

# Vendor News Analytics

Anup Bera<sup>1\*</sup> and Sujaya Kundu<sup>2</sup>

<sup>1,2</sup> Accenture, India

## Abstract

Vendor management plays a pivotal role in the seamless functioning of modern businesses. In today's dynamic landscape, organizations heavily rely on external vendors to meet their diverse needs. However, effectively monitoring vendor performance, evaluating associated risks, and staying abreast of market dynamics can be daunting tasks, given the deluge of information available from various news sources. To address these challenges, this research proposes a holistic solution that harnesses the power of natural language processing techniques, Generative AI, and machine learning algorithms. By employing these advanced technologies, the proposed solution aims to gather, analyze, and present real-time news data relevant to vendors. This initiative seeks to empower organizations with a cutting-edge Vendor News Analytics platform that offers timely and relevant insights. The envisioned outcome of this project is multifaceted. Firstly, it will facilitate informed decision-making by providing stakeholders with up-to-the-minute intelligence regarding vendor-related developments. Moreover, the solution will enable early identification of potential risks, allowing proactive mitigation measures to be implemented. Additionally, it is anticipated that the platform will foster stronger vendor relationships through enhanced communication and transparency. Furthermore, by bolstering risk management strategies, organizations can better safeguard their interests while assessing overall vendor performance. The results are available at <https://github.com/anupb08/vendor-news-analytics>.

## Keywords

Business events, vendor news analytics, business event extraction, summarization, LLM, NLP.

## 1. Introduction

Large corporations rely significantly on a substantial multitude of external vendors to facilitate their daily operational activities. They face challenges on finding the right vendors, negotiating contracts, managing relationships, ensuring compliance, and monitoring performance. Online news is very important source of information to track the business activities such as acquisition, investment, trade result, data security breach etc. that may impact future direction of the related enterprises. Organizations rely significantly on news articles to monitor events related to their vendors which may impact the existing relationship or aid in finding suitable prospective vendors. However, effectively monitoring of vendor performance, assessing risks, and staying informed about market

---

*Baltic DB&IS Conference Forum and Doctoral Consortium 2024*

\* Corresponding author.

✉ anupbera@gmail.com (A. Bera); sujaya.kundu@gmail.com (S. Kundu)

🆔 0000-0001-6493-293X (A. Bera); 0009-0007-0214-3952 (S. Kundu)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

dynamics can be challenging due to the vast amount of information available from news sources. It's evident that vendor managers need enhanced visibility to optimize vendor risks, whether in terms of operations, cost implications, or regulatory compliance. It has become essential to have enablement in place for ongoing, real-time automated monitoring of vendor news and extracting insights through advanced analytics. Manually curating news articles and analyzing vendor critical events is inconsistent, inefficient and time consuming. Having automatic tool streamlines the entire process which saves cost, improve efficiency, and reduce decision-making time. Although summary report generation of news documents is studied intensively, few works have been conducted on event-based news summary report generation. As the summarization techniques aim to extract the main ideas of documents, analyzing an event and its context elements from news articles is not far from finding the most important information conveyed in the text. Also incorporating sentiment of the news event in the summary report will provide indication of the business trend of vendor. Therefore, in this study, we propose summary report by combining of event-based summary with business event sentiment. We propose an enhanced event extraction technique using generative AI to filter relevant events. These selected events will then be combined to generate a comprehensive summary report also using generative AI. We offer a fully customizable end-to-end framework, beginning with data collection, proceed through data processing, and generation of final reports.

## **2. Literature Review**

Vendor news analysis refers to the process of extracting valuable insights and intelligence from news articles and publications concerning vendors or companies. To achieve this, researchers have delved into various methodologies including text mining, topic modelling, natural language processing (NLP), sentiment analysis, named entity recognition (NER) and event extraction. For example, Khedr et al. [1] and Khant et al. [2] utilized NLP and AI techniques to analyze vendor news, thereby uncovering emerging trends and market sentiments. Sentiment analysis plays a vital role in understanding the perception of vendors in the market. Patric et al. [3] used several text mining methods for sentiment analysis in financial markets by integrating word association and lexical resources to analyze stock market news reports. Moreover, researchers have employed diverse machine learning models such as Support Vector Machines (SVM), Recurrent Neural Networks (RNN), and Transformer-based models like BERT for sentiment analysis of vendor news [4] [5]. In addition to sentiment analysis, topic modeling techniques such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) have proven instrumental in discerning market trends through the lens of vendor news data [6] [7]. These techniques help in identifying key topics and themes prevalent in news articles related to vendors. Business events play a pivotal role in assessing the performance and charting the future trajectory of associated vendors [8].

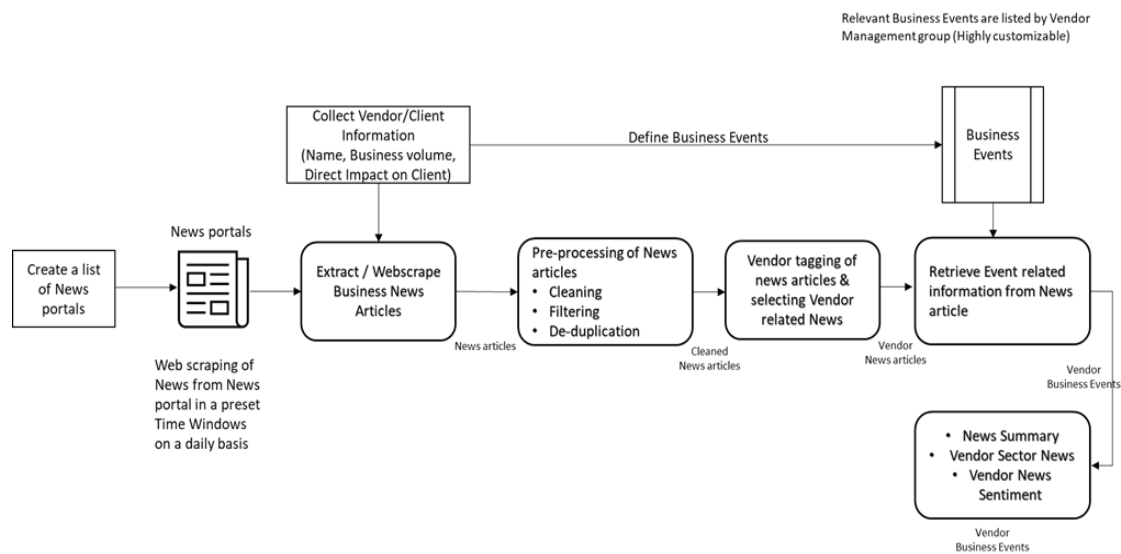
Although there are some works available for news event extraction methods but very few are related to business event extraction space. Earlier works primarily relied on handcrafted features and rule-based systems [9] to identify events and their associated information. However, with the advent of deep learning techniques, the field has

