Ex8 - SVM vs Logistic Regression

Register Number:19BLC1186

Name:Tarun Sidhu

Lab Exercise No:9

Date:12/4/2021

Dataset:Cleaned.Data

Task:Compare the performance of the best SVM model from Ex7 with the performance of a Logistic Regression model for the same task. Present the performance metrics and ROC curves.

```
#Logistic Regression for Cleaned.Data
library(tidyverse)
```

```
## ── Attaching packages ──────────────────────────────── tidyverse 1.3.0 ─
─
```

```
## ✓ ggplot2 3.3.3     ✓ purrr   0.3.4
## ✓ tibble  3.1.0     ✓ dplyr   1.0.5
## ✓ tidyr   1.1.3     ✓ stringr 1.4.0
## ✓ readr   1.4.0     ✓ forcats 0.5.1
```

```
## ── Conflicts ──────────────────────────────── tidyverse_conflicts() ─
─
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
Cleaned.Data=read.csv("/Users/tarunsidhu/Desktop/Sem 4/ML/ML(Lab)/Data Sets/Cleane
d-Data.csv")
Cleaned.Data <- na.omit(Cleaned.Data)
sample_n(Cleaned.Data, 3)
```

```
##   Fever Tiredness Dry.Cough Difficulty.in.Breathing Sore.Throat None_Sympton
## 1     1         1         1                       1           0            0
## 2     1         0         0                       0           0            0
## 3     1         1         1                       1           1            0
##   Pains Nasal.Congestion Runny.Nose Diarrhea None_Experiencing Age_0.9
## 1     0                1          1        0                 0       0
## 2     0                1          0        0                 0       0
## 3     1                0          0        0                 0       0
##   Age_10.19 Age_20.24 Age_25.59 Age_60. Gender_Female Gender_Male
## 1         0         0         0       1             0           0
## 2         0         0         1       0             1           0
## 3         0         1         0       0             0           0
##   Gender_Transgender Severity_Mild Severity_Moderate Severity_None
## 1                  1             1                 0             0
## 2                  0             0                 0             0
## 3                  1             0                 0             0
##   Severity_Severe Contact_Dont.Know Contact_No Contact_Yes Country
## 1               0                 0          1           0 Germany
## 2               1                 1          0           0     UAE
## 3               1                 1          0           0    Iran
```

```
set.seed(123)
training.samples <- Cleaned.Data$Severity_Mild %>%
  createDataPartition(p = 0.8, list = FALSE)
train.data  <- Cleaned.Data[training.samples, ]
test.data <- Cleaned.Data[-training.samples, ]
model <- glm( Severity_Mild ~., data = train.data, family = binomial)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
summary(model)
```

```
