COMSATS University Islamabad (Lahore Campus)



Assignment <5> FALL 2022

| Student Name: | | | | Registration N | | SP20-BCS-093 | | | |
|--------------------|-----------------|-----------------------------|---------|--------------------------|------|--------------|---------------|---|--|
| Due Date: | 30-12-2022 | | | Maximum Ma | rks: | 10 | | | |
| Semester: | 6 th | 6 th Batch: SP20 | | Section: | C | Date: | 25-12-2022 | 2 | |
| Course Instructor: | Muhammad S | Sharjeel | | Programme Name: BS Compu | | | ter Science | | |
| Course Title: | Introduction | to Data S | Science | Course Code: | | CSC461 | Credit Hours: | 3 | |

Solve the following two questions manually as well as implement the solution using Python. Submit both solutions.

- Q1. Compute the BoW model, TF model, and IDF model for each of the terms in the following three sentences. Then calculate the TF.IDF values.
- S1 "sunshine state enjoy sunshine"
- S2 "brown fox jump high, brown fox run"
- S3 "sunshine state fox run fast"

Answer:

MANUALLY

| | | | 17. | 22 11 1 | | | | | | an or decide | |
|--|--|--|--|--|-------------------------------------|--|-------|--------|---------------------|--------------|--------|
| | | | AND A STATE OF THE RESIDENCE | 10 | 0# | 1):- | | | | | - |
| | | described and the second secon | | | | California September | | | | | |
| => | RO | K= W | | and the contract of the contra | | and adult of the last of the l | | | | | |
| | -00 | | entariaren eta esta entre entr | CONTRACTOR OF STREET | or to accord depth personnel office | war and the state of the sequence of | | | 1 | | |
| | | pronr | chion | 1 | fast | Cox | high | jump | run | 1 | |
| | Si | 0 | | | 0 | 0 | 0 | 0 | 0 | - | |
| | 52 | 2 | 0 | | 0 | 2 | 1 | 1 | | | - |
| Marie Company of the | 53 | 0 | 0 | | | | 0 | 0 | | | |
| | | CONTRACTOR OF THE PROPERTY OF | and the same are mounted as a second | | | , | | | | | |
| | | state | sw | rsh | ine | | | | mark wheeler Amplem | | ****** |
| | Sı | 1 | | 2 | | | | - 11 | | _ | |
| | Sz | 0 | | 0 | | | | | |) | |
| | 53 | l l | | 1 | | | 111 | | | | |
| | | | | | | | | | | | |
| => | Term- | requen | CH => | | | | | | | | |
| THE REAL PROPERTY AND ADDRESS OF THE PARTY AND | - | | 4 | -, | 4. | i xii | | | | | |
| - | brown | enjoy | fast | fox | high | Link | 00 YU | in sto | te | w | shine |
| S, | 0 | 1/4 | 0 | 0 | 0 | 0 | C |) 1/1 | 1 | 1 | 12 |
| Sz | 2/7 | 0 | 0 | 2/7 | 117 | 1/7 | 1/ | 7 0 |) | | 0 |
| S3 | 0 | 0 | 115 | 1/5 | 0 | 0 | . 1/ | 5 1/3 | 5 | | 1/5 |
| Commission of the Party of the | MANUFACTURE OF THE RESIDENCE OF THE PERSON O | Manager of the State of the Sta | STATE OF THE PARTY | | | | | | | | - |

| =77 | IDF: | > | | - | | | | | |
|----------------|-------|-------|-------|-------|-------|---|-------|-----------|-----------|
| | brown | enion | fast | Cox | high | iumo | vun | State | Sunchine |
| IOF | | | | | | | | log (3/2) | log (3/2) |
| | 0.47 | 6.47 | 0.43 | 70.17 | 0.47 | 0.47 | 0.17 | 0.17 | 0.17 |
| -> | TF-IC |)e =5 | | | | | | | |
| install | T. T. | T - | | | | *************************************** | | | |
| | brown | Crion | fast | fox | high | jungo | run | state | sunshine |
| SI | O | 0.119 | 0 | 0 | 6 | 0 | 0 | 0.044 | 0.088 |
| Sz | 0.23 | O | 0 | 0.08 | 0-119 | 0.119 | 0.044 | 0 | 0 |
| S ₃ | 0 | 0 | 0.119 | 0.04 | 0 | .0 | 0.044 | o our | 0.044 |
| F. S. | | | | | | | | | |

BY CODE

BOW:

| | brown | enjoy | fast | fox | high | jump | run | state | sunshine |
|----|-------|-------|------|-----|------|------|-----|-------|----------|
| s1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| s2 | 2 | 0 | 0 | 2 | 1 | 1 | 1 | 0 | 0 |
| s3 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| | | | | | | | | | |

TF:

| | brown | enjoy | fast | fox | high | jump | run | state | sunshine | |
|----|-------|-------|------|-----|------|------|-----|-------|----------|--|
| s1 | 0 | 1/4 | 0 | 0 | 0 | 0 | 0 | 1/4 | 1/2 | |
| s2 | 2/7 | 0 | 0 | 2/7 | 1/7 | 1/7 | 1/7 | 0 | 0 | |
| s3 | 0 | 0 | 1/5 | 1/5 | 0 | 0 | 1/5 | 1/5 | 1/5 | |
| | | | | | | | | | | |

IDF:

| | brown | enjoy | fast | fox | high | jump | run | state | sunshine |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| IDF's | 0.477121 | 0.477121 | 0.477121 | 0.176091 | 0.477121 | 0.477121 | 0.176091 | 0.176091 | 0.176091 |

TF-IDF:

| | brown | enjoy | fast | fox | high | jump | run | state | sunshine |
|---|----------|---------|----------|----------|----------|----------|----------|----------|----------|
| 0 | 0.000000 | 0.50689 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.385503 | 0.771006 |
| 1 | 0.670703 | 0.00000 | 0.000000 | 0.510087 | 0.335352 | 0.335352 | 0.255044 | 0.000000 | 0.000000 |
| 2 | 0.000000 | 0.00000 | 0.549351 | 0.417796 | 0.000000 | 0.000000 | 0.417796 | 0.417796 | 0.417796 |
| | | | | | | | | | |

<u>Q2.</u> Compute the cosine similarity between S1 and S3. <u>Answer:</u>

MANUALLY

| The state of the s | | and a service constant of the same |
|--|--|--|
| 1. | -i(Q#2):- | |
| | The state of the s | |
| => | Cosine similiarity=> | No. of Street, and Street, Str |
| | | |
| | Si = 0.1546 | |
| | 53 = 0.1482 | and a Company of the |
| | | |
| | 5, 53 2 0+0+0+0+0+0+(0.044)*(6.044)+ | EMPLOY AND A CONTRACT AND |
| | (0.088)*(6.044) | END PROPERTY OF |
| | (0.088)*(0.044) = (0.0058) | Barya estendistri reccus Fishin |
| | | |
| | Cos (Si, S3) = 0.00581 | |
| | (0.1546)(0.1482) | of anythernomination of Market |
| | = [0.2535] | No. of Particular States |
| | | SERVICE STREET, SECTION OF |
| AND DESCRIPTION OF THE PARTY OF | NAMES OF THE PROPERTY OF THE P | - |

BY CODE

COSINE: