Taylor Herb
INFSCI2480 Final Project
Spring 2021

Content Based News Recommender System

https://github.com/t-4-h/infsci2480finalProject

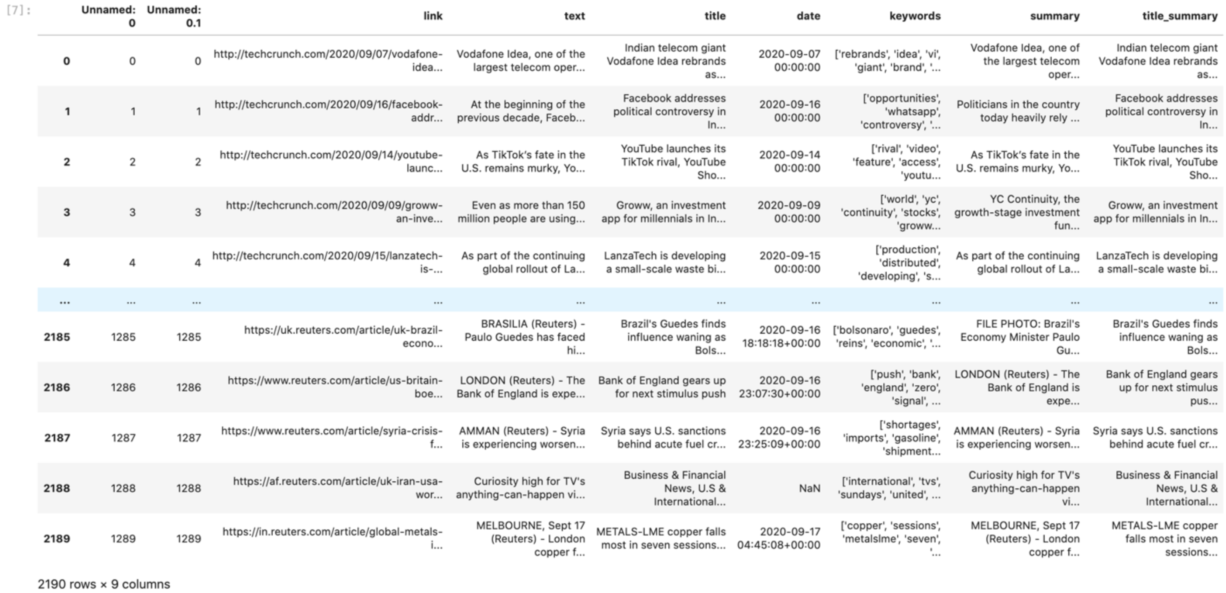
I. Introduction

News websites and blogs are a highly popular aspect of the internet we use today. With a decrease in printed newspapers, online news websites have become an increased source for people to access local, domestic and international news. Since these news sources are being used by people so frequently, they would likely benefit from including a recommender system to personalize the user's experience. For example, if a user reads an article about the stock market in Europe, they may want to read additional articles on the topic. This could mean they want to see more articles about the stock market in general, Europe, the stock market in China, or maybe they aren't sure where to look next. In this situation, a recommender system would be useful to show them other relevant articles on the site that they may be interested in reading.

In this project, I will begin the process of designing a news recommender system by first using content-based recommendation methods. Since there is no user data on this dataset yet, the recommender system must rely on the information about the articles themselves rather than the user. The goal of the current system is to allow the user to search for recommendations based on the article's text body or the article's tagged keywords.

II. Dataset and Pre-processing

The original dataset was pre-processed in order to remove unnecessary columns, missing data and duplicate entries. The final dataset included the five columns to be used in the recommender system (id, title, text, keyword, link) and 1720 rows that did not include any missing data or duplicates.

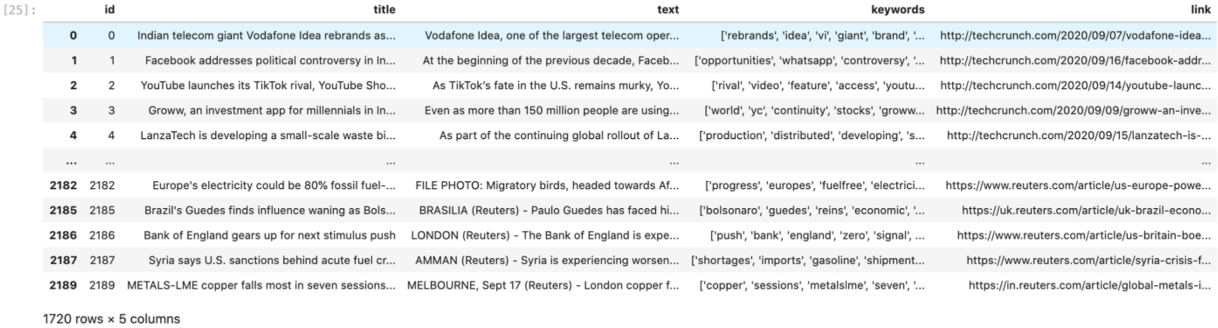


[7]:

