**Tanmay Gupta**

**Project R-code**

**Covid-19 Data Analysis**

**#loading the required packages**

**library(tidyverse)**

**library(plyr)**

**library(tidyr)**

**library(rpart)**

**library(DMwR)**

**library(corrplot)**

**library(leaps)**

**library(UBL)**

**library(e1071)**

**library(class)**

**library(caret)**

**#Loading the dataset**

**covid\_USA <- read.csv('national-history.csv')**

**#having a look at the first few rows of data**

**head(covid\_USA)**

**Date Total\_death Today\_death In\_ICU\_cummalative Total\_in\_ICU Today\_hospitalized**

**1 12/6/2020 273374 1138 31946 20145 2256**

**2 12/5/2020 272236 2445 31831 19950 3316**

**3 12/4/2020 269791 2563 31608 19858 4652**

**4 12/3/2020 267228 2706 31276 19723 5331**

**5 12/2/2020 264522 2733 31038 19680 5028**

**6 12/1/2020 261789 2473 30749 19295 5222**

**Total\_hospitalized Hospitalized\_cummalative Total\_negative Today\_negative Ventilator\_cummalative**

**1 101487 585676 161986294 1172590 3322**

**2 101190 583420 160813704 1526995 3321**

**3 101276 580104 159286709 1260657 3305**

**4 100755 575452 158026052 1238465 3280**

**5 100322 570121 156787587 982032 3252**

**6 98777 565093 155805555 1941714 3223**

**Total\_on\_ventilator Total\_positive\_cases Today\_positive\_cases Total\_people\_recovered states**

**1 7094 14534035 176771 5624444 56**

**2 7005 14357264 211073 5576026 56**

**3 6999 14146191 224831 5470389 56**

**4 6867 13921360 210204 5404018 56**

**5 6855 13711156 195796 5322128 56**

**6 6649 13515360 176753 5226581 56**

**Total\_tests\_results Total\_tests\_results\_today**

**1 204063869 1634532**

**2 202429337 2169756**

**3 200259581 1854869**

**4 198404712 1828230**

**5 196576482 1459202**

**6 195117280 2340996**

**#getting variable names**

**colnames(covid\_USA)**

**[1] "Date" "Total\_death" "Today\_death" "In\_ICU\_cummalative"**

**[5] "Total\_in\_ICU" "Today\_hospitalized" "Total\_hospitalized" "Hospitalized\_cummalative"**

**[9] "Total\_negative" "Today\_negative" "Ventilator\_cummalative" "Total\_on\_ventilator"**

**[13] "Total\_positive\_cases" "Today\_positive\_cases" "Total\_people\_recovered" "states"**

**[17] "Total\_tests\_results" "Total\_tests\_results\_today"**

**#Renaming the Variables for easy understanding**

**names(covid\_USA[1])**

**colnames(covid\_USA)[1]<- 'Date'**

**names(covid\_USA[2])**

**colnames(covid\_USA)[2]<- 'Total\_death'**

**names(covid\_USA[3])**

**colnames(covid\_USA)[3]<- 'Today\_death'**

**names(covid\_USA[4])**

**colnames(covid\_USA)[4]<- 'In\_ICU\_cummalative'**

**names(covid\_USA[5])**

**colnames(covid\_USA)[5]<- 'Total\_in\_ICU'**

**names(covid\_USA[6])**

**colnames(covid\_USA)[6]<- 'Today\_hospitalized'**

**names(covid\_USA[7])**

**colnames(covid\_USA)[7]<- 'Total\_hospitalized'**

**names(covid\_USA[8])**

**colnames(covid\_USA)[8]<- 'Hospitalized\_cummalative'**

**names(covid\_USA[9])**

**colnames(covid\_USA)[9]<- 'Total\_negative'**

**names(covid\_USA[10])**

**colnames(covid\_USA)[10]<- 'Today\_negative'**

**names(covid\_USA[11])**

**colnames(covid\_USA)[11]<- 'Ventilator\_cummalative'**

**names(covid\_USA[12])**

**colnames(covid\_USA)[12]<- 'Total\_on\_ventilator'**

**names(covid\_USA[13])**

**colnames(covid\_USA)[13]<- 'Total\_positive\_cases'**

**names(covid\_USA[14])**

**colnames(covid\_USA)[14]<- 'Today\_positive\_cases'**

**names(covid\_USA[15])**

**colnames(covid\_USA)[15]<- 'Total\_people\_recovered'**

**names(covid\_USA[17])**

**colnames(covid\_USA)[17]<- 'Total\_tests\_results'**

**names(covid\_USA[18])**

**colnames(covid\_USA)[18]<- 'Total\_tests\_results\_today'**

**#taking a look at the renamed data**

**glimpse(covid\_USA)**

**Rows: 320**

**Columns: 18**

**$ Date <fct> 12/6/2020, 12/5/2020, 12/4/2020, 12/3/2020, 12/2/2020, 12/1/2020, 11/30/2020, 1...**

**$ Total\_death <dbl> 273374, 272236, 269791, 267228, 264522, 261789, 259316, 258180, 257377, 256132,...**

**$ Today\_death <int> 1138, 2445, 2563, 2706, 2733, 2473, 1136, 803, 1245, 1372, 1336, 2289, 2066, 95...**

**$ In\_ICU\_cummalative <dbl> 31946, 31831, 31608, 31276, 31038, 30749, 30469, 30274, 30109, 29858, 29673, 29...**

**$ Total\_in\_ICU <dbl> 20145, 19950, 19858, 19723, 19680, 19295, 18801, 18437, 18249, 18061, 18019, 17...**

**$ Today\_hospitalized <int> 2256, 3316, 4652, 5331, 5028, 5222, 3394, 2429, 3404, 3499, 2247, 4568, 4591, 2...**

**$ Total\_hospitalized <dbl> 101487, 101190, 101276, 100755, 100322, 98777, 96149, 93357, 91762, 89950, 9056...**

**$ Hospitalized\_cummalative <dbl> 585676, 583420, 580104, 575452, 570121, 565093, 559871, 556477, 554048, 550644,...**

**$ Total\_negative <int> 161986294, 160813704, 159286709, 158026052, 156787587, 155805555, 153863841, 15...**

**$ Today\_negative <int> 1172590, 1526995, 1260657, 1238465, 982032, 1941714, 1219808, 883148, 1276935, ...**

**$ Ventilator\_cummalative <dbl> 3322, 3321, 3305, 3280, 3252, 3223, 3205, 3184, 3179, 3171, 3153, 3147, 3123, 3...**

**$ Total\_on\_ventilator <dbl> 7094, 7005, 6999, 6867, 6855, 6649, 6520, 6245, 6148, 6030, 5986, 5990, 5630, 5...**

**$ Total\_positive\_cases <int> 14534035, 14357264, 14146191, 13921360, 13711156, 13515360, 13338607, 13191020,...**

**$ Today\_positive\_cases <int> 176771, 211073, 224831, 210204, 195796, 176753, 147587, 135242, 154522, 193725,...**

**$ Total\_people\_recovered <dbl> 5624444, 5576026, 5470389, 5404018, 5322128, 5226581, 5146643, 5065148, 5024447...**

**$ states <int> 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 56,...**

**$ Total\_tests\_results <int> 204063869, 202429337, 200259581, 198404712, 196576482, 195117280, 192776284, 19...**

**$ Total\_tests\_results\_today <int> 1634532, 2169756, 1854869, 1828230, 1459202, 2340996, 1603253, 1289970, 1709566...**