witnessed a paradigm shift towards more data-driven approaches [10] [11] [12] and knowledge based methods [13] [14]. Recent research focuses on employing pre-trained language models, such as BERT [15] and ELMo [16], to capture contextual information and improve event extraction performance. However, these methods delve into very generic kind of event extraction and not any pre-defined event type. In other words, they tried to identify the possible events that may occur in the news article. Liu et al. [17] presented a survey paper where it extensively discussed various event extraction methods and its evaluation techniques. Among these methods, several have been proposed specifically for business event extraction. These include rule-based approaches, machine learning techniques, and deep learning models. Rule-based systems rely on predefined patterns and linguistic rules to extract business events from text [18]. These approaches often require domain-specific knowledge and manual crafting of rules. Machine learning methods, such as Support Vector Machines (SVM) [19] and Conditional Random Fields (CRF), are widely utilized in the domain of business event extraction [20]. Effective entity recognition is essential for accurate event extraction. Named Entity Recognition (NER) systems are often integrated into event extraction pipelines to identify relevant entities [21]. Event classification involves categorizing extracted events into predefined classes or types. This task can be tackled as either a multi-label classification problem or a sequence labeling task [22]. These models learn patterns and features from annotated data. Recent advancements in deep learning have led to the development of neural network-based models for event extraction, such as Recurrent Neural Networks (RNNs) and Transformer-based architectures [23] [24]. These models have shown promising results in capturing complex event structures. The ML approaches have not achieved very high-precision performance due to the lack of enough annotated corpus content for news and documents. To address this, Han et al. [8] adopted a combination of patterns, ML algorithms and word embedding technology. Most recent event extraction techniques are based on supervised methods which heavily depend on manually labeled training data. However, in particular areas, like financial, medical and judicial face a shortage of labeled data due to the high cost involved in data labeling process. The unsupervised approaches such as, topic modelling, graph-based method, clustering techniques are suffered from subpar performance. There is no noticeable work dedicated to generating vendor insight report that systematically analyze the sequence of business events associated with vendors.

### **3. Proposed Methodology**

The research aims to address above mentioned limitations of existing vendor news analytics solutions by developing an innovative and customizable framework specifically designed for vendor management. The primary challenge lies in effectively acquiring, analyzing, and mining key insights from news articles and other textual data sources related to vendors and events of interest.

This involves a series of process steps to be followed utilizing data science methodologies to identify relevant business events from trusted news portals, screen, analyze, deduplicate news stories, detect sentiment and finally, curate summary and detailed insight automatically from the digital news sources. The aim is to generate comprehensive summaries using generative AI techniques and detailed reports that present the analyzed news data, empowering stakeholders to make informed decisions and track the trends relevant to vendors. Additionally, the project aims to tackle the challenge of customizability by developing a domain-specific approach that allows organizations to tailor analytics capabilities to their unique vendor management requirements.



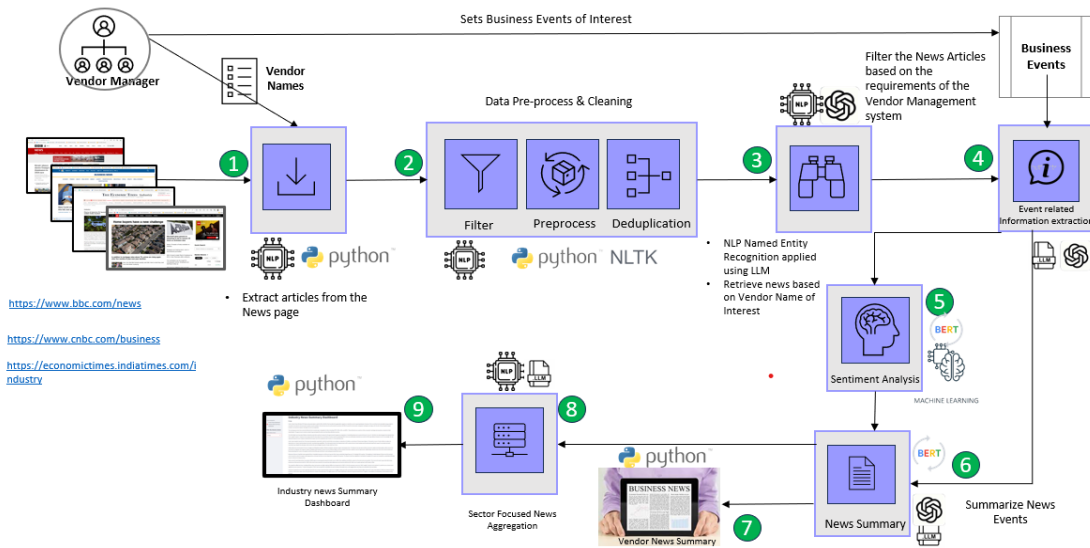
**Figure 1:** Proposed framework for vendor analytics report generation.

The proposed methodology is a framework comprising several processes. As shown in the Figure 1, starting with curating trusted and reliable news articles followed by vendor information such as vendor names, business domain, relationship type, impact level. Prepare a business event taxonomy in which concern party is interested, as vendors may involve in multiple business activities simultaneously. A web crawler module effectively gathers news data from various trusted news sources. Subsequently, this news data undergoes some preprocessing and filtration based on related domain and vendor identities. Next, the relevant business events will be extracted from news data using list of business events of interest. The final summary report will be generated using relevant business events, its associated information and sentiment of the events.