##
## Call:
## glm(formula = Severity_Mild ~ ., family = binomial, data = train.data)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -2.409e-06  -2.409e-06  -2.409e-06  -1.204e-06   2.409e-06
##
## Coefficients: (3 not defined because of singularities)
```

```
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              2.657e+01  3.831e+03   0.007    0.994
## Fever                   -2.835e-11  1.712e+03   0.000    1.000
## Tiredness               -9.100e-12  1.713e+03   0.000    1.000
## Dry.Cough               -1.165e-11  1.714e+03   0.000    1.000
## Difficulty.in.Breathing -8.587e-12  1.714e+03   0.000    1.000
## Sore.Throat             -3.166e-11  1.714e+03   0.000    1.000
## None_Sympton            -3.347e-11  3.225e+03   0.000    1.000
## Pains                   -2.857e-11  1.626e+03   0.000    1.000
## Nasal.Congestion        -9.934e-12  1.626e+03   0.000    1.000
## Runny.Nose              -8.456e-12  1.626e+03   0.000    1.000
## Diarrhea                -2.654e-11  1.626e+03   0.000    1.000
## None_Experiencing       -2.350e-11  2.872e+03   0.000    1.000
## Age_0.9                 -4.113e-11  2.238e+03   0.000    1.000
## Age_10.19               -1.772e-12  2.237e+03   0.000    1.000
## Age_20.24               -8.467e-13  2.239e+03   0.000    1.000
## Age_25.59               -5.599e-12  2.237e+03   0.000    1.000
## Age_60.                         NA         NA      NA       NA
## Gender_Female            3.834e-13  1.732e+03   0.000    1.000
## Gender_Male             -2.137e-11  1.734e+03   0.000    1.000
## Gender_Transgender              NA         NA      NA       NA
## Severity_Moderate       -5.313e+01  2.001e+03  -0.027    0.979
## Severity_None           -5.313e+01  2.001e+03  -0.027    0.979
## Severity_Severe         -5.313e+01  2.000e+03  -0.027    0.979
## Contact_Dont.Know        2.310e-11  1.733e+03   0.000    1.000
## Contact_No               2.310e-11  1.733e+03   0.000    1.000
## Contact_Yes                     NA         NA      NA       NA
## CountryFrance            7.232e-11  3.165e+03   0.000    1.000
## CountryGermany           7.161e-11  3.163e+03   0.000    1.000
## CountryIran              7.133e-11  3.166e+03   0.000    1.000
## CountryItaly             7.202e-11  3.156e+03   0.000    1.000
## CountryOther             7.180e-11  3.160e+03   0.000    1.000
## CountryOther-EUR         7.138e-11  3.165e+03   0.000    1.000
## CountryRepublic of Korean 7.096e-11  3.163e+03   0.000    1.000
## CountrySpain             7.149e-11  3.160e+03   0.000    1.000
## CountryUAE               7.131e-11  3.164e+03   0.000    1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.8504e+05  on 253439  degrees of freedom
## Residual deviance: 1.4704e-06  on 253408  degrees of freedom
## AIC: 64
##
## Number of Fisher Scoring iterations: 25
```

```