**#removing date and states for analysis**

**covid\_USA\_upd<- covid\_USA[-c(1,16)]**

**#replacing na with 0**

**covid\_USA\_upd[is.na(covid\_USA\_upd)]<-0**

**#correlation between variables**

**round(cor(covid\_USA\_upd[,-1]),digits=2)**

**Today\_death In\_ICU\_cummalative Total\_in\_ICU Today\_hospitalized Total\_hospitalized**

**Today\_death 1.00 0.31 0.84 0.60 0.77**

**In\_ICU\_cummalative 0.31 1.00 0.60 0.28 0.68**

**Total\_in\_ICU 0.84 0.60 1.00 0.55 0.96**

**Today\_hospitalized 0.60 0.28 0.55 1.00 0.55**

**Total\_hospitalized 0.77 0.68 0.96 0.55 1.00**

**Hospitalized\_cummalative 0.36 0.99 0.64 0.30 0.71**

**Total\_negative 0.27 0.98 0.55 0.27 0.63**

**Today\_negative 0.36 0.96 0.62 0.32 0.73**

**Ventilator\_cummalative 0.29 1.00 0.57 0.27 0.64**

**Total\_on\_ventilator 0.82 0.35 0.91 0.49 0.84**

**Total\_positive\_cases 0.33 0.98 0.61 0.30 0.68**

**Today\_positive\_cases 0.50 0.83 0.73 0.44 0.82**

**Total\_people\_recovered 0.28 0.97 0.57 0.27 0.63**

**Total\_tests\_results 0.28 0.97 0.56 0.27 0.63**

**Total\_tests\_results\_today 0.37 0.97 0.63 0.34 0.73**

**Hospitalized\_cummalative Total\_negative Today\_negative Ventilator\_cummalative**

**Today\_death 0.36 0.27 0.36 0.29**

**In\_ICU\_cummalative 0.99 0.98 0.96 1.00**

**Total\_in\_ICU 0.64 0.55 0.62 0.57**

**Today\_hospitalized 0.30 0.27 0.32 0.27**

**Total\_hospitalized 0.71 0.63 0.73 0.64**

**Hospitalized\_cummalative 1.00 0.95 0.96 0.98**

**Total\_negative 0.95 1.00 0.92 0.99**

**Today\_negative 0.96 0.92 1.00 0.94**

**Ventilator\_cummalative 0.98 0.99 0.94 1.00**

**Total\_on\_ventilator 0.39 0.28 0.39 0.30**

**Total\_positive\_cases 0.96 1.00 0.93 0.98**

**Today\_positive\_cases 0.81 0.85 0.86 0.82**

**Total\_people\_recovered 0.94 1.00 0.91 0.98**

**Total\_tests\_results 0.95 1.00 0.92 0.98**

**Total\_tests\_results\_today 0.96 0.95 0.99 0.96**

**Total\_on\_ventilator Total\_positive\_cases Today\_positive\_cases Total\_people\_recovered**

**Today\_death 0.82 0.33 0.50 0.28**

**In\_ICU\_cummalative 0.35 0.98 0.83 0.97**

**Total\_in\_ICU 0.91 0.61 0.73 0.57**

**Today\_hospitalized 0.49 0.30 0.44 0.27**

**Total\_hospitalized 0.84 0.68 0.82 0.63**

**Hospitalized\_cummalative 0.39 0.96 0.81 0.94**

**Total\_negative 0.28 1.00 0.85 1.00**

**Today\_negative 0.39 0.93 0.86 0.91**

**Ventilator\_cummalative 0.30 0.98 0.82 0.98**

**Total\_on\_ventilator 1.00 0.35 0.51 0.29**

**Total\_positive\_cases 0.35 1.00 0.88 1.00**

**Today\_positive\_cases 0.51 0.88 1.00 0.86**

**Total\_people\_recovered 0.29 1.00 0.86 1.00**

**Total\_tests\_results 0.29 1.00 0.86 1.00**

**Total\_tests\_results\_today 0.39 0.96 0.90 0.94**

**Total\_tests\_results Total\_tests\_results\_today**

**Today\_death 0.28 0.37**

**In\_ICU\_cummalative 0.97 0.97**

**Total\_in\_ICU 0.56 0.63**

**Today\_hospitalized 0.27 0.34**

**Total\_hospitalized 0.63 0.73**

**Hospitalized\_cummalative 0.95 0.96**

**Total\_negative 1.00 0.95**

**Today\_negative 0.92 0.99**

**Ventilator\_cummalative 0.98 0.96**

**Total\_on\_ventilator 0.29 0.39**

**Total\_positive\_cases 1.00 0.96**

**Today\_positive\_cases 0.86 0.90**

**Total\_people\_recovered 1.00 0.94**

**Total\_tests\_results 1.00 0.95**

**Total\_tests\_results\_today 0.95 1.00**

**#plotting all the variables against our response variable**

**covid\_USA\_upd %>%**

**gather(key, val, -Today\_positive\_cases) %>%**

**ggplot(aes(x = val, y = Today\_positive\_cases)) +**

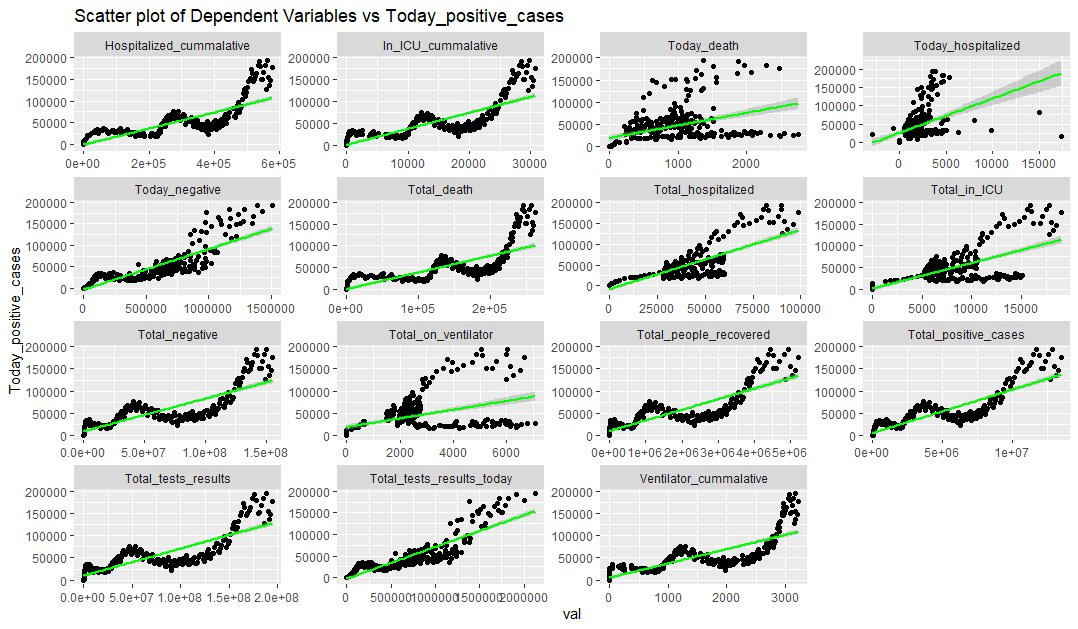
