Financial Sentiment Analysis: Techniques and Applications

KELVIN DU, School of Computer Science and Engineering, Nanyang Technological University, Singapore
FRANK XING, Department of Information Systems and Analytics, National University of Singapore, Singapore
RUI MAO*, School of Computer Science and Engineering, Nanyang Technological University, Singapore
ERIK CAMBRIA, School of Computer Science and Engineering, Nanyang Technological University, Singapore

Financial Sentiment Analysis (FSA) is an important domain application of sentiment analysis that has gained increasing attention in the past decade. FSA research falls into two main streams. The *first stream* focuses on defining tasks and developing techniques for FSA, and its main objective is to improve the performances of various FSA tasks by advancing methods and using/curating human-annotated datasets. The *second stream* of research focuses on using financial sentiment, implicitly or explicitly, for downstream applications on financial markets, which has received more research efforts. The main objective is to discover appropriate market applications for existing techniques. More specifically, the application of FSA mainly includes hypothesis testing and predictive modeling in financial markets. This survey conducts a comprehensive review of FSA research in both the technique and application areas and proposes several frameworks to help understand the two areas' interactive relationship. This article defines a clearer scope for FSA studies and conceptualizes the FSA-investor sentiment-market sentiment relationship. Major findings, challenges, and future research directions for both FSA techniques and applications have also been summarized and discussed.

CCS Concepts: • **Computing methodologies** → **Natural language processing**; **Neural networks**; **Machine learning algorithms**; • **Information systems** → *Information retrieval*; **Sentiment analysis**; • **Applied computing** → **Law, social and behavioral sciences**.

Additional Key Words and Phrases: Financial Sentiment Analysis, Financial Forecasting, Natural Language Processing, Information System, Machine Learning, Deep Learning

ACM Reference Format:
Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria. 2024. Financial Sentiment Analysis: Techniques and Applications. 1, 1 (February 2024), 40 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Sentiment analysis is a field of study that analyzes people's sentiments, attitudes, opinions, emotions, evaluations, and appraisals towards various entities such as events, topics, services, products, individuals, organizations, issues, and their attributes [92]. Financial Sentiment Analysis (FSA), which in broad terms studies investor sentiment and financial textual sentiment [78], is an important domain application for sentiment analysis. Given the intricate nature of the

*Corresponding author.

Authors' addresses: Kelvin Du, School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore, zidong001@e.ntu.edu.sg; Frank Xing, Department of Information Systems and Analytics, National University of Singapore, Singapore, Singapore, xing@nus.edu.sg; Rui Mao, School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore, rui.mao@ntu.edu.sg; Erik Cambria, School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore, cambria@ntu.edu.sg.

financial market, individuals involved in varying market conditions exhibit diverse cognitive patterns [111], rendering it challenging to dynamically comprehend and analyze the market for robust financial decision-making. To address the challenge posed by the market's active shifts, automated FSA has gained increasing attention in the past decade [170]. It is proven to be a powerful tool to support business decision-making and perform financial forecasting [101, 102]. The application scenarios include corporate disclosures, annual reports, earning calls, financial news, social media interactions, and more [170, 183]. Sentiment analysis is a suitcase problem and domain-dependent. The phenomenon of domain-dependence is more pronounced in the finance domain [106] because of both topic concentration and the use of highly professional language. For example, a word such as "liability" and "debt" is considered negative in general-purpose sentiment analysis, whereas it is frequent and has a neutral meaning in the financial context.

In terms of where the ground truth comes from, financial sentiment indicators are categorized into market-derived [84] and human-annotated sentiments [99]. The market-derived sentiments are computed proxies from market dynamics, such as price movement and trading volume, thus, may include noise from other sources. For example, generally positive news is related to large changes in price for a short time, while the effect of negative news lasts longer [35]. The subjective human-annotated sentiments, however, are specifically labeled by professionals [106] or investors themselves [181]. FSA has received great attention from researchers and investors and has become a prominent and interesting research topic in recent years [5, 158] mainly due to the increase in online materials such as digital news, World Wide Web, and social media. FSA research is shifting from human-annotated to market-derived sentiment. More specifically, the application of FSA in financial forecasting has become more popular in recent years.

In the realm of FSA research, Kearney and Liu [78] conducted a comprehensive survey in 2014, focusing primarily on FSA techniques rooted in dictionaries and machine learning. [107] presented a brief review of various FSA methodologies in 2019, encompassing data sources, lexicon-based approaches, traditional machine learning, and deep learning techniques. While prior reviews have tended to be skewed towards either FSA techniques or applications [78, 107, 127, 183], this survey aims to provide a comprehensive review of the most recent FSA research bridging two aspects of this spectrum. We believe that linking both techniques and applications can enable researchers to have an overarching understanding of FSA studies and more importantly, facilitate better adoption of FSA in downstream applications to generate more promising results. Our work entails an extensive examination of the most recent FSA studies, offering a dual perspective from both technical and applied standpoints. Notably, our investigation extends beyond the confines of computer science literature, establishing connections with other disciplines such as information systems and finance. In particular, we delve into the foundational principles of financial forecasting, lending support to the market predictability of financial sentiment from a financial theory standpoint. We have meticulously defined the scope of FSA research, reaffirming the intricate relationship between FSA, investor sentiment, and market sentiment. Furthermore, we have scrutinized the genesis of financial sentiment, whether implicit or explicit, in its applications within financial markets. This analysis sheds light on the dynamic interplay between FSA techniques and their practical applications, ultimately facilitating a more seamless integration of financial sentiment in downstream tasks. Besides, we deliver structured summaries for different technical trends, tasks, features, and applications. Finally, building upon the most recent brief survey on FSA [107], we additionally review FSA tasks with more recent benchmark datasets, learning approaches, pre-trained language models, word representation techniques, and evaluation methods. We also highlight FSA applications, including data sources, hypothesis testing, and predictive modeling.

