CSI – 695 Final Project

PRESENTED BY: AJAY KULKARNI (G01024139)

Outline

- Data import from MySQL to R using "RMySQL".
- Data cleaning and processing.
- Results of kNN and kd-tree.
- Comparison of kNN and kd-tree based on execution time.

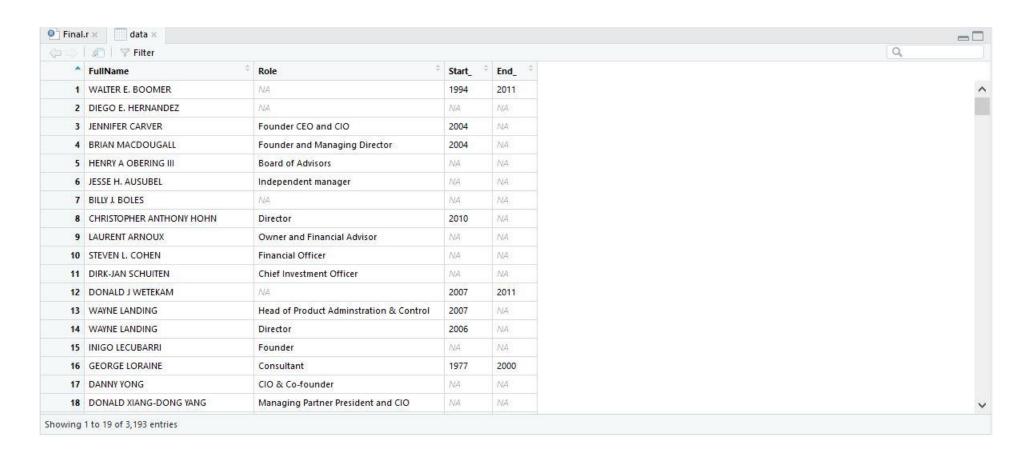
Data import

- The important attributes considered for finding similarity are
 - Role
 - Start Year
 - End Year

• The data is imported into R using the following query

"SELECT people.FullName, peopleandcompanies.Role, peopleandcompanies.Start_, peopleandcompanies.End_ FROM people INNER JOIN peopleandcompanies ON people.Dir_ID=peopleandcompanies.People_ID"

Data import



Data cleaning and processing

• NAs in the data are removed using na.omit() function in R and "id" column has been added as an identification for each data point.

	20	8	Filter				Q
*	id		FullName	Role	Start_		
16		1	GEORGE LORAINE	Consultant	1977	2000	
27		2	PETER SANDYS	Managing Director	1992	1998	
28		3	PETER SANDYS	Director	1989	1998	
35		4	PATRICK LITTAYE	Managing Director	1977	1985	
58		5	LISA ALEXANDER	CFO	2004	2006	
60		6	DAVID J TEAL	Director of Business Development	1992	2011	
73		7	MATTHEW GOLDSTEIN	President	1998	1999	
112		8	MICHAEL M DUNN	President	2006	2011	
121		9	JOHN A. DUBIA	EVP	1999	2011	
133		10	JOHN D HOPPER JR	President	2005	2011	
140		11	JAMES J. LINDSAY	Board of Advisors	1990	2011	
144		12	PATRICK K GAMBLE	President	2001	2011	
151		13	WALTER E BUCHANAN III	SVP - Engineering and IT	2006	2011	
159		14	RONALD H. GRIFFITH	Board of Advisors	1997	2011	
165		15	JEAN KELLER	CEO	2005	2011	
168		16	TED UEMAE	Chairman	2002	2006	
172		17	PETER BEGLINGER	President. Chairman. Director	2001	2011	
182		18	COLBY M. BROADWATER	president	2006	2011	

Data cleaning and processing

- After performing data cleaning data contains 461 rows and 5 columns.
- One hot encoding has been performed on all the selected attributes.



Results of kNN and kd-tree

- For kNN, the binary similarity is used to find the similarity between different objects.
- KNN and kd-tree are tested on following examples.