### 3.1. Solution Architecture Design

The proposed solution design adheres to a modular framework, allowing for the independent tuning and enhancement of each component with minimal repercussions on the overall system. By integrating cutting-edge methodologies and technologies, we aim to ensure the delivery of accurate report. Throughout the implementation, multiple methodologies will be considered and rigorously evaluated, with the goal of selecting the

most effective approaches to achieve superior output. This approach promotes flexibility, optimization and employs the latest advancements in this field. The framework begins with news data collection, followed by preprocessing and cleaning the data for event extraction, as illustrated in step 3, utilizing LLM. Subsequently, in step 4, the extracted events are then filtered based on business event taxonomy. The sentiment analysis for each event conducted at step 5 and step 6 involves generating summaries using LLM. The final report is prepared by combining event sentiment with its corresponding summary.



**Figure 2:** Different components of the solution architecture.

### 3.2. News Article Extraction

News articles to be extracted from primary digital news portals using appropriate NLP libraries; the critical information e.g., news header, news body, news URL, author, and news date-time etc. to be extracted from the news pages related to each news article. Custom functions to handle the extraction process across different sources with modularized configuration.

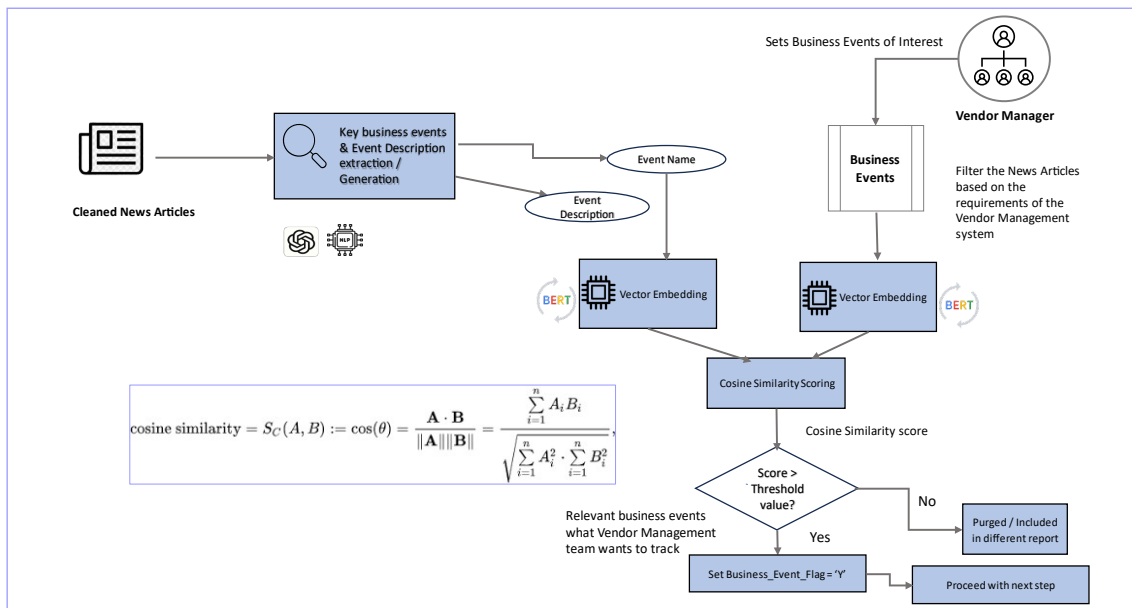
### 3.3. Preprocessing and Cleaning

News article filtration relevant to vendor identities using entity extraction techniques, preprocess, de-duplication. A pivotal aspect of vendor management involves maintaining a dynamic list of stakeholders for monitoring purposes. This necessitates the extraction of pertinent information from news articles, particularly focusing on vendors of interest. In this regard, entity extraction methodologies, such as Named Entity Recognition (NER), play a pivotal role. Our initiative entailed compiling a list of 200 vendor names relevant to a specific software consultancy enterprise. Traditional tools like Spacy [25], NLTK [26], Stanford CoreNLP [27] were initially explored for entity extraction. However, these tools fell short in efficiently identifying business entities within news articles. In our pursuit of an effective solution, we turned to Large Language Models (LLMs) and, specifically, employed GPT-3.5. Leveraging the multifaceted capabilities of LLMs, we aimed to enhance

the precision of vendor name extraction from news articles. Our findings revealed that GPT-3.5 exhibited superior accuracy in identifying key vendor names compared to conventional libraries such as Spacy, NLTK, and CoreNLP. To optimize the extraction process, we devised a meticulously crafted prompt to guide GPT-3.5. This prompt serves as a directive for the model, ensuring optimal performance in identifying and extracting vendor names from news articles. In summary, our approach underscores the efficacy of employing cutting-edge language models like GPT-3.5 for vendor identity extraction from news articles.

### 3.4. Extracting Relevant Business Events

Vendor manager has compiled a list of business events of interest such as, acquisition, investment, security breach, revenue growth etc. which need to be studied from news articles. However, identifying and extracting business events from unstructured data, like news article is a challenging task. Most of the previous event-extraction techniques have only been aimed at general events rather than business events. For instance, automatic content extraction (ACE), which has been studied in a research program for developing advanced information extraction technologies [8]. Existing event - extraction methods,



**Figure 3:** Process for business event extraction from news articles.

including natural language processing (NLP), machine learning (ML) and pattern-based techniques, have struggled to attain satisfactory results when applied to event extraction task. Machine learning approaches, in particular, have not delivered highly precise results due to limited availability of well-annotated corpus data for news and documents. On the other hand, pattern-based methods have not achieved robust recall, which measures the scope of the approach, largely because of incomplete event patterns and dictionaries [8]. The NLP techniques aims to discover event triggers with specific types and their

arguments from unstructured text and save them in a structured format. But in business domains, this is still a lack of effective event-extraction approaches [28] due to some problems that still need to be solved. Lu et al. [29] proposed Text2Event, a sequence-to-structure generation paradigm that can directly extract events from the text. However, it is crucial to note that Text2Event operates at the sentence level, rather than the broader scope of document-level event extraction. In the context of business event extraction, which often demands the comprehension of multiple paragraphs or even entire documents, this limitation becomes apparent.

