

Search Retrieval System of Articles Related to Volatile Stocks

The following python scripts were developed by me to help separate the different aspects of this project. I have included comments in the files themselves to help explain what is happening within the code but I will go through an overview of the functionality here to explain what needs to happen before any document retrieval can be done.

The first file “Robinhood API.py” is how I connect to the Robinhood API and download a list of stocks that can be used as queries for the purpose of this project. Originally I wanted to use the collection of stocks that I own on Robinhood as the stocks to query, but the market has gone crazy due to the COVID situation which caused me to liquidate my portfolio. So, instead of using the API to get the stocks that I own, I will use it to query the top moving (up or down) stocks of the day. For the purpose of this assignment, I decided to get two days of the top movers, resulting in 32 total stocks. I had hoped since they are the top moving stocks, there will be documents associated with them, but as I will touch on, this was optimistic. The python file outputs a .json file containing the stock ticker and company name as strings for all 32 stocks. This code is commented out because the file has already been generated at ‘stock_portfolio.json’.

The second file “News API.py” is code for accessing the News API. This is a free API that aggregates news sources and returns a list of all url’s of articles contained in the API request. The request can be specified to only return articles of a certain category, so for this project I decided to only return articles related to ‘business’ and ‘technology’. I also limited the articles returned to be created in the last 3 weeks. This created a list of 2074 articles. I also used a python library called ‘newspaper’ to handle the web scraping of the url’s for me. The code from this file ultimately outputs a .json file with article urls and the text of their contents. The code is commented out here because the files have already been read and downloaded to the file ‘newsapi_articles.json’.

The third file “Inverted Index Creation.py” is where I create the inverted index using the text from the articles. I utilize a python library ‘nltk’ to handle the stop words and provide a Porter Stemmer. The inverted index takes advantage of pointers to decrease the size of the data structure. There is an ‘inverted index’ dictionary with the stemmed terms as keys, and the document python id’s (lookup locations) as the values. The ‘pointer index’ is how the inverted index finds which articles the terms actually belong to, by using the document python id. In addition to creating an inverted index of the document terms, I also used the stemmer to update the query terms. This code also handles creation of the TFxIDF matrix by doing the calculation when iterating over the inverted index. First it calculates the IDF of each term by looking at how often the term is used in all the documents contained in the matrix. Next, it calculates TFxIDF by multiplying the IDF with the term counts. This outputs a new dictionary with term strings as keys, and TFxIDF values. The outputted .json files are the TFxIDF dictionary and the stemmed query terms dictionary. This code is commented out because the files have already been generated at ‘article_tf_idf.json’ and ‘query_terms.json’.

The fourth file “Read Cranfield Collection.py” is code for reading in the Cranfield Test Collection. The text files from which to read from have different methods of separating document, document ID, query and query ID so must be handled in a different manner. This collection also includes data on which documents are relevant to which query so that is also downloaded and handled. Using a similar method to the above file, I created an inverted index using the terms contained within the documents, and created a separate dictionary for the queries and the relevancy results. This file ultimately outputs .json files of the tf*idf dictionary, the query terms, and the relevance results. This code is commented out because the files have already been generated at ‘/cran.tar/cranfield_document_tf_idf.json’, ‘/cran.tar/query_terms.json’ and ‘/cran.tar/cranfield_query_rel.json’.

The last file “Final Project Functions.py” is where I put the functions I need to load the documents, create dataframes, expand the queries, and get results from the queries themselves. The functions are designed to apply to both the Cranfield dataset as well as my own, but with the exception of the query results. The query result function needs to be different, because for the Cranfield dataset, I know the relevance results. My own dataset needs human input to tell it which documents are relevant and which aren’t. I will explain more about the functions as I call them in the script below.

```
[1]: %run -i "Robinhood API"
      %run -i "News API"
      %run -i "Inverted Index Creation"
      %run -i "Read Cranfield Collection"
      %run -i "Final Project Functions"
```

Cranfield Dataset Test

First test the Cranfield Dataset to see how my algorithm does by computing metrics such as precision and recall. The code will load in the files that have been generated previously to create a tf*idf dictionary, stemmed query term dictionary, and query relevance dictionary

```
[2]: # Main Code
      # start by testing algorithm vs Cranfield Test Dataset
      document_tf_idf_dict = {}
      stemmed_query_terms = {}
      query_rel = {}

      filenames = loadFiles(True)
      # import json files with all neccessary information
      with open(filenames[0], 'r') as json_file:
          document_tf_idf_dict = json.load(json_file)

      with open(filenames[1], 'r') as json_file:
          stemmed_query_terms = json.load(json_file)

      with open(filenames[2], 'r') as json_file:
          query_rel = json.load(json_file)
```

Make calls to create pandas dataframe objects of the tf*idf. Then create a term co-occurrence matrix using numpy. After making a term co-occurrence matrix I can expand the query to use the n = 3 closest related terms (via Cosine Similarity) to each of the query terms.

```
[3]: # global analysis results for test data
      # get top n related terms to each query term and run query with all those terms

      # get entire df and co-occurrence matrix
      test_term_doc_df = getTermDocDF(document_tf_idf_dict)
      test_term_cooccur = getTermCoOccurMatrix(test_term_doc_df)

      # expand query to include all related terms
      test_expanded_query_dict = expandQuery(test_term_cooccur, test_term_doc_df,
      ↪stemmed_query_terms, 4)
```

Combine the document dataframe with the query dataframe to create a matrix with which I can create a

cosine similarity matrix out of. After getting the cosine similarities for each document, I can get the results of the queries.

```
[4]: # get a numpy array of all documents + queries and their associated terms
test_term_doc_queries_array = queryDocuments(test_expanded_query_dict,
↳test_term_doc_df)

# get results of expanded query
test_query_results = queryResults(test_expanded_query_dict,
↳test_term_doc_queries_array)

# output results of query
test_query_top_docs = outputQueryResultsForTest(test_query_results, test_term_doc_df, .
↳26)
```

Final Project Functions.py:100: RuntimeWarning: invalid value encountered in double_scalars

```
np.dot(term_doc_queries_array[query_loc], row) /
(norm(term_doc_queries_array[query_loc]) * norm(row))
```

Cosine Similarity Threshold = 0.26

I set a threshold of .26 (best estimate) for which the cosine similarity needs to be greater than for the document to be relevant to the query. This threshold seems to be the point at which precision and relevance are maximized, given a few test cases. The scores aren't great, it appears too many documents are returned that are irrelevant, so maybe the local analysis can do better.

```
[5]: # output precision scores of query
outputPrecisionResultsForTest(test_query_top_docs, query_rel, 'global analysis for
↳cranfield')
```