Id	K
5	1
30	5
100	10
243	20
343	50

Example – 1 (id = 5, k = 1)

• KNN result and similarity plot

```
Console Terminal ×

> myknn(5,1)

[1] "Execution time"

8.69 sec elapsed

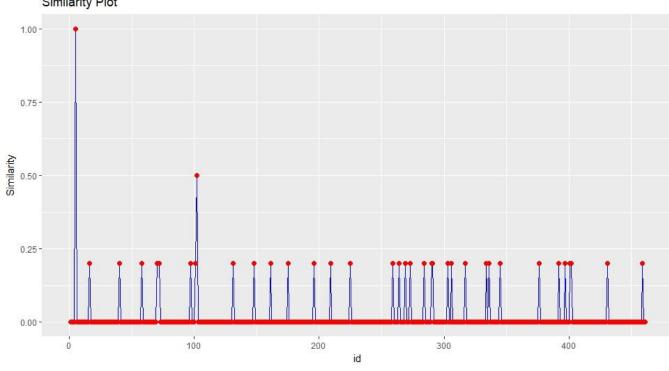
[1] "Binary Similarity"

[1] 5 102

[1] "LISA ALEXANDER" "RICHARD BATTEY"

[1] "CFO" "CFO"

[1] 1.0 0.5
```



Example – 1 (id = 5, k = 1)

KNN

```
Console Terminal x

> myknn(5,1)
[1] "Execution time"
8.69 sec elapsed
[1] "Binary Similarity"
[1] 5 102
[1] "LISA ALEXANDER" "RICHARD BATTEY"
[1] "CFO" "CFO"
[1] 1.0 0.5
>
```

Kd-tree

```
Console Terminal ×

C:/Users/srishti/Desktop/ajay/ > mykdtree(5,1)

[1] "kd-tree results"

[1] "Execution Time"

0.12 sec elapsed

[1] 5 102

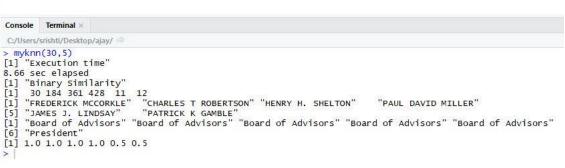
[1] "LISA ALEXANDER" "RICHARD BATTEY"

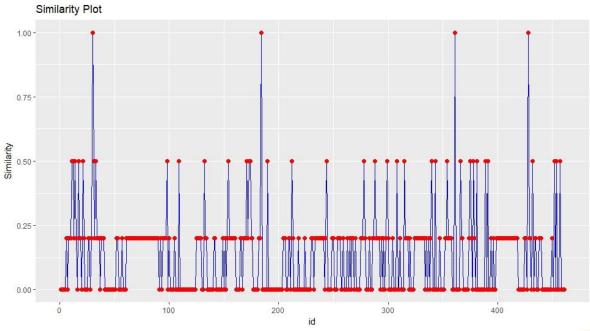
[1] "CFO" "CFO"

[1] 0 2
```

Example – 2 (id = 30, k = 5)

• KNN result and similarity plot





Example – 2 (id = 30, k = 5)

KNN

```
C:/Users/srishti/Desktop/ajay/ > myknn(30,5)
[1] "Execution time"
8.66 sec elapsed
[1] "Binary Similarity"
[1] 30 184 361 428 11 12
[1] "FREDERICK MCCORKLE" "CHARLES T ROBERTSON" "HENRY H. SHELTON" "PAUL DAVID MILLER"
[5] "JAMES J. LINDSAY" "PATRICK K GAMBLE"
[1] "Board of Advisors" "Board of Advisors" "Board of Advisors" "Board of Advisors" "Board of Advisors"
[6] "President"
[1] 1.0 1.0 1.0 1.0 0.5 0.5
```

Kd-tree

```
Console Terminal ×

C:/Users/srishti/Desktop/ajay/ >> mykdtree(30,5)

[1] "kd-tree results"
[1] "Execution Time"

0.13 sec elapsed
[1] 30 428 361 184 173 381

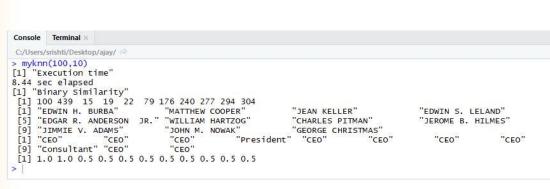
[1] "FREDERICK MCCORKLE" "PAUL DAVID MILLER" "HENRY H. SHELTON" "CHARLES T ROBERTSON"

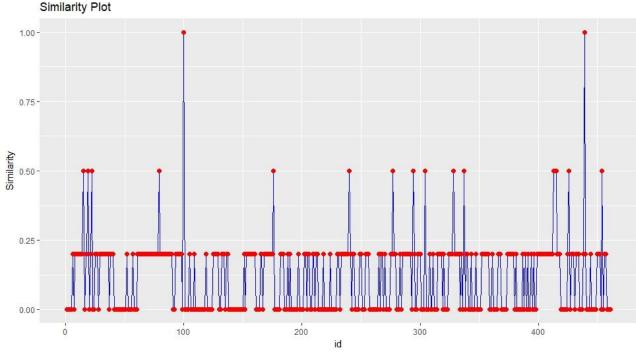
[5] "ARTHUR J LICHTE" "HENRY A OBERING III"

[1] "Board of Advisors" "Board of A
```

