Algorithm

Anandu R

5/31/2020

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 = rpois(1000,25)
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,dist_prev,dist_next,speed,schd_time,arr_time,on_time)
head(data)
```

```
##   crowd dist_prev dist_next speed           schd_time            arr_time on_time
## 1    28      3.13      1.11 40.60 2018-04-30 11:00:00 2018-04-30 10:56:24       1
## 2    26      3.11      0.08 52.75 2018-01-12 22:00:00 2018-01-12 21:51:50       1
## 3    31      1.13      3.62 37.47 2018-06-28 01:00:00 2018-06-28 00:41:40       1
## 4    20      2.21      2.52 73.93 2018-02-24 22:00:00 2018-02-24 21:50:30       1
## 5    27      2.07      1.94 54.94 2018-12-07 10:00:00 2018-12-07 09:56:16       1
## 6    23      0.34      2.70 37.69 2018-01-10 19:00:00 2018-01-10 18:58:13       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