Precision score overall for global analysis for cranfield is 0.19071644803229063

Recall score overall for global analysis for cranfield is 0.20577027762656505

Local analysis is different than global because a query is run before the query expansion. This means that documents only related to the original query are used to determine terms that are most similar to the query term. This should get rid of terms that aren't relevant to the document, but happen to be relevant to the query term.

```
[6]: # local analysis results for test data
# get top n related terms for only documents that are returned on the initial query

local_test_term_doc_df = getTermDocDF(document_tf_idf_dict)

# get a numpy array of relevant docs + queries and their associated terms
test_term_doc_queries_array = queryDocuments(stemmed_query_terms, test_term_doc_df)

# get results of local query
test_query_results = queryResults(stemmed_query_terms, test_term_doc_queries_array)

# get only documents applicable to local query
```

```
test_local_query_documents = getRelevantDocuments(test_query_results)
```

Final Project Functions.py:100: RuntimeWarning: invalid value encountered in double_scalars

```
np.dot(term_doc_queries_array[query_loc], row) /  
(norm(term_doc_queries_array[query_loc]) * norm(row)))
```

The results of each queries individual score needs to be combined to get a full picture of the algorithm. Use a for loop to run through each query and add their results to a dictionary,

```
[7]: # for each query do a local expansion  
test_local_query_results_combined = {}  
  
for name, indices in test_local_query_documents.items():  
    # get entire df and co-occurrence matrix  
    test_local_term_doc_df = getTermDocDF(document_tf_idf_dict, indices)  
    test_local_term_cooccur = getTermCoOccurMatrix(test_local_term_doc_df)  
  
    # expand query to include all related terms  
    test_local_expanded_query_dict = expandQuery(test_local_term_cooccur,   
↳ test_local_term_doc_df,  
                                                {name: stemmed_query_terms[name]}, 4)  
  
    # get a numpy array of all documents + queries and their associated terms  
    test_local_term_doc_queries_array = queryDocuments(test_local_expanded_query_dict,   
↳ test_local_term_doc_df)  
  
    # get results of expanded query  
    test_local_query_results = queryResults(test_local_expanded_query_dict,   
↳ test_local_term_doc_queries_array)  
  
    # get list of local query values, filter out documents with none  
    local_val_list = list(test_local_query_results.values())[0]  
    if not any(x > 0 for x in local_val_list):  
        continue  
  
    # set results in an array that spans the entire term collection  
    full_array_query_results = expandToFullArray(local_val_list,   
↳ test_local_term_doc_df, indices)  
  
    test_local_query_results_combined[name] = full_array_query_results
```

Output results using a new threshold to maximize precision and recall (.24). Surprisingly, the results are worse with the local analysis. I'm somewhat suprised, but with a smallish (1400) corpus of documents it probably shouldn't be that surprising. This means there are less terms for the global analysis to get confused, and a smaller sample size for the local analysis to draw from.

```
[8]: # output results of query  
local_query_rel_docs = outputQueryResultsForTest(test_local_query_results_combined,   
↳ test_term_doc_df, .24)  
  
# output precision scores of query
```

```
outputPrecisionResultsForTest(local_query_rel_docs, query_rel, 'local analysis for_
↳recall')
```

Cosine Similarity Threshold = 0.24

Precision score overall for local analysis for recall is 0.1461211477151966

Recall score overall for local analysis for recall is 0.1497005988023952

Querying Stock Names with Set of Online News Articles

Knowing that the algorithm runs pretty well I will test using my own dataset. I will once again look at both global and local analysis methods of query expansion. However, it is now up to the user to determine if the articles are relevant or not to the query.

```
[9]: # now use own dataset to run application

inverted_index = {}
pointer_index = {}
document_tf_idf_dict = {}
stemmed_query_terms = {}

filenames = loadFiles()
# import json files with all neccessary information
with open(filenames[0], 'r') as json_file:
    document_tf_idf_dict = json.load(json_file)

with open(filenames[1], 'r') as json_file:
    stemmed_query_terms = json.load(json_file)
```

The stemmed query terms are the tokenized names of each stock plus the stock ticker. Use the same sequence of calls as from the test set to get the global analysis expansion of the query terms. Then get the results from the query by using cosine similarity.

```
[10]: # global analysis results
# get top n related terms to each query term and run query with all those terms

# get entire df and co-occurrence matrix
global_term_doc_df = getTermDocDF(document_tf_idf_dict)
global_term_cooccur = getTermCoOccurMatrix(global_term_doc_df)

# expand query to include all related terms
global_expanded_query_dict = expandQuery(global_term_cooccur, global_term_doc_df,
↳stemmed_query_terms)

# get a numpy array of all documents + queries and their associated terms
global_term_doc_queries_array = queryDocuments(global_expanded_query_dict,
↳global_term_doc_df)

# get results of expanded query
global_query_results = queryResults(global_expanded_query_dict,
↳global_term_doc_queries_array)
```

Final Project Functions.py:100: RuntimeWarning: invalid value encountered in double_scalars

```
np.dot(term_doc_queries_array[query_loc], row) /  
(norm(term_doc_queries_array[query_loc]) * norm(row)))
```

For functional output purposes, the top 3 articles related to each stock are listed. It is unfortunate that there are 7 companies with no articles that mention them at all by name. Maybe this says something about our online media and which companies they choose to write about.

```
[11]: # output results of query  
global_query_top_n = outputQueryResults(global_query_results, global_term_doc_df)
```

Company lincoln financial group top 3 articles are:

- 1: <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/>
- 2: <https://www.cnn.com/2020/03/03/ford-confirms-us-built-electric-cargo-van-under-its-11point5-billion-plan.html>
- 3: <https://www.theverge.com/2020/3/3/21163669/ford-all-electric-transit-cargo-van-2021>

Company coty top 3 articles are:

No articles relating to coty

Company ameriprise top 3 articles are:

- 1: <https://www.bloomberg.com/news/articles/2020-03-07/u-s-pro-sports-leagues-plan-to-ban-outsiders-from-locker-rooms>
- 2: <https://www.wired.com/2020/03/warp-drive-movies/>
- 3: <https://www.wsj.com/articles/a-bloomberg-business-manifesto-11582586849>

Company state street top 3 articles are:

- 1: <https://www.businessinsider.com/coronavirus-testing-covid-19-tests-per-capita-chart-us-behind-2020-3>
- 2: <https://arstechnica.com/science/2020/03/118-us-coronavirus-cases-9-deaths-as-ramped-up-testing-uncovers-hidden-spread/>
- 3: <https://www.businessinsider.com/us-cdc-conducted-fewer-than-6000-coronavirus-tests-2020-3>

Company ventas top 3 articles are:

No articles relating to ventas

Company northern trust top 3 articles are:

- 1: <https://www.theverge.com/2020/3/7/21169637/italy-lockdown-lombardy-coronavirus-outbreak>
- 2: <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3>
- 3: <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/>

Company oracle top 3 articles are:

- 1: <https://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-black-steps-down/>
- 2: <http://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-black-steps-down/>
- 3: <https://www.businessinsider.com/salesforce-ceo-keith-black-operator-sudden->

exit-oracle-2020-2

Company occidental petroleum top 3 articles are:

- 1: <https://www.cnn.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html>
- 2: <https://www.wsj.com/articles/carl-icahn-boosts-occidental-stake-to-almost-10-as-shares-plummet-11583969050>
- 3: <https://www.cnn.com/the-u-s-stock-market-crash-will-bankrupt-these-3-companies/>