Example -3 (id = 100, k = 10)

• KNN result and similarity plot





Example -3 (id = 100, k = 10)

KNN

```
Console Terminal ×
C:/Users/srishti/Desktop/ajay/
> myknn(100,10)
[1] "Execution time"
8.44 sec elapsed
[1] "Binary Similarity"
[1] 100 439 15 19 22 79 176 240 277 294 304
[1] "EDWIN H. BURBA"
                               "MATTHEW COOPER"
                                                         "JEAN KELLER"
                                                                                   "EDWIN S. LELAND"
[5] "EDGAR R. ANDERSON JR." "WILLIAM HARTZOG"
                                                         "CHARLES PITMAN"
                                                                                  "JEROME B. HILMES"
    "JIMMIE V. ADAMS"
                               "JOHN M. NOWAK"
                                                         "GEORGE CHRISTMAS"
                                                                                     "CEO"
[1] "CEO"
                  "CEO"
                                "CEO"
                                             "President" "CEO"
                                                                        "CEO"
                                                                                                   "CEO"
[9] "Consultant" "CEO"
                               "CEO"
[1] 1.0 1.0 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
```

Kd-tree

```
Console Terminal ×
C:/Users/srishti/Desktop/ajay/
> mykdtree(100,10)
[1] "kd-tree results"
[1] "Execution Time"
0.17 sec elapsed
 [1] 100 439 328 15 454 79 337 22 304 294 240
 [1] "EDWIN H. BURBA"
                               "MATTHEW COOPER"
                                                         "WILLIAM J. FALLON"
                                                                                   "JEAN KELLER"
 [5] "JOHN G. COBURN"
                               "WILLIAM HARTZOG"
                                                         "SAMUEL E. EBBESEN"
                                                                                   "EDGAR R. ANDERSON JR."
 [9] "GEORGE CHRISTMAS"
                               "JOHN M. NOWAK"
                                                         "JEROME B. HILMES"
 [1] "CEO" "CEO"
 [1] 0 0 2 2 2 2 2 2 2 2 2 2
```

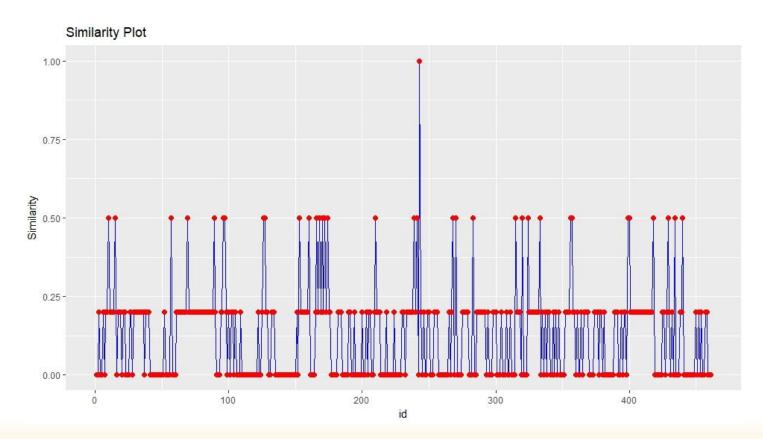
Example – 4 (id = 243, k = 20)