**geom\_point() +**

**stat\_smooth(method = "lm", se = TRUE, col = "green") +**

**facet\_wrap(~key, scales = "free") +**

**theme\_gray() +**

**ggtitle("Scatter plot of Dependent Variables vs Today\_positive\_cases")**

****

**#SmoteR for tackling imbalance**

**smote\_covid\_2<- SmoteRegress(Today\_positive\_cases~., covid\_USA\_upd, C.perc = list(0.5,2.5))**

**#scaling data**

**scaled\_data<- scale(smote\_covid\_2)**

**#converting into a data frame**

**scaled\_data\_1<- as.data.frame(scaled\_data)**

**#train-test split of the data**

**set.seed(2020)**

**training\_ind<- sample(nrow(scaled\_data), nrow(scaled\_data)\* 0.8)**

**train<- scaled\_data[training\_ind, ]**

**train[is.na(train)]<- 0**

**train\_1<- as.data.frame(train)**

**test<- scaled\_data[-training\_ind, ]**

**test\_1<- as.data.frame(test)**

**#Linear Regression model**

**reg<- lm(Today\_positive\_cases~., data= train\_1)**

**info\_reg<-summary(reg)**

**#predicting on test set**

**reg\_test<- predict(reg, newdata = test\_1)**

**head(reg\_test)**

**246 199 76 252 138 122**

**-0.57446361 -0.83456098 -0.69932503 -0.91999596 -0.05229056 -0.26444104**

**#model with subset**

**model\_1 <- regsubsets(Today\_positive\_cases ~., data= train\_1, nbest=1, nvmax= 15)**

**summary(model\_1)**

**Subset selection object**

**Call: regsubsets.formula(Today\_positive\_cases ~ ., data = train\_1,**

**nbest = 1, nvmax = 15)**

**15 Variables (and intercept)**

**Forced in Forced out**

**Total\_death FALSE FALSE**

**Today\_death FALSE FALSE**

**In\_ICU\_cummalative FALSE FALSE**

**Total\_in\_ICU FALSE FALSE**

**Today\_hospitalized FALSE FALSE**

**Total\_hospitalized FALSE FALSE**

**Hospitalized\_cummalative FALSE FALSE**

**Total\_negative FALSE FALSE**

**Today\_negative FALSE FALSE**

**Ventilator\_cummalative FALSE FALSE**

**Total\_on\_ventilator FALSE FALSE**

**Total\_positive\_cases FALSE FALSE**

**Total\_people\_recovered FALSE FALSE**

**Total\_tests\_results FALSE FALSE**

**Total\_tests\_results\_today FALSE FALSE**

**1 subsets of each size up to 15**

**Selection Algorithm: exhaustive**

**Total\_death Today\_death In\_ICU\_cummalative Total\_in\_ICU Today\_hospitalized Total\_hospitalized**

**1 ( 1 ) " " " " " " " " " " " "**

**2 ( 1 ) " " " " " " " " " " "\*"**

**3 ( 1 ) " " " " " " " " " " "\*"**

**4 ( 1 ) "\*" " " " " " " " " "\*"**

**5 ( 1 ) "\*" " " "\*" " " " " "\*"**

**6 ( 1 ) "\*" " " "\*" " " " " "\*"**

**7 ( 1 ) "\*" " " "\*" " " " " "\*"**

**8 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**9 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**10 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**11 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**12 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**13 ( 1 ) "\*" "\*" "\*" " " " " "\*"**

