Prediction Model

Voting Prediction Model

Loading required libraries

```
library(tidyverse)
## -- Attaching packages ------ tidyverse 1.3.1 --
## v ggplot2 3.3.3
                v purrr
                            0.3.4
## v tibble 3.1.2
                  v dplyr 1.0.6
## v tidyr 1.1.3
                  v stringr 1.4.0
## v readr 1.4.0 v 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
##
```

Loading 2014 and 2019 Loksabha Data

```
lok_2019 <- read.csv("D:/Janta ka mood Intern/Model/loksabha_2019.csv")
lok_2014 <- read.csv("D:/Janta ka mood Intern/Model/loksabha_2014 new.csv")
glimpse(lok_2019)</pre>
```

```
## Rows: 609
## Columns: 11
## $ Constituency <chr> "ARARIA", "A
## $ Candidate
                                                              <chr> "Pradeep Kumar Singh Â  Winner", "Abdul Wahid Khan", ~
## $ Partv
                                                              <chr> "BJP", "IND", "IND", "IND", "IND", "IND", "BSP", "IND",~
## $ Criminal.Cases <int> 3, 1, 1, 0, 0, 0, 0, 0, 6, 1, 2, 2, 1, 0, 0, 0, 1, 0, 0~
## $ Education
                                                             <chr> "10th Pass", "10th Pass", "Literate", "10th Pass", "Pos~
## $ Age
                                                              <int> 52, 69, 31, 33, 46, 36, 46, 66, 50, 28, 54, 60, 66, 66,~
## $ Total.Assets <chr> "Rs 50,10,577", "Rs 15,45,000", "Rs 6,02,510", "Rs ~
                                                             <chr> "Rs 11,59,200", "Rs 1,19,028", "Rs 0", "Rs 0", "RsÂ~
## $ Liabilities
                                                              <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0~
## $ Winner
                                                              ## $ Surname
                                                              ## $ Caste
```

```
glimpse(lok_2014)
```

```
## Rows: 610
## Columns: 11
## $ Constituency
                <chr> "ARARIA", "ARARIA", "ARARIA", "ARARIA", "ARARIA", "ARAR
                <chr> "Taslim Uddin Â  Winner", "Abdul Rahman", "Bidya Nand~
## $ Candidate
## $ Party
                 <chr> "RJD", "BSP", "Bahujan Mukti Party", "AAP", "Janta Dal ~
## $ Criminal.Cases <int> 4, 0, 0, 0, 1, 2, 4, 0, 0, 1, 0, 0, 1, 4, 0, 0, 3, 2, 0~
                <chr> "Literate", "Literate", "12th Pass", "Graduate Professi~
## $ Education
## $ Age
                 <int> 72, 29, 61, 48, 35, 34, 48, 46, 63, 32, 37, 38, 51, 49,~
## $ Total.Assets
                <chr> "Rs 4,23,68,468", "Rs 8,13,848", "Rs 8,56,973", "RsÂ~
## $ Liabilities
                <chr> "Rs 10,26,919", "Rs 0", "Rs 7,000", "Rs 50,25,161",~
## $ Winner
                <chr> "", "", "Paswan", "", "", "", "", "", "", "", "~
## $ Surname
                ## $ Caste
```

Selecting Feature Columns

```
loksabha_2019 <- lok_2019 %>% select(Constituency,Party,Criminal.Cases,Education,Age,Winner,Tota
l.Assets,Liabilities)
loksabha_2014 <- lok_2014 %>% select(Constituency,Party,Criminal.Cases,Education,Age,Winner,Tota
l.Assets,Liabilities)
```