KNN result

```
Console Terminal ×
C:/Users/srishti/Desktop/ajay/
> myknn(243,20)
[1] "Execution time"
8.23 sec elapsed
[1] "Binary Similarity"
 [1] 243 10 15 57 69
                            96 97 126 127 153 160 166 168 170 172 174 210 239 241 268
 [1] "WILLIAM L. NYLAND"
                           "JOHN D HOPPER JR"
                                                  "JEAN KELLER"
                                                                        "CARL G. O'BERRY"
 [5] "JOHN P. JUMPER"
                           "DONALD G COOK"
                                                  "WILLIAM S. WALLACE"
                                                                        "GREGORY G. JOHNSON"
 [9] "HARRY RADUEGE"
                           "PAUL J. KERN"
                                                  "JOHN H. TILELLI"
                                                                        "DENNIS J. REIMER"
                           "PHILIP M. BALISLE"
                                                  "CHARLES MCCAUSLAND"
[13] "MARVIN D. BRAILSFORD"
                                                                        "BARRY R. MCCAFFREY"
[17] "MICHAEL A HOUGH"
                           "CARL E. MUNDY"
                                                  "JOHN J. GROSSENBACHER" "RICHARD D. HEARNEY"
[21] "RICHARD A. HACK"
[1] "Director"
                         "President"
                                            "CEO"
                                                                "Director"
 [5] "Associate"
                         "Board of Directors"
                                            "Director"
                                                                "Director"
[9] "Senior Counselor"
                        "Senior Counselor"
                                                                "Director"
                                            "COO"
[13] "Director"
                                            "Director"
                                                                "Director"
[17] "Board of Advisors"
                        "Director"
                                            "Director"
                                                                "Director"
[21] "SVP - Operations"
>
```

Example – 4 (id = 243, k = 20)

• KNN similarity plot



Example – 4 (id = 243, k = 20)

KNN

```
Console Terminal
C:/Users/srishti/Desktop/ajay/
> myknn(243,20)
[1] "Execution time"
8.23 sec elapsed
[1] "Binary Similarity"
 [1] 243 10 15 57 69 89 96 97 126 127 153 160 166 168 170 172 174 210 239 241 268
 [1] "WILLIAM L. NYLAND"
                           "JOHN D HOPPER JR"
                                                  "JEAN KELLER"
                                                                        "CARL G. O'BERRY"
    "JOHN P. JUMPER"
                           "DONALD G COOK"
                                                  "WILLIAM S. WALLACE"
                                                                        "GREGORY G. JOHNSON"
    "HARRY RADUEGE"
                                                  "JOHN H. TILELLI"
                                                                        "DENNIS J. REIMER"
                           "PAUL J. KERN"
[13] "MARVIN D. BRAILSFORD"
                           "PHILIP M. BALISLE"
                                                  "CHARLES MCCAUSLAND"
                                                                        "BARRY R. MCCAFFREY"
[17] "MICHAEL A HOUGH"
                           "CARL E. MUNDY"
                                                  "JOHN J. GROSSENBACHER" "RICHARD D. HEARNEY"
[21] "RICHARD A. HACK"
 [1] "Director"
                        "President"
                                            "CEO"
                                                                "Director"
    "Associate"
                        "Board of Directors" "Director"
                                                                "Director"
                                                                "Director"
     "Senior Counselor"
                        "Senior Counselor"
                                            "coo"
    "Director"
                                            "Director"
                                                                "Director"
                        "VP"
[17] "Board of Advisors" "Director"
                                            "Director"
                                                                "Director"
[21] "SVP - Operations"
```

Kd-tree

```
Console Terminal
C:/Users/srishti/Desktop/ajay/
> mykdtree(243,20)
[1] "kd-tree results"
[1] "Execution Time"
0.14 sec elapsed
 [1] 243 333 96 283 324 97 239 270 160 241 172 210 57 320 170 166 168 268 357 126 127
[1] "WILLIAM L. NYLAND"
                             "BRUCE CARLSON"
                                                     "WILLIAM S. WALLACE"
                                                                             "MARTIN PAUL TOLCHER"
 [5] "GERALD L. HOEWING"
                             "GREGORY G. JOHNSON"
                                                                            "MAXWELL C BAILEY"
                                                     "JOHN J. GROSSENBACHER"
 [9] "DENNIS J. REIMER"
                             "RICHARD D. HEARNEY"
                                                     "BARRY R. MCCAFFREY"
                                                                             "CARL E. MUNDY"
[13] "CARL G. O'BERRY"
                             "JOHN R. DAILEY"
                                                     "CHARLES MCCAUSLAND"
                                                                            "MARVIN D. BRAILSFORD"
[17] "PHILIP M. BALISLE"
                            "RICHARD A. HACK"
                                                     "BRIAN A ARNOLD"
                                                                             "HARRY RADUEGE"
[21] "PAUL J. KERN"
 [1] "Director"
                        "Director"
                                          "Director"
                                                              "Director"
                                                                                "Director"
[6] "Director"
                        "Director"
                                          "Director"
                                                             "Director"
                                                                                "Director"
[11] "Director"
                        "Director"
                                          "Director"
                                                              "Director"
                                                                                "Director"
                        "VP"
                                          "SVP - Operations" "Space Systems"
[16] "Director"
                                                                                "Senior Counselor"
[21] "Senior Counselor"
```