The emergence of large language models such as GPT-3.5 provides an opportunity to solve language tasks with simple prompts without the need for task-specific datasets and fine-tuning procedures. Our innovative approach proposes leveraging Large Language Models (LLMs) for event extraction through prompt engineering. LLMs are trained on huge and diverse range of textual data and capable of completing tasks that needs broad understanding of human language, common sense reasoning, and the ability to generate contextually appropriate responses. Our event extraction methodology involves a systematic two-step process, as depicted in Figure 3. Initially, in the first step, we utilized LLM GPT-3.5 model [30] to identify and extract event names. Subsequently, in the second step, we focus on extracting event descriptions, which provide associated information about the business event. Following the extraction of events from the news article, we proceed to identify the events of interest by cross-referencing them with an event taxonomy. The cross-reference process aims to find the events which are belonging to business events taxonomy. We achieve this through calculating the cosine similarity score between the word embeddings of the extracted event names and the event names in the taxonomy. To obtain embedding vectors of event names, we utilize the pretrained BERT embedding model [15]. Subsequently, we calculate the cosine similarity score among the extracted event names and the events from the taxonomy. The events with the scores higher than a threshold value 0.7, are then selected as the final list of business events. Threshold value used based on manually evaluate the similarity between actual events and extracted events. Consequently, we filter out the corresponding event descriptions based on these selected events.

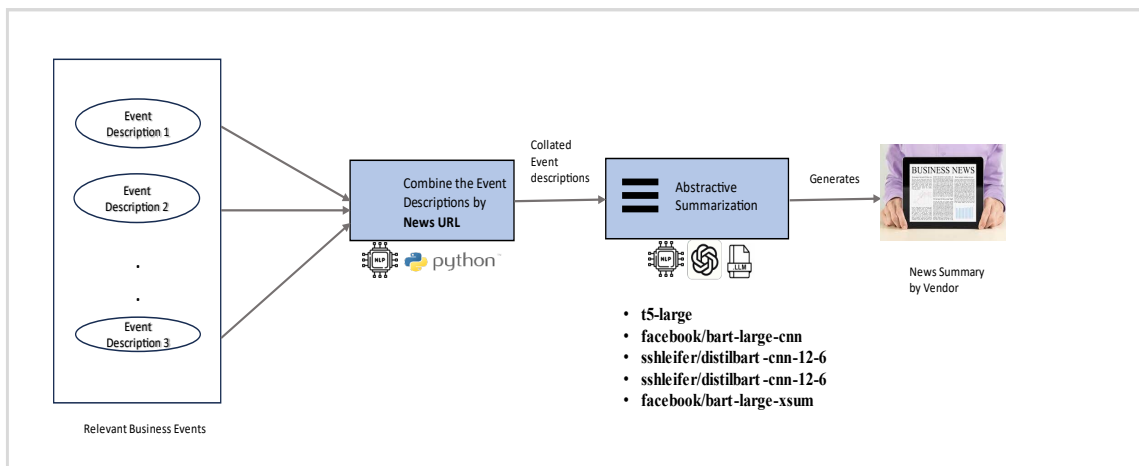
### **3.5. Vendor News Sentiment Analysis**

Vendor sentiment analysis will enable organizations to enhance their vendor relationships by providing near real-time insights into market trends and risk indicators. This analysis involves assessing the sentiment of each business event and its corresponding description, capturing both positive and negative sentiments to compile a comprehensive summary report. There are various techniques, such as lexicon-based approaches, machine learning models, deep learning neural networks and hybrid approaches are available for sentiment detection in news articles. Among these techniques, pre-trained language models are increasingly favored for their ability to capture contextual nuances effectively. In their work, Sinha et al. [31] has demonstrated the effectiveness of pre-trained language models. Specifically, RoBERTa and FinBERT achieve the highest average accuracy of 94.29% and F1-score of 93.27% respectively. The RoBERTa [32] is a robustly optimized BERT pretraining model and FinBERT [33] is domain specific model which is trained on financial

corpus. In our study to detect sentiment of business events, we employed two pre-trained models – distilRoBERTa [34] which has been fine-tuned on financial data specifically for sentiment analysis, and FinBERT. Following manual verification of the outcomes, it became evident that distilRoBERTa outperformed FinBERT in terms of performance. Consequently, the sentiment analysis results obtained from distilRoBERTa were included in the final report.

### 3.6. Vendor Wise News Summary

Following the extraction of business events and their corresponding description, we meticulously compiled all pertinent information, including associated sentiments. Subsequently, we generated a comprehensive abstractive summary reports using 4 tools, GPT3.5 [35], T5 [36], BART [37] and distilbert [38] as showed in Figure 4. To assess the quality of summarization, we employ reference-less measurement techniques. Reference-less measurement is a mechanism that does not rely on the availability of ground truth data. Given the cost and subjectivity associated with manually creating summaries, we opt for reference-less measuring as a more practical approach. In this method, we gauge the quality of generated summaries without relying on human-written reference summaries. Our evaluation involves the application of three distinct metrics: SummaQA [39], BLANC [40], and SUPERT [41], all of which are designed to assess the quality of abstractive summaries. SummaQA operates on a question-answering basis, scrutinizing the summary text to ascertain how well it matches the content of the source document in terms of providing answers to relevant questions. On the other hand, the BLANC method adopts a strategy where it randomly masks certain tokens within the summary text and then endeavors to fill these blanks using information extracted from the source document. Lastly, the SUPERT method employs a process wherein it extracts salient sentences from the source document and subsequently compares these with the generated summary. Our analysis reveals that across all three metrics, the overall scores are notably impressive. This suggests that our abstractive summary generation process is robust and effective in producing high-quality summaries.



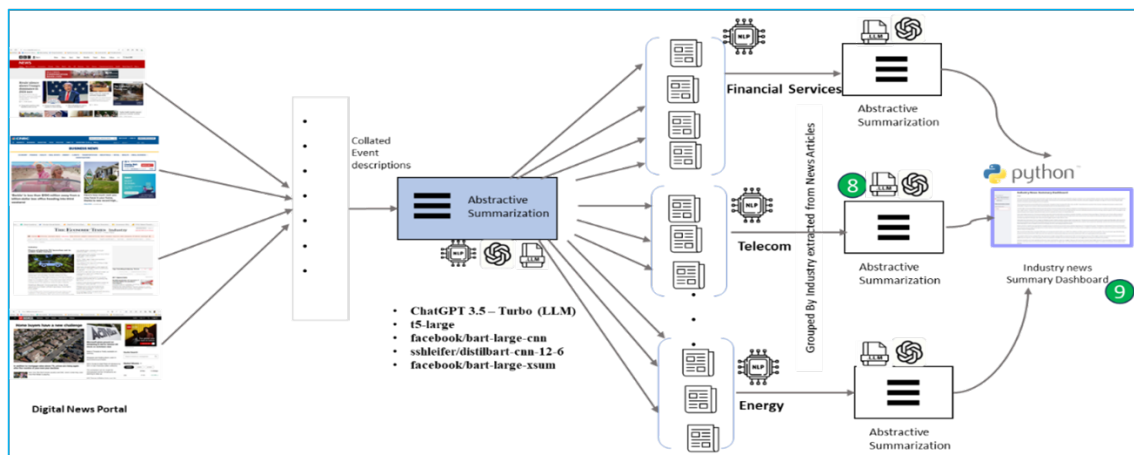
**Figure 4:** News summary generation process.