Company morgan stanley top 3 articles are:

- 1: <https://www.cnn.com/video/2020/03/03/this-is-an-opportunity-to-upgrade-your-portfolio-morgan-stanleys-wilson.html>
- 2: <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3>
- 3: <https://www.businessinsider.com/us-coronavirus-cases-linked-quarantined-nile-river-cruise-ship-2020-3>

Company devon energy top 3 articles are:

- 1: <https://www.businessinsider.com/how-to-make-a-slack-channel-private>
- 2: <https://business.financialpost.com/commodities/alberta-eyes-curtailment-to-ensure-a-survival-price-as-province-stares-down-recession>
- 3: <https://www.businessinsider.com/how-to-add-emojis-to-discord>

Company newmont top 3 articles are:

- 1: <https://www.businessinsider.com/de-blasio-blasts-buttigieg-after-his-third-place-nevada-finish-2020-2>
- 2: <https://www.infomoney.com.br/mercados/track-field-prima-foods-e-alphaville-urbanismo-entram-com-pedidos-de-ipo/>
- 3: <https://www.infomoney.com.br/politica/bolsonaro-sugere-adiamento-de-manifestacoes-de-15-de-marco/>

Company zimmer biomet top 3 articles are:

No articles relating to zimmer biomet

Company capri holdings top 3 articles are:

- 1: <http://techcrunch.com/2020/03/13/sponsors-could-be-key-to-solving-techs-lack-of-gender-diversity/>
- 2: <https://www.cnn.com/2020/02/26/mike-bloombergs-campaign-holding-talks-with-unions.html>
- 3: <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2>

Company hp top 3 articles are:

- 1: <https://www.cnn.com/video/2020/02/26/coronavirus-outbreak-will-dent-earnings-for-current-quarter-hp-ceo.html>
- 2: <https://www.cnn.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html>
- 3: <https://www.cnn.com/2020/02/26/hp-ceo-forecasts-coronavirus-will-dent-quarterly-earnings-by-8-cents.html>

Company lennar top 3 articles are:

- 1: <https://www.businessinsider.com/amazon-just-walk-out-technology-questions-remain-2020-3>
- 2: <https://thenextweb.com/plugged/2020/03/10/leak-pixel-4a-hands-on-video-leaves-little-for-google-to-announce/>
- 3: <https://www.businessinsider.com/samsung-galaxy-s20-plus-review>

Company public service enterprise group top 3 articles are:

- 1: <https://www.cnbc.com/2020/03/06/coronavirus-live-updates-sxsw-canceled-grand-princess-cruise-ship-has-21-cases.html>
- 2: <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere>
- 3: <https://www.cnbc.com/2020/02/23/saudi-arabia-will-support-lebanon-if-it-sees-viable-reform-plan.html>

Company mohawk top 3 articles are:

No articles relating to mohawk

Company nvr top 3 articles are:

No articles relating to nvr

Company edwards lifesciences top 3 articles are:

- 1: <https://thenextweb.com/neural/2020/03/07/the-case-for-an-ai-that-puts-nature-and-ethics-first-not-humans-syndication/>
- 2: <https://thenextweb.com/neural/2020/03/01/ai-will-dominate-the-2020s-governments-must-ensure-its-ethical/>
- 3: <https://www.businessinsider.com/how-to-shop-for-perfume>

Company fox (class b) top 3 articles are:

- 1: <https://www.businessinsider.com/where-are-firefox-bookmarks-stored>
- 2: <https://www.businessinsider.com/coronavirus-school-closures-have-left-kids-feeling-scared-confused-2020-3>
- 3: <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere>

Company h&r block top 3 articles are:

- 1: <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2>
- 2: <https://www.businessinsider.com/salesforce-keith-block-marc-benioff-salary-2020-2>
- 3: <https://www.techradar.com/news/best-soundbar-for-under-dollar200pound200>

Company charles schwab top 3 articles are:

- 1: <https://techcrunch.com/2020/03/03/los-angeles-gets-a-new-consumer-fund-as-founders-of-sprinkles-seek-25-million-for-cn2-ventures/>
- 2: <https://www.techradar.com/news/f1-live-stream-how-to-watch-the-australian-grand-prix-2020-online-from-anywhere>
- 3: <https://www.cnbc.com/2020/02/27/dow-futures-fall-100-points-after-another-massive-rout-amid-coronavirus-fears.html>

Company jpmorgan chase top 3 articles are:

- 1: <https://www.cnbc.com/select/using-a-credit-card-to-donate-to-political-campaigns/>

2: <https://www.businessinsider.com/how-to-get-rid-of-metal-credit-cards>

3: <https://www.cnbc.com/select/best-credit-cards-for-home-improvements/>

Company raymond james top 3 articles are:

1: <https://www.ccn.com/virtual-racers-like-james-baldwin-might-soon-replace-traditional-racing-drivers/>

2: <https://www.ccn.com/zion-williamson-could-surpass-lebron-james-if-he-fixes-these-three-things/>

3: <https://www.cnbc.com/2020/03/05/jpmorgan-says-ceo-jamie-dimon-is-recuperating-after-emergency-heart-surgery.html>

Company etrade top 3 articles are:

No articles relating to etrade

Company bank of america top 3 articles are:

1: <https://business.financialpost.com/news/economy/canadian-recession-is-likely-without-fiscal-stimulus-scotiabank>

2: <https://www.afr.com/companies/financial-services/cba-ready-to-defer-small-business-loan-repayments-fees-20200311-p548va>

3: <https://business.financialpost.com/news/fp-street/rbc-to-cut-canada-prime-rate-to-3-45-rivals-follow>

Company apache top 3 articles are:

1: <https://business.financialpost.com/commodities/alberta-eyes-curtailement-to-ensure-a-survival-price-as-province-stares-down-recession>

2: <https://www.cnbc.com/2020/03/09/cramer-9-or-10-oil-companies-may-go-bankrupt-amid-crude-declines.html>

3: <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html>

Company expedia top 3 articles are:

1: <https://fortune.com/2020/02/29/coronavirus-travel-wealthy-avoid-airlines-flying-private-jets/>

2: <https://www.cnbc.com/2020/03/11/coronavirus-travel-industry-could-lose-24-billion-in-tourism-from-outside-us.html>

3: <https://www.businessinsider.com/private-jet-travel-surge-evacuate-coronavirus-outbreak-covid19-travel-aviation-2020>

Company united airlines top 3 articles are:

1: <https://www.cnbc.com/2020/02/28/coronavirus-latest-updates-china-hubei.html>

2: <https://www.businessinsider.com/coronavirus-changing-canceling-flights-united-jetblue-american-delta-southwest-2020-3>

3: <https://www.businessinsider.com/coronavirus-united-airlines-cost-cutting-cancellations-hiring-freeze-unpaid-leave-2020-3>

Company broadcom top 3 articles are:

1: <https://www.cnbc.com/2020/02/27/stocks-making-the-biggest-moves-after-hours-beyond-meat-autodesk-amc-and-more.html>