Data Labeling

```
education_level <- c( "10th Pass", "Literate", "Post Graduate", "8th Pass", "12th Pass", "Graduate",
"Others", "Graduate Professional", "Doctorate", "5th Pass", "Not Given", "Illiterate", "")
education_label <- c(1:13)</pre>
loksabha_2019$Education <- as.integer(factor(loksabha_2019$Education,levels = education_level, l</pre>
abels = education_label))
loksabha_2014$Education <- as.integer(factor(loksabha_2014$Education,levels = education_level,la</pre>
bels = education_label))
constituency_level <- unique(loksabha_2019$Constituency)</pre>
loksabha_2019$Constituency <- as.integer(factor(loksabha_2019$Constituency,levels = constituency</pre>
level, labels = c(1:40))
loksabha_2014$Constituency <- as.integer(factor(loksabha_2014$Constituency,levels = constituency</pre>
level, labels = c(1:40))
x <- unique(loksabha 2019$Party)</pre>
y <- unique(loksabha 2014$Party)</pre>
z \leftarrow c(x,y)
part level <- unique(z)</pre>
party label <- c(1:length(part level))</pre>
print(part_level)
```

```
[1] "BJP"
##
     [2] "IND"
##
##
     [3] "BSP"
     [4] "RJD"
##
     [5] "Bihar Lok Nirman Dal"
##
##
     [6] "Bahujan Mukti Party"
##
     [7] "Pragatishil Samajwadi Party (Lohia)"
##
     [8] "Akhil Bharatiya Jan Sangh"
     [9] "Bhartiya Kranti Vir Party"
##
    [10] "Shoshit Samaj Dal"
##
##
    [11] "CPI(ML)(L)"
    [12] "Peoples Party of India (Democratic)"
##
    [13] "Akhil Hind Forward Bloc (Krantikari)"
##
    [14] "Swaraj Party (Loktantrik)"
##
    [15] "Hindustani Awam Morcha (Secular)"
##
   [16] "JD(U)"
##
    [17] "Bhartiya Dalit Party"
##
    [18] "Bharatiya Momin Front"
    [19] "JMM"
##
##
    [20] "Bhartiya Lokmat Rashtrwadi Party"
##
   [21] "CPI"
##
    [22] "SUCI(C)"
   [23] "AAP"
##
##
   [24] "Jantantrik Vikas Party"
    [25] "Rashtriya Dal United"
##
   [26] "Voters Party International"
##
    [27] "Suheldev Bhartiya Samaj Party"
##
    [28] "Mithilanchal Mukti Morcha"
    [29] "Moolniwasi Samaj Party"
##
   [30] "NCP"
##
##
    [31] "Bharatiya Rashtriya Morcha"
    [32] "Public Mission Party"
##
##
    [33] "Aam Janta Party Rashtriya"
##
    [34] "Moulik Adhikar Party"
    [35] "Ambedkarite Party of India"
    [36] "Bharat Bhrashtachar Mitao Party"
    [37] "Janta Dal Rashtravadi"
##
    [38] "SHS"
##
##
    [39] "Jai Prakash Janata Dal"
    [40] "LJP"
##
    [41] "Sathi Aur Aapka Faisala Party"
##
   [42] "Bajjikanchal Vikas Party"
##
    [43] "Rajnaitik Vikalp Party"
##
   [44] "Rashtriya Samta Party (Secular)"
    [45] "Bharatiya Bahujan Congress"
##
   [46] "Lok Jan Vikas Morcha"
##
   [47] "Rashtriya mahan Gantantra Party"
##
   [48] "Hindusthan Nirman Dal"
##
   [49] "Rashtriya Lok Samta Party"
##
   [50] "Bhartiya Mitra Party"
##
   [51] "Samajwadi Janata Dal Democratic"
##
    [52] "RPI(A)"
##
    [53] "Aam Adhikar Morcha"
```

```
##
    [54] "Aadarsh Mithila Party"
    [55] "AIFB"
##
    [56] "SP"
##
##
    [57] "Rashtra Sewa Dal"
    [58] "Ambedkar National Congress"
##
    [59] "Asli Deshi Party"
##
    [60] "Rashtriya Jansambhavna Party"
##
##
    [61] "INC"
    [62] "Vikassheel Insaan Party"
##
    [63] "Garib Janshakti Party"
##
    [64] "Proutist Sarva Samaj"
##
    [65] "Janhit Kisan Party"
##
##
    [66] "All India Majlis-E-Ittehadul Muslimeen"
    [67] "AITC"