**14 ( 1 ) "\*" "\*" "\*" " " "\*" "\*"**

**15 ( 1 ) "\*" "\*" "\*" "\*" "\*" "\*"**

**Hospitalized\_cummalative Total\_negative Today\_negative Ventilator\_cummalative Total\_on\_ventilator**

**1 ( 1 ) " " " " " " " " " "**

**2 ( 1 ) " " " " " " " " " "**

**3 ( 1 ) " " " " "\*" " " " "**

**4 ( 1 ) " " " " "\*" " " " "**

**5 ( 1 ) " " " " "\*" " " " "**

**6 ( 1 ) " " " " "\*" " " " "**

**7 ( 1 ) " " "\*" "\*" " " "\*"**

**8 ( 1 ) " " "\*" "\*" " " "\*"**

**9 ( 1 ) " " "\*" "\*" " " "\*"**

**10 ( 1 ) " " "\*" "\*" " " "\*"**

**11 ( 1 ) "\*" "\*" "\*" " " "\*"**

**12 ( 1 ) "\*" "\*" "\*" "\*" "\*"**

**13 ( 1 ) "\*" "\*" "\*" "\*" "\*"**

**14 ( 1 ) "\*" "\*" "\*" "\*" "\*"**

**15 ( 1 ) "\*" "\*" "\*" "\*" "\*"**

**Total\_positive\_cases Total\_people\_recovered Total\_tests\_results Total\_tests\_results\_today**

**1 ( 1 ) "\*" " " " " " "**

**2 ( 1 ) " " "\*" " " " "**

**3 ( 1 ) " " " " " " "\*"**

**4 ( 1 ) " " " " " " "\*"**

**5 ( 1 ) " " " " " " "\*"**

**6 ( 1 ) " " " " "\*" "\*"**

**7 ( 1 ) " " " " " " "\*"**

**8 ( 1 ) " " " " " " "\*"**

**9 ( 1 ) " " " " "\*" "\*"**

**10 ( 1 ) "\*" " " "\*" "\*"**

**11 ( 1 ) "\*" " " "\*" "\*"**

**12 ( 1 ) "\*" " " "\*" "\*"**

**13 ( 1 ) "\*" "\*" "\*" "\*"**

**14 ( 1 ) "\*" "\*" "\*" "\*"**

**15 ( 1 ) "\*" "\*" "\*" "\*"**

**#Variable selection using stepwise selection**

**#Null linear model**

**nullmodel\_1 <- lm(Today\_positive\_cases ~ 1, data = scaled\_data\_1)**

**#Full linear model**

**fullmodel\_1 <- lm(Today\_positive\_cases ~., data = scaled\_data\_1)**

**#forward selection**

**model\_forward<-step(nullmodel\_1, scope = list(lower = nullmodel\_1, upper = fullmodel\_1), direction = "forward")**

**#backward model**

**model.step.b <- step(fullmodel\_1, direction = "backward")**

**#stepwise selection**

**model\_3\_1<- step(nullmodel\_1, scope = list(lower = nullmodel\_1, upper = fullmodel\_1), direction = "both")**

**Start: AIC=1**

**Today\_positive\_cases ~ 1**

**Df Sum of Sq RSS AIC**

**+ Total\_positive\_cases 1 215.144 32.856 -500.30**

**+ Total\_tests\_results\_today 1 213.652 34.348 -489.25**

**+ Total\_people\_recovered 1 209.369 38.631 -459.99**

**+ Total\_tests\_results 1 206.285 41.715 -440.86**

**+ Total\_negative 1 202.985 45.015 -421.91**

**+ In\_ICU\_cummalative 1 197.660 50.340 -394.06**

**+ Today\_negative 1 197.238 50.762 -391.99**

**+ Total\_hospitalized 1 197.024 50.976 -390.94**

**+ Hospitalized\_cummalative 1 189.655 58.345 -357.32**

**+ Ventilator\_cummalative 1 189.449 58.551 -356.44**

**+ Total\_death 1 180.697 67.303 -321.75**

**+ Total\_in\_ICU 1 173.956 74.044 -297.98**

**+ Total\_on\_ventilator 1 102.446 145.554 -129.69**

**+ Today\_death 1 83.765 164.235 -99.62**

**+ Today\_hospitalized 1 68.691 179.309 -77.76**

**<none> 248.000 1.00**

**Step: AIC=-500.3**

**Today\_positive\_cases ~ Total\_positive\_cases**

**Df Sum of Sq RSS AIC**

**+ Total\_hospitalized 1 13.569 19.288 -630.94**

**+ Ventilator\_cummalative 1 10.906 21.950 -598.74**

**+ Total\_negative 1 9.255 23.601 -580.68**

**+ Total\_in\_ICU 1 9.164 23.692 -579.73**

**+ Today\_death 1 8.875 23.981 -576.71**

**+ Total\_on\_ventilator 1 7.562 25.294 -563.43**

**+ Total\_people\_recovered 1 7.303 25.553 -560.90**

**+ Total\_tests\_results 1 6.418 26.438 -552.42**

**+ In\_ICU\_cummalative 1 5.824 27.032 -546.89**

**+ Total\_death 1 5.521 27.335 -544.11**

**+ Hospitalized\_cummalative 1 3.843 29.013 -529.28**

**+ Total\_tests\_results\_today 1 3.342 29.515 -525.01**

**+ Today\_hospitalized 1 3.245 29.611 -524.19**

**+ Today\_negative 1 0.666 32.190 -503.40**

**<none> 32.856 -500.30**

**- Total\_positive\_cases 1 215.144 248.000 1.00**

**Step: AIC=-630.94**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized**

**Df Sum of Sq RSS AIC**

**+ Total\_death 1 4.421 14.866 -693.77**

**+ Hospitalized\_cummalative 1 4.131 15.157 -688.95**

**+ Ventilator\_cummalative 1 3.139 16.148 -673.17**

**+ In\_ICU\_cummalative 1 2.896 16.391 -669.46**

**+ Total\_in\_ICU 1 1.691 17.596 -651.79**

**+ Total\_tests\_results\_today 1 1.291 17.997 -646.18**

**+ Total\_on\_ventilator 1 1.224 18.064 -645.26**

**+ Total\_people\_recovered 1 0.443 18.845 -634.72**

**+ Today\_hospitalized 1 0.319 