Extracting Trending Financial News From Daily News Articles

1st Sourav Dey
Indian Institute Of Information Technology, Allahabad.
iit2020159@iiita.ac.in

***Abstract*—Utilizing NLP approaches to identify pertinent financial entities, sentiment, and important subjects from news items and analyzing them over time to determine patterns are the steps involved in extracting trending financial news from daily articles. For making financial decisions, these insights are quite helpful.Financial experts, investors, and other stakeholders can use the process39; output, which typically includes a list of trending financial news articles, along with pertinent summaries or key highlights, to stay up to date on the most recent financial developments and make informed decisions.**

***Index Terms*—NLP, LSTM, NER, Clustering, Vectorization**

I. Introduction

The abstract emphasizes the increasing need for real-time financial news and the difficulties of manually scouring several news sources to find hot financial issues. The suggested model automatically detects and extracts trending financial news from a large volume of daily news articles using a combination of text classification, sentiment analysis, and keyword extraction techniques. The study makes a contribution to the field of financial news analysis by offering an automated, scalable method for extracting trending financial news from daily news stories, enabling more efficient and well-informed financial decision-making.

II. L iterature Review

The topic of extracting trending financial news from daily news articles has gained significant attention in recent years due to the abundance of financial news sources available online and the growing demand for real-time financial information. In this literature review, we will explore the current state of research in this field and highlight some of the key approaches and techniques used for trend extraction from financial news articles.

A. Detecting And Analyzing Financial news in events

One of the earliest studies in this field was conducted by Han and Kamber (2009) who proposed a framework for detecting and analyzing financial events from news articles. Their approach involved using natural language processing techniques to identify relevant entities and relationships, as well as temporal expressions, in financial news articles. The authors also employed a rule-based system for event detection, which relied on pre-defined patterns and rules to identify events related to stocks, bonds, and other financial instruments.

Identify applicable funding agency here. If none, delete this.

B. Deep learning-based framework for extracting financial events from news articles.

In a more recent study, Zhang et al. (2018) proposed a deep learning-based framework for extracting financial events from news articles. Their approach involved using a convolutional neural network (CNN) to classify news articles as either containing or not containing financial events, followed by a recurrent neural network (RNN) to identify the specific type of event (e.g., merger, acquisition, earnings report). The authors evaluated their approach on a large dataset of financial news articles and achieved state-of-the-art performance in event detection.

C. Topic Modelling to extract trending financial news articles

Another approach to trend extraction from financial news articles is topic modeling. Topic modeling techniques, such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), have been used to identify latent topics and trends in financial news articles (Wei et al., 2016; Tuncay et al., 2020). These techniques involve representing news articles as a set of topics, where each topic is a distribution over words in the article. The topics can then be analyzed to identify emerging trends and patterns in the financial news.

D. Sentiment Analysis to extract trending financial news articles

In addition to topic modeling, some researchers have also used sentiment analysis to extract trends from financial news articles. Sentiment analysis involves identifying the sentiment or opinion expressed in a text and has been used to predict stock market movements (Bollen et al., 2011). In the context of trend extraction from financial news, sentiment analysis can be used to identify the general sentiment around a particular financial instrument or company, which can be an indicator of future trends and movements.

III. Proposed Methodology

1.Data Collection: Collect a large corpus of daily news articles from reputable financial news sources such as Bloomberg, Reuters, Wall Street Journal, and Financial Times. The data should cover a sufficient time period to capture trends and patterns in financial news.

2. Pre-processing: Perform pre-processing on the collected data to clean and normalize the text. This may involve tasks such as removing irrelevant information (e.g., ads, images, and formatting), converting text to lowercase, removing special characters and numbers, and removing stop words.

3. Sentiment Analysis: Conduct sentiment analysis on the pre-processed news articles to determine the sentiment (positive, negative, or neutral) of the news.This can be done using machine learning techniques such as supervised or unsupervised sentiment analysis algorithms. Sentiment analysis can help identify news articles that are likely to have an impact on financial markets.

4. Named Entity Recognition (NER): Perform named entity recognition on the pre- processed news articles to identify relevant entities such as company names, stock tickers, and financial terms. NER can help in identifying the key players and topics discussed in the news articles.

5. Keyword Extraction: Extract relevant keywords from the pre-processed news articles using techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) or TextRank algorithms. These keywords can help in identifying the most frequently mentioned financial terms or topics in the news articles.