probabilities <- model %>% predict(test.data, type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from a rank-deficient fit may be misleading
```

```
predicted.classes <- ifelse(probabilities > 0.5, "pos", "neg")
mean(predicted.classes == test.data$Severity_Mild)
```

```
## [1] 0
```

```
model <- glm( Severity_Mild ~ Fever, data = train.data, family = binomial)
summary(model)$coef
```

```
##                   Estimate   Std. Error      z value   Pr(>|z|)
## (Intercept) -1.100228807 0.005536898 -198.7085344 0.0000000
## Fever        0.005159614 0.009887642    0.5218245 0.6017925
```

```
newdata <- data.frame(Fever = c(20,  180))
probabilities <- model %>% predict(newdata, type = "response")
predicted.classes <- ifelse(probabilities > 0.5, "pos", "neg")
predicted.classes
```
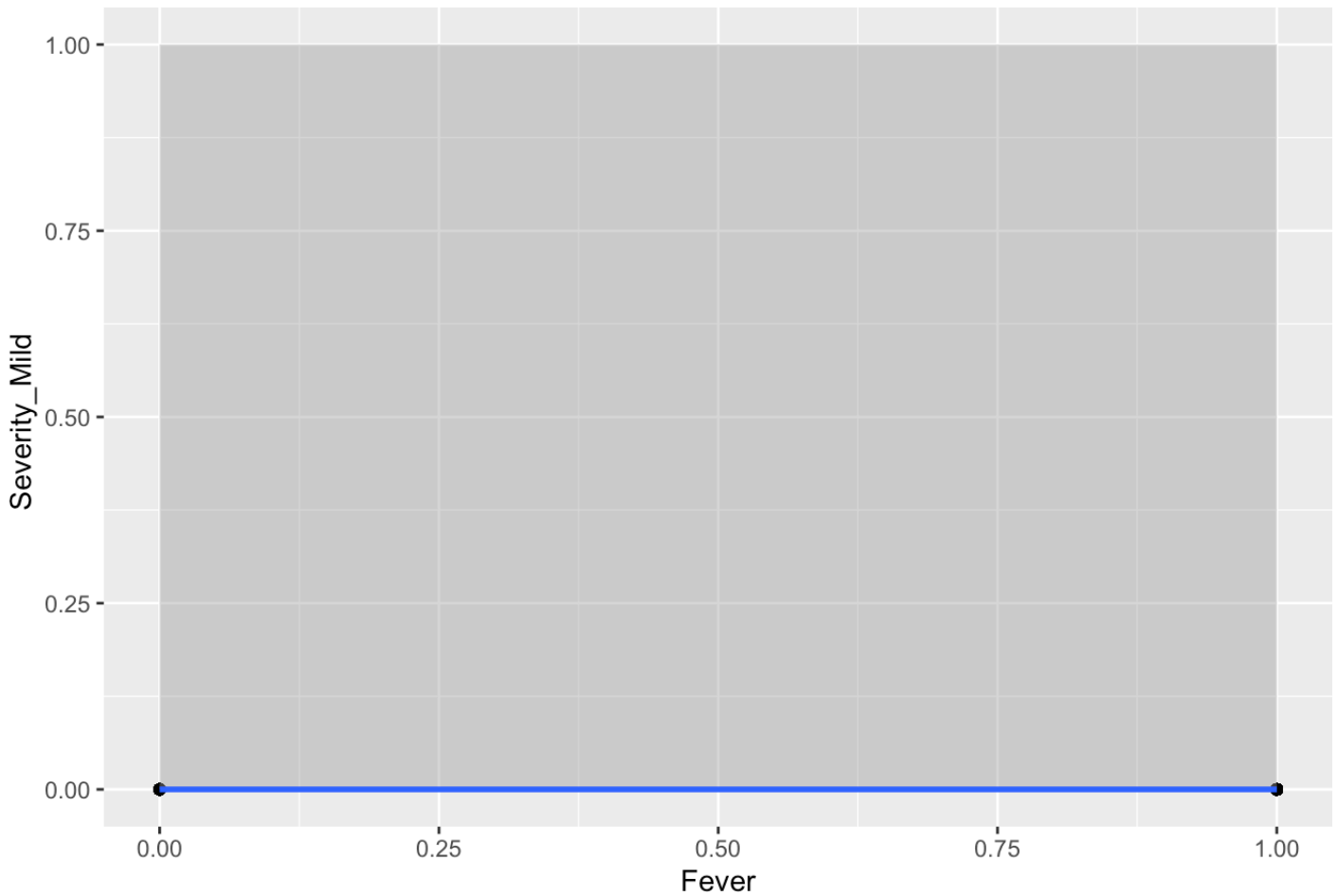
```
##     1     2
## "neg" "neg"
```

```
train.data %>%
  mutate(prob = ifelse(Severity_Mild == "pos", 1, 0)) %>%
  ggplot(aes(Fever, prob)) +
  geom_point(alpha = 0.2) +
  geom_smooth(method = "glm", method.args = list(family = "binomial")) +
  labs(
    title = "Logistic Regression Model",
    x = "Fever",
    y = "Severity_Mild "
  )
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: glm.fit: algorithm did not converge
```

Logistic Regression Model


```
model <- glm( Severity_Mild ~ Fever+Sore.Throat,
             data = train.data, family = binomial)
summary(model)$coef
```

```
##                 Estimate  Std. Error      z value  Pr(>|z|)
## (Intercept) -1.098917084 0.006628773 -165.7798718 0.0000000
## Fever        0.004565438 0.010024571    0.4554248 0.6488037
## Sore.Throat -0.003613004 0.010044562   -0.3596975 0.7190733
```

```
model <- glm( Severity_Mild ~., data = train.data, family = binomial)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
summary(model)$coef
```

```
##                          Estimate  Std. Error         z value  Pr(>|z|)
## (Intercept)          2.656606e+01    3831.079   6.934355e-03  0.9944672
## Fever               -2.834899e-11    1712.241  -1.655666e-14  1.0000000
## Tiredness           -9.099505e-12    1713.002  -5.312022e-15  1.0000000
## Dry.Cough           -1.165313e-11    1714.381  -6.797282e-15  1.0000000
## Difficulty.in.Breathing  -8.586718e-12  1714.068  -5.009556e-15  1.0000000
## Sore.Throat         -3.165622e-11    1713.610  -1.847341e-14  1.0000000
## None_Sympton        -3.347069e-11    3224.916  -1.037878e-14  1.0000000
## Pains               -2.856694e-11    1626.117  -1.756759e-14  1.0000000
## Nasal.Congestion    -9.933926e-12    1625.700  -6.110552e-15  1.0000000
## Runny.Nose          -8.456039e-12    1625.676  -5.201552e-15  1.0000000
## Diarrhea            -2.653532e-11    1626.142  -1.631796e-14  1.0000000
## None_Experiencing   -2.349686e-11    2871.525  -8.182709e-15  