### 3.7. Sector Focus News Summary

The Sector Focus Summary equips organizations with the insights needed to make strategic decisions by leveraging trends and risk indicators specific to their industry. Sector Focused News Summarization is a method that involves collecting, compiling, and condensing pertinent business events from news articles that pertain to a particular industry or sector. This summarization process entails aggregating news articles relevant to a specific industry or sector. To evaluate the quality of sector-wise summaries, we apply the same measurement techniques used for vendor-wise summary data. As depicted in Figure 5, three key processes are undertaken to summarize sector-wise news events.

- News Aggregation: This collates news content from various news websites or digital news platforms. The aggregation process combines business news articles or information related to a vendor listed events.
- Summarization: After aggregating the news content, it is condensed into crisp /shorter form. These summaries usually provide key information, highlights or insights from the original news articles.
- Sector Focused: Aggregation and summaries are done around a specific industry or sector.



**Figure 5:** Sector wise news events summary.

Aggregated News summary generated for the total 20 industries. The list of industries is flexible and will depend on the vendors available in the extracted News. These are the some of the industry sectors that we considered for analysis:

- energy,
- financial services,
- retail,
- automotive industry,
- consumer goods,
- customer services,
- health care,
- media & entertainment.

## 4. Result of the Proposed Solution

We have collected around 250 news articles from 4 trusted news portals. Following the extraction of events from these articles, we identified 164 events relevant to vendors. These articles encompass a total of 106 vendors, which we will analyze as part of our proposed analysis. In these news articles, we covered total 20 industry sectors in our analysis.

### 4.1. Result on Business Event Extraction

We chose Open AI GPT-3.5 as LLM for business event extraction from vendor news articles during our research. Crafting effective input instructions is crucial for obtaining the desired output from the Language Model. Through numerous experiments, we developed a carefully constructed prompt and rigorously tested the output generated by GPT-3.5. The event extraction prompt follows the format of [system\_msg & query prompt]. This format prompts users to provide specific details for extracting events. Below, we elaborate on the prompting process and present examples of output events in Table 1 for better comprehension.

system\_msg = 'You are a business news summarization assistant. Your task is to identify the top 3 to 5 main business events from the given news articles and provide detailed descriptions for each event.'

query = "'Identify and describe the top 3 to 5 main business "Events" and their comprehensive description "Event description" that includes relevant details in 3-4 sentences for each in JSON structure from the News Article given.

The tags should be "Business\_Events" and attributes "Event" and "Description".

```
Sample output from zero-shot LLM call:
json_response: json
{
  "Business_Events": [
    {
      "Event": "Apple to Update iPhone 12 in France Due to Radiation Concerns",
      "Description": "Apple will deliver a software update for iPhone 12 users in France following fears about radiation levels detected by French regulators. Sales of iPhone 12 were halted in France after the regulator found excessive electromagnetic radiation, prompting Apple to address the issue with a software update. The update, specific to France, aims to comply with the testing protocol used by French regulators and ensure the device meets emissions rules in the country."
    },
    {
      "Event": "ANFR Testing and Compliance for iPhone 12 in France",
      "Description": "The radio frequency regulator ANFR in France will test the new software update provided by Apple to ensure compliance before allowing iPhone 12 sales to resume in the country. The regulator had previously warned Apple of a potential recall if the radiation issue was not resolved. Regulators from other EU countries like Belgium, the Netherlands, and Germany are also investigating the matter, potentially leading to broader measures across the EU."
    },
    {
      "Event": "Global Response to iPhone 12 Radiation Concerns",
      "Description": "While France has taken action to halt iPhone 12 sales over radiation levels, the UK and US have not announced any measures. The World Health Organization has stated that low-level electromagnetic fields from mobile phones are not harmful to humans. Apple has contested the findings in France, emphasizing compliance with emissions rules worldwide and providing lab results to support its claims."
    }
  ]
}
```

**Table 1.** Example of extracted events with source URL

News Portal	News URL	News Heading	News Datetime	Event	Event Description
BBC	<a href="https://www.bbc.com/news/technology-66795175">https://www.bbc.com/news/technology-66795175</a>	Apple to update iPhone 12 in France over radiation	9/15/2023 10:06	Apple to update iPhone 12 in France	Apple will deliver a software update for iPhone 12 users in France after sales were halted due to concerns about electromagnetic radiation. The update will only apply to users in France, where a specific testing protocol exists. The radio frequency regulator ANFR will test the update for compliance before allowing iPhone 12 back on sale in the country.
CNBC	<a href="https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html">https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html</a>	GM sweetens UAW offer to include 20 wage increase other benefits as it tries to avoid strike	9/14/2023 19:57	General Motors' new offer to avoid strike	General Motors (GM) has made a new offer to the United Auto Workers (UAW) union in an attempt to avoid a strike. The offer includes significant pay increases, more vacation days, and better benefits for retirees. However, it falls short of the union's demand for a 40% pay increase over the four years of the deal.
CNBC	<a href="https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html">https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html</a>		9/14/2023 19:57	GM's offer details released before strike deadline	GM released details of its new offer just hours before the UAW's strike deadline. The UAW had the option to initiate targeted strikes against GM, Ford Motor, and Stellantis if deals were not reached by 11:59 p.m. ET on Thursday. The release of the offer puts pressure on the UAW to consider the proposal.
CNBC	<a href="https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html">https://www.cnn.com/2023/09/14/uaw-strike-gm-sweetens-wage-and-benefits-offer-to-workers.html</a>		9/14/2023 19:57	Ford warns of potential financial impact of meeting UAW's demands	Ford has warned about the potential financial impact of meeting the UAW's demands. Sources from Ford stated that the automaker would have lost \$14.4 billion over the last four years if the current demands had been in effect, instead of recording nearly \$30 billion in profits. This highlights the significant financial implications of the UAW's demands for the automakers.
CNN	<a href="https://www.cnn.com/2023/09/15/business/mary-barra-gm-ceo-uaw-strike-interview/index.html">https://www.cnn.com/2023/09/15/business/mary-barra-gm-ceo-uaw-strike-interview/index.html</a>	GM CEO Mary Barra tells CNN she's frustrated about the strike	9/15/2023 9:35	General Motors CEO frustrated with autoworker strike	General Motors CEO Mary Barra expressed frustration with the autoworker strike and stated that the company has a compelling offer for the union, including pay raises, job security, and healthcare. Barra emphasized the need for UAW

	x.html				leadership to return to the negotiation table to resolve the issues and get people back to work. She also pushed back against the union's demand for a fair pay increase, pointing to record levels of profit sharing for employees.
CNN	<a href="https://www.cnn.com/2023/09/15/business/mary-barragm-ceo-uaw-strike-interview/index.html">https://www.cnn.com/2023/09/15/business/mary-barragm-ceo-uaw-strike-interview/index.html</a>		9/15/2023 9:35	United Auto Workers union goes on strike against automakers	The United Auto Workers (UAW) union went on strike against General Motors, Ford, and Stellantis, marking the first time in history that all three of America's unionized automakers were struck simultaneously. Workers walked out of three plants in Missouri, Michigan, and Ohio. The UAW referred to this targeted strike as a Stand Up Strike, a new strategic approach to walking off the job.
CNN	<a href="https://www.cnn.com/2023/09/15/business/mary-barragm-ceo-uaw-strike-interview/index.html">https://www.cnn.com/2023/09/15/business/mary-barragm-ceo-uaw-strike-interview/index.html</a>		9/15/2023 9:35	Smaller than expected UAW strike	The UAW strike involved fewer than 13,000 of its 145,000 members, which was less extensive than initially anticipated. The strike was expected to involve all 145,000 UAW members at the three companies, potentially becoming the largest strike of active workers in 25 years. The strike was a response to the automakers' dismissal of the union's ambitious demands for increased wages, benefits, and job protections.

#### 4.2. Result on Sentiment Analysis

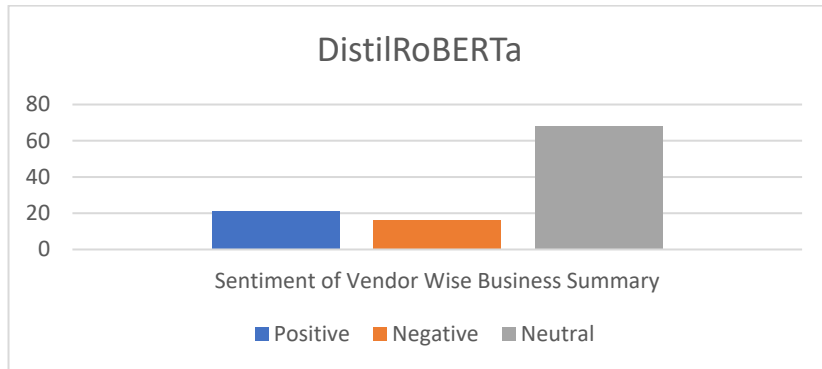
The news sentiment will be included in final summary report generation. We employed 2 different sentiment analysis methods to drill out the business sentiment for the vendor news.

**DistilRoBERTa** – This model is a distilled version of the RoBERTa-base model [32]. Further this model is finetuned on Financial PhraseBank [42]. further lexicon-based sentiment analysis tool. This is a rule-based method to classify the sentiments. It's a pre-trained model on social media and news articles.

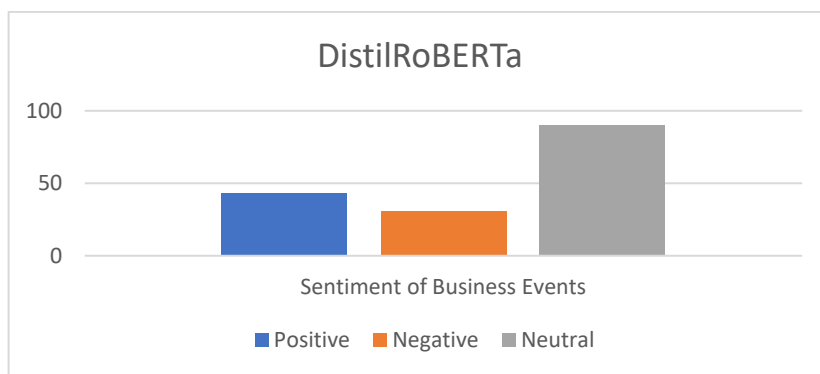
**FinBERT**– This is a language model based on BERT for financial NLP tasks. This is also a powerful pre-trained language model developed by Araci [33]. They pre-train the model on financial corpus TRC2-financial which consists of 1.8M news articles that were published by Reuters between 2008 and 2010 and then further finetune on Financial PhraseBank.

In some cases, these two methods have shown varying sentiment detection results for the same news text. However, in the majority of scenarios, the former approach has

demonstrated superior sentiment analysis outcomes. We conducted sentiment analysis on both event descriptions and vendor-specific summaries, and the sentiment results are accessible in our Github repository. The distribution of sentiment based on the DistilRoBERTa model is illustrated in Figures 6 and 7.