##
##
   [68] "Rashtravadi Janata Party"
    [69] "Jan Adhikar Party (Loktantrik)"
##
   [70] "Baliraja Party"
##
   [71] "Akhil Bhartiya Mithila Party"
##
    [72] "Purvanchal Janta Party (Secular)"
##
    [73] "Jago Hindustan Party"
##
##
    [74] "Maanavvaadi Janta Party"
    [75] "Bharatiya Jan Kranti Dal (Democratic)"
##
##
    [76] "Bhartiya New Sanskar Krantikari Party"
   [77] "Rashtriya Hind Sena"
##
    [78] "Jan Adhikar Party"
##
    [79] "Sankhyanupati Bhagidari Party"
##
    [80] "Revolutionary Socialist Party of India(Marxis"
##
    [81] "Rashtriya Rashtrawadi Party"
    [82] "Janata Party"
##
    [83] "Aap Aur Hum Party"
##
##
    [84] "Bharat Nirman Party"
##
    [85] "Yuva Krantikari Party"
##
    [86] "National Jagaran Party"
    [87] "Samagra Utthan Party"
##
    [88] "Purvanchal Mahapanchayat"
##
##
    [89] "Bahujan Nyay Dal"
    [90] "RPI"
##
    [91] "Bharat Prabhat Party"
##
    [92] "Rashtriya Ulama Council"
##
   [93] "Loktantrik Jan Swaraj Party"
##
    [94] "Rashtriya Janvikas Party (Democratic)"
##
   [95] "Rashtrawadi Chetna Party"
##
   [96] "Bharatiya Aam Awam Party"
##
   [97] "Apna Kisan Party"
##
   [98] "Vanchit Samaj Party"
    [99] "Rashtriya Sarvjan Vikas Party"
## [100] "Janvadi Party(Socialist)"
## [101] "Akhil Bhartiya Apna Dal"
## [102] "Wazib Adhikar Party"
## [103] "Bhartiya Insan Party"
## [104] "Bahujan Azad Party"
## [105] "Kisan Party of India"
## [106] "Proutist Bloc, India"
## [107] "Bharatiya Samta Samaj Party"
```

```
## [108] "Swatantra Samaj Party"
## [109] "Sanyukt Vikas Party"
## [110] "Rashtriya Sahyog Party"
## [111] "Hind Samrajya Party"
## [112] "Jammu & Kashmir National Panthers Party"
## [113] "Jai Hind Party"
## [114] "Lok Sewa Dal"
## [115] "CPI(M)"
## [116] "Janta Raj Vikas Party"
## [117] "Lok Chetna Dal"
## [118] "Rashtriya Pragati Party"
## [119] "Sapaks Party"
## [120] "Janata Congress"
## [121] "Bhartiya Panchyat Party"
## [122] "Bharat Vikas Morcha"
## [123] "Lok Dal"
## [124] "CPI(ML)L"
## [125] "Samajwadi Janata Party (Rashtriya)"
## [126] "JAP(L)"
## [127] "Samajwadi Forward Bloc"
## [128] "Hindustan Vikas Dal"
## [129] "National Loktantrik Party"
## [130] "Bhartiya Ekta Manch Party"
## [131] "National Tiger Party"
## [132] "Pragatisheel Manav Samaj Party"
## [133] "Rashtriya Ahinsa Manch"
## [134] "Rashtriya Jan-Jagram Morcha"
## [135] "Rashtriya Bahujan Congress Party"
## [136] "Bharatiya Ekta Dal"
## [137] "Krantikari Vikas Dal"
## [138] "VANCHITSAMAJ INSAAF PARTY"
## [139] "Rashtriya Naujawan Dal"
## [140] "Sarvajan Kalyan Loktantrik Party"
## [141] "Loktantrik Janata Party (Secular)"
## [142] "IUML"
## [143] "Peace Party"
## [144] "Jai Maha Bharath Party"
## [145] "Rashtriya Krantikari Samajwadi Party"
## [146] "Ati Picchara party"
## [147] "Bharatiya Inqalab Party"
## [148] "Jharkhand Disom Party"
## [149] "Bhartiya Jantantrik Janata Dal"
## [150] "JKNPP"
## [151] "Kalinga Sena"
## [152] "Bihar Janta Party"
## [153] "National Lokmat Party"
## [154] "GGP"
## [155] "Naya Daur Party"
## [156] "Mahamukti Dal"
## [157] "Lokpriya Samaj Party"
```

```
loksabha_2019$Party <- strtoi(factor(loksabha_2019$Party,levels = part_level,labels = party_labe
l))
loksabha_2014$Party <- strtoi(factor(loksabha_2014$Party,levels = part_level,labels = party_labe
l))</pre>
```

Data Cleaning

```
loksabha_2014$Total.Assets[is.na(loksabha_2014$Total.Assets)] <- "0"
loksabha_2019$Total.Assets[is.na(loksabha_2019$Total.Assets)] <- "0"
loksabha_2014$Liabilities[is.na(loksabha_2014$Liabilities)] <- "0"
loksabha_2019$Liabilities[is.na(loksabha_2019$Liabilities)] <- "0"

loksabha_2019$Total.Assets <- strtoi(gsub("[RsÂ,]","",loksabha_2019$Total.Assets))
loksabha_2019$Liabilities <- strtoi(gsub("[RsÂ,]","",loksabha_2019$Liabilities))
loksabha_2014$Total.Assets <- strtoi(gsub("[RsÂ,]","",loksabha_2014$Total.Assets))
loksabha_2014$Liabilities <- strtoi(gsub("[RsÂ,]","",loksabha_2014$Liabilities))

str(loksabha_2019)
```

```
## 'data.frame':
                  609 obs. of 8 variables:
## $ Constituency : int 1 1 1 1 1 1 1 1 1 ...
## $ Party
                  : int 1 2 2 2 2 2 3 2 4 2 ...
## $ Criminal.Cases: int 3 1 1 0 0 0 0 0 6 1 ...
  $ Education : int 1 1 2 1 3 4 3 5 6 7 ...
                  : int 52 69 31 33 46 36 46 66 50 28 ...
##
   $ Age
## $ Winner
              : int 1000000000...
   $ Total.Assets : int 5010577 1545000 602510 1110500 572000 1240800 6916000 50000 76618366
12421636 ...
## $ Liabilities
                  : int 1159200 119028 0 0 0 0 1000000 0 8566199 300000 ...
```

```
str(loksabha_2014)
```

```
## 'data.frame':
                  610 obs. of 8 variables:
## $ Constituency : int 1 1 1 1 1 1 1 1 1 ...
## $ Party
                  : int 4 3 6 23 37 2 1 122 123 124 ...
## $ Criminal.Cases: int 4000124001...
  $ Education
                 : int 2 2 5 8 5 1 1 1 5 2 ...
##
   $ Age
                  : int 72 29 61 48 35 34 48 46 63 32 ...
  $ Winner
                 : int 1000000000...
##
  $ Total.Assets : int 42368468 813848 856973 14088067 15000 202412 4291479 1475000 10000 14
175 ...
## $ Liabilities
                  : int 1026919 0 7000 5025161 63733 50000 802182 50000 0 0 ...
```