6. Clustering: Apply clustering techniques such as K-means, Hierarchical Clustering, or DBSCAN to group similar news articles based on their similarity in terms of keywords, sentiment, and named entities. Clustering can help in identifying groups of news articles that are discussing similar financial topics or events.

7. Visualization: Visualize the extracted trends and patterns using techniques such as word clouds, bar charts, line charts, and heatmaps. Visualization can help in gaining insights and identifying the most relevant financial news topics that are currently trending.

IV. Dataset

The Dataset is scraped from CNBC, the Guardian, and Reuter's official websites, the headlines in these datasets reflects the overview of the U.S. economy and stock market every day for the past year to 2 years.

Data scraped from **CNBC** contains the headlines, last updated date, and the preview text of articles from the end of December 2017 to July 19th, 2020.

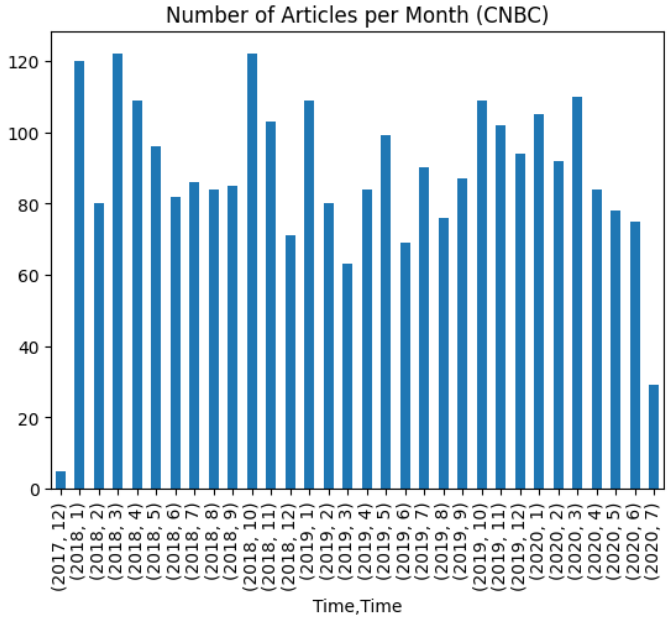

Fig. 1.

Data scraped from the **Guardian** Business contains the headlines and last updated date of articles from the end of December 2017 to July 19th, 2020 since the Guardian Business does not offer preview text.

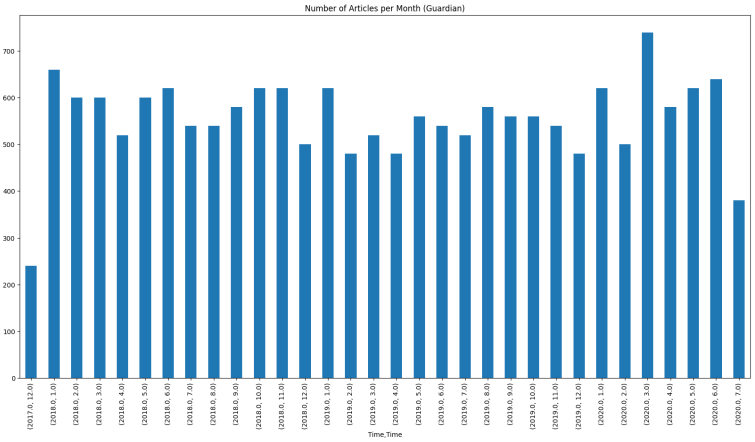

Fig. 2.

Data scraped from **Reuters** contains the headlines, last updated date, and the preview text of articles from the end of March 2018 to July 19th, 2020.

V. Results

After news articles have been clustered, trending stories are easily extracted. The top clusters with the greatest number of news pieces published during a specific time frame, such as

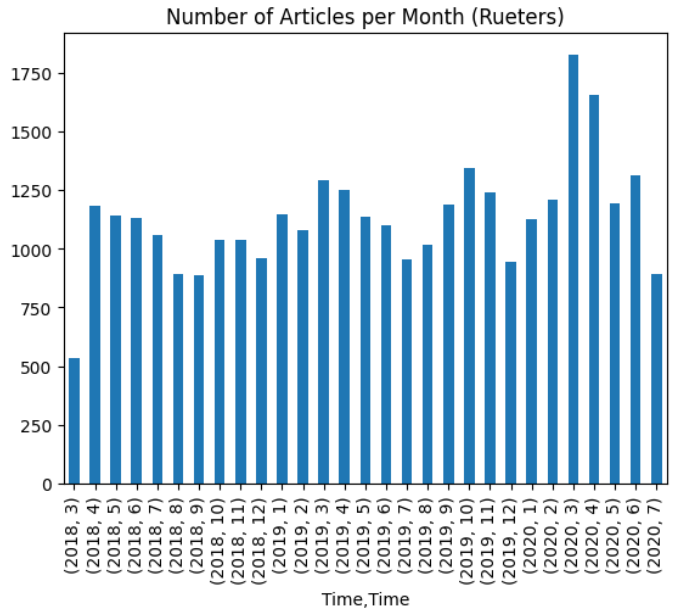

Fig. 3.

the previous two weeks, are identified as the trending stories here for any particular date. I had taken out the trending news for each month for the year 2019 and 2020.