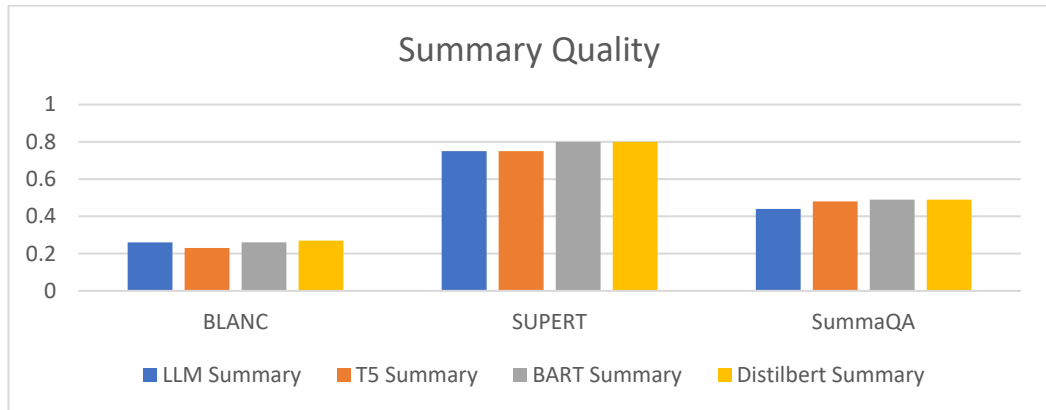
**Figure 6:** Vendor wise sentiment distribution.



**Figure 7:** Sentiment distribution on event description.

### 4.3. Result on Vendor Wise Summary

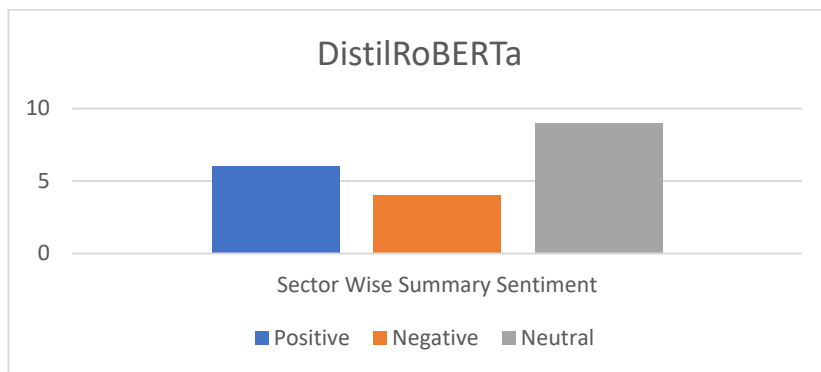
In our study, we undertook a comparative analysis of summary generation across four distinct tools: GPT3.5, T5, BART and Distilbert. To assess the quality of these summaries, we employed four metrics: BLANC [43] SummaQA [44], SUPERT [45]. The findings depicted in Figure 8 indicate a noteworthy similarity in summary quality between BART and DistilBERT. Consequently, for the purpose of our report, we opted to utilize DistilBERT for generating summaries pertaining to business events.



**Figure 8:** Quality measurement of summary generated by different tools.

#### 4.4. Result on Sector Focus Summary

We specialize in aggregating business events within specific sectors to provide comprehensive insights. As an illustration, we have generated a summary report for the Financial Performance sector by consolidating all relevant events. There is total 20 sectors identified from the news corpus data. Our analysis demonstrates that DistilBert consistently produces summaries of the highest quality. Therefore, we leverage this capability to generate sector-specific summaries efficiently and effectively.



**Figure 9:** Business sector wise summary sentiment distribution

## 5. Conclusion

The research aims at creating a News Analytics Solution using data science methodologies that can help the organizations to retrieve insights of their vendor or partner information from trusted news portals. In this research, we have extracted relevant Vendor News Articles from renowned news channels & portals and generated summary news on the business events of interest – it will simplify the decision making for vendor managers. This approach can be extended to other data sources such as blog posts, social media updates and other textual or multimedia sources from a wide range of digital platforms.

The goal of this research is to revolutionize vendor management practices within organizations by harnessing the power of news analytics. By leveraging NLP techniques, near real-time data analysis, machine learning and Generative AI, the system will provide organizations with insights and can produce real-time alerts to optimize vendor management strategies. Ultimately, this research aims to enhance operational efficiency, reduce risks, and foster stronger and more productive vendor relationships, leading to improved organizational performance and competitive advantage in the marketplace.

## References

- [1] A. E. Khedr and Y. Nagwa, "Predicting stock market behavior using data mining technique and news sentiment analysis," *International Journal of Intelligent Systems and Applications*, pp. 9(7):22-30, 2017.
- [2] A. Khant and M. Mehta, "Analysis of Financial News Using Natural Language Processing and Artificial Intelligence.," in *1st INTERNATIONAL CONFERENCE ON BUSINESS INNOVATION*, 2018.
- [3] P. Uhr, J. Zenkert and M. Fathi, "Sentiment analysis in financial markets A framework to utilize the human ability of word association for analyzing stock market news reports.," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017.
- [4] M. Vicari and M. Gaspari., "Analysis of news sentiments using natural language processing and deep learning.," *AI & society*, vol. 36, pp. 931-937, 2021.
- [5] W. Souma, I. Vodenska and H. Aoyama, "Enhanced news sentiment analysis using deep learning methods.," *Journal of Computational Social Science* , vol. 2, no. 1, pp. 33-46, 2019.
- [6] J. Wörner, D. Konadl, I. Schmid and S. Leist., "COMPARISON OF TOPIC MODELLING TECHNIQUES IN MARKETING-RESULTS FROM AN ANALYSIS OF DISTINCTIVE USE CASES.," in *European Conference on Information Systems (ECIS)*, 2021.
- [7] J. Wörner, D. Konadl, S. Leist and I. Schmid., "MANTRA: A Topic Modeling-Based Tool to Support Automated Trend Analysis on Unstructured Social Media Data.," in *International Conference on Information Systems (ICIS 2023)*, Hyderabad, India, 2023.
- [8] S. Han, X. Hao and H. Huang, "An event-extraction approach for business analysis from online Chinese news.," *Electronic Commerce Research and Applications* , vol. 28, pp. 244-260, 2018.
- [9] R. Kalaiselvi, "Hot News Detection System Using Rule-Based Technique.," *Materials Today: Proceedings*, vol. 24, pp. 2330-2338., 2020.
- [10] Z. Lei, L.-d. Wu, Y. Zhang and Y.-c. Liu, "A system for detecting and tracking internet news event," in *In Advances in Multimedia Information Processing-PCM 2005: 6th Pacific Rim Conference on Multimedia*, Jeju Island, Korea, 2005.
- [11] M. Liu, Y. Liu, L. Xiang, X. Chen and Q. Yang, "Extracting key entities and significant