	Unnamed: 0	Unnamed: 0.1	link	text	title	date	keywords	summary	title_summary
0	0	0	http://techcrunch.com/2020/09/07/vodafone-idea...	Vodafone Idea, one of the largest telecom oper...	Indian telecom giant Vodafone Idea rebrands as...	2020-09-07 00:00:00	['rebrands', 'idea', 'vi', 'giant', 'brand', '...	Vodafone Idea, one of the largest telecom oper...	Indian telecom giant Vodafone Idea rebrands as...
1	1	1	http://techcrunch.com/2020/09/16/facebook-addr...	At the beginning of the previous decade, Faceb...	Facebook addresses political controversy in In...	2020-09-16 00:00:00	['opportunities', 'whatsapp', 'controversy', '...	Politicians in the country today heavily rely ...	Facebook addresses political controversy in In...
2	2	2	http://techcrunch.com/2020/09/14/youtube-launc...	As TikTok's fate in the U.S. remains murky, Yo...	YouTube launches its TikTok rival, YouTube Sho...	2020-09-14 00:00:00	['rival', 'video', 'feature', 'access', 'youtu...	As TikTok's fate in the U.S. remains murky, Yo...	YouTube launches its TikTok rival, YouTube Sho...
3	3	3	http://techcrunch.com/2020/09/09/groww-an-inve...	Even as more than 150 million people are using...	Groww, an investment app for millennials in In...	2020-09-09 00:00:00	['world', 'yc', 'continuity', 'stocks', 'groww...	YC Continuity, the growth-stage investment fun...	Groww, an investment app for millennials in In...
4	4	4	http://techcrunch.com/2020/09/15/lanzatech-is-...	As part of the continuing global rollout of La...	LanzaTech is developing a small-scale waste bi...	2020-09-15 00:00:00	['production', 'distributed', 'developing', 's...	As part of the continuing global rollout of La...	LanzaTech is developing a small-scale waste bi...
...	...	...	...	...	...	...	...	...	...
2185	1285	1285	https://uk.reuters.com/article/uk-brazil-econo...	BRASILIA (Reuters) - Paulo Guedes has faced hi...	Brazil's Guedes finds influence waning as Bols...	2020-09-16 18:18:18+00:00	['bolsonaro', 'guedes', 'reins', 'economic', '...	FILE PHOTO: Brazil's Economy Minister Paulo Gu...	Brazil's Guedes finds influence waning as Bols...
2186	1286	1286	https://www.reuters.com/article/us-britain-boe...	LONDON (Reuters) - The Bank of England is expe...	Bank of England gears up for next stimulus push	2020-09-16 23:07:30+00:00	['push', 'bank', 'england', 'zero', 'signal', ...	LONDON (Reuters) - The Bank of England is expe...	Bank of England gears up for next stimulus pus...
2187	1287	1287	https://www.reuters.com/article/syria-crisis-f...	AMMAN (Reuters) - Syria is experiencing worsen...	Syria says U.S. sanctions behind acute fuel cr...	2020-09-16 23:25:09+00:00	['shortages', 'imports', 'gasoline', 'shipment...	AMMAN (Reuters) - Syria is experiencing worsen...	Syria says U.S. sanctions behind acute fuel cr...
2188	1288	1288	https://af.reuters.com/article/uk-iran-usa-wor...	Curiosity high for TV's anything-can-happen vi...	Business & Financial News, U.S & International...	NaN	['international', 'tvs', 'sundays', 'united', ...	Curiosity high for TV's anything-can-happen vi...	Business & Financial News, U.S & International...
2189	1289	1289	https://in.reuters.com/article/global-metals-i...	MELBOURNE, Sept 17 (Reuters) - London copper f...	METALS-LME copper falls most in seven sessions...	2020-09-17 04:45:08+00:00	['copper', 'sessions', 'metalslme', 'seven', '...	MELBOURNE, Sept 17 (Reuters) - London copper f...	METALS-LME copper falls most in seven sessions...

2190 rows × 9 columns

Fig 1: Original data (2190 rows, 9 columns)



[25]:

	id	title	text	keywords	link
0	0	Indian telecom giant Vodafone Idea rebrands as...	Vodafone Idea, one of the largest telecom oper...	['rebrands', 'idea', 'vi', 'giant', 'brand', '...	http://techcrunch.com/2020/09/07/vodafone-idea...
1	1	Facebook addresses political controversy in In...	At the beginning of the previous decade, Faceb...	['opportunities', 'whatsapp', 'controversy', '...	http://techcrunch.com/2020/09/16/facebook-addr...
2	2	YouTube launches its TikTok rival, YouTube Sho...	As TikTok's fate in the U.S. remains murky, Yo...	['rival', 'video', 'feature', 'access', 'youtu...	http://techcrunch.com/2020/09/14/youtube-launc...
3	3	Groww, an investment app for millennials in In...	Even as more than 150 million people are using...	['world', 'yc', 'continuity', 'stocks', 'groww...	http://techcrunch.com/2020/09/09/groww-an-inve...
4	4	LanzaTech is developing a small-scale waste bi...	As part of the continuing global rollout of La...	['production', 'distributed', 'developing', 's...	http://techcrunch.com/2020/09/15/lanzatech-is-...
...	...	...	...	...	...
2182	2182	Europe's electricity could be 80% fossil fuel-...	FILE PHOTO: Migratory birds, headed towards Af...	['progress', 'europes', 'fuelfree', 'electrici...	https://www.reuters.com/article/us-europe-powe...
2185	2185	Brazil's Guedes finds influence waning as Bols...	BRASILIA (Reuters) - Paulo Guedes has faced hi...	['bolsonaro', 'guedes', 'reins', 'economic', '...	https://uk.reuters.com/article/uk-brazil-econo...
2186	2186	Bank of England gears up for next stimulus push	LONDON (Reuters) - The Bank of England is expe...	['push', 'bank', 'england', 'zero', 'signal', ...	https://www.reuters.com/article/us-britain-boe...
2187	2187	Syria says U.S. sanctions behind acute fuel cr...	AMMAN (Reuters) - Syria is experiencing worsen...	['shortages', 'imports', 'gasoline', 'shipment...	https://www.reuters.com/article/syria-crisis-f...
2189	2189	METALS-LME copper falls most in seven sessions...	MELBOURNE, Sept 17 (Reuters) - London copper f...	['copper', 'sessions', 'metalslme', 'seven', '...	https://in.reuters.com/article/global-metals-i...

1720 rows × 5 columns

Fig 2: Final dataset after initial pre-processing (1720 rows, 5 columns)

Additionally, text cleaning methods were used to clean the article text and keywords data so that the text could be implemented efficiently within the recommender system. The two text columns were cleaned by converting the text to lowercase, removing special characters, and removing stopwords. In combination with cleaning, the text from the keywords column were stored in a string array.

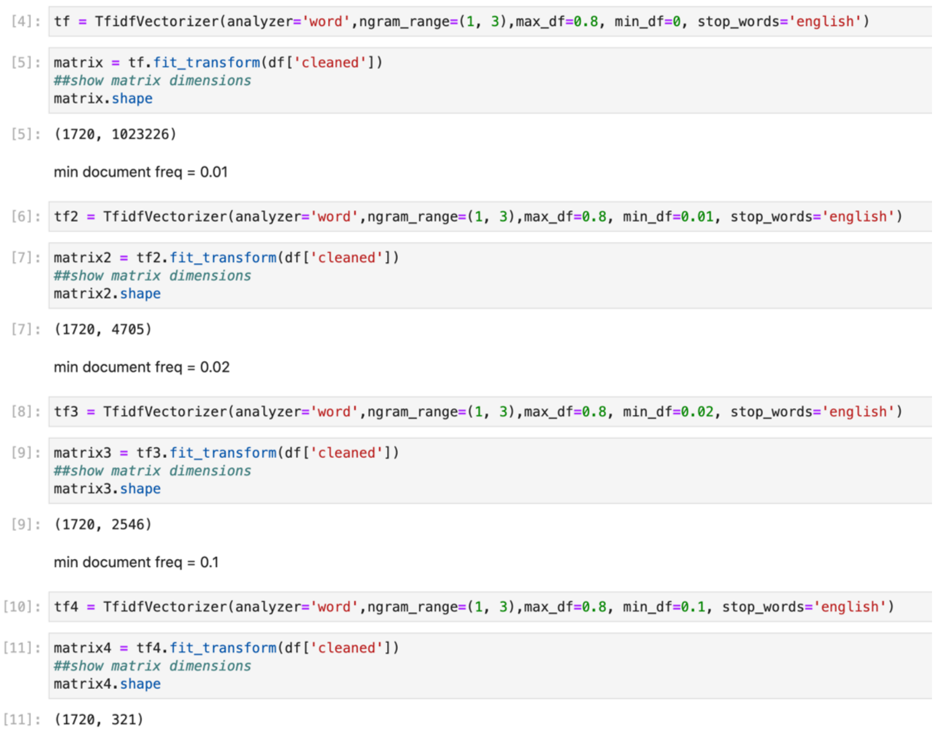
III. Technologies

For this project I have primarily used Python and Jupyter lab. Since the database is small and static, the .csv database is stored within the Jupyter notebook. The interactive portion works using the ipywidgets interact library. A pseudo front-end is created using the Voila server extension, which creates a simple web application that allows the user to enter an article title and receive the top 20 recommendations based on either the article text or article keywords. The user can view the recommended article titles and then click links to visit the article online.

IV. Recommender Design

This recommender will use content based filtering methods to give recommendations based on the articles text or the articles keywords. The reason for doing both is to understand which system performs more accurately in general or which system may perform best under certain conditions. Logically, it would seem that keywords may be helpful when a user is interested in a broad topic while text may be more helpful for more specific topics (including multiple terms). Therefore, I decided it would be useful to build and test both in order to develop a more well-rounded system.