1.0000000
## Age_0.9             -4.112961e-11    2238.427  -1.837434e-14  1.0000000
## Age_10.19           -1.772394e-12    2236.878  -7.923515e-16  1.0000000
## Age_20.24           -8.466565e-13    2238.713  -3.781890e-16  1.0000000
## Age_25.59           -5.598877e-12    2236.504  -2.503406e-15  1.0000000
## Gender_Female        3.834451e-13    1732.163   2.213678e-16  1.0000000
## Gender_Male         -2.137028e-11    1733.744  -1.232609e-14  1.0000000
## Severity_Moderate   -5.313213e+01    2001.004  -2.655274e-02  0.9788165
## Severity_None       -5.313213e+01    2001.113  -2.655129e-02  0.9788176
## Severity_Severe     -5.313213e+01    2000.442  -2.656019e-02  0.9788105
## Contact_Dont.Know    2.310094e-11    1733.461   1.332648e-14  1.0000000
## Contact_No           2.310135e-11    1732.623   1.333317e-14  1.0000000
## CountryFrance        7.231626e-11    3164.549   2.285200e-14  1.0000000
## CountryGermany       7.160902e-11    3163.358   2.263703e-14  1.0000000
## CountryIran          7.133118e-11    3165.502   2.253392e-14  1.0000000
## CountryItaly         7.202028e-11    3156.092   2.281945e-14  1.0000000
## CountryOther         7.179578e-11    3159.935   2.272065e-14  1.0000000
## CountryOther-EUR     7.137971e-11    3164.704   2.255494e-14  1.0000000
## CountryRepublic of Korean  7.095567e-11  3162.669  2.243537e-14  1.0000000
## CountrySpain         7.148842e-11    3159.955   2.262324e-14  1.0000000
## CountryUAE           7.131168e-11    3164.062   2.253802e-14  1.0000000
```

```
coef(model)
```

```
##              (Intercept)                    Fever                 Tiredness
##             2.656606e+01            -2.834899e-11             -9.099505e-12
##                Dry.Cough    Difficulty.in.Breathing               Sore.Throat
##            -1.165313e-11            -8.586718e-12             -3.165622e-11
##             None_Sympton                    Pains           Nasal.Congestion
##            -3.347069e-11            -2.856694e-11             -9.933926e-12
##                Runny.Nose                 Diarrhea           None_Experiencing
##            -8.456039e-12            -2.653532e-11             -2.349686e-11
##                  Age_0.9                 Age_10.19                 Age_20.24
##            -4.112961e-11            -1.772394e-12             -8.466565e-13
##                Age_25.59                   Age_60.             