18.968 -633.10**

**+ Total\_negative 1 0.177 19.110 -631.24**

**<none> 19.288 -630.94**

**+ Total\_tests\_results 1 0.096 19.191 -630.19**

**+ Today\_death 1 0.091 19.197 -630.12**

**+ Today\_negative 1 0.039 19.249 -629.44**

**- Total\_hospitalized 1 13.569 32.856 -500.30**

**- Total\_positive\_cases 1 31.688 50.976 -390.94**

**Step: AIC=-693.77**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death**

**Df Sum of Sq RSS AIC**

**+ Total\_tests\_results\_today 1 5.5204 9.346 -807.35**

**+ In\_ICU\_cummalative 1 4.4013 10.465 -779.19**

**+ Today\_negative 1 2.7596 12.107 -742.90**

**+ Total\_in\_ICU 1 2.5718 12.294 -739.07**

**+ Total\_on\_ventilator 1 1.5197 13.347 -718.62**

**+ Today\_hospitalized 1 0.4472 14.419 -699.38**

**+ Hospitalized\_cummalative 1 0.3153 14.551 -697.11**

**+ Total\_tests\_results 1 0.2911 14.575 -696.69**

**+ Total\_negative 1 0.2243 14.642 -695.56**

**+ Total\_people\_recovered 1 0.1810 14.685 -694.82**

**<none> 14.866 -693.77**

**+ Ventilator\_cummalative 1 0.1150 14.751 -693.70**

**+ Today\_death 1 0.0644 14.802 -692.85**

**- Total\_death 1 4.4214 19.288 -630.94**

**- Total\_hospitalized 1 12.4684 27.335 -544.11**

**- Total\_positive\_cases 1 18.6769 33.543 -493.15**

**Step: AIC=-807.35**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today**

**Df Sum of Sq RSS AIC**

**+ Today\_negative 1 2.8827 6.4632 -897.18**

**+ In\_ICU\_cummalative 1 2.2784 7.0675 -874.92**

**+ Total\_in\_ICU 1 0.7464 8.5995 -826.07**

**+ Total\_on\_ventilator 1 0.3254 9.0205 -814.17**

**+ Hospitalized\_cummalative 1 0.1483 9.1977 -809.33**

**+ Total\_tests\_results 1 0.1029 9.2430 -808.10**

**<none> 9.3459 -807.35**

**+ Today\_death 1 0.0673 9.2786 -807.15**

**+ Total\_negative 1 0.0518 9.2941 -806.73**

**+ Today\_hospitalized 1 0.0365 9.3094 -806.32**

**+ Ventilator\_cummalative 1 0.0270 9.3189 -806.07**

**+ Total\_people\_recovered 1 0.0051 9.3408 -805.48**

**- Total\_tests\_results\_today 1 5.5204 14.8663 -693.77**

**- Total\_positive\_cases 1 8.0315 17.3774 -654.91**

**- Total\_hospitalized 1 8.0527 17.3986 -654.60**

**- Total\_death 1 8.6511 17.9970 -646.18**

**Step: AIC=-897.18**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative**

**Df Sum of Sq RSS AIC**

**+ In\_ICU\_cummalative 1 1.3839 5.0793 -955.18**

**+ Total\_in\_ICU 1 1.0083 5.4549 -937.42**

**+ Total\_on\_ventilator 1 0.3529 6.1103 -909.16**

**+ Total\_people\_recovered 1 0.0929 6.3703 -898.79**

**+ Ventilator\_cummalative 1 0.0793 6.3839 -898.25**

**<none> 6.4632 -897.18**

**+ Total\_negative 1 0.0247 6.4384 -896.14**

**+ Today\_death 1 0.0153 6.4478 -895.77**

**+ Total\_tests\_results 1 0.0122 6.4509 -895.65**

**+ Hospitalized\_cummalative 1 0.0087 6.4545 -895.52**

**+ Today\_hospitalized 1 0.0004 6.4628 -895.20**

**- Total\_positive\_cases 1 1.1379 7.6011 -858.80**

**- Today\_negative 1 2.8827 9.3459 -807.35**

**- Total\_death 1 3.2691 9.7323 -797.26**

**- Total\_tests\_results\_today 1 5.6435 12.1067 -742.90**

**- Total\_hospitalized 1 8.3619 14.8251 -692.46**

**Step: AIC=-955.18**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative**

**Df Sum of Sq RSS AIC**

**+ Ventilator\_cummalative 1 0.6502 4.4290 -987.29**

**+ Total\_in\_ICU 1 0.5782 4.5011 -983.27**

**+ Total\_tests\_results 1 0.4471 4.6322 -976.12**

**+ Total\_negative 1 0.4085 4.6707 -974.06**

**+ Total\_on\_ventilator 1 0.2051 4.8741 -963.45**

**+ Hospitalized\_cummalative 1 0.0556 5.0236 -955.92**

**+ Total\_people\_recovered 1 0.0553 5.0239 -955.91**

**+ Today\_death 1 0.0495 5.0297 -955.62**

**<none> 5.0793 -955.18**

**+ Today\_hospitalized 1 0.0010 5.0783 -953.23**

**- Total\_positive\_cases 1 0.5372 5.6165 -932.15**

**- In\_ICU\_cummalative 1 1.3839 6.4632 -897.18**

**- Today\_negative 1 1.9883 7.0675 -874.92**

**- Total\_death 1 2.3344 7.4137 -863.02**

**- Total\_tests\_results\_today 1 3.7467 8.8260 -819.60**

**- Total\_hospitalized 1 8.6808 13.7601 -709.02**

**Step: AIC=-987.29**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative**

**Df Sum of Sq RSS AIC**

**+ Total\_on\_ventilator 1 0.4255 4.0035 -1010.44**

**+ Total\_in\_ICU 1 0.2957 4.1333 -1002.49**

**+ Today\_death 1 0.2366 4.1925 -998.96**

**+ Total\_negative 1 0.1839 4.2451 -995.85**

**+ Total\_people\_recovered 1 0.0603 4.3688 -988.70**

**+ Today\_hospitalized 1 0.0440 4.3850 -987.77**

**+ Total\_tests\_results 1 0.0404 4.3886 -987.57**

**<none> 4.4290 -987.29**

**+ Hospitalized\_cummalative 1 0.0006 4.4284 -985.32**

**- Total\_positive\_cases 1 0.4603 4.8894 -964.67**

**- Ventilator\_cummalative 1 0.6502 5.0793 -955.18**

**- In\_ICU\_cummalative 1 1.9549 6.3839 -898.25**

**- Total\_death 1 2.3072 6.7362 -884.88**

**- Today\_negative 1 2.3623 6.7913 -882.85**

**- Total\_hospitalized 1 4.0691 8.4981 -827.02**

**- Total\_tests\_results\_today 1 4.1665 8.5956 -824.19**

**Step: AIC=-1010.44**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative + Total\_on\_ventilator**

**Df Sum of Sq RSS AIC**

**+ Today\_death 1 0.5315 3.4720 -1043.91**

**+ Total\_people\_recovered 1 0.2261 3.7774 -1022.92**

**+ Total\_tests\_results 1 0.0681 3.9355 -1012.71**

**+ Total\_negative 1 0.0397 3.9639 -1010.92**

**+ Today\_hospitalized 1 0.0372 3.9663 -1010.76**

**<none> 4.0035 -1010.44**

**+ Hospitalized\_cummalative 1 0.0107 3.9929 -1009.10**

**+ Total\_in\_ICU 1 0.0009 4.0026 -1008.50**

**- Total\_positive\_cases 1 0.3736 4.3771 -990.22**

**- Total\_on\_ventilator 1 0.4255 4.4290 -987.29**

**- Ventilator\_cummalative 1 0.8706 4.8741 -963.45**

**- In\_ICU\_cummalative 1 1.9063 5.9098 -915.47**

**- Total\_death 1 1.9887 5.9922 -912.02**

**- Today\_negative 1 2.5162 6.5197 -891.01**

**- Total\_hospitalized 1 3.3838 7.3873 -859.90**

**- Total\_tests\_results\_today 1 4.1357 8.1393 -835.77**

**Step: AIC=-1043.91**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative + Total\_on\_ventilator +**

**Today\_death**

**Df Sum of Sq RSS AIC**

**+ Total\_tests\_results 1 0.15969 3.3123 -1053.63**

**+ Total\_people\_recovered 1 0.12536 3.3467 -1051.06**

**+ Total\_negative 1 0.10016 3.3719 -1049.20**

**+ Hospitalized\_cummalative 1 0.06358 3.4084 -1046.51**

**<none> 3.4720 -1043.91**

**+ Total\_in\_ICU 1 0.01499 3.4570 -1042.98**

**+ Today\_hospitalized 1 0.00254 3.4695 -1042.09**

**- Total\_positive\_cases 1 0.50034 3.9724 -1012.39**

**- Today\_death 1 0.53152 4.0035 -1010.44**

**- Total\_on\_ventilator 1 0.72044 4.1925 -998.96**

**- Ventilator\_cummalative 1 1.29118 4.7632 -967.18**

**- Today\_negative 1 1.90642 5.3784 -936.93**

**- Total\_hospitalized 1 2.37077 5.8428 -916.31**

**- In\_ICU\_cummalative 1 2.42056 5.8926 -914.20**

**- Total\_death 1 2.42583 5.8978 -913.97**

**- Total\_tests\_results\_today 1 3.10682 6.5788 -886.77**

**Step: AIC=-1053.63**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative + Total\_on\_ventilator +**

**Today\_death + Total\_tests\_results**

**Df Sum of Sq RSS AIC**

**+ Hospitalized\_cummalative 1 0.08945 3.2229 -1058.45**

**+ Total\_negative 1 0.07973 3.2326 -1057.70**

**<none> 3.3123 -1053.63**

**+ Total\_in\_ICU 1 0.02649 3.2858 -1053.63**

**+ Total\_people\_recovered 1 0.00692 3.3054 -1052.15**

**+ Today\_hospitalized 1 0.00257 3.3098 -1051.82**

**- Total\_tests\_results 1 0.15969 3.4720 -1043.91**

**- Total\_positive\_cases 1 0.51178 3.8241 -1019.86**

**- Total\_death 1 0.62311 3.9354 -1012.71**

**- Today\_death 1 0.62314 3.9355 -1012.71**

**- Ventilator\_cummalative 1 0.66558 3.9779 -1010.04**

**- Total\_on\_ventilator 1 0.80538 4.1177 -1001.44**

**- Today\_negative 1 1.85302 5.1653 -945.00**

**- In\_ICU\_cummalative 1 2.28767 5.6000 -924.88**

**- Total\_hospitalized 1 2.41343 5.7257 -919.35**

**- Total\_tests\_results\_today 1 2.99852 6.3108 -895.12**

**Step: AIC=-1058.45**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative + Total\_on\_ventilator +**

**Today\_death + Total\_tests\_results + Hospitalized\_cummalative**

**Df Sum of Sq RSS AIC**

**+ Total\_negative 1 0.04666 3.1762 -1060.08**

**<none> 3.2229 -1058.45**

**+ Total\_people\_recovered 1 0.00667 3.2162 -1056.96**

**+ Total\_in\_ICU 1 0.00135 3.2215 -1056.55**

**+ Today\_hospitalized 1 0.00038 3.2225 -1056.48**

**- Hospitalized\_cummalative 1 0.08945 3.3123 -1053.63**

**- Total\_death 1 0.10324 3.3261 -1052.60**

**- Total\_tests\_results 1 0.18557 3.4084 -1046.51**

**- Total\_positive\_cases 1 0.55875 3.7816 -1020.64**

**- Ventilator\_cummalative 1 0.67673 3.8996 -1012.99**

**- Today\_death 