To build both recommender types (text-based and keyword-based), Term Frequency-Inverse Document Frequency (TF-IDF) was used on the text to evaluate term importance to the document within the corpus. To accomplish this, I used the TfidfVectorizer method from the sklearn.feature library. For the keyword-based recommender an ngram range of 1 to 2 was used (because keywords typically exist of a maximum of 2 terms) with a min_df of 0 (to include all keywords regardless of document frequency). For the text-based recommender, an ngram range of 1 to 3 was used to allow for longer combinations of terms within the article. Although stopwords were removed, a max_df of 0.8 was also selected to ignore terms that occur too often in the corpus because they are unlikely useful for recommendations (i.e. words similar to stop words that do not add meaning). Additionally, multiple min_df values were tested (0.0, 0.01, 0.02, 0.1) which yielded different matrix dimensions based on the amount of terms used [(1720, 1023226), (1720, 4705), (1720, 2546), (1720, 321)]. The results from min_df 0.1 were not tested due to such a small number of terms. However, the other 3 matrices all yielded seemingly accurate and relevant results in the recommendation system. It was impossible to know which performed best without user feedback. So, min_df = 0 was ultimately chosen as it contains the most terms but user feedback should be included to understand which parameters perform best.



```
[4]: tf = TfidfVectorizer(analyzer='word',ngram_range=(1, 3),max_df=0.8, min_df=0, stop_words='english')

[5]: matrix = tf.fit_transform(df['cleaned'])
     ##show matrix dimensions
     matrix.shape

[5]: (1720, 1023226)

     min document freq = 0.01

[6]: tf2 = TfidfVectorizer(analyzer='word',ngram_range=(1, 3),max_df=0.8, min_df=0.01, stop_words='english')

[7]: matrix2 = tf2.fit_transform(df['cleaned'])
     ##show matrix dimensions
     matrix2.shape

[7]: (1720, 4705)

     min document freq = 0.02

[8]: tf3 = TfidfVectorizer(analyzer='word',ngram_range=(1, 3),max_df=0.8, min_df=0.02, stop_words='english')

[9]: matrix3 = tf3.fit_transform(df['cleaned'])
     ##show matrix dimensions
     matrix3.shape

[9]: (1720, 2546)

     min document freq = 0.1

[10]: tf4 = TfidfVectorizer(analyzer='word',ngram_range=(1, 3),max_df=0.8, min_df=0.1, stop_words='english')

[11]: matrix4 = tf4.fit_transform(df['cleaned'])
      ##show matrix dimensions
      matrix4.shape

[11]: (1720, 321)
```

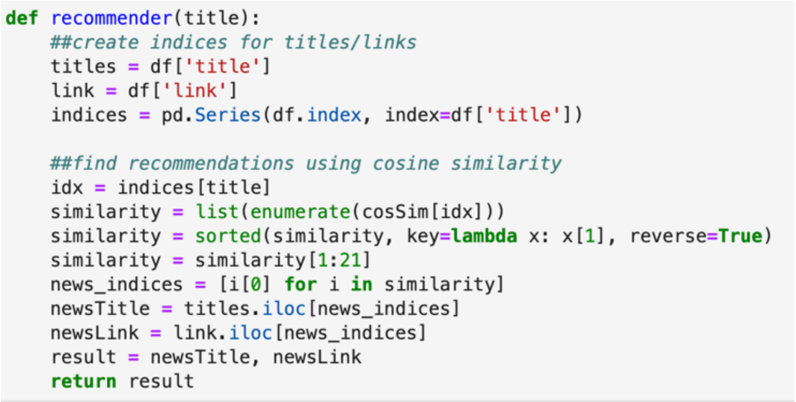
Fig. 3: Matrix dimensions with different min_df values for text-based recommender

Once the article text or keywords were encoded into tf-idf matrices, I used cosine similarity (sklearn cosine_similarity method) to obtain a similarity measure based on proximity between the two matrices. Cosine similarity is useful here because it is a measure of distance and orientation rather than magnitude (i.e. Euclidean distance).

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}},$$

Fig. 4: Cosine similarity equation showing similarity ranging from -1 (opposite) to 1 (equal) where 0 indicates no relationship.

