```
/Users/vedantmahajan/PycharmProjects/news_segregator/venv/bin/python /Users/vedantmahajan/PycharmProjects/news_segregator/Main.py


-----------------------------------WORD COUNT FOR EACH ARTICLE----------------------------


            Article 1  Article 2  Article 3  ...  Article 13  Article 14  Article 15
about               0          0          0  ...           0           1           1
accelerate          0          0          0  ...           0           0           0
accessory           0          0          0  ...           0           0           0
accord              1          0          0  ...           0           0           0
achieve             0          0          0  ...           0           1           0
...               ...        ...        ...  ...         ...         ...         ...
would               0          0          0  ...           0           0           0
write               0          0          0  ...           0           1           0
year                0          1          0  ...           0           0           1
yet                 0          0          0  ...           0           0           0
you                 0          0          0  ...           0           1           0

[690 rows x 15 columns]


---------------------- TDIDF SCORES TRANSFORMED FROM BAG-OF-WORDS MODEL----------------------


            Article 1  Article 2  Article 3  ...  Article 13  Article 14  Article 15
about        0.000000   0.000000        0.0  ...         0.0    1.980829    1.980829
accelerate   0.000000   0.000000        0.0  ...         0.0    0.000000    0.000000
```

```
31 accessory    0.000000   0.000000         0.0  ...          0.0    0.000000    0.000000
32 accord       3.079442   0.000000         0.0  ...          0.0    0.000000    0.000000
33 achieve      0.000000   0.000000         0.0  ...          0.0    3.079442    0.000000
34 ...               ...        ...         ...  ...          ...         ...         ...
35 would        0.000000   0.000000         0.0  ...          0.0    0.000000    0.000000
36 write        0.000000   0.000000         0.0  ...          0.0    3.079442    0.000000
37 year         0.000000   2.386294         0.0  ...          0.0    0.000000    2.386294
38 yet          0.000000   0.000000         0.0  ...          0.0    0.000000    0.000000
39 you          0.000000   0.000000         0.0  ...          0.0    3.079442    0.000000
40
41 [690 rows x 15 columns]
42
43
44
45 ---------------------- DIRECTLY OBTAINING TDIDF SCORES USING TfidfVectorize----------------------
46
47
48             Article 1  Article 2  Article 3  ...  Article 13  Article 14  Article 15
49 about         0.000000   0.000000        0.0  ...         0.0    1.980829    1.980829
50 accelerate    0.000000   0.000000        0.0  ...         0.0    0.000000    0.000000
51 accessory     0.000000   0.000000        0.0  ...         0.0    0.000000    0.000000
52 accord        3.079442   0.000000        0.0  ...         0.0    0.000000    0.000000
53 achieve       0.000000   0.000000        0.0  ...         0.0    3.079442    0.000000
54 ...                ...        ...        ...  ...         ...         ...         ...
55 would         0.000000   0.000000        0.0  ...         0.0    0.000000    0.000000
56 write         0.000000   0.000000        0.0  ...         0.0    3.079442    0.000000
57 year          0.000000   2.386294        0.0  ...         0.0    0.000000    2.386294
58 yet           0.000000   0.000000        0.0  ...         0.0    0.000000    0.000000
59 you           0.000000   0.000000        0.0  ...         0.0    3.079442    0.000000
60
61 [690 rows x 15 columns]
```

```
  \n\nAre the tf-idf scores the same?
0                                  YES


Actual topic of the Article 1 is -> Music news .
Actual topic of the Article 2 is -> hong kong news .
Actual topic of the Article 3 is -> Education news .
Actual topic of the Article 4 is -> Black lives matter news .
Actual topic of the Article 5 is -> railway news .
Actual topic of the Article 6 is -> australia cricket tourism news .
Actual topic of the Article 7 is -> apple tech car news .
Actual topic of the Article 8 is -> Fb Deepfakes news .
Actual topic of the Article 9 is -> Ai and climate change news .
Actual topic of the Article 10 is -> Quantum computing in business news .
Actual topic of the Article 11 is -> Rayshard Brooks case .
Actual topic of the Article 12 is -> US open news .
Actual topic of the Article 13 is -> Covid 19 news .
Actual topic of the Article 14 is -> spaceX news .
Actual topic of the Article 15 is -> Sushant singh rajput news .


Topics of the news articles detected using TFIDF scores :

Article 1    billboard
dtype: object
Article 2    hong
dtype: object
Article 3    university
dtype: object
Article 4    life
```

```
 93 dtype: object
 94 Article 5    engine
 95 dtype: object
 96 Article 6    australia
 97 dtype: object
 98 Article 7    car
 99 dtype: object
100 Article 8    deepfakes
101 dtype: object
102 Article 9    ai
103 dtype: object
104 Article 10    business
105 dtype: object
106 Article 11    brook
107 dtype: object
108 Article 12    tennis
109 dtype: object
110 Article 13    disease
111 dtype: object
112 Article 14    software
113 dtype: object
114 Article 15    actor
115 dtype: object
116
117 Process finished with exit code 0
118
```