Gender_Female
##            -5.598877e-12                       NA              3.834451e-13
##              Gender_Male        Gender_Transgender         Severity_Moderate
##            -2.137028e-11                       NA             -5.313213e+01
##            Severity_None          Severity_Severe         Contact_Dont.Know
##            -5.313213e+01            -5.313213e+01              2.310094e-11
##               Contact_No              Contact_Yes             CountryFrance
##             2.310135e-11                       NA              7.231626e-11
##            CountryGermany              CountryIran              CountryItaly
##             7.160902e-11             7.133118e-11              7.202028e-11
##             CountryOther         CountryOther-EUR CountryRepublic of Korean
##             7.179578e-11             7.137971e-11              7.095567e-11
##             CountrySpain               CountryUAE
##             7.148842e-11             7.131168e-11
```

```
summary(model )$coef
```

```
##                              Estimate Std. Error       z value Pr(>|z|)
## (Intercept)               2.656606e+01   3831.079  6.934355e-03 0.9944672
## Fever                    -2.834899e-11   1712.241 -1.655666e-14 1.0000000
## Tiredness                -9.099505e-12   1713.002 -5.312022e-15 1.0000000
## Dry.Cough                -1.165313e-11   1714.381 -6.797282e-15 1.0000000
## Difficulty.in.Breathing  -8.586718e-12   1714.068 -5.009556e-15 1.0000000
## Sore.Throat              -3.165622e-11   1713.610 -1.847341e-14 1.0000000
## None_Sympton             -3.347069e-11   3224.916 -1.037878e-14 1.0000000
## Pains                    -2.856694e-11   1626.117 -1.756759e-14 1.0000000
## Nasal.Congestion         -9.933926e-12   1625.700 -6.110552e-15 1.0000000
## Runny.Nose               -8.456039e-12   1625.676 -5.201552e-15 1.0000000
## Diarrhea                 -2.653532e-11   1626.142 -1.631796e-14 1.0000000
## None_Experiencing        -2.349686e-11   2871.525 -8.182709e-15 1.0000000
## Age_0.9                  -4.112961e-11   2238.427 -1.837434e-14 1.0000000
## Age_10.19                -1.772394e-12   2236.878 -7.923515e-16 1.0000000
## Age_20.24                -8.466565e-13   2238.713 -3.781890e-16 1.0000000
## Age_25.59                -5.598877e-12   2236.504 -2.503406e-15 1.0000000
## Gender_Female             3.834451e-13   1732.163  2.213678e-16 1.0000000
## Gender_Male              -2.137028e-11   1733.744 -1.232609e-14 1.0000000
## Severity_Moderate        -5.313213e+01   2001.004 -2.655274e-02 0.9788165
## Severity_None            -5.313213e+01   2001.113 -2.655129e-02 0.9788176
## Severity_Severe          -5.313213e+01   2000.442 -2.656019e-02 0.9788105
## Contact_Dont.Know         2.310094e-11   1733.461  1.332648e-14 1.0000000
## Contact_No                2.310135e-11   1732.623  1.333317e-14 1.0000000
## CountryFrance             7.231626e-11   3164.549  2.285200e-14 1.0000000
## CountryGermany            7.160902e-11   3163.358  2.263703e-14 1.0000000
## CountryIran               7.133118e-11   3165.502  2.253392e-14 1.0000000
## CountryItaly              7.202028e-11   3156.092  2.281945e-14 1.0000000
## CountryOther              7.179578e-11   3159.935  2.272065e-14 1.0000000
## CountryOther-EUR          7.137971e-11   3164.704  2.255494e-14 1.0000000
## CountryRepublic of Korean 7.095567e-11   3162.669  2.243537e-14 1.0000000
## CountrySpain              7.148842e-11   3159.955  2.262324e-14 1.0000000
## CountryUAE                7.131168e-11   3164.062  2.253802e-14 1.0000000
```

```
#SVM for Cleaned.Data
Cleaned.Data=read.csv("/Users/tarunsidhu/Desktop/Sem 4/ML/ML(Lab)/Data Sets/Cleane
d-Data.csv")
summary(Cleaned.Data)
```
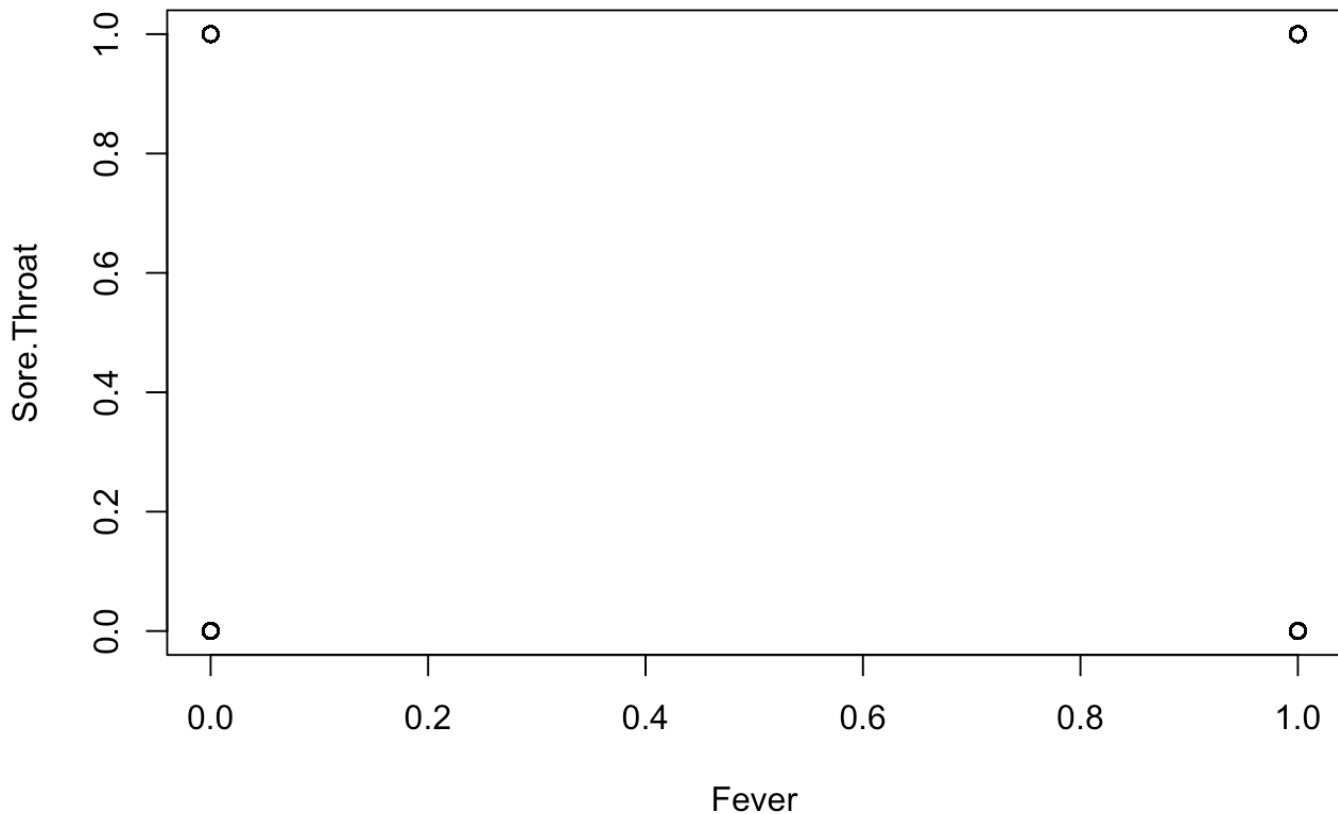
```
##       Fever            Tiredness         Dry.Cough         Difficulty.in.Breathing
##   Min.    :0.0000    Min.    :0.0     Min.    :0.0000    Min.    :0.0
##   1st Qu.:0.0000    1st Qu.:0.0     1st Qu.:0.0000    1st Qu.:0.0
##   Median :0.0000    Median :0.5     Median :1.0000    Median :0.5
##   Mean    :0.3125    Mean    :0.5     Mean    :0.5625    Mean    :0.5
##   3rd Qu.:1.0000    3rd Qu.:1.0     3rd Qu.:1.0000    3rd Qu.:1.0
##   Max.    :1.0000    Max.    :1.0     Max.    :1.0000    Max.    :1.0
##    Sore.Throat        None_Sympton        Pains          Nasal.Congestion
##   Min.    :0.0000    Min.    :0.0000    Min.    :0.0000    Min.    :0.0000
##   1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000
##   Median :0.0000    Median :0.0000    Median :0.0000    Median :1.0000
##   Mean    :0.3125    Mean    :0.0625    Mean    :0.3636    Mean    :0.5455
##   3rd Qu.:1.0000    3rd Qu.:0.0000    3rd Qu.:1.0000    3rd Qu.:1.0000
##   Max.    :1.0000    Max.    :1.0000    Max.    :1.0000    Max.    :1.0000
##     Runny.Nose          Diarrhea        None_Experiencing   Age_0.9
##   Min.    :0.0000    Min.    :0.0000    Min.    :0.00000    Min.    :0.0
##   1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:0.0
##   Median :1.0000    Median :0.0000    Median :0.00000    Median :0.0
##   Mean    :0.5455    Mean    :0.3636    Mean    :0.09091    Mean    :0.2
##   3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:0.00000    3rd Qu.:0.0
##   Max.    :1.0000    Max.    :1.0000    Max.    :1.00000    Max.    :1.0
##     Age_10.19        Age_20.24        Age_25.59        Age_60.        Gender_Female
##   Min.    :0.0     Min.    :0.0     Min.    :0.0     Min.    :0.0     Min.    :0.0000
##   1st Qu.:0.0     1st Qu.:0.0     1st Qu.:0.0     1st Qu.:0.0     1st Qu.:0.0000
##   Median :0.0     Median :0.0     Median :0.0     Median :0.0     Median :0.0000
##   Mean    :0.2     Mean    :0.2     Mean    :0.2     Mean    :0.2     Mean    :0.3333
##   3rd Qu.:0.0     3rd Qu.:0.0     3rd Qu.:0.0     3rd Qu.:0.0     3rd Qu.:1.0000
##   Max.    :1.0     Max.    :1.0     Max.    :1.0     Max.    :1.0     Max.    :1.0000
##    Gender_Male       Gender_Transgender  Severity_Mild   Severity_Moderate
##   Min.    :0.0000    Min.    :0.0000    Min.    :0.00    Min.    :0.00
##   1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.00    1st Qu.:0.00
##   Median :0.0000    Median :0.0000    Median :0.00    Median :0.00
##   Mean    :0.3333    Mean    :0.3333    Mean    :0.25    Mean    :0.25
##   3rd Qu.:1.0000    3rd Qu.:1.0000    3rd Qu.:0.25    3rd Qu.:0.25
##   Max.    :1.0000    Max.    :1.0000    Max.    :1.00    Max.    :1.00
##   Severity_None    Severity_Severe  Contact_Dont.Know   Contact_No
##   Min.    :0.00    Min.    :0.00    Min.    :0.0000    Min.    :0.0000
##   1st Qu.:0.00    1st Qu.:0.00    1st Qu.:0.0000    1st Qu.:0.0000
##   Median :0.00    Median :0.00    Median :0.0000    Median :0.0000
##   Mean    :0.25    Mean    :0.25    Mean    :0.3333    Mean    :0.3333
##   3rd Qu.:0.25    3rd Qu.:0.25    3rd Qu.:1.0000    3rd Qu.:1.0000
##   Max.    :1.00    Max.    :1.00    Max.    :1.0000    Max.    :1.0000
##    Contact_Yes         Country
##   Min.    :0.0000    Length:316800
##   1st Qu.:0.0000    Class :character
##   Median :0.0000    Mode  :character
##   Mean    :0.3333
##   3rd Qu.:1.0000
##   Max.    :1.0000
```

```
Cleaned_Data_1 <- Cleaned.Data[1:10000,]
attach(Cleaned_Data_1)
plot(Fever,Sore.Throat)
```



```
library(e1071)
svm.fit <- svm(Severity_Mild ~ Fever+Sore.Throat, data = Cleaned_Data_1, type='C-c
lassification', kernel='linear', cost=10, scale=FALSE)
plot(svm.fit, Cleaned_Data_1)
```

```
## Error in plot.svm(svm.fit, Cleaned_Data_1): missing formula.
```

```
summary(svm.fit)
```

```
##
## Call:
## svm(formula = Severity_Mild ~ Fever + Sore.Throat, data = Cleaned_Data_1,
##     type = "C-classification", kernel = "linear", cost = 10, scale = FALSE)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  10
##
## Number of Support Vectors:  5004
##
##  ( 2502 2502 )
##
##
## Number of Classes:  2
##
## Levels:
##  0 1
```

```
yhat <- predict(svm.fit, Cleaned_Data_1)
table(predict=yhat, truth=Severity_Mild)
```

```
##        truth
## predict    0    1
##       0 7498 2502
##       1    0    0
```

```
library(caret)
confusionMatrix(yhat, as.factor(Cleaned_Data_1$Severity_Mild))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 7498 2502
##          1    0    0
##
##                Accuracy : 0.7498
##                  95% CI : (0.7412, 0.7583)
##     No Information Rate : 0.7498
##     P-Value [Acc > NIR] : 0.5054
##
##                   Kappa : 0
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 1.0000
##             Specificity : 0.0000
##          Pos Pred Value : 0.7498
##          Neg Pred Value :    NaN
##              Prevalence : 0.7498
##          Detection Rate : 0.7498
##    Detection Prevalence : 1.0000
##       Balanced Accuracy : 0.5000
##
##        'Positive' Class : 0
##
```