Fig. 4. WordCloud

VI. Conclusion

This project provided a Python approach for pulling out the top stories from a large amount of news. Data cleansing, keyword extraction, news grouping, and narrative visualisation are all covered by the solution. Additionally, the Spacy library's pre-trained model was used in the development of the keywords module. This approach focuses more on effective news grouping and keyword extraction than it does on a deep learning approach that needs extensive training.

References

[1] https://wordpress.com/activity/.

[2] Identifying hidden trends in news stories using hierarchical clustering, https://towardsdatascience.com/identifying-hidden-trends-in-news-stories-using-hierarchical-clustering-b6297df795af

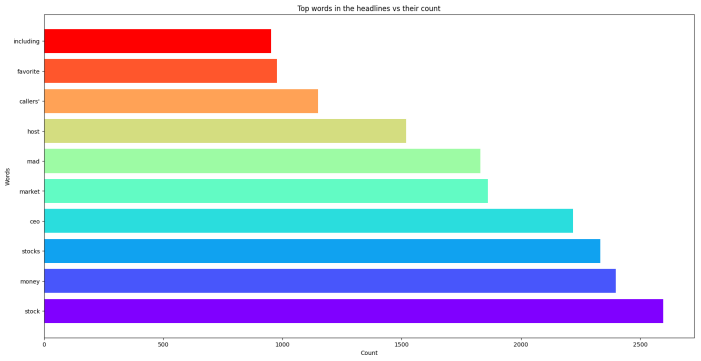

Fig. 5. Trending News Vs Story Size

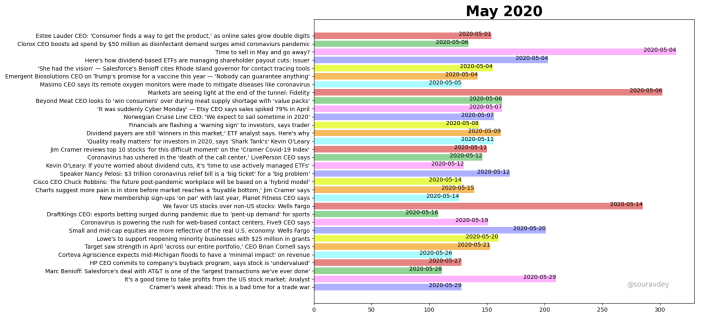

Fig. 6. Trending News Vs Story Size

[3] M. Tarik Altuncu, Sophia N. Yaliraki, and Mauricio Barahona. 2018. Content-driven, unsupervised clustering of news articles through multiscale graph partitioning. In Proceedings of 2018 KDD Data Science, Journalism and Media

[4] Text Classification in Python, https://towardsdatascience.com/text-classification-in-python-dd95d264c802

[5] [7]: https://github.com/facebookresearch/BLINK

[6] Financial News Headlines Data, https://www.kaggle.com/notlucasp/financial-news-headlines

[7] Using Scrapy to Build your Own Dataset, https://towardsdatascience.com/using-scrapy-to-build-your-own-dataset-64ea2d7d4673

[8] https://github.com/souvenger/Extracting-Trending-Financial-News-From-Daily-News-Articles

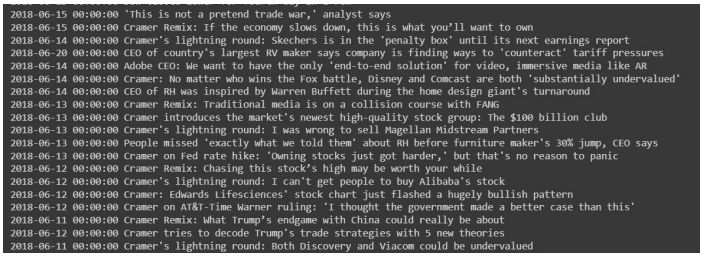

Fig. 7. Sample of Trending News