- events from online daily news.," in In Intelligent Data Engineering and Automated Learning–IDEAL 2008: 9th International Conference , Daejeon, South Korea, 2008.
- [12] M. Okamoto and M. Kikuchi., "Discovering volatile events in your neighborhood: Local-area topic extraction from blog entries.," in In Information Retrieval Technology: 5th Asia Information Retrieval Symposium, AIRS 2009, Sapporo, Japan, 2009.
- [13] F. Xu, H. Uszkoreit and H. Li., "Automatic event and relation detection with seeds of varying complexity," in In Proceedings of the AAAI workshop event extraction and synthesis, 2006.
- [14] S.-H. Hung, C.-H. Lin and J.-S. Hong, "Web mining for event-based commonsense knowledge using lexico-syntactic pattern matching and semantic role labeling," *Expert Systems with Applications*, vol. 37, no. 1, pp. 341-347, 2010.
- [15] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [16] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.
- [17] J. Liu, L. Min and X. Huang, "An overview of event extraction and its applications.," *arXiv preprint arXiv:2111.03212*, 2021.
- [18] J. Borsje, F. Hogenboom and a. F. Frasincar, "Semi-automatic financial events discovery based on lexico-semantic patterns," *International Journal of Web Engineering and Technology*, vol. 6, no. 2, pp. 115-140., 2010.
- [19] G. Jacobs, E. Lefever and V. Hoste, "Economic event detection in company-specific news text," in In 1st Workshop on Economics and Natural Language Processing (ECONLP) at Meeting of the Association-for-Computational-Linguistics (ACL), 2018.
- [20] B. Yang and T. Mitchell., "Joint extraction of events and entities within a document context," *arXiv preprint arXiv:1609.03632*, 2016.
- [21] B. Drury and J. J. Almeida., "Identification, extraction and population of collective named entities from business news.," in In Entity2010–Workshop on Resources and Evaluation for Entity Resolution and Entity Management, 2010.
- [22] A.-L. M. Minard, M. Speranza, E. Agirre, I. Aldabe, M. V. Erp, B. Magnini, G. Rigau and R. Urizar, "Semeval-2015 task 4: Timeline: Cross-document event ordering," in In proceedings of the 9th International Workshop on Semantic Evaluation, 2015.
- [23] D. Wadden, Y. L. Ulme Wennberg and H. Hajishirzi, "Entity, relation, and event extraction with contextualized span representations," *arXiv preprint arXiv:1909.03546*, 2019.
- [24] Q. Li, J. Li, J. Sheng, S. Cui, J. Wu, Y. Hei, H. Peng, S. Guo, L. Wang, A. Beheshti and S. Philip, "A survey on deep learning event extraction: Approaches and applications.," *IEEE Transactions on Neural Networks and Learning Systems.*, vol. 14, no. 9, 2022.
- [25] "Named Entity Recognition (NER) using spaCy," [Online]. Available: <https://spacy.io/universe/project/video-spacys-ner-model-alt>. [Accessed 7 March 2024].



- [26] "Extracting Information from Text," [Online]. Available: <https://www.nltk.org/book/ch07.html>. [Accessed 7 March 2024].
- [27] "Named Entity Recognition," Stanford, [Online]. Available: <https://stanfordnlp.github.io/CoreNLP/ner.html>. [Accessed 7 March 2024].
- [28] E. Arendarenko and T. Kakkonen, "Ontology-based information and event extraction for business intelligence," in In Artificial Intelligence: Methodology, Systems, and Applications: 15th International Conference, AIMS 2012, Varna, Bulgaria, 2012.
- [29] Y. Lu, H. Lin, J. Xu, X. Han, J. Tang, A. Li, L. Sun, M. Liao and S. Chen, "ext2Event: Controllable sequence-to-structure generation for end-to-end event extraction," arXiv preprint arXiv:2106.09232, 2021.
- [30] "GPT-3.5 Turbo," OpenAI, [Online]. Available: <https://platform.openai.com/docs/models/gpt-3-5-turbo>. [Accessed 7 March 2024].
- [31] A. Sinha, S. Kedas, R. Kumar and P. Malo, "SEntFiN 1.0: Entity-aware sentiment analysis for financial news," Journal of the Association for Information Science and Technology, vol. 73, no. 9, pp. 1314-1335, 2022.
- [32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
- [33] D. Araci, "Finbert: Financial sentiment analysis with pre-trained language models," arXiv preprint arXiv:1908.10063, 2019.
- [34] "DistilRoberta-financial-sentiment," Huggingface, [Online]. Available: <https://huggingface.co/mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis>. [Accessed 7 March 2024].
- [35] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan and P. Dhariwal, "Language models are few-shot learners," in Advances in neural information processing systems, 2020.
- [36] C. N. S. Raffel, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," The Journal of Machine Learning Research, vol. 21, no. 1, pp. 5485-5551, 2020.
- [37] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," arXiv preprint arXiv:1910.13461, 2019.
- [38] V. Sanh, L. Debut, J. Chaumond and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv preprint arXiv:1910.01108, 2019.
- [39] T. Scialom, S. Lamprier, B. Piwowarski and J. Staiano, "Answers unite! unsupervised metrics for reinforced summarization models," arXiv preprint arXiv:1909.01610, 2019.
- [40] O. Vasilyev, V. Dharnidharka and J. Bohannon, "Fill in the BLANC: Human-free quality estimation of document summaries," arXiv preprint arXiv:2002.09836, 2020.

- [41] Y. Gao, W. Zhao and S. Eger., "SUPERT: Towards new frontiers in unsupervised evaluation metrics for multi-document summarization," arXiv preprint arXiv:2005.03724, 2020.
- [42] P. Malo, A. Sinha, P. Korhonen, J. Wallenius and P. Takala, "Good debt or bad debt: Detecting semantic orientations in economic texts," *Journal of the Association for Information Science and Technology*, vol. 65, no. 4, pp. 782-796., 2014.
- [43] "Evaluation measures," Primer.ai r, [Online]. Available: <https://github.com/PrimerAI/blanc>. [Accessed 7 March 2024].
- [44] "SummaQA," [Online]. Available: <https://github.com/ThomasScialom/summa-qa>. [Accessed 7 March 2024].
- [45] "SUPERT: Unsupervised Multi-Document Summarization Evaluation & Generation," [Online]. Available: <https://github.com/yg211/acl20-ref-free-eval>. [Accessed 7 March 2024].