monthlyincome : Factor w/ 2900 levels "?","0","1","100",..: 1546 2549 1381 50 1 1899 1358 1 5 2602 ...

$ no\_opencreditll : int 9 4 10 13 15 7 12 4 1 16 ...

$ no\_nightydayslate: int 0 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int 2 0 0 4 1 0 2 0 0 0 ...

$ no\_60\_89 : int 0 0 0 0 0 0 0 0 0 0 ...

$ dependents : Factor w/ 11 levels "?","0","1","10",..: 2 2 2 5 2 3 2 2 2 3 ...

> credittest$seriousd = as.factor(credittest$seriousd)

> credittest$monthlyincome = as.integer(credittest$monthlyincome)

> credittest$dependents = as.integer(credittest$dependents)

> str(credittest)

'data.frame': 9000 obs. of 11 variables:

$ seriousd : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

$ revolvingul : num 0.866 0.145 0.422 0.301 0.835 ...

$ age : int 58 29 33 38 54 41 56 67 22 66 ...

$ no\_30\_59 : int 0 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num 0.475 0.235 0.13 0.472 4642 ...

$ monthlyincome : int 1546 2549 1381 50 1 1899 1358 1 5 2602 ...

$ no\_opencreditll : int 9 4 10 13 15 7 12 4 1 16 ...

$ no\_nightydayslate: int 0 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int 2 0 0 4 1 0 2 0 0 0 ...

$ no\_60\_89 : int 0 0 0 0 0 0 0 0 0 0 ...

$ dependents : int 2 2 2 5 2 3 2 2 2 3 ...

> predicttest = predict(creditlog,type="response",newdata=credittest)

> summary(predicttest)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0001014 0.0330700 0.0507000 0.0623600 0.0749000 0.9981000

> summary(predicttrain)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00007 0.03525 0.05404 0.06686 0.07997 0.99310

> table(credittest$seriousd,predicttest>0.5)

FALSE TRUE

0 8387 11

1 583 19

> table(newcredit$seriousd,predicttrain>0.5)

FALSE TRUE

0 19551 44

1 1354 50

> 19551+50

[1] 19601

> 19551+44+1354+50

[1] 20999

> 19601/20999

[1] 0.9334254

> 8387+19

[1] 8406

> 8387+11+583+19

[1] 9000

> 8406/9000

[1] 0.934

**Computed SVM model on R**

> credittrain <- read.csv("C:/Users/anup/Desktop/credittrain.csv")

> View(credittrain)

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : int 0 0 0 0 0 0 0 0 0 0 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : Factor w/ 5272 levels "","?","0","1",..: 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : Factor w/ 13 levels "","?","0","1",..: 1 3 6 3 3 8 3 4 3 3 ...

> credittrain$seriousd = as.factor(credittrain$seriousd)

> credittrain$monthlyincome = as.integer(credittrain$monthlyincome)

> credittrain$dependents = as.integer(credittrain$dependents)

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : int 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : int 1 3 6 3 3 8 3 4 3 3 ...

> table(is.na(credittrain))

FALSE TRUE

230992 8

> newcredit = na.omit(credittrain)

> table(is.na(newcredit))

FALSE

230989

> library(e1071)

Warning message:

package ‘e1071’ was built under R version 3.0.3

> model = svm(seriousd ~.,data=newcredit)

> summary(model)

Call:

svm(formula = seriousd ~ ., data = newcredit)

Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

cost: 1

gamma: 0.1

Number of Support Vectors: 3810

( 2410 1400 )

Number of Classes: 2

Levels:

0 1

> pred <- predict(model, newcredit)

> summary(pred)

0 1

20959 40

> table(pred,newcredit$seriousd)

pred 0 1

0 19591 1368

1 4 36

> credittest <- read.csv("C:/Users/anup/Desktop/credittest.csv")

> View(credittest)

> str(credittest)

'data.frame': 9000 obs. of 11 variables:

$ seriousd : int 0 0 0 0 0 0 0 0 0 0 ...

$ revolvingul : num 0.866 0.145 0.422 0.301 0.835 ...

$ age : int 58 29 33 38 54 41 56 67 22 66 ...

$ no\_30\_59 : int 0 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num 0.475 0.235 0.13 0.472 4642 ...

$ monthlyincome : Factor w/ 2900 levels "?","0","1","100",..: 1546 2549 1381 50 1 1899 1358 1 5 2602 ...

$ no\_opencreditll : int 9 4 10 13 15 7 12 4 1 16 ...

$ no\_nightydayslate: int 0 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int 2 0 0 4 1 0 2 0 0 0 ...

