# Naïve Bayes on Colic Dataset

horse<-read.csv("../input/horse.csv")

str(horse)

'data.frame': 299 obs. of 28 variables:

$ surgery : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 2 2 1 1 ...

$ age : Factor w/ 2 levels "adult","young": 1 1 1 2 1 1 1 1 1 2 ...

$ hospital\_number : int 530101 534817 530334 5290409 530255 528355 526802 529607 530051 5299629 ...

$ rectal\_temp : num 38.5 39.2 38.3 39.1 37.3 NA 37.9 NA NA 38.3 ...

$ pulse : int 66 88 40 164 104 NA 48 60 80 90 ...

$ respiratory\_rate : int 28 20 24 84 35 NA 16 NA 36 NA ...

$ temp\_of\_extremities : Factor w/ 4 levels "cold","cool",..: 2 NA 3 1 NA 4 3 2 2 3 ...

$ peripheral\_pulse : Factor w/ 4 levels "absent","increased",..: 4 NA 3 3 NA 3 3 NA 1 NA ...

$ mucous\_membrane : Factor w/ 6 levels "bright\_pink",..: NA 5 6 3 3 6 4 NA 6 4 ...

$ capillary\_refill\_time: Factor w/ 3 levels "3","less\_3\_sec",..: 3 2 2 3 3 2 2 2 2 2 ...

$ pain : Factor w/ 5 levels "alert","depressed",..: 3 4 4 2 NA 2 4 NA 5 3 ...

$ peristalsis : Factor w/ 4 levels "absent","hypermotile",..: 1 1 3 1 NA 3 3 1 1 3 ...

$ abdominal\_distention : Factor w/ 4 levels "moderate","none",..: 3 4 2 3 NA 4 1 4 3 2 ...

$ nasogastric\_tube : Factor w/ 3 levels "none","significant",..: NA NA NA 1 NA 3 1 3 3 3 ...

$ nasogastric\_reflux : Factor w/ 3 levels "less\_1\_liter",..: NA NA NA 1 NA 3 3 3 3 3 ...

$ nasogastric\_reflux\_ph: num NA NA NA 5 NA NA NA NA NA NA ...

$ rectal\_exam\_feces : Factor w/ 4 levels "absent","decreased",..: 2 1 4 2 NA 2 2 2 2 2 ...

$ abdomen : Factor w/ 5 levels "distend\_large",..: 1 5 4 NA NA 3 1 2 1 NA ...

$ packed\_cell\_volume : num 45 50 33 48 74 NA 37 44 38 40 ...

$ total\_protein : num 8.4 85 6.7 7.2 7.4 NA 7 8.3 6.2 6.2 ...

$ abdomo\_appearance : Factor w/ 3 levels "clear","cloudy",..: NA 2 NA 3 NA NA NA NA NA 1 ...

$ abdomo\_protein : num NA 2 NA 5.3 NA NA NA NA NA 2.2 ...

$ outcome : Factor w/ 3 levels "died","euthanized",..: 1 2 3 1 1 3 3 1 2 3 ...

$ surgical\_lesion : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 2 2 2 1 ...

$ lesion\_1 : int 11300 2208 0 2208 4300 0 3124 2208 3205 0 ...

$ lesion\_2 : int 0 0 0 0 0 0 0 0 0 0 ...

$ lesion\_3 : int 0 0 0 0 0 0 0 0 0 0 ...

$ cp\_data : Factor w/ 2 levels "no","yes": 1 1 2 2 1 1 1 1 1 2 ...

library(caret)

train\_ind <- createDataPartition(horse$outcome, p= .8, list=FALSE)

training\_set<-horse[train\_ind,]

testing\_set<-horse[-train\_ind,]

Loading required package: lattice

Loading required package: ggplot2

library(e1071)

model<-naiveBayes(outcome~., data = training\_set, laplace = 1)

prop.table(table(horse$outcome))

died euthanized lived

0.2575251 0.1471572 0.5953177

model$tables

$surgery

surgery

Y no yes

died 0.2031250 0.7968750

euthanized 0.4210526 0.5789474

lived 0.4620690 0.5379310

$age

age

Y adult young

died 0.82812500 0.17187500

euthanized 0.94736842 0.05263158

lived 0.93793103 0.06206897

$hospital\_number

hospital\_number

Y [,1] [,2]

died 1603525.3 2006451

euthanized 796466.5 1103547

lived 896958.4 1271879

$rectal\_temp

rectal\_temp

Y [,1] [,2]

died 38.11957 0.8820732

euthanized 38.02500 1.0178499

lived 38.17642 0.5669998

$pulse

pulse

Y [,1] [,2]

died 86.24528 33.98580

euthanized 84.60000 25.84820

lived 62.98485 24.41248

$respiratory\_rate

respiratory\_rate

Y [,1] [,2]