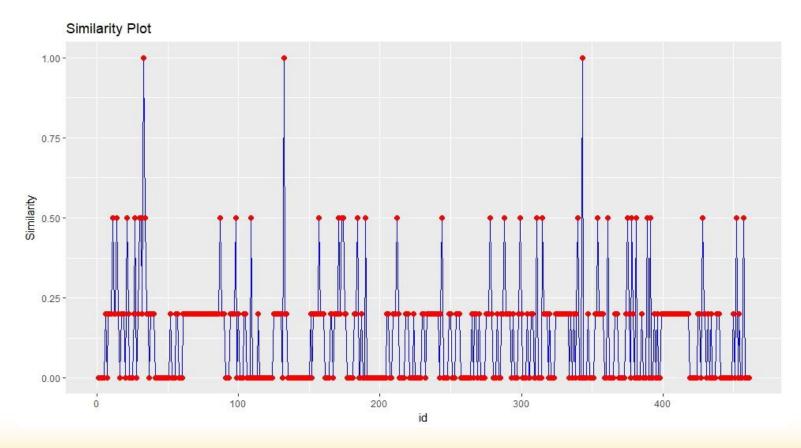
Example -5 (id = 343, k = 50)

KNN result

```
Console Terminal
C:/Users/srishti/Desktop/ajay/
> myknn(343,50)
[1] "Execution time"
8.51 sec elapsed
[1] "Binary Similarity"
[1] 33 132 343 11 14 21 27 30 31 34 87 98 109 157 171 173 174 184 190 212 244 278 288 299 311 315 340 354 361 375 378 381 389 391
[35] 428 452 457 6 8 9 10 12 13 15 17 18 19 22 26 29 32
[1] "HARRY E. SOYSTER"
                                "ALFRED M. GRAY"
                                                            "CHARLES B. EICHELBERGER"
                                                                                       "JAMES J. LINDSAY"
                                                                                       "FREDERICK MCCORKLE"
 [5] "RONALD H. GRIFFITH"
                                "JOHN A GORDON'
                                                           "CARL SMITH"
 [9] "TERRANCE T. ETNYRE"
                                "DONALD W. JONES"
                                                           "LEN WISHART"
                                                                                       "MICHAEL J. WILLIAMS"
                                                           "RICHARD E HAWLEY"
                                "DIEGO E. HERNANDEZ"
                                                                                       "ARTHUR J LICHTE"
[13] "WILLIAM C. BOWES"
[17] "MICHAEL A HOUGH"
                                "CHARLES T ROBERTSON"
                                                           "CHARLES R. LARSON"
                                                                                       "HENRY H. MAUZ"
[21] "EMIL BEDARD"
                                "JOHN M. SHALIKASHVILI"
                                                                                       "HAROLD W. GEHMAN"
                                                           "CARLISLE TROST"
[25] "\"EG \"\"BUCK\"\" SHUBER\""
                                "FRANK L. BOWMAN"
                                                                                       "MICHAEL W. HAGEE'
                                                           "FREDERICK M. FRANKS"
[29] "HENRY H. SHELTON"
                                "STEVEN W. BOUTELLE"
                                                           "GARY MCKISSOCK"
                                                                                       "HENRY A OBERING III"
[33] "LYLE G. BIEN"
                                "GORDON R. SULLIVAN"
                                                           "PAUL DAVID MILLER"
                                                                                       "THOMAS H. COLLINS"
    "OTTO J. GUENTHER"
                                "DAVID J TEAL"
                                                           "MICHAEL M DUNN"
                                                                                       "JOHN A. DUBIA"
    "JOHN D HOPPER JR"
                                "PATRICK K GAMBLE"
                                                           "WALTER E BUCHANAN III"
                                                                                       "JEAN KELLER"
                                "COLBY M. BROADWATER"
    "PETER BEGLINGER"
                                                           "EDWIN S. LELAND"
                                                                                       "EDGAR R. ANDERSON JR."
[49] "MICHAEL W PETERSON"
                                "THOMAS R. WILSON"
                                                           "RICHARD I. NEAL"
 [1] "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
 [4] "Board of Advisors"
                                     "Board of Advisors"
                                                                       "Board of Advisors"
    "Chairman"
                                     "Board of Advisors"
                                                                       "Board of Advisors"
[10] "Regional directors"
                                      "Associate"
                                                                       "Board of Advisors"
[13] "Board of Advisors"
                                      "Principle"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
    "Chair of trustees"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
    "Board of Advisors"
                                      "Board of Advisors"
                                                                       "Board of Advisors"
    "Board of Advisors"
                                      "Director of Business Development"
                                                                      "President"
[40] "EVP"
                                      "President"
                                                                       "President"
                                      "CEO"
    "SVP - Engineering and IT"
                                                                       "President. Chairman. Director"
                                                                       "CEO"
    "president"
                                      "President"
[49]
    "VP"
                                      "SVP"
                                                                       "President"
```