This study aims to answer the following four groups of research questions. For easy navigation, our findings are elaborated in Section 6.

(1) FSA studies have evolved over the years with more data available. What is the *scope* of FSA in today's context, and what is the relationship among FSA, investor sentiment, and market sentiment?
(2) What *trends* are emerging from the latest tasks, benchmark datasets, and methods in the FSA technique studies?
(3) FSA has been widely used in financial applications since Engle and Ng suggested the asymmetric and affective impact of news on market volatility in 1993. *How many* data sources, tasks, methods, and financial markets have been researched in FSA applications?
(4) How financial sentiment is involved in financial *forecasting* and the focus is on FSA techniques or applications?

Our contributions can be summarized from the following four aspects:

(1) We have conducted a comprehensive review of the latest FSA studies from both the technique and application perspective. This effort fills the gap in the literature by having a detailed and referential anchoring point for FSA research.
(2) Our field of investigation goes beyond the computer science literature and links to other disciplines, such as information systems and finance. Specifically, we review the underlying principles of financial forecasting and provide support to the market predictability of financial sentiment from a financial theory perspective.
(3) We have defined the scope of FSA research and re-confirmed the relationship among FSA, investor sentiment, and market sentiment.
(4) We have reviewed how financial sentiment is generated, implicitly or explicitly, during its applications in financial markets, and the interactive relationship between FSA techniques and applications, which will facilitate a better adoption of financial sentiment in downstream application tasks.

The remainder of this article is organized as follows: Section 2 provides the background of FSA, including its definition, motivation and importance; Section 3 provides the literature review framework; Section 4 and Section 5 review existing studies on FSA techniques and applications, respectively; Section 6 demonstrates the research findings of this survey; Section 7 lists challenges and future directions; finally, Section 8 offers concluding remarks.

2 BACKGROUND

The term "sentiment" is used in the context of analyzing evaluative texts automatically and detecting predictive judgments from negatively and positively opinionated texts [15]. This term first appeared in the studies by [28] and [165], where researchers were interested in market sentiment analysis [158]. Traditionally, investor sentiment is collected via surveys which ask for opinions on the markets regularly. With the advancement in textual data such as news texts, social media collections, and automatic processing technologies, the media has become an important source of investor sentiments. The task of FSA is to perform sentiment analysis from financial texts.

In computational finance, the adoption of robo-readers to process and analyze texts are emerging technology trends [106]. It is an area of knowledge that emerged in the 1980s that uses computational methods to solve problems in finance. From this perspective, FSA is also a research area under computational finance. Generally, FSA techniques refer to the methods to perform sentiment analysis (e.g., extraction of sentiment polarities or intensities) from financial texts, which could be categorized into the lexicon, machine learning, deep learning, hybrid, and pre-trained language model approaches. In terms of the applications of FSA, which refers to the adoption of financial sentiment in downstream tasks such as hypothesis testing and predictive modeling, the most important application of FSA is the forecasting of financial markets. Efficient Market Hypothesis (EMH), proposed by Fama in the 1970s [47], is a critical foundation of modern financial market analysis. It hypothesizes that financial markets are efficient and the price has incorporated all available market information. To tie the concept to reality, Fama classified the efficient market into three forms, namely strong, semi-strong, and weak form. Under weak form, it is assumed that the information set is merely historical prices, and that any current or private information will not influence the market. The weak form test has been renamed to tests for return predictability in [48] by Fama. The semi-strong form, which is changed to event studies in [48], states that stock prices reflect all public and historical information, while private information fails to influence market movements. The strong form describes a market condition, in which the price of securities reflects all information including public, private, and historical price information with a presumption that it is free to trade and access information, which is rarely the case. It is worth noting that Fama acknowledged that such a market is rarely seen in reality, but is useful for theoretical purposes only. However, the intellectual dominance of EMH had become less universal by the start of 21^{st} century. A large number of statisticians and financial economists began to believe that stock prices could be predicted at least partially [105]. Under behavioral finance, investors make decisions from a psychological perspective [3], and their state of mind, or sentiment, influences them when making that decision [11, 98]. Today, the debate on market predictability is not about whether or not investor sentiment affects markets anymore but how to measure the sentiment and quantify its effect [7].