Preparing Data for Training and Normalizing

```
loksabha_2014$Winner <- as.factor(loksabha_2014$Winner)
loksabha_2019$Winner <- as.factor(lok_2019$Winner)
str(loksabha_2014)</pre>
```

```
## 'data.frame': 610 obs. of 8 variables:
## $ Constituency : int 1 1 1 1 1 1 1 1 1 1 1 ...
## $ Party : int 4 3 6 23 37 2 1 122 123 124 ...
## $ Criminal.Cases: int 4 0 0 0 1 2 4 0 0 1 ...
## $ Education : int 2 2 5 8 5 1 1 1 5 2 ...
## $ Age : int 72 29 61 48 35 34 48 46 63 32 ...
## $ Winner : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
## $ Total.Assets : int 42368468 813848 856973 14088067 15000 202412 4291479 1475000 10000 14
175 ...
## $ Liabilities : int 1026919 0 7000 5025161 63733 50000 802182 50000 0 0 ...
```

str(loksabha_2019)

```
## 'data.frame': 609 obs. of 8 variables:
## $ Constituency : int 1 1 1 1 1 1 1 1 1 1 1 ...
## $ Party : int 1 2 2 2 2 2 3 2 4 2 ...
## $ Criminal.Cases: int 3 1 1 0 0 0 0 0 6 1 ...
## $ Education : int 1 1 2 1 3 4 3 5 6 7 ...
## $ Age : int 52 69 31 33 46 36 46 66 50 28 ...
## $ Winner : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
## $ Total.Assets : int 5010577 1545000 602510 1110500 572000 1240800 6916000 50000 76618366
12421636 ...
## $ Liabilities : int 1159200 119028 0 0 0 0 10000000 0 8566199 3000000 ...
```

```
dim(loksabha_2014)
```

```
## [1] 610 8
```

```
dim(loksabha_2019)
```

```
## [1] 609 8
```

```
loksabha_2014 <- na.omit(loksabha_2014)
loksabha_2019 <- na.omit(loksabha_2019)

train_y <- loksabha_2014$Winner
train_x <- subset(loksabha_2014,select=-Winner)

test_y <- loksabha_2019$Winner
test_x <- subset(loksabha_2019,select=-Winner)

q <- preProcess(train_x[3:7],method=c("center", "scale"))
train_x[3:7] <- predict(q,train_x[3:7])
head(train_x)</pre>
### Constituency Party (riminal Cases Education Age Total Assets
```

```
Constituency Party Criminal.Cases Education
##
                                                         Age Total.Assets
## 1
               1
                     4
                            1.51344387 -1.1270267 1.9774464
                                                               0.1553505
## 2
               1
                      3
                           -0.38559578 -1.1270267 -1.3294045
                                                               -0.2296146
## 3
               1
                     6
                          -0.38559578 0.1367144 1.1315078
                                                              -0.2292151
## 4
               1
                    23
                           -0.38559578 1.4004556 0.1317622
                                                               -0.1066412
## 5
                     37
                            0.08916413 0.1367144 -0.8679834
                                                               -0.2370152
                     2
## 6
                            0.56392405 -1.5482738 -0.9448869
                                                               -0.2352790
##
    Liabilities
## 1 -0.04827611
## 2 -0.09043718
## 3 -0.09014979
## 4 0.11587524
## 5 -0.08782056
## 6 -0.08838438
```

```
q <- preProcess(test_x[3:7],method=c("center", "scale"))
test_x[3:7] <- predict(q,test_x[3:7])
head(test_x)</pre>
```

```
Constituency Party Criminal.Cases Education
##
                                                           Age Total.Assets
## 1
                1
                      1
                             1.0575241 -1.4958207 0.46379046
                                                                 -0.2004068
## 2
                1
                      2
                             0.1018825 -1.4958207 1.85327047
                                                                 -0.2346979
## 3
                1
                      2
                             0.1018825 -1.0776269 -1.25262603
                                                                 -0.2440236
## 4
                1
                      2
                            -0.3759384 -1.4958207 -1.08915779
                                                                 -0.2389972
## 5
                      2
                            -0.3759384 -0.6594330 -0.02661425
                1
                                                                 -0.2443255
## 6
                            -0.3759384 -0.2412391 -0.84395543
                                                                 -0.2377079
##
    Liabilities
## 1 -0.05641842
## 2 -0.12314226
## 3 -0.13077754
## 4 -0.13077754
## 5 -0.13077754
## 6 -0.13077754
```

```
train_x[is.na(train_x)==TRUE]
```

```
## numeric(0)

test_x[is.na(test_x)==TRUE]

## numeric(0)

train_y[is.na(train_y)==TRUE]

## factor(0)
## Levels: 0 1

test_y[is.na(test_y)==TRUE]

## factor(0)
## factor(0)
## Levels: 0 1
```

Counting Actual Sits

```
Actual_sits <- loksabha_2019 %>% filter(Winner==1) %>% select(Party)
Actual_sits <- Actual_sits %>% group_by(Party) %>% mutate(count=n())
Actual_sits <- unique(Actual_sits)
Actual_sits$Party <- factor(Actual_sits$Party,levels = party_label,labels = part_level)
Actual_sits
```

KNN Model Implementaion

```
library(caTools)
library(class)
classifier_knn <- knn(train = train_x, test = test_x, cl = train_y, k = 2)
classifier_knn</pre>
```

```
##
             \begin{smallmatrix} 1 \end{smallmatrix} ] \hspace{.1cm} 0 \hspace{.1cm} 1 \hspace{.1cm} 0 \hspace{.1c
          ##
         ## [186] 0 0 0 0 0 0 1 0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0
## [593] 0 1 0 0 0 0 0 0 0 0 1 1 0
## Levels: 0 1
```

Counting Predicted Sits

```
knn_pred_sits <- mutate(test_x,classifier_knn)
knn_pred_sits <- knn_pred_sits %>% filter(classifier_knn==1) %>% select(Party)
knn_pred_sits <- knn_pred_sits %>% group_by(Party) %>% mutate(count=n())
knn_pred_sits <- unique(knn_pred_sits)
knn_pred_sits$Party <- factor(knn_pred_sits$Party,levels = party_label,labels = part_level)
knn_pred_sits</pre>
```

```
## # A tibble: 28 x 2
## # Groups:
               Party [28]
##
      Party
                                     count
##
      <fct>
                                     <int>
   1 Bihar Lok Nirman Dal
##
                                         1
##
   2 BJP
                                         7
   3 BSP
                                         1
   4 RJD
##
                                         1
##
   5 IND
                                         4
   6 Jai Prakash Janata Dal
##
                                         2
##
   7 LJP
                                         3
##
   8 Sathi Aur Aapka Faisala Party
                                         1
                                         2
## 9 Bajjikanchal Vikas Party
## 10 Bharatiya Bahujan Congress
                                         2
## # ... with 18 more rows
```

```
# Confusion Matrix
confusionMatrix(table(test_y, classifier_knn))
```

```
## Confusion Matrix and Statistics
##
##
         classifier_knn
## test_y
            0
                1
##
        0 523 42
##
        1 29 11
##
##
                  Accuracy : 0.8826
                    95% CI: (0.8543, 0.9072)
##
##
       No Information Rate : 0.9124
       P-Value [Acc > NIR] : 0.9947
##
##
##
                     Kappa : 0.1743
##
##
   Mcnemar's Test P-Value: 0.1544
##
               Sensitivity: 0.9475
##
##
               Specificity: 0.2075
            Pos Pred Value : 0.9257
##
            Neg Pred Value : 0.2750
##
                Prevalence: 0.9124
##
            Detection Rate: 0.8645
##
##
      Detection Prevalence : 0.9339
##
         Balanced Accuracy: 0.5775
##
          'Positive' Class : 0
##
##
```