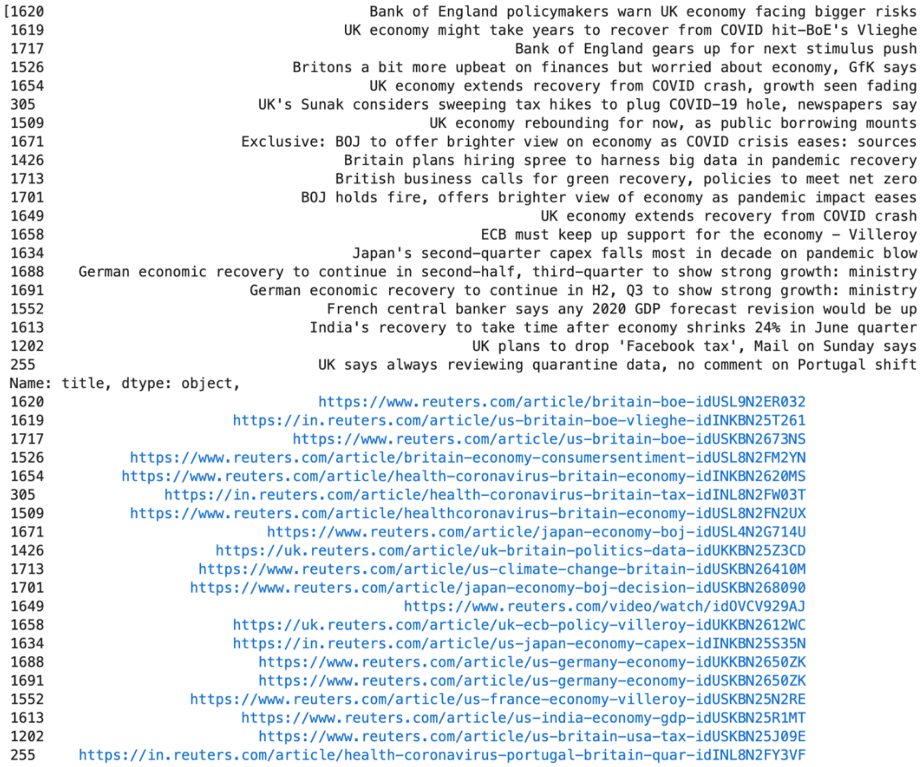
The recommender function was built to create indices for titles then find recommendations based on the cosine similarity (cosSim). This function returns the top 20 most relevant items and shows the user the title of the article as well as the link to the article.



```python
def recommender(title):
    ##create indices for titles/links
    titles = df['title']
    link = df['link']
    indices = pd.Series(df.index, index=df['title'])

    ##find recommendations using cosine similarity
    idx = indices[title]
    similarity = list(enumerate(cosSim[idx]))
    similarity = sorted(similarity, key=lambda x: x[1], reverse=True)
    similarity = similarity[1:21]
    news_indices = [i[0] for i in similarity]
    newsTitle = titles.iloc[news_indices]
    newsLink = link.iloc[news_indices]
    result = newsTitle, newsLink
    return result
```