Example -5 (id = 343, k = 50)

• KNN similarity plot



Example -5 (id = 343, k = 50)

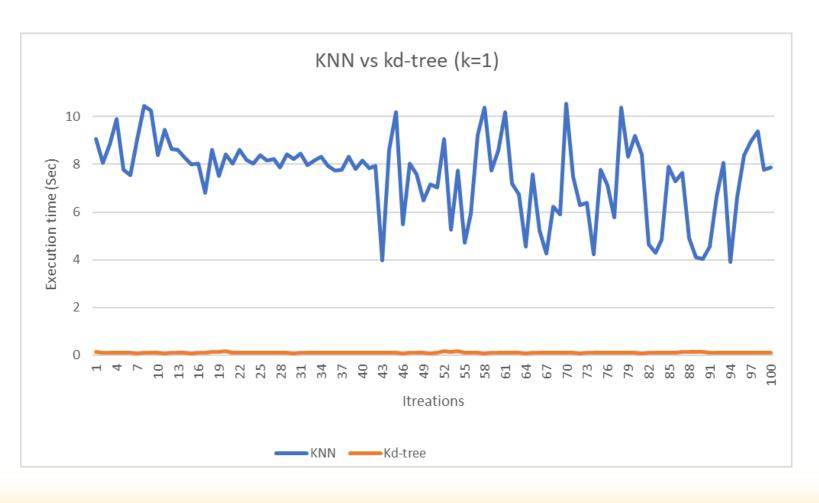
• Kd-tree result

```
Console Terminal
C:/Users/srishti/Desktop/ajay/
> mykdtree(343,50)
[1] "kd-tree results"
[1] "Execution Time"
0.16 sec elapsed
[1] 343 132 33 173 381 31 354 452 174 315 244 109 378 98 30 428 361 184 21 299 389 171 190 14 278 457 375 391 340 212 11 288 34 157
[35] 27 311 87 275 250 330 292 287 158 86 154 26 276 347 331 286 168
[1] "CHARLES B. EICHELBERGER"
                                  "ALFRED M. GRAY"
                                                               "HARRY E. SOYSTER"
                                                                                           "ARTHUR J LICHTE"
[5] "HENRY A OBERING III"
                                  "TERRANCE T. ETNYRE'
                                                               "MICHAEL W. HAGEE"
                                                                                           "THOMAS H. COLLINS"
[9] "MICHAEL A HOUGH"
                                  "FRANK L. BOWMAN"
                                                               "EMIL BEDARD"
                                                                                           "WILLIAM C. BOWES"
[13] "GARY MCKISSOCK"
                                                               "FREDERICK MCCORKLE'
                                                                                           "PAUL DAVID MILLER"
                                  "MICHAEL J. WILLIAMS"
                                  "CHARLES T ROBERTSON"
[17] "HENRY H. SHELTON"
                                                               "JOHN A GORDON"
                                                                                           "HAROLD W. GEHMAN"
[21] "LYLE G. BIEN"
                                  "RICHARD E HAWLEY"
                                                               "CHARLES R. LARSON"
                                                                                           "RONALD H. GRIFFITH"
[25] "JOHN M. SHALIKASHVILI"
                                  "OTTO J. GUENTHER"
                                                               "STEVEN W. BOUTELLE"
                                                                                           "GORDON R. SULLIVAN"
                                                               "JAMES J. LINDSAY"
[29] "FREDERICK M. FRANKS"
                                  "HENRY H. MAUZ"
                                                                                           "CARLISLE TROST'
                                                                                           "\"EG \"\"BUCK\"\" SHUBER\""
[33] "DONALD W. JONES"
                                  "DIEGO E. HERNANDEZ"
                                                               "CARL SMITH"
[37] "LEN WISHART"
                                  "JAMES J. LOVELACE"
                                                               "JOHN H CAMPBELL"
                                                                                           "JOHN J. MAZACH"
[41] "JAMES L JAMERSON"
                                                               "MALCOM B. ARMSTRONG"
                                  "MARVIN R ESMOND"
                                                                                           "CARLES OTSTOTT"
[45] "RICHARD T. SWOPE"
                                  "MICHAEL W PETERSON"
                                                              "RICHARD A. CODY"
                                                                                           "JOHN CASTELLAW"
[49] "EVAN M. CHANIK"
                                  "DAVID J. KELLEY"
                                                              "PHILIP M. BALISLE"
[1] "Board of Advisors"
                                              "Board of Advisors"
                                                                                      "Board of Advisors"
    "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[7] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
                                              "Board of Advisors"
                                                                                      "Board of Advisors"
[10] "Board of Advisors"
[13] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[16] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[19] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[22] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[25] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[28] "Board of Advisors"
                                             "Board of Advisors"
                                                                                      "Board of Advisors"
[31] "Board of Advisors"
                                              "Board of Advisors"
                                                                                      "Regional directors"
[34] "Principle"
                                             "Chairman"
                                                                                      "Chair of trustees"
[37] "Associate"
                                             "VP International Programs"
                                                                                      "VP Government Relations"
[40] "VP Business Development"
                                             "VP Business Development"
                                                                                      "VP Air force programs"
[43] "VP - Operations"
                                             "VP - Ground Systems and Transportation'
                                                                                      "VP - Air Force Programs"
[46] "VP"
                                                                                      "VP"
[49] "VP"
                                             "VP"
```

KNN and kd-tree comparison

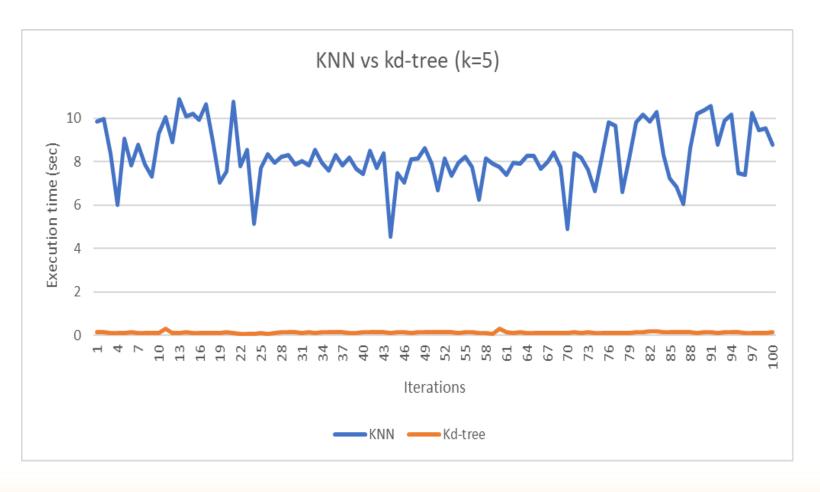
- KNN and kd-tree are compared based on execution time.
- KNN and kd-tree are compared for k = 1, 5, 10, 20 and 50.
- 100 queries were executed for every value of k and average execution time was calculated.

KNN and kd-tree comparison (k = 1)



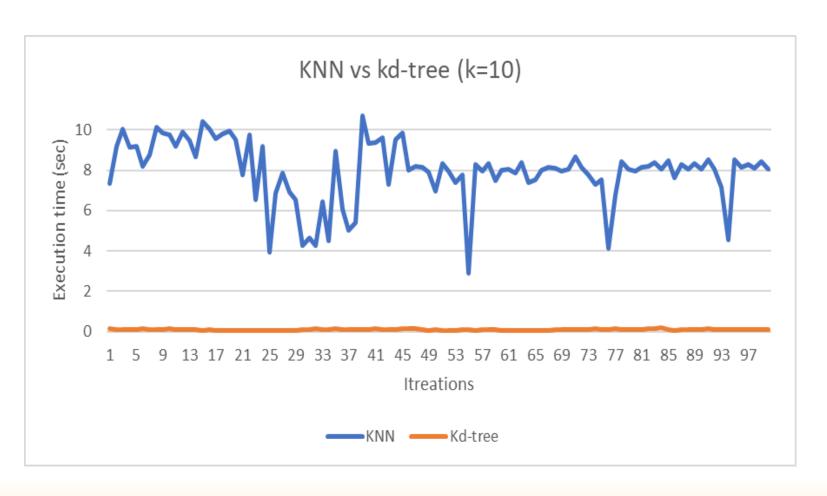
- Average execution time
 - KNN = 7.55 sec
 - Kd-tree = 0.10 sec

