**kmeans**



**KNN**

> library(readxl)

Warning message:

package ‘readxl’ was built under R version 4.3.3

> data=read\_excel("C:\\Users\\Dell\\Documents\\for knn1.xlsx",col\_names = TRUE)

-

/

/

-

>

> head(data) #Top observations present in the data

# A tibble: 6 × 10

overall total score of individual (ou…¹ physical health scor…² `mental health`

<dbl> <dbl> <dbl>

1 49 60 60

2 60 60 55

3 45 60 50

4 61 20 50

5 78 100 60

6 69 73.3 55

# ℹ abbreviated names: ¹​`overall total score of individual (out of 100)`,

# ²​`physical health score(out of 100)`

# ℹ 7 more variables: `commercial(effect on businesses)` <dbl>,

# education <dbl>, `career opportunities(positive impact)` <dbl>,

# `cyber crimes` <dbl>, `communication skills` <dbl>,

# `political decisions` <dbl>, Gender <chr>

> dim(data) #Check the dimensions of the data

[1] 460 10

> summary(data) #Summarise the data

overall total score of individual (out of 100)

Min. :21.00

1st Qu.:56.00

Median :61.00

Mean :61.19

3rd Qu.:66.00

Max. :93.00

physical health score(out of 100) mental health

Min. : 20.00 Min. :25.00

1st Qu.: 40.00 1st Qu.:50.00

Median : 53.33 Median :55.00

Mean : 55.43 Mean :56.14

3rd Qu.: 66.67 3rd Qu.:65.00

Max. :100.00 Max. :90.00

commercial(effect on businesses) education

Min. : 20.00 Min. : 20.0

1st Qu.: 50.00 1st Qu.: 60.0

Median : 60.00 Median : 70.0

Mean : 59.17 Mean : 68.3

3rd Qu.: 70.00 3rd Qu.: 80.0

Max. :100.00 Max. :100.0

career opportunities(positive impact) cyber crimes communication skills

Min. : 20.00 Min. : 20.00 Min. : 20.00

1st Qu.: 53.33 1st Qu.: 60.00 1st Qu.: 60.00

Median : 60.00 Median : 70.00 Median : 70.00

Mean : 63.87 Mean : 67.93 Mean : 65.02

3rd Qu.: 73.33 3rd Qu.: 80.00 3rd Qu.: 80.00

Max. :100.00 Max. :100.00 Max. :100.00

political decisions Gender

Min. : 20.00 Length:460

1st Qu.: 50.00 Class :character

Median : 60.00 Mode :character

Mean : 60.26

3rd Qu.: 70.00

Max. :100.00

> install.packages("class")

Installing package into ‘C:/Users/Dell/AppData/Local/R/win-library/4.3’

(as ‘lib’ is unspecified)

trying URL 'https://cloud.r-project.org/bin/windows/contrib/4.3/class\_7.3-22.zip'

Content type 'application/zip' length 97449 bytes (95 KB)

downloaded 95 KB

package ‘class’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\Dell\AppData\Local\Temp\Rtmp4wIysp\downloaded\_packages

> library(class)

Warning message:

package ‘class’ was built under R version 4.3.3

>

> #Normalization

> normalize=function(x) {

+ return ((x - min(x)) / (max(x) - min(x))) }

> norm=as.data.frame(lapply(data[,1:9], normalize))

>

> set.seed(123)

> data\_spl=sample(1:nrow(norm),size=nrow(norm)\*0.7,replace = FALSE)

> data\_spl

[1] 415 179 14 195 426 306 118 299 229 244 458 374 153 90 91 256 197 446

[19] 441 348 137 355 328 26 7 456 440 254 211 420 78 81 43 359 373 332

[37] 143 32 109 263 393 330 23 411 309 135 416 417 224 166 217 290 69 72

[55] 76 63 141 210 353 347 448 294 277 41 447 316 223 16 116 94 262 235

[73] 86 342 39 159 240 209 449 455 34 4 13 408 243 308 278 89 25 291

[91] 368 286 385 121 110 158 64 199 67 151 335 85 165 136 51 74 178 236

[109] 98 419 214 127 212 174 273 302 310 232 376 360 280 113 107 361 154 102

[127] 255 160 155 5 326 272 340 288 398 55 238 252 386 344 434 226 48 77

[145] 83 184 322 196 257 168 369 391 329 20 457 164 52 22 177 42 84 11

[163] 367 427 194 338 317 198 249 200 333 250 431 33 40 10 371 125 265 421

[181] 186 61 323 152 54 437 389 185 253 115 285 276 205 267 372 304 215 409

[199] 57 105 134 129 218 106 339 438 429 29 275 27 213 436 188 245 336 397

[217] 442 260 242 222 370 93 145 192 358 390 148 163 410 161 66 379 439 225

[235] 133 117 266 357 325 343 297 45 146 170 384 387 363 189 261 176 324 173

[253] 191 53 104 207 405 350 239 403 2 345 130 404 24 203 293 402 208 406

[271] 298 378 221 312 413 112 364 80 36 35 87 362 318 111 418 320 279 314

[289] 31 73 182 303 122 247 92 202 283 428 383 295 287 395 140 352 21 237

[307] 414 181 30 454 459 271 190 82 96 341 258 382 444 246 394 433

> train2=data[data\_spl,] # 70% training data

> test2=data[-data\_spl,] # remaining 30% test data

> View(train2)

>

> train\_labels=data[data\_spl,]

> test\_labels=data[-data\_spl,]

> knn\_pred=knn(train=train2[,-10], test=test2[,-10],train\_labels$Gender,k=2)

>

>

> confusionmatrix=table(knn\_pred,test\_labels$Gender)

> confusionmatrix

knn\_pred Female Male

Female 42 29

Male 36 30

Other 1 0

> accuracy=sum(diag(confusionmatrix)/nrow(test\_labels))

> accuracy

[1] 0.5217391