1 0.69754 3.9204 -1011.66**

**- Total\_on\_ventilator 1 0.89177 4.1146 -999.62**

**- Today\_negative 1 1.87220 5.0951 -946.41**

**- Total\_hospitalized 1 2.08942 5.3123 -936.01**

**- In\_ICU\_cummalative 1 2.37712 5.6000 -922.88**

**- Total\_tests\_results\_today 1 2.99185 6.2147 -896.94**

**Step: AIC=-1060.08**

**Today\_positive\_cases ~ Total\_positive\_cases + Total\_hospitalized +**

**Total\_death + Total\_tests\_results\_today + Today\_negative +**

**In\_ICU\_cummalative + Ventilator\_cummalative + Total\_on\_ventilator +**

**Today\_death + Total\_tests\_results + Hospitalized\_cummalative +**

**Total\_negative**

**Df Sum of Sq RSS AIC**

**<none> 3.1762 -1060.08**

**- Total\_negative 1 0.04666 3.2229 -1058.45**

**+ Today\_hospitalized 1 0.00139 3.1748 -1058.19**

**+ Total\_in\_ICU 1 0.00137 3.1748 -1058.19**

**+ Total\_people\_recovered 1 0.00043 3.1758 -1058.11**

**- Hospitalized\_cummalative 1 0.05638 3.2326 -1057.70**

**- Ventilator\_cummalative 1 0.09610 3.2723 -1054.66**

**- Total\_tests\_results 1 0.09937 3.2756 -1054.41**

**- Total\_death 1 0.11198 3.2882 -1053.45**

**- Total\_positive\_cases 1 0.22143 3.3976 -1045.30**

**- Total\_on\_ventilator 1 0.41383 3.5900 -1031.58**

**- Today\_death 1 0.70741 3.8836 -1012.01**

**- In\_ICU\_cummalative 1 1.18293 4.3591 -983.25**

**- Today\_negative 1 1.60192 4.7781 -960.40**

**- Total\_hospitalized 1 1.96157 5.1378 -942.33**

**- Total\_tests\_results\_today 1 2.58880 5.7650 -913.65**

**#Linear Regression model with best parameters**

**best\_reg <- lm(Today\_positive\_cases ~ . -Today\_hospitalized -Total\_negative -Hospitalized\_cummalative**

**-Total\_in\_ICU -Total\_people\_recovered , data = train\_1)**

**summary(best\_reg)**

**Call:**

**lm(formula = Today\_positive\_cases ~ . - Today\_hospitalized -**

**Total\_negative - Hospitalized\_cummalative - Total\_in\_ICU -**

**Total\_people\_recovered, data = train\_1)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-0.32061 -0.08224 0.00098 0.06291 0.34144**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 0.001872 0.008808 0.213 0.832**

**Total\_death -1.714851 0.279906 -6.127 5.15e-09 \*\*\***

**Today\_death 0.106832 0.017336 6.162 4.27e-09 \*\*\***

**In\_ICU\_cummalative 4.090333 0.372538 10.980 < 2e-16 \*\*\***

**Total\_hospitalized 0.433717 0.038692 11.209 < 2e-16 \*\*\***

**Today\_negative -0.964211 0.095417 -10.105 < 2e-16 \*\*\***

**Ventilator\_cummalative -2.129461 0.381649 -5.580 8.31e-08 \*\*\***

**Total\_on\_ventilator -0.172722 0.025943 -6.658 2.97e-10 \*\*\***

**Total\_positive\_cases -1.264371 0.276012 -4.581 8.42e-06 \*\*\***

**Total\_tests\_results 1.054091 0.415747 2.535 0.012 \***

**Total\_tests\_results\_today 1.507529 0.120417 12.519 < 2e-16 \*\*\***

**---**

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

**Residual standard error: 0.1233 on 188 degrees of freedom**

**Multiple R-squared: 0.825, Adjusted R-squared: 0.816**

**F-statistic: 1350 on 10 and 188 DF, p-value: < 2.2e-16**

**#predicting on test set using the best linear regression model**

**best\_reg\_test<- predict(best\_reg, newdata = test\_1)**

**regr.eval(test\_1$Today\_positive\_cases,best\_reg\_test)**

**mae mse rmse mape**

**0.073883442 0.010586182 0.107909048 0.274716355**

**#Plot of performance of Linear Regression model**

**x<- 1:length(test\_1$Today\_positive\_cases)**

**plot(x,test\_1$Today\_positive\_cases, pch=20, col= 'black', main='Actual to predicted plot for Linear Regression',**

**xlab='Index of the test value',ylab= 'Actual\_Test\_value' )**

**points(x, best\_reg\_test, lwd='2', col='red')**

**Chart, scatter chart

Description automatically generated**

**#Linear SVM model**

**set.seed(2020)**

**model\_svm<- svm(Today\_positive\_cases~ .,data=train\_1, kernel= 'linear')**

**summary(model\_svm)**

**Call:**

**svm(formula = Today\_positive\_cases ~ ., data = train\_1, kernel = "linear")**

**Parameters:**

**SVM-Type: eps-regression**

**SVM-Kernel: linear**

**cost: 1**

**gamma: 0.06666667**

**epsilon: 0.1**

**Number of Support Vectors: 104**

**#tuning parameters**

**tune\_svm\_lin<- tune(svm, Today\_positive\_cases~ .,data=train\_1, kernel= "linear", ranges=list(cost=c(0.01,.1,1,5,10,100,500)))**