$ no\_60\_89 : int 0 0 0 0 0 0 0 0 0 0 ...

$ dependents : Factor w/ 11 levels "?","0","1","10",..: 2 2 2 5 2 3 2 2 2 3 ...

> credittest$seriousd = as.factor(credittest$seriousd)

> credittest$monthlyincome = as.integer(credittest$monthlyincome)

> credittest$dependents = as.integer(credittest$dependents)

> table(is.na(credittest))

FALSE

99000

> predicttest = predict(model,newdata=credittest,type="response")

> summary(predicttest)

0 1

8988 12

> table(predicttest,credittest$seriousd)

predicttest 0 1

0 8392 596

1 6 6

> table(pred,newcredit$seriousd)

pred 0 1

0 19591 1368

1 4 36

> 19591+36

[1] 19627

> 19591+1368+4+36

[1] 20999

> 19627/20999

[1] 0.9346636

> 8392+6

[1] 8398

> 8392+596+6+6

[1] 9000

> 8398/9000

[1] 0.9331111

**Computed Naïve Bayes Model on R**

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : int 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : int 1 3 6 3 3 8 3 4 3 3 ...

> credittrain$seriousd = as.factor(credittrain$seriousd)

> credittrain$monthlyincome = as.integer(credittrain$monthlyincome)

> credittrain$dependents = as.integer(credittrain$dependents)

> str(credittrain)

'data.frame': 21000 obs. of 11 variables:

$ seriousd : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

$ revolvingul : num NA 0.0709 0.0575 0.7065 0.1352 ...

$ age : int NA 68 46 45 63 62 60 53 40 30 ...

$ no\_30\_59 : int NA 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num NA 1.44e-02 3.20e-01 9.79e-01 5.49e+03 ...

$ monthlyincome : int 1 4583 5030 1687 2 4541 247 2 2 786 ...

$ no\_opencreditll : int NA 11 11 15 8 21 12 4 2 10 ...

$ no\_nightydayslate: int NA 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int NA 1 2 1 3 1 1 0 0 0 ...

$ no\_60\_89 : int NA 0 0 0 0 0 0 0 0 0 ...

$ dependents : int 1 3 6 3 3 8 3 4 3 3 ...

> table(is.na(credittrain))

FALSE TRUE

230992 8

> newcredit = na.omit(credittrain)

> table(is.na(newcredit))

FALSE

230989

> library(e1071)

> model = naiveBayes(seriousd ~ ., data = newcredit)

> pred <- predict(model, newcredit)

> table(pred, newcredit$seriousd)

pred 0 1

0 19162 1288

1 433 116

> credittest <- read.csv("C:/Users/anup/Desktop/credittest.csv")

> View(credittest)

> str(credittest)

'data.frame': 9000 obs. of 11 variables:

$ seriousd : int 0 0 0 0 0 0 0 0 0 0 ...

$ revolvingul : num 0.866 0.145 0.422 0.301 0.835 ...

$ age : int 58 29 33 38 54 41 56 67 22 66 ...

$ no\_30\_59 : int 0 0 0 0 0 0 0 0 0 0 ...

$ debtratio : num 0.475 0.235 0.13 0.472 4642 ...

$ monthlyincome : Factor w/ 2900 levels "?","0","1","100",..: 1546 2549 1381 50 1 1899 1358 1 5 2602 ...

$ no\_opencreditll : int 9 4 10 13 15 7 12 4 1 16 ...

$ no\_nightydayslate: int 0 0 0 0 0 0 0 0 0 0 ...

$ no\_realestatell : int 2 0 0 4 1 0 2 0 0 0 ...

$ no\_60\_89 : int 0 0 0 0 0 0 0 0 0 0 ...

$ dependents : Factor w/ 11 levels "?","0","1","10",..: 2 2 2 5 2 3 2 2 2 3 ...

> credittest$seriousd = as.factor(credittest$seriousd)

> credittest$monthlyincome = as.integer(credittest$monthlyincome)

> credittest$dependents = as.integer(credittest$dependents)

> table(is.na(credittest))

FALSE

99000

> predicttest = predict(model,newdata=credittest,type="class")

> summary(predicttest)

0 1

8810 190

> table(predicttest,credittest$seriousd)

predicttest 0 1

0 8250 560

1 148 42

> table(pred, newcredit$seriousd)

pred 0 1

0 19162 1288

1 433 116

> 19162+116

[1] 19278

> 19162+1288+433+116

[1] 20999

> 19278/20999

[1] 0.9180437

> 8250+42

[1] 8292

> 8250+560+148+42

[1] 9000

> 8292/9000

[1] 0.9213333