2: <https://www.cnbc.com/2020/03/12/stocks-making-the-biggest-moves-after-hours-slack-gap-broadcom-and-more.html>

3: <https://www.cnbc.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html>

Company d. r. horton top 3 articles are:

- 1: <http://techcrunch.com/2020/03/02/apple-agrees-to-settlement-of-up-to-500-million-from-lawsuit-alleging-it-throttled-older-phones/>
- 2: <https://business.financialpost.com/news/retail-marketing/tim-hortons-parent-sued-by-u-s-franchisee-group-in-florida-for-price-gouging>
- 3: <https://www.businessinsider.com/princess-cruises-sued-after-coronavirus-lockdown-near-california-2020-3>

Company lyondellbasell top 3 articles are:

No articles relating to lyondellbasell

Check precision results by asking user to determine if listed article is relevant or not. With a dataset of over 2000 news articles I didn't want to label each one as relevant or not in order to determine recall. Maybe for another time...

```
[12]: # output precision scores of query
      outputPrecisionResults(global_query_top_n, 'global analysis')
```

Is <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/> relevant to company lincoln financial group? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/03/03/ford-confirms-us-built-electric-cargo-van-under-its-11point5-billion-plan.html> relevant to company lincoln financial group? 1 for yes 0 for no0

Is <https://www.theverge.com/2020/3/3/21163669/ford-all-electric-transit-cargo-van-2021> relevant to company lincoln financial group? 1 for yes 0 for no0

Is <https://www.bloomberg.com/news/articles/2020-03-07/u-s-pro-sports-leagues-plan-to-ban-outsiders-from-locker-rooms> relevant to company ameriprise? 1 for yes 0 for no0

Is <https://www.wired.com/2020/03/warp-drive-movies/> relevant to company ameriprise? 1 for yes 0 for no0

Is <https://www.wsj.com/articles/a-bloomberg-business-manifesto-11582586849> relevant to company ameriprise? 1 for yes 0 for no1

Is <https://www.businessinsider.com/coronavirus-testing-covid-19-tests-per-capita-chart-us-behind-2020-3> relevant to company state street? 1 for yes 0 for no0

Is <https://arstechnica.com/science/2020/03/118-us-coronavirus-cases-9-deaths-as-ramped-up-testing-uncovers-hidden-spread/> relevant to company state street? 1 for yes 0 for no0

Is <https://www.businessinsider.com/us-cdc-conducted-fewer-than-6000-coronavirus-tests-2020-3> relevant to company state street? 1 for yes 0 for no0

Is <https://www.theverge.com/2020/3/7/21169637/italy-lockdown-lombardy-coronavirus-outbreak> relevant to company northern trust? 1 for yes 0 for no0

Is <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3> relevant to company northern trust? 1 for yes 0 for no0

Is <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/> relevant to company northern trust? 1 for yes 0 for no0

Is <https://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/> relevant to company oracle? 1 for yes 0 for no1

Is <http://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/> relevant to company oracle? 1 for yes 0 for no1

Is <https://www.businessinsider.com/salesforce-ceo-keith-stong-operator-sudden-exit-oracle-2020-2> relevant to company oracle? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch->

fix-vail-resorts-marathon-oil-and-more.html relevant to company occidental petroleum? 1 for yes 0 for no1

Is <https://www.wsj.com/articles/carl-icahn-boosts-occidental-stake-to-almost-10-as-shares-plummet-11583969050> relevant to company occidental petroleum? 1 for yes 0 for no1

Is <https://www.ccn.com/the-u-s-stock-market-crash-will-bankrupt-these-3-companies/> relevant to company occidental petroleum? 1 for yes 0 for no1

Is <https://www.cnbc.com/video/2020/03/03/this-is-an-opportunity-to-upgrade-your-portfolio-morgan-stanleys-wilson.html> relevant to company morgan stanley? 1 for yes 0 for no1

Is <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3> relevant to company morgan stanley? 1 for yes 0 for no1

Is <https://www.businessinsider.com/us-coronavirus-cases-linked-quarantined-nile-river-cruise-ship-2020-3> relevant to company morgan stanley? 1 for yes 0 for no0

Is <https://www.businessinsider.com/how-to-make-a-slack-channel-private> relevant to company devon energy? 1 for yes 0 for no0

Is <https://business.financialpost.com/commodities/alberta-eyes-curtailment-to-ensure-a-survival-price-as-province-stares-down-recession> relevant to company devon energy? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-add-emojis-to-discord> relevant to company devon energy? 1 for yes 0 for no0

Is <https://www.businessinsider.com/de-blasio-blasts-buttigieg-after-his-third-place-nevada-finish-2020-2> relevant to company newmont? 1 for yes 0 for no0

Is <https://www.infomoney.com.br/mercados/track-field-prima-foods-e-alphaville-urbanismo-entram-com-pedidos-de-ipo/> relevant to company newmont? 1 for yes 0 for no0

Is <https://www.infomoney.com.br/politica/bolsonaro-sugere-adiamento-de-manifestacoes-de-15-de-marco/> relevant to company newmont? 1 for yes 0 for no0

Is <http://techcrunch.com/2020/03/13/sponsors-could-be-key-to-solving-techs-lack-of-gender-diversity/> relevant to company capri holdings? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/26/mike-bloombergs-campaign-holding-talks-with-unions.html> relevant to company capri holdings? 1 for yes 0 for no0

Is <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2> relevant to company capri holdings? 1 for yes 0 for no0

Is <https://www.cnbc.com/video/2020/02/26/coronavirus-outbreak-will-dent-earnings-for-current-quarter-hp-ceo.html> relevant to company hp? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html> relevant to company hp? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/26/hp-ceo-forecasts-coronavirus-will-dent-quarterly-earnings-by-8-cents.html> relevant to company hp? 1 for yes 0 for no1

Is <https://www.businessinsider.com/amazon-just-walk-out-technology-questions-remain-2020-3> relevant to company lennar? 1 for yes 0 for no1

Is <https://thenextweb.com/plugged/2020/03/10/leak-pixel-4a-hands-on-video-leaves-little-for-google-to-announce/> relevant to company lennar? 1 for yes 0 for no0

Is <https://www.businessinsider.com/samsung-galaxy-s20-plus-review> relevant to company lennar? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/03/06/coronavirus-live-updates-sxsw-canceled-grand-princess-cruise-ship-has-21-cases.html> relevant to company public service enterprise group? 1 for yes 0 for no0

Is <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere> relevant to company public service enterprise group? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/23/saudi-arabia-will-support-lebanon-if-it-sees-viable-reform-plan.html> relevant to company public service enterprise group? 1 for yes 0 for no0

Is <https://thenextweb.com/neural/2020/03/07/the-case-for-an-ai-that-puts-nature-and-ethics-first-not-humans-syndication/> relevant to company edwards lifesciences? 1 for yes 0 for no1

Is <https://thenextweb.com/neural/2020/03/01/ai-will-dominate-the-2020s-governments-must-ensure-its-ethical/> relevant to company edwards lifesciences? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-shop-for-perfume> relevant to company edwards lifesciences? 1 for yes 0 for no0