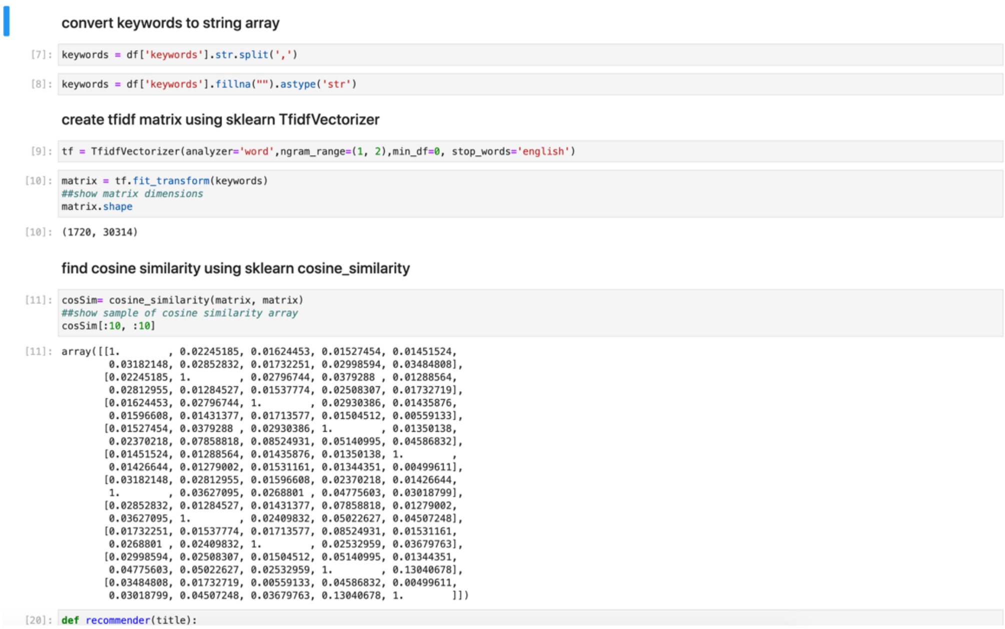
Fig. 5: Recommender function based on cosine similarity



```
[1620                                      Bank of England policymakers warn UK economy facing bigger risks
 1619                                   UK economy might take years to recover from COVID hit-BoE's Vlieghe
 1717                                                       Bank of England gears up for next stimulus push
 1526                             Britons a bit more upbeat on finances but worried about economy, GfK says
 1654                                      UK economy extends recovery from COVID crash, growth seen fading
 305                          UK's Sunak considers sweeping tax hikes to plug COVID-19 hole, newspapers say
 1509                                             UK economy rebounding for now, as public borrowing mounts
 1671                       Exclusive: BOJ to offer brighter view on economy as COVID crisis eases: sources
 1426                                   Britain plans hiring spree to harness big data in pandemic recovery
 1713                                  British business calls for green recovery, policies to meet net zero
 1701                              BOJ holds fire, offers brighter view of economy as pandemic impact eases
 1649                                                          UK economy extends recovery from COVID crash
 1658                                                   ECB must keep up support for the economy – Villeroy
 1634                                    Japan's second-quarter capex falls most in decade on pandemic blow
 1688    German economic recovery to continue in second-half, third-quarter to show strong growth: ministry
 1691                       German economic recovery to continue in H2, Q3 to show strong growth: ministry
 1552                                 French central banker says any 2020 GDP forecast revision would be up
 1613                               India's recovery to take time after economy shrinks 24% in June quarter
 1202                                                  UK plans to drop 'Facebook tax', Mail on Sunday says
 255                                 UK says always reviewing quarantine data, no comment on Portugal shift
 Name: title, dtype: object,
 1620                                https://www.reuters.com/article/britain-boe-idUSL9N2ER032
 1619                      https://in.reuters.com/article/us-britain-boe-vlieghe-idINKBN25T261
 1717                             https://www.reuters.com/article/us-britain-boe-idUSKBN2673NS
 1526          https://www.reuters.com/article/britain-economy-consumersentiment-idUSL8N2FM2YN
 1654         https://www.reuters.com/article/health-coronavirus-britain-economy-idINKBN2620MS
 305              https://in.reuters.com/article/health-coronavirus-britain-tax-idINL8N2FW03T
 1509          https://www.reuters.com/article/healthcoronavirus-britain-economy-idUSL8N2FN2UX
 1671                          https://www.reuters.com/article/japan-economy-boj-idUSL4N2G714U
 1426                    https://uk.reuters.com/article/uk-britain-politics-data-idUKKBN25Z3CD
 1713                  https://www.reuters.com/article/us-climate-change-britain-idUSKBN26410M
 1701                 https://www.reuters.com/article/japan-economy-boj-decision-idUSKBN268090
 1649                                          https://www.reuters.com/video/watch/idOVCV929AJ
 1658                      https://uk.reuters.com/article/uk-ecb-policy-villeroy-idUKKBN2612WC
 1634                      https://in.reuters.com/article/us-japan-economy-capex-idINKBN25S35N
 1688                         https://www.reuters.com/article/us-germany-economy-idUKKBN2650ZK
 1691                         https://www.reuters.com/article/us-germany-economy-idUSKBN2650ZK
 1552                 https://www.reuters.com/article/us-france-economy-villeroy-idUSKBN25N2RE
 1613                       https://www.reuters.com/article/us-india-economy-gdp-idUSKBN25R1MT
 1202                         https://www.reuters.com/article/us-britain-usa-tax-idUSKBN25J09E
 255     https://in.reuters.com/article/health-coronavirus-portugal-britain-quar-idINL8N2FY3VF
```

Fig 6: Results for text-based recommender (title “Bank of England policymakers warn of bigger risks for UK economy”

```
(1620                         Bank of England policymakers warn UK economy facing bigger risks
 1619                       UK economy might take years to recover from COVID hit-BoEs Vlieghe
 1717                                          Bank of England gears up for next stimulus push
 1196                             JPMorgan to launch UK digital lender in early 2021: Sky News
 1300                     Experts warn: High-tech tools to fight COVID-19 pose their own risks
 1658                                      ECB must keep up support for the economy - Villeroy
 1654                         UK economy extends recovery from COVID crash, growth seen fading
 178                     COVID generation risks child marriage, forced labour, ex-leaders warn
 1552                    French central banker says any 2020 GDP forecast revision would be up
 1604             Australias central bank has limited options as economy sinks into steep slump
 1671          Exclusive: BOJ to offer brighter view on economy as COVID crisis eases: sources
 1560  Indian government consumption key to growth in economy amid pandemic, central bank says
 1629                             Feds Mester says central bank wont let inflation run rampant
 1618               Brazil economy back to 2009 size after record 9.7% slump in second quarter
 1652                                           ECB sees "strong rebound" signs, monitoring FX
 1097                                                 Britains Tesco to trial drone deliveries
 1642                Climate change may wreck economy unless we act soon, federal report warns
 142                                                      Column: Megacities after coronavirus
 769                 The GoPro Hero9 Is a Little Bigger and a Lot Better in Every Possible Way
 1660    Global economy seeing sharper V recovery, raising case for inflation - Morgan Stanley
 Name: title, dtype: object,
 1620                                                              https://www.reuters.com/article/britain-boe-idUSL9N2ER032
 1619                                                     https://in.reuters.com/article/us-britain-boe-vlieghe-idINKBN25T261
 1717                                                           https://www.reuters.com/article/us-britain-boe-idUSKBN2673NS
 1196                                                following https://www.reuters.com/article/jp-morgan-digital-banking-idUSKBN25H1H0
```



convert keywords to string array

```
[7]: keywords = df['keywords'].str.split(',')
```

```
[8]: keywords = df['keywords'].fillna("").astype('str')
```

create tfidf matrix using sklearn TfidfVectorizer

```
[9]: tf = TfidfVectorizer(analyzer='word',ngram_range=(1, 2),min_df=0, stop_words='english')
```

```
[10]: matrix = tf.fit_transform(keywords)
      ##show matrix dimensions
      matrix.shape

[10]: (1720, 30314)
```

