Algorithm

Anandu R

5/31/2020

Importing necessary packages/Libraries

```
invisible(library(dplyr))
invisible(library(lubridate))
invisible(library(caTools))
invisible(library(data.table))
```

Generating the dataset

```
set.seed(1)
speed = round(rnorm(1000,50,15),2)
dist_prev = abs(round(rnorm(1000,2,1),2))
dist_next = abs(round(rnorm(1000,2,1),2))
crowd_curr = rpois(1000,25)
crowd_next = rpois(1000,25)
booked = rpois(1000,40)
schd_time = sample(seq(strptime('01/01/2018',format = "%d/%m/%Y"),
                       strptime('01/01/2019',format = "%d/%m/%Y"),
                       by="hour"), 1000, replace = T)
arr_time = schd_time+(rnorm(1000,300,350)*-1)
on_time = ifelse(difftime(arr_time,schd_time)<=0,1,0)
data = data.frame(crowd_curr,crowd_next,booked,
                  dist_prev, dist_next,speed,
                  schd_time,arr_time,on_time)
head(select(data,crowd_curr,crowd_next,booked,on_time))
```

```
##   crowd_curr crowd_next booked on_time
## 1         28         27     39       1
## 2         26         24     38       1
## 3         31         21     28       1
## 4         20         28     41       1
## 5         27         21     36       1
## 6         23         21     43       1
```

Generating an algorithm to label the datasets

Each record is considered as a bus and the label is the indication given to the bus driver whether to maintain speed, decrease speed, or to increase represented by 0,1,2 respectively

```
indicate = ifelse((
    (crowd_curr<28)&
    (crowd_next>28)&
    (booked>30)&
    (on_time==0)),2,ifelse((
      (crowd_curr>28)&
      (crowd_next<28)&
      (booked<30)&
      (on_time==1)),1,0))
data$indicate = indicate
head(select(data,crowd_curr,booked,on_time,indicate))
```

```
##   crowd_curr booked on_time indicate
## 1         28     39       1        0
## 2         26     38       1        0
## 3         31     28       1        1
## 4         20     41       1        0
## 5         27     36       1        0
## 6         23     43       1        0
```