Is <https://www.businessinsider.com/where-are-firefox-bookmarks-stored> relevant to company fox (class b)? 1 for yes 0 for no0

Is <https://www.businessinsider.com/coronavirus-school-closures-have-left-kids-feeling-scared-confused-2020-3> relevant to company fox (class b)? 1 for yes 0 for no1

Is <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere> relevant to company fox (class b)? 1 for yes 0 for no1

Is <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2> relevant to company h&r block? 1 for yes 0 for no0

Is <https://www.businessinsider.com/salesforce-keith-block-marc-benioff-salary-2020-2> relevant to company h&r block? 1 for yes 0 for no1

Is <https://www.techradar.com/news/best-soundbar-for-under-dollar200pound200> relevant to company h&r block? 1 for yes 0 for no0

Is <https://techcrunch.com/2020/03/03/los-angeles-gets-a-new-consumer-fund-as-founders-of-sprinkles-seek-25-million-for-cn2-ventures/> relevant to company charles schwab? 1 for yes 0 for no1

Is <https://www.techradar.com/news/f1-live-stream-how-to-watch-the-australian-grand-prix-2020-online-from-anywhere> relevant to company charles schwab? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/27/dow-futures-fall-100-points-after-another-massive-rout-amid-coronavirus-fears.html> relevant to company charles schwab? 1 for yes 0 for no1

Is <https://www.cnbc.com/select/using-a-credit-card-to-donate-to-political-campaigns/> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-get-rid-of-metal-credit-cards> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.cnbc.com/select/best-credit-cards-for-home-improvements/> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.ccn.com/virtual-racers-like-james-baldwin-might-soon-replace-traditional-racing-drivers/> relevant to company raymond james? 1 for yes 0 for no0

Is <https://www.ccn.com/zion-williamson-could-surpass-lebron-james-if-he-fixes-these-three-things/> relevant to company raymond james? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/03/05/jpmorgan-says-ceo-jamie-dimon-is-recuperating-after-emergency-heart-surgery.html> relevant to company raymond james? 1 for yes 0 for no1

Is <https://business.financialpost.com/news/economy/canadian-recession-is-likely-without-fiscal-stimulus-scotiabank> relevant to company bank of america? 1 for yes 0 for no1

Is <https://www.afr.com/companies/financial-services/cba-ready-to-defer-small-business-loan-repayments-fees-20200311-p548va> relevant to company bank of america? 1 for yes 0 for no1

Is <https://business.financialpost.com/news/fp-street/rbc-to-cut-canada-prime-rate-to-3-45-rivals-follow> relevant to company bank of america? 1 for yes 0 for no1

Is <https://business.financialpost.com/commodities/alberta-eyes-curtailment-to-ensure-a-survival-price-as-province-stares-down-recession> relevant to company apache? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/09/cramer-9-or-10-oil-companies-may-go-bankrupt-amid-crude-declines.html> relevant to company apache? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html> relevant to company apache? 1 for yes 0 for

no1

Is <https://fortune.com/2020/02/29/coronavirus-travel-wealthy-avoid-airlines-flying-private-jets/> relevant to company expedia? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/11/coronavirus-travel-industry-could-lose-24-billion-in-tourism-from-outside-us.html> relevant to company expedia? 1 for yes 0 for no1

Is <https://www.businessinsider.com/private-jet-travel-surge-evacuate-coronavirus-outbreak-covid19-travel-aviation-2020> relevant to company expedia? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/28/coronavirus-latest-updates-china-hubei.html> relevant to company united airlines? 1 for yes 0 for no1

Is <https://www.businessinsider.com/coronavirus-changing-canceling-flights-united-jetblue-american-delta-southwest-2020-3> relevant to company united airlines? 1 for yes 0 for no1

Is <https://www.businessinsider.com/coronavirus-united-airlines-cost-cutting-cancellations-hiring-freeze-unpaid-leave-2020-3> relevant to company united airlines? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/27/stocks-making-the-biggest-moves-after-hours-beyond-meat-autodesk-amc-and-more.html> relevant to company broadcom? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/12/stocks-making-the-biggest-moves-after-hours-slack-gap-broadcom-and-more.html> relevant to company broadcom? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html> relevant to company broadcom? 1 for yes 0 for no1

Is <http://techcrunch.com/2020/03/02/apple-agrees-to-settlement-of-up-to-500-million-from-lawsuit-alleging-it-throttled-older-phones/> relevant to company d. r. horton? 1 for yes 0 for no1

Is <https://business.financialpost.com/news/retail-marketing/tim-hortons-parent-sued-by-u-s-franchisee-group-in-florida-for-price-gouging> relevant to company d. r. horton? 1 for yes 0 for no0

Is <https://www.businessinsider.com/princess-cruises-sued-after-coronavirus-lockdown-near-california-2020-3> relevant to company d. r. horton? 1 for yes 0 for no0

Precision score overall for global analysis is 0.5466666666666666

The precision score is OK, seems to do well for stocks the general public knows, but not so well for the lesser known ones. Maybe local analysis will be better. Use same sequence of calls as for the test data set to get the results of a local analysis expansion to the query terms.

```
[15]: # local analysis results
# get top n related terms for only documents that are returned on the initial query

# get a numpy array of relevant docs + queries and their associated terms
term_doc_queries_array = queryDocuments(stemmed_query_terms, global_term_doc_df)

# get results of local query
query_results = queryResults(stemmed_query_terms, term_doc_queries_array)

# get only documents applicable to local query
local_query_documents = getRelevantDocuments(query_results)

# for each query do a local expansion
local_query_results_combined = {}

for name, indices in local_query_documents.items():
    # get entire df and co-occurrence matrix
```

```

local_term_doc_df = getTermDocDF(document_tf_idf_dict, indices)
local_term_cooccur = getTermCoOccurMatrix(local_term_doc_df)

# expand query to include all related terms
local_expanded_query_dict = expandQuery(local_term_cooccur, local_term_doc_df,
                                         {name: stemmed_query_terms[name]})

# get a numpy array of all documents + queries and their associated terms
local_term_doc_queries_array = queryDocuments(local_expanded_query_dict,
↪local_term_doc_df)

# get results of expanded query
local_query_results = queryResults(local_expanded_query_dict,
↪local_term_doc_queries_array)

# get list of local query values, filter out companies with none
local_val_list = list(local_query_results.values())[0]
if not any(x > 0 for x in local_val_list):
    continue

# set results in an array that spans the entire term collection
full_array_query_results = expandToFullArray(local_val_list, global_term_doc_df,
↪indices)

local_query_results_combined[name] = full_array_query_results

```

Final Project Functions.py:100: RuntimeWarning: invalid value encountered in double_scalars

```

np.dot(term_doc_queries_array[query_loc], row) /
(norm(term_doc_queries_array[query_loc]) * norm(row))

```

[16]:

```

# output results of query
local_query_top_n = outputQueryResults(local_query_results_combined,
↪global_term_doc_df)

# output precision scores of query
outputPrecisionResults(local_query_top_n, 'local analysis')

```

Company lincoln financial group top 3 articles are:

- 1: <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2>
- 2: <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/>
- 3: <https://www.theverge.com/2020/3/3/21163669/ford-all-electric-transit-cargo-van-2021>

Company ameriprise top 3 articles are:

- 1: <https://www.cnbc.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response-pelosi-says.html>
- 2: <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management->

platform-swiftstack/

3: <https://www.wsj.com/articles/a-bloomberg-business-manifesto-11582586849>

Company state street top 3 articles are:

1: <https://www.businessinsider.com/coronavirus-testing-covid-19-tests-per-capita-chart-us-behind-2020-3>

2: <https://arstechnica.com/science/2020/03/118-us-coronavirus-cases-9-deaths-as-ramped-up-testing-uncovers-hidden-spread/>

3: <https://www.businessinsider.com/us-cdc-conducted-fewer-than-6000-coronavirus-tests-2020-3>

Company northern trust top 3 articles are:

1: <https://www.theverge.com/2020/3/7/21169637/italy-lockdown-lombardy-coronavirus-outbreak>

2: <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/>

3: <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3>

Company oracle top 3 articles are:

1: <https://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/>

2: <http://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/>

3: <https://www.businessinsider.com/salesforce-ceo-keith-stong-operator-sudden-exit-oracle-2020-2>

Company occidental petroleum top 3 articles are:

1: <https://www.cnn.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html>

2: <https://www.wsj.com/articles/carl-icahn-boosts-occidental-stake-to-almost-10-as-shares-plummet-11583969050>

3: <https://www.cnn.com/the-u-s-stock-market-crash-will-bankrupt-these-3-companies/>

Company morgan stanley top 3 articles are:

1: <https://www.cnn.com/video/2020/03/03/this-is-an-opportunity-to-upgrade-your-portfolio-morgan-stanleys-wilson.html>

2: <https://www.businessinsider.com/us-coronavirus-cases-linked-quarantined-nile-river-cruise-ship-2020-3>

3: <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3>

Company devon energy top 3 articles are:

1: <https://www.businessinsider.com/how-to-make-a-slack-channel-private>

2: <https://business.financialpost.com/commodities/alberta-eyes-curtailed-to-ensure-a-survival-price-as-province-stares-down-recession>

3: <https://www.businessinsider.com/how-to-add-emojis-to-discord>

Company newmont top 3 articles are:

1: <https://www.cnn.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response-pelosi-says.html>

2: <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management-platform-swiftstack/>

3: <https://www.infomoney.com.br/mercados/track-field-prima-foods-e-alphaville-urbanismo-entram-com-pedidos-de-ipo/>

Company capri holdings top 3 articles are:

1: <http://techcrunch.com/2020/03/13/sponsors-could-be-key-to-solving-techs-lack-of-gender-diversity/>

2: <https://www.cnbc.com/2020/02/26/mike-bloombergs-campaign-holding-talks-with-unions.html>

3: <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2>

Company hp top 3 articles are:

1: <https://www.cnbc.com/video/2020/02/26/coronavirus-outbreak-will-dent-earnings-for-current-quarter-hp-ceo.html>

2: <https://www.cnbc.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html>

3: <https://www.cnbc.com/2020/02/26/hp-ceo-forecasts-coronavirus-will-dent-quarterly-earnings-by-8-cents.html>

Company lennar top 3 articles are:

1: <http://techcrunch.com/2020/02/25/samsungs-galaxy-s20-ultra-is-a-lot-of-phone-for-a-lot-of-money/>

2: <https://www.businessinsider.com/samsung-galaxy-s20-plus-review>

3: <https://www.businessinsider.com/amazon-just-walk-out-technology-questions-remain-2020-3>

Company public service enterprise group top 3 articles are:

1: <https://www.cnbc.com/2020/03/06/coronavirus-live-updates-sxsw-canceled-grand-princess-cruise-ship-has-21-cases.html>

2: <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere>

3: <https://www.cnbc.com/2020/02/23/saudi-arabia-will-support-lebanon-if-it-sees-viable-reform-plan.html>

Company edwards lifesciences top 3 articles are:

1: <https://www.cnbc.com/2020/03/08/plummeting-oil-prices-and-mortgage-rates-could-boost-consumers.html>

2: <https://www.businessinsider.com/how-to-shop-for-perfume>

3: <https://thenextweb.com/neural/2020/03/07/the-case-for-an-ai-that-puts-nature-and-ethics-first-not-humans-syndication/>

Company fox (class b) top 3 articles are:

1: <https://www.businessinsider.com/where-are-firefox-bookmarks-stored>

2: <https://www.businessinsider.com/coronavirus-school-closures-have-left-kids-feeling-scared-confused-2020-3>

3: <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere>

Company h&r block top 3 articles are:

1: <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2>

2: <https://www.businessinsider.com/salesforce-keith-block-marc-benioff->

salary-2020-2

3: <https://www.techradar.com/news/best-soundbar-for-under-dollar200pound200>

Company charles schwab top 3 articles are:

1: <https://techcrunch.com/2020/03/03/los-angeles-gets-a-new-consumer-fund-as-founders-of-sprinkles-seek-25-million-for-cn2-ventures/>

2: <https://www.techradar.com/news/f1-live-stream-how-to-watch-the-australian-grand-prix-2020-online-from-anywhere>

3: <https://www.cnbc.com/2020/02/27/dow-futures-fall-100-points-after-another-massive-rout-amid-coronavirus-fears.html>

Company jpmorgan chase top 3 articles are:

1: <https://www.cnbc.com/select/using-credit-cards-during-coronavirus/>

2: <https://www.businessinsider.com/how-to-get-rid-of-metal-credit-cards>

3: <https://www.cnbc.com/select/best-credit-cards-for-home-improvements/>

Company raymond james top 3 articles are:

1: <https://www.ccn.com/virtual-racers-like-james-baldwin-might-soon-replace-traditional-racing-drivers/>

2: <https://www.cnbc.com/2020/03/05/jpmorgan-says-ceo-jamie-dimon-is-recuperating-after-emergency-heart-surgery.html>

3: <https://www.ccn.com/zion-williamson-could-surpass-lebron-james-if-he-fixes-these-three-things/>

Company bank of america top 3 articles are:

1: <https://business.financialpost.com/news/economy/canadian-recession-is-likely-without-fiscal-stimulus-scotiabank>

2: <https://www.afr.com/companies/financial-services/cba-ready-to-defer-small-business-loan-repayments-fees-20200311-p548va>

3: <https://business.financialpost.com/news/fp-street/rbc-to-cut-canada-prime-rate-to-3-45-rivals-follow>

Company apache top 3 articles are:

1: <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management-platform-swiftstack/>

2: <https://www.cnbc.com/2020/03/02/google-says-cloud-next-will-be-held-virtually-as-coronavirus-spreads.html>

3: <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html>

Company expedia top 3 articles are:

1: <https://www.cnbc.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response-pelosi-says.html>

2: <https://www.cnbc.com/2020/03/11/coronavirus-travel-industry-could-lose-24-billion-in-tourism-from-outside-us.html>

3: <https://www.businessinsider.com/private-jet-travel-surge-evacuate-coronavirus-outbreak-covid19-travel-aviation-2020>

Company united airlines top 3 articles are:

1: <https://www.cnbc.com/2020/02/28/coronavirus-latest-updates-china-hubei.html>

2: <https://www.businessinsider.com/coronavirus-changing-canceling-flights-united-jetblue-american-delta-southwest-2020-3>

3: <https://www.businessinsider.com/coronavirus-united-airlines-cost-cutting-cancellations-hiring-freeze-unpaid-leave-2020-3>

Company broadcom top 3 articles are:

1: <https://www.wsj.com/articles/tech-companies-see-mixed-impact-from-coronavirus-11584050206>

2: <https://www.theverge.com/2020/3/2/21162061/waymo-outside-investment-amount-autonation-magna-silver-lake>

3: <https://www.cnbc.com/2020/03/12/stocks-making-the-biggest-moves-after-hours-slack-gap-broadcom-and-more.html>

Company d. r. horton top 3 articles are:

1: <https://business.financialpost.com/news/ottawa-to-fly-home-canadians-stranded-on-cruise-ship-as-number-of-cases-across-country-increases>

2: <https://business.financialpost.com/news/canadian-coronavirus-infections-rising-as-alberta-ontario-and-quebec-announce-confirmed-and-possible-new-cases>

3: <https://business.financialpost.com/news/retail-marketing/tim-hortons-parent-sued-by-u-s-franchisee-group-in-florida-for-price-gouging>

Is <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2> relevant to company lincoln financial group? 1 for yes 0 for no

Is <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/> relevant to company lincoln financial group? 1 for yes 0 for no

Is <https://www.theverge.com/2020/3/3/21163669/ford-all-electric-transit-cargo-van-2021> relevant to company lincoln financial group? 1 for yes 0 for no

Is <https://www.cnbc.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response-pelosi-says.html> relevant to company ameriprise? 1 for yes 0 for no

Is <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management-platform-swiftstack/> relevant to company ameriprise? 1 for yes 0 for no

Is <https://www.wsj.com/articles/a-bloomberg-business-manifesto-11582586849> relevant to company ameriprise? 1 for yes 0 for no

Is <https://www.businessinsider.com/coronavirus-testing-covid-19-tests-per-capita-chart-us-behind-2020-3> relevant to company state street? 1 for yes 0 for no

Is <https://arstechnica.com/science/2020/03/118-us-coronavirus-cases-9-deaths-as-ramped-up-testing-uncovers-hidden-spread/> relevant to company state street? 1 for yes 0 for no

Is <https://www.businessinsider.com/us-cdc-conducted-fewer-than-6000-coronavirus-tests-2020-3> relevant to company state street? 1 for yes 0 for no

Is <https://www.theverge.com/2020/3/7/21169637/italy-lockdown-lombardy-coronavirus-outbreak> relevant to company northern trust? 1 for yes 0 for no

Is <https://thenextweb.com/syndication/2020/03/08/why-we-need-more-women-to-build-real-world-ai-products-explained-by-science/> relevant to company northern trust? 1 for yes 0 for no

Is <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3> relevant to company northern trust? 1 for yes 0 for no

Is <https://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/> relevant to company oracle? 1 for yes 0 for no

Is <http://techcrunch.com/2020/02/25/salesforce-co-ceo-keith-block-steps-down/> relevant to company oracle? 1 for yes 0 for no

Is <https://www.businessinsider.com/salesforce-ceo-keith-stong-operator-sudden-exit-oracle-2020-2> relevant to company oracle? 1 for yes 0 for no

Is <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html> relevant to company occidental petroleum? 1

for yes 0 for no1

Is <https://www.wsj.com/articles/carl-icahn-boosts-occidental-stake-to-almost-10-as-shares-plummet-11583969050> relevant to company occidental petroleum? 1 for yes 0 for no1

Is <https://www.ccn.com/the-u-s-stock-market-crash-will-bankrupt-these-3-companies/> relevant to company occidental petroleum? 1 for yes 0 for no1

Is <https://www.cnbc.com/video/2020/03/03/this-is-an-opportunity-to-upgrade-your-portfolio-morgan-stanleys-wilson.html> relevant to company morgan stanley? 1 for yes 0 for no1

Is <https://www.businessinsider.com/us-coronavirus-cases-linked-quarantined-nile-river-cruise-ship-2020-3> relevant to company morgan stanley? 1 for yes 0 for no0

Is <https://www.businessinsider.com/the-points-guy-founder-brian-kelly-networth-extravagant-life-spending-2020-3> relevant to company morgan stanley? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-make-a-slack-channel-private> relevant to company devon energy? 1 for yes 0 for no0

Is <https://business.financialpost.com/commodities/alberta-eyes-curtailement-to-ensure-a-survival-price-as-province-stares-down-recession> relevant to company devon energy? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-add-emojis-to-discord> relevant to company devon energy? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response-pelosi-says.html> relevant to company newmont? 1 for yes 0 for no0

Is <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management-platform-swiftstack/> relevant to company newmont? 1 for yes 0 for no0

Is <https://www.infomoney.com.br/mercados/track-field-prima-foods-e-alphaville-urbanismo-entram-com-pedidos-de-ipo/> relevant to company newmont? 1 for yes 0 for no0

Is <http://techcrunch.com/2020/03/13/sponsors-could-be-key-to-solving-techs-lack-of-gender-diversity/> relevant to company capri holdings? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/26/mike-bloombergs-campaign-holding-talks-with-unions.html> relevant to company capri holdings? 1 for yes 0 for no1

Is <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2> relevant to company capri holdings? 1 for yes 0 for no0

Is <https://www.cnbc.com/video/2020/02/26/coronavirus-outbreak-will-dent-earnings-for-current-quarter-hp-ceo.html> relevant to company hp? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/24/stocks-making-the-biggest-moves-after-hours-shake-shack-hp-moderna-and-more.html> relevant to company hp? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/26/hp-ceo-forecasts-coronavirus-will-dent-quarterly-earnings-by-8-cents.html> relevant to company hp? 1 for yes 0 for no1

Is <http://techcrunch.com/2020/02/25/samsungs-galaxy-s20-ultra-is-a-lot-of-phone-for-a-lot-of-money/> relevant to company lennar? 1 for yes 0 for no0

Is <https://www.businessinsider.com/samsung-galaxy-s20-plus-review> relevant to company lennar? 1 for yes 0 for no0

Is <https://www.businessinsider.com/amazon-just-walk-out-technology-questions-remain-2020-3> relevant to company lennar? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/06/coronavirus-live-updates-sxsw-canceled-grand-princess-cruise-ship-has-21-cases.html> relevant to company public service enterprise group? 1 for yes 0 for no0

Is <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere> relevant to company public service enterprise group? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/02/23/saudi-arabia-will-support-lebanon-if-it-sees-viable-reform-plan.html> relevant to company public service enterprise group? 1 for yes 0 for no0