died 33.79592 17.28677

euthanized 32.36667 18.02390

lived 29.40496 18.54031

$temp\_of\_extremities

temp\_of\_extremities

Y cold cool normal warm

died 0.16363636 0.56363636 0.20000000 0.07272727

euthanized 0.25000000 0.59375000 0.12500000 0.03125000

lived 0.08196721 0.31967213 0.40983607 0.18852459

$peripheral\_pulse

peripheral\_pulse

Y absent increased normal reduced

died 0.07692308 0.01923077 0.26923077 0.63461538

euthanized 0.13793103 0.10344828 0.17241379 0.58620690

lived 0.01694915 0.03389831 0.65254237 0.29661017

$mucous\_membrane

mucous\_membrane

Y bright\_pink bright\_red dark\_cyanotic normal\_pink pale\_cyanotic

died 0.13559322 0.20338983 0.15254237 0.06779661 0.18644068

euthanized 0.08108108 0.21621622 0.13513514 0.08108108 0.32432432

lived 0.13385827 0.03937008 0.03937008 0.45669291 0.10236220

mucous\_membrane

Y pale\_pink

died 0.25423729

euthanized 0.16216216

lived 0.22834646

$capillary\_refill\_time

capillary\_refill\_time

Y 3 less\_3\_sec more\_3\_sec

died 0.03508772 0.54385965 0.42105263

euthanized 0.02631579 0.39473684 0.57894737

lived 0.01515152 0.86363636 0.12121212

$pain

pain

Y alert depressed extreme\_pain mild\_pain severe\_pain

died 0.05263158 0.22807018 0.33333333 0.12280702 0.26315789

euthanized 0.05882353 0.32352941 0.14705882 0.17647059 0.29411765

lived 0.24590164 0.21311475 0.08196721 0.36885246 0.09016393

$peristalsis

peristalsis

Y absent hypermotile hypomotile normal

died 0.43396226 0.07547170 0.47169811 0.01886792

euthanized 0.47058824 0.05882353 0.38235294 0.08823529

lived 0.17557252 0.22137405 0.51145038 0.09160305

$abdominal\_distention

abdominal\_distention

Y moderate none severe slight

died 0.3818182 0.1272727 0.2545455 0.2363636

euthanized 0.3870968 0.1935484 0.2903226 0.1290323

lived 0.1983471 0.4297521 0.0661157 0.3057851

$nasogastric\_tube

nasogastric\_tube

Y none significant slight

died 0.3863636 0.1363636 0.4772727

euthanized 0.3200000 0.1600000 0.5200000

lived 0.3404255 0.1063830 0.5531915

$nasogastric\_reflux

nasogastric\_reflux

Y less\_1\_liter more\_1\_liter none

died 0.2857143 0.2857143 0.4285714

euthanized 0.2000000 0.3333333 0.4666667

lived 0.1630435 0.1195652 0.7173913

$nasogastric\_reflux\_ph

nasogastric\_reflux\_ph

Y [,1] [,2]

died 5.653333 1.376780

euthanized 3.285714 1.997022

lived 4.927778 1.683066

$rectal\_exam\_feces

rectal\_exam\_feces

Y absent decreased increased normal

died 0.56818182 0.18181818 0.11363636 0.13636364

euthanized 0.59090909 0.09090909 0.09090909 0.22727273

lived 0.29702970 0.27722772 0.06930693 0.35643564

$abdomen

abdomen

Y distend\_large distend\_small firm normal other

died 0.4750000 0.3500000 0.0250000 0.1250000 0.0250000

euthanized 0.5312500 0.2187500 0.0937500 0.0625000 0.0937500

lived 0.2873563 0.2068966 0.1264368 0.2068966 0.1724138

$packed\_cell\_volume

packed\_cell\_volume

Y [,1] [,2]

died 52.12727 11.452081

euthanized 53.20690 10.768271

lived 42.59470 7.774771

$total\_protein

total\_protein

Y [,1] [,2]

died 6.805882 1.244413

euthanized 40.900000 31.605549

lived 27.081955 28.203975

$abdomo\_appearance

abdomo\_appearance

Y clear cloudy serosanguious

died 0.2258065 0.2580645 0.5161290

euthanized 0.1363636 0.4545455 0.4090909

lived 0.4262295 0.3606557 0.2131148

$abdomo\_protein

abdomo\_protein

Y [,1] [,2]

died 4.110526 2.347077

euthanized 2.435000 1.479785

lived 2.965000 2.041938

$surgical\_lesion

surgical\_lesion

Y no yes

died 0.1093750 0.8906250

euthanized 0.3684211 0.6315789

lived 0.4965517 0.5034483

$lesion\_1

lesion\_1

Y [,1] [,2]

died 4484.226 5399.777

euthanized 2767.861 1264.384

lived 3342.406 5897.819

$lesion\_2

lesion\_2

Y [,1] [,2]

died 0.00000 0.0000

euthanized 38.88889 233.3333

lived 179.49650 927.0393

$lesion\_3

lesion\_3

Y [,1] [,2]

died 0.00000 0.0000

euthanized 0.00000 0.0000

lived 15.44755 184.7259

$cp\_data

cp\_data

Y no yes

died 0.5312500 0.4687500

euthanized 0.7631579 0.2368421

lived 0.6896552 0.3103448

library(ROCR)

predicted<-predict(model, testing\_set)

confusionMatrix(data=predicted, reference = testing\_set$outcome)

Loading required package: gplots

Attaching package: ‘gplots’

The following object is masked from ‘package:stats’:

lowess

Confusion Matrix and Statistics

Reference

Prediction died euthanized lived

died 15 7 34

euthanized 0 1 1

lived 0 0 0

Overall Statistics

Accuracy : 0.2759

95% CI : (0.1666, 0.409)

No Information Rate : 0.6034

P-Value [Acc > NIR] : 1

Kappa : 0.0287

Mcnemar's Test P-Value : 4.012e-09

Statistics by Class:

Class: died Class: euthanized Class: lived

Sensitivity 1.00000 0.12500 0.0000

Specificity 0.04651 0.98000 1.0000

Pos Pred Value 0.26786 0.50000 NaN

Neg Pred Value 1.00000 0.87500 0.3966

Prevalence 0.25862 0.13793 0.6034

Detection Rate 0.25862 0.01724 0.0000

Detection Prevalence 0.96552 0.03448 0.0000

Balanced Accuracy 0.52326 0.55250 0.5000