KNN and kd-tree comparison (k = 5)



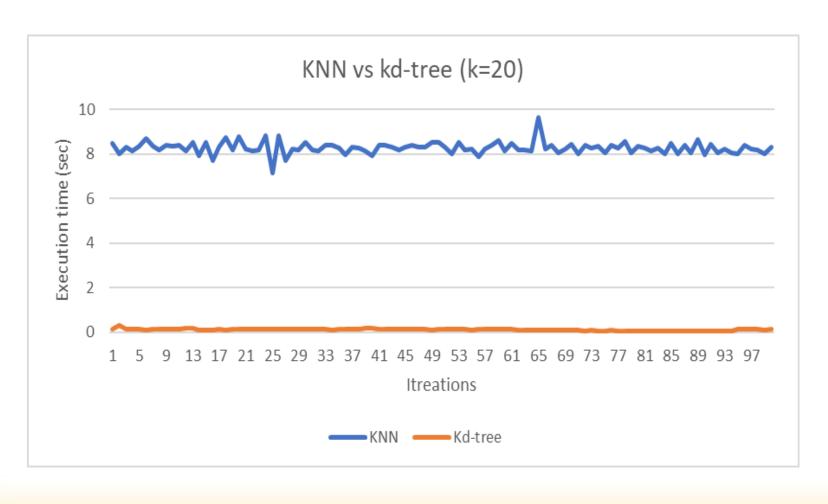
- Average execution time
 - KNN = 8.30 sec
 - Kd-tree = 0.12 sec

KNN and kd-tree comparison (k = 10)



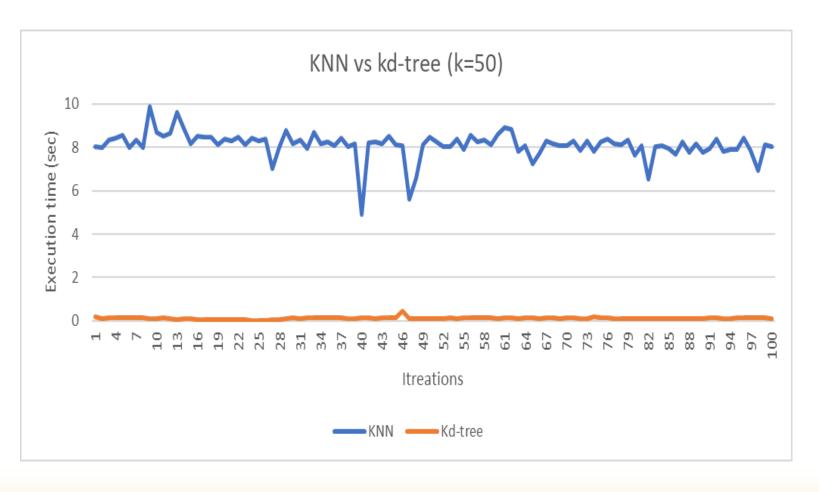
- Average execution time
 - KNN = $7.96 \sec$
 - Kd-tree = 0.11 sec

KNN and kd-tree comparison (k = 20)



- Average execution time
 - KNN = 8.27 sec
 - Kd-tree = 0.11 sec

KNN and kd-tree comparison (k = 50)



- Average execution time
 - KNN = 8.12 sec
 - Kd-tree = 0.12 sec

Conclusion

- Kd-tree is more accurate as compared to kNN.
- Kd-tree is efficient for execution as compared to kNN.
 - For k = 1 kd-tree is 74.5% faster than kNN
 - For k = 5 kd-tree is 68.16% faster than kNN
 - For k = 10 kd-tree is 71.36% faster than kNN
 - For k = 20 kd-tree is 74.18% faster than kNN
 - For k = 50 kd-tree is 66.66% faster than kNN

References

- https://cran.r-project.org/web/packages/RMySQL/RMySQL.pdf
- https://cran.r-project.org/web/packages/proxy/proxy.pdf
- https://cran.r-project.org/web/packages/tictoc/tictoc.pdf
- https://cran.r-project.org/web/packages/rflann/rflann.pdf
- https://www.r-bloggers.com/accessing-mysql-through-r/
- https://stackoverflow.com/questions/24142576/one-hot-encoding-in-r-categorical-to-dummy-variables
- https://stackoverflow.com/questions/6262203/measuring-function-execution-time-in-r

Thank You