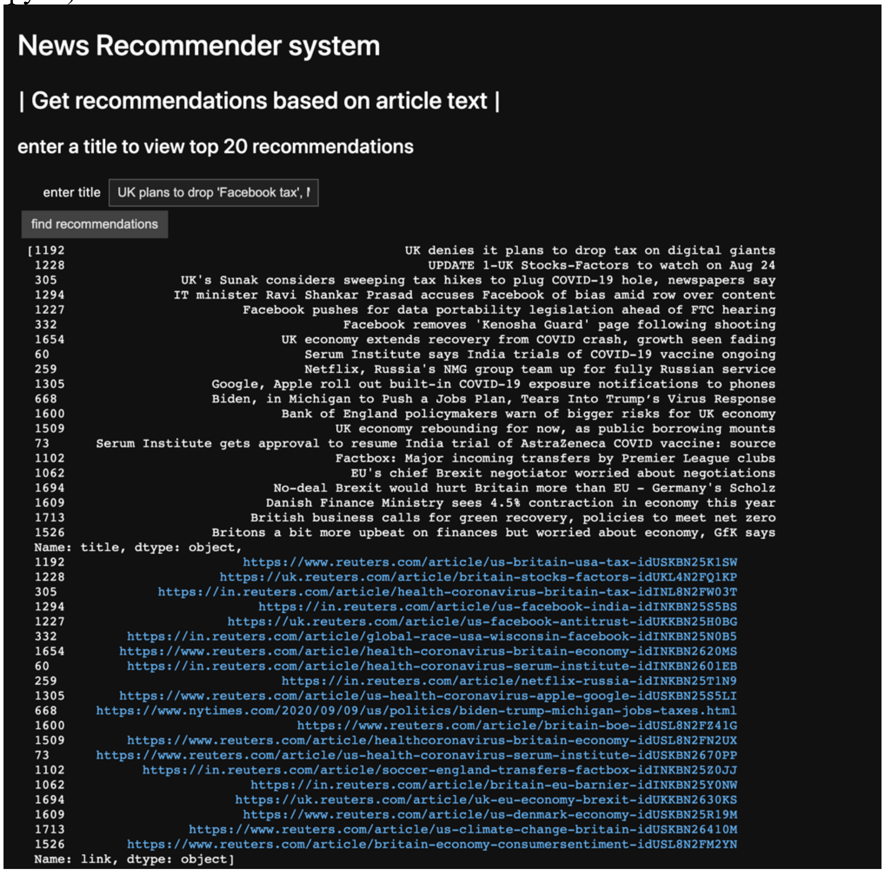
find cosine similarity using sklearn cosine_similarity

```
[11]: cosSim= cosine_similarity(matrix, matrix)
      ##show sample of cosine similarity array
      cosSim[:10, :10]

[11]: array([[1.        , 0.02245185, 0.01624453, 0.01527454, 0.01451524,
              0.03182148, 0.02852832, 0.01732251, 0.02998594, 0.03484808],
             [0.02245185, 1.        , 0.02796744, 0.0379288 , 0.01288564,
              0.02812955, 0.01284527, 0.01537774, 0.02508307, 0.01732719],
             [0.01624453, 0.02796744, 1.        , 0.02930386, 0.01435876,
              0.01596608, 0.01431377, 0.01713577, 0.01504512, 0.00559133],
             [0.01527454, 0.0379288 , 0.02930386, 1.        , 0.01350138,
              0.02370218, 0.07858818, 0.08524931, 0.05140995, 0.04586832],
             [0.01451524, 0.01288564, 0.01435876, 0.01350138, 1.        ,
              0.01426644, 0.01279002, 0.01531161, 0.01344351, 0.00499611],
             [0.03182148, 0.02812955, 0.01596608, 0.02370218, 0.01426644,
              1.        , 0.03627095, 0.0268801 , 0.04775603, 0.03018799],
             [0.02852832, 0.01284527, 0.01431377, 0.07858818, 0.01279002,
              0.03627095, 1.        , 0.02409832, 0.05022627, 0.04507248],
             [0.01732251, 0.01537774, 0.01713577, 0.08524931, 0.01531161,
              0.0268801 , 0.02409832, 1.        , 0.02532959, 0.03679763],
             [0.02998594, 0.02508307, 0.01504512, 0.05140995, 0.01344351,
              0.04775603, 0.05022627, 0.02532959, 1.        , 0.13040678],
             [0.03484808, 0.01732719, 0.00559133, 0.04586832, 0.00499611,
              0.03018799, 0.04507248, 0.03679763, 0.13040678, 1.        ]])
```

```
[20]: def recommender(title):
```