**best\_para<- tune\_svm\_lin$best.parameters**

**summary(tune\_svm\_lin)**

**Parameter tuning of ‘svm’:**

**- sampling method: 10-fold cross validation**

**- best parameters:**

**cost**

**10**

**- best performance: 0.017202**

**- Detailed performance results:**

**cost error dispersion**

**1 1e-02 0.08338001 0.021394202**

**2 1e-01 0.03744080 0.009059482**

**3 1e+00 0.02297057 0.005512584**

**4 5e+00 0.01865929 0.005623620**

**5 1e+01 0.01720200 0.006189526**

**6 1e+02 0.01728017 0.007649078**

**7 5e+02 0.01769542 0.008689245**

**#plot for best cost**

**plot(tune\_svm\_lin, main='Best cost for linear SVM')**

**Chart, histogram

Description automatically generated**

**#Model with tunes parameters**

**model\_svm\_1<- svm(Today\_positive\_cases~ .,data=train\_1, kernel= 'linear',cost= best\_para$cost)**

**summary(model\_svm\_1)**

**Call:**

**svm(formula = Today\_positive\_cases ~ ., data = train\_1, kernel = "linear", cost = best\_para$cost)**

**Parameters:**

**SVM-Type: eps-regression**

**SVM-Kernel: linear**

**cost: 10**

**gamma: 0.06666667**

**epsilon: 0.1**

**Number of Support Vectors: 87**

**#prediction on test set**

**test\_svm\_lin<- predict(model\_svm\_1,test\_1)**

**head(test\_svm\_lin)**

**246 199 76 252 138 122**

**-0.64343329 -0.88443243 -0.59911662 -0.95211924 -0.06826764 -0.22235887**

**#performance check**

**regr.eval(test\_1$Today\_positive\_cases,test\_svm\_lin)**

**mae mse rmse mape**

**0.08857549 0.097984 0.31212508 0.22202950**

**#Plot of performance**

**x<- 1:length(test\_1$Today\_positive\_cases)**

**plot(x,test\_1$Today\_positive\_cases, pch=20, col= 'black', main='Actual to predicted plot for Linear Kernel SVM',**

**xlab='Index of the test value',ylab= 'Actual\_Test\_value' )**

**points(x,test\_svm\_lin, lwd='2', col='blue')**

**Chart, scatter chart

Description automatically generated**

**set.seed(2020)**

**Radial\_KernelSVM\_tune<- tune(svm,Today\_positive\_cases~ ., data= train\_1, kernel= 'radial', ranges= list(cost=c(0.01,.1,1,5), gamma=c(.01,.02)))**

**summary(Radial\_KernelSVM\_tune)**

**Parameter tuning of ‘svm’:**

**- sampling method: 10-fold cross validation**

**- best parameters:**

**cost gamma**

**5 0.02**

**- best performance: 0.01518118**

**- Detailed performance results:**

**cost gamma error dispersion**

**1 0.01 0.01 0.72370963 0.260325962**

**2 0.10 0.01 0.07955981 0.031054916**

**3 1.00 0.01 0.03059068 0.008082804**

**4 5.00 0.01 0.02048386 0.006418007**

**5 0.01 0.02 0.60167795 0.233275100**

**6 0.10 0.02 0.05731602 0.025246736**

**7 1.00 0.02 0.02401548 0.006970768**

**8 5.00 0.02 0.01518118 0.006121776**

**#plot for parameters**

**plot(Radial\_KernelSVM\_tune, main='Best Parameters for Radial SVM Kernel')**

**Chart

Description automatically generated**

**#Radial SVM kernel model with best parameters**

**Kernel\_svm\_model<- svm(Today\_positive\_cases~., data=train\_1, kernel= 'radial',**

**cost= Radial\_KernelSVM\_tune$best.parameters$cost,**

**gamma= Radial\_KernelSVM\_tune$best.parameters$gamma)**

**summary(Kernel\_svm\_model)**

**Call:**

**svm(formula = Today\_positive\_cases ~ ., data = train\_1, kernel = "radial", cost = Radial\_KernelSVM\_tune$best.parameters$cost,**

**gamma = Radial\_KernelSVM\_tune$best.parameters$gamma)**

**Parameters:**

**SVM-Type: eps-regression**

**SVM-Kernel: radial**

**cost: 5**

**gamma: 0.02**

**epsilon: 0.1**

**Number of Support Vectors: 71**

**#prediction on test set**

**pred\_kernel\_svm<- predict(Kernel\_svm\_model, test\_1)**

**head(pred\_kernel\_svm)**

**246 199 76 252 138 122**

**-0.8268279 -0.8164904 -0.6829424 -0.9860113 -0.1887864 -0.2519124**

**#performance evaluation**

**regr.eval(test\_1$Today\_positive\_cases,pred\_kernel\_svm)**

**mae mse rmse mape**

**0.08890176 0.01843159 0.13576299 0.40394466**

**#plot for kernel svm performance**

**x<- 1:length(test\_1$Today\_positive\_cases)**

**plot(x,test\_1$Today\_positive\_cases, pch=20, col= 'black', main='Actual to predicted plot for Radial Kernel SVM',**

**xlab='Index of the test value',ylab= 'Actual\_Test\_value')**

**points(x, pred\_kernel\_svm, lwd='2', col='green')**

**#KNN algorith**

**#model with best parameters**

**set.seed(2020)**

**ctrl <- trainControl(method="repeatedcv",repeats = 3) #,classProbs=TRUE,summaryFunction = twoClassSummary)**

**knnFit <- train(Today\_positive\_cases ~ ., data = train\_1, method = "knn", trControl = ctrl, preProcess = c("center","scale"), tuneLength = 8)**

**knnFit**

**plot(knnFit)**

**Chart, line chart

Description automatically generated**

**#predicting on test set**

**knnpredict<- predict(knnFit, newdata= test\_1)**

**head(knnpredict)**

**[1] -0.8391491 -0.8876670 -0.5011509 -1.1870069 -0.2244641 -0.2114759**

**regr.eval(test\_1$Today\_positive\_cases,knnpredict)**

**mae mse rmse mape**

**0.07400351 0.01086222 0.10422197 0.63775659**

**#Performance of KNN**

**x<- 1:length(test\_1$Today\_positive\_cases)**

**plot(x,test\_1$Today\_positive\_cases, pch=20, col= 'black',main='Actual to predicted plot for K-nearest Neighbour',**

**xlab='Index of the test value',ylab= 'Actual\_Test\_value')**

**points(x, knnpredict, lwd='2', col='purple')**

**Chart, scatter chart

Description automatically generated**