Is [19](https://www.cnbc.com/2020/03/08/plummeting-oil-prices-and-mortgage-rates-could-boost-</p>
</div>
<div data-bbox=)

consumers.html relevant to company edwards lifesciences? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-shop-for-perfume> relevant to company edwards lifesciences? 1 for yes 0 for no0

Is <https://thenextweb.com/neural/2020/03/07/the-case-for-an-ai-that-puts-nature-and-ethics-first-not-humans-syndication/> relevant to company edwards lifesciences? 1 for yes 0 for no1

Is <https://www.businessinsider.com/where-are-firefox-bookmarks-stored> relevant to company fox (class b)? 1 for yes 0 for no0

Is <https://www.businessinsider.com/coronavirus-school-closures-have-left-kids-feeling-scared-confused-2020-3> relevant to company fox (class b)? 1 for yes 0 for no1

Is <https://www.techradar.com/news/nba-live-stream-how-to-watch-every-game-online-in-2020-from-anywhere> relevant to company fox (class b)? 1 for yes 0 for no1

Is <https://www.businessinsider.com/bloomberg-created-culture-of-cruelty-and-harassment-2020-2> relevant to company h&r block? 1 for yes 0 for no0

Is <https://www.businessinsider.com/salesforce-keith-block-marc-benioff-salary-2020-2> relevant to company h&r block? 1 for yes 0 for no1

Is <https://www.techradar.com/news/best-soundbar-for-under-dollar200pound200> relevant to company h&r block? 1 for yes 0 for no0

Is <https://techcrunch.com/2020/03/03/los-angeles-gets-a-new-consumer-fund-as-founders-of-sprinkles-seek-25-million-for-cn2-ventures/> relevant to company charles schwab? 1 for yes 0 for no1

Is <https://www.techradar.com/news/f1-live-stream-how-to-watch-the-australian-grand-prix-2020-online-from-anywhere> relevant to company charles schwab? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/27/dow-futures-fall-100-points-after-another-massive-rout-amid-coronavirus-fears.html> relevant to company charles schwab? 1 for yes 0 for no1

Is <https://www.cnbc.com/select/using-credit-cards-during-coronavirus/> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.businessinsider.com/how-to-get-rid-of-metal-credit-cards> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.cnbc.com/select/best-credit-cards-for-home-improvements/> relevant to company jpmorgan chase? 1 for yes 0 for no1

Is <https://www.ccn.com/virtual-racers-like-james-baldwin-might-soon-replace-traditional-racing-drivers/> relevant to company raymond james? 1 for yes 0 for no0

Is <https://www.cnbc.com/2020/03/05/jpmorgan-says-ceo-jamie-dimon-is-recuperating-after-emergency-heart-surgery.html> relevant to company raymond james? 1 for yes 0 for no0

Is <https://www.ccn.com/zion-williamson-could-surpass-lebron-james-if-he-fixes-these-three-things/> relevant to company raymond james? 1 for yes 0 for no0

Is <https://business.financialpost.com/news/economy/canadian-recession-is-likely-without-fiscal-stimulus-scotiabank> relevant to company bank of america? 1 for yes 0 for no1

Is <https://www.afr.com/companies/financial-services/cba-ready-to-defer-small-business-loan-repayments-fees-20200311-p548va> relevant to company bank of america? 1 for yes 0 for no1

Is <https://business.financialpost.com/news/fp-street/rbc-to-cut-canada-prime-rate-to-3-45-rivals-follow> relevant to company bank of america? 1 for yes 0 for no1

Is <http://techcrunch.com/2020/03/05/nvidia-acquires-data-storage-and-management-platform-swiftstack/> relevant to company apache? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/02/google-says-cloud-next-will-be-held-virtually-as-coronavirus-spreads.html> relevant to company apache? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/03/09/stocks-making-the-biggest-moves-after-hours-stitch-fix-vail-resorts-marathon-oil-and-more.html> relevant to company apache? 1 for yes 0 for no1

Is <https://www.cnbc.com/2020/02/29/house-to-vote-on-funding-for-coronavirus-response->

pelosi-says.html relevant to company expedia? 1 for yes 0 for no1
 Is <https://www.cnbc.com/2020/03/11/coronavirus-travel-industry-could-lose-24-billion-in-tourism-from-outside-us.html> relevant to company expedia? 1 for yes 0 for no1
 Is <https://www.businessinsider.com/private-jet-travel-surge-evacuate-coronavirus-outbreak-covid19-travel-aviation-2020> relevant to company expedia? 1 for yes 0 for no1
 Is <https://www.cnbc.com/2020/02/28/coronavirus-latest-updates-china-hubei.html> relevant to company united airlines? 1 for yes 0 for no1
 Is <https://www.businessinsider.com/coronavirus-changing-canceling-flights-united-jetblue-american-delta-southwest-2020-3> relevant to company united airlines? 1 for yes 0 for no1
 Is <https://www.businessinsider.com/coronavirus-united-airlines-cost-cutting-cancellations-hiring-freeze-unpaid-leave-2020-3> relevant to company united airlines? 1 for yes 0 for no1
 Is <https://www.wsj.com/articles/tech-companies-see-mixed-impact-from-coronavirus-11584050206> relevant to company broadcom? 1 for yes 0 for no1
 Is <https://www.theverge.com/2020/3/2/21162061/waymo-outside-investment-amount-autonation-magna-silver-lake> relevant to company broadcom? 1 for yes 0 for no1
 Is <https://www.cnbc.com/2020/03/12/stocks-making-the-biggest-moves-after-hours-slack-gap-broadcom-and-more.html> relevant to company broadcom? 1 for yes 0 for no1
 Is <https://business.financialpost.com/news/ottawa-to-fly-home-canadians-stranded-on-cruise-ship-as-number-of-cases-across-country-increases> relevant to company d. r. horton? 1 for yes 0 for no0
 Is <https://business.financialpost.com/news/canadian-coronavirus-infections-rising-as-alberta-ontario-and-quebec-announce-confirmed-and-possible-new-cases> relevant to company d. r. horton? 1 for yes 0 for no0
 Is <https://business.financialpost.com/news/retail-marketing/tim-hortons-parent-sued-by-u-s-franchisee-group-in-florida-for-price-gouging> relevant to company d. r. horton? 1 for yes 0 for no0

Precision score overall for local analysis is 0.56

Unfortunately local analysis doesn't do that much better, with only a handful of different top 3 articles for each of the stocks. It seems that the issue is with the corpus of news articles and not the algorithm itself. Many stocks don't have one article that mentions them by name, which is surprising for a group of stocks that either went up or down the most in a trading day! I think online article writers will always tend to write about stocks the general public knows regardless of if they were one of the most volatile stocks on that day. I'm not surprised that queries for Oracle, United, JP Morgan, etc.. did well but queries for stocks like COTY, Ameriprise, Ventas, etc.. did poorly. I think if I did this project again with the old stocks from my portfolio, which was mainly 'blue chip' well-known stocks; it would perform better. However, I am pleased with how I managed to combine all the moving parts to make a functional algorithm that lets users know the best articles related to the stocks from their Robinhood portfolios.