Fig. 8: Screenshot of part of the keyword-based development notebook (recommender-keywordBased.ipynb)



News Recommender system

| Get recommendations based on article text |

enter a title to view top 20 recommendations

enter title UK plans to drop 'Facebook tax', I

find recommendations

```
[1192                                            UK denies it plans to drop tax on digital giants
 1228                                               UPDATE 1-UK Stocks-Factors to watch on Aug 24
 305                UK's Sunak considers sweeping tax hikes to plug COVID-19 hole, newspapers say
 1294              IT minister Ravi Shankar Prasad accuses Facebook of bias amid row over content
 1227                       Facebook pushes for data portability legislation ahead of FTC hearing
 332                                     Facebook removes 'Kenosha Guard' page following shooting
 1654                             UK economy extends recovery from COVID crash, growth seen fading
 60                                  Serum Institute says India trials of COVID-19 vaccine ongoing
 259                                 Netflix, Russia's NMG group team up for fully Russian service
 1305                   Google, Apple roll out built-in COVID-19 exposure notifications to phones
 668                    Biden, in Michigan to Push a Jobs Plan, Tears Into Trump’s Virus Response
 1600                             Bank of England policymakers warn of bigger risks for UK economy
 1509                                    UK economy rebounding for now, as public borrowing mounts
 73      Serum Institute gets approval to resume India trial of AstraZeneca COVID vaccine: source
 1102                                    Factbox: Major incoming transfers by Premier League clubs
 1062                                      EU's chief Brexit negotiator worried about negotiations
 1694                            No-deal Brexit would hurt Britain more than EU - Germany's Scholz
 1609                           Danish Finance Ministry sees 4.5% contraction in economy this year
 1713                         British business calls for green recovery, policies to meet net zero
 1526                   Britons a bit more upbeat on finances but worried about economy, GfK says
 Name: title, dtype: object,
 1192                       https://www.reuters.com/article/us-britain-usa-tax-idUSKBN25K1SW
 1228                    https://uk.reuters.com/article/britain-stocks-factors-idUKL4N2FQ1KP
 305             https://in.reuters.com/article/health-coronavirus-britain-tax-idINL8N2FW03T
 1294                     https://in.reuters.com/article/us-facebook-india-idINKBN25S5BS
 1227                 https://uk.reuters.com/article/us-facebook-antitrust-idUKKBN25H0BG
 332         https://in.reuters.com/article/global-race-usa-wisconsin-facebook-idINKBN25N0B5
 1654       https://www.reuters.com/article/health-coronavirus-britain-economy-idINKBN2620MS
 60          https://in.reuters.com/article/health-coronavirus-serum-institute-idINKBN2601EB
 259                            https://in.reuters.com/article/netflix-russia-idINKBN25T1N9
 1305       https://www.reuters.com/article/us-health-coronavirus-apple-google-idUSKBN25S5LI
 668     https://www.nytimes.com/2020/09/09/us/politics/biden-trump-michigan-jobs-taxes.html
 1600                             https://www.reuters.com/article/britain-boe-idUSL8N2FZ41G
 1509        https://www.reuters.com/article/healthcoronavirus-britain-economy-idUSL8N2FN2UX
 73      https://www.reuters.com/article/us-health-coronavirus-serum-institute-idUSKBN2670PP
 1102          https://in.reuters.com/article/soccer-england-transfers-factbox-idINKBN25Z0JJ
 1062                    https://in.reuters.com/article/britain-eu-barnier-idINKBN25Y0NW
 1694                  https://uk.reuters.com/article/uk-eu-economy-brexit-idUKKBN2630KS
 1609                   https://www.reuters.com/article/us-denmark-economy-idUSKBN25R19M
 1713            https://www.reuters.com/article/us-climate-change-britain-idUSKBN26410M
 1526        https://www.reuters.com/article/britain-economy-consumersentiment-idUSL8N2FM2YN
 Name: link, dtype: object]
```

Fig. 9: Screenshot of part of the production notebook (FULL-RECOMMENDER.ipynb)

VI. Conclusions and Future Work

In this project, I have developed a news recommender system based on article text and article keywords. Since there is not yet any user data, content based methods were used to provide the recommendations about the articles. TF-IDF feature extraction and cosine similarity methods were used to create similarity measures for the recommendations.

Currently, it seems that both recommendation systems perform quite well despite the absence of any additional user data. At first glance it seems as though the text based recommender is more consistently accurate, however, it is impossible to guess which model performs better in the absence of user feedback. As the project develops, collaborative filtering methods can be employed based on user behavior data that is gathered while they interact with the system. Ideally, future users will be able to give feedback about their recommendations in order to fine tune the algorithms and personalize their experience. Additional features such as creating profiles, "favoriting" articles and searching by other terms (i.e. search by keyword or author) will be implemented during continued front-end development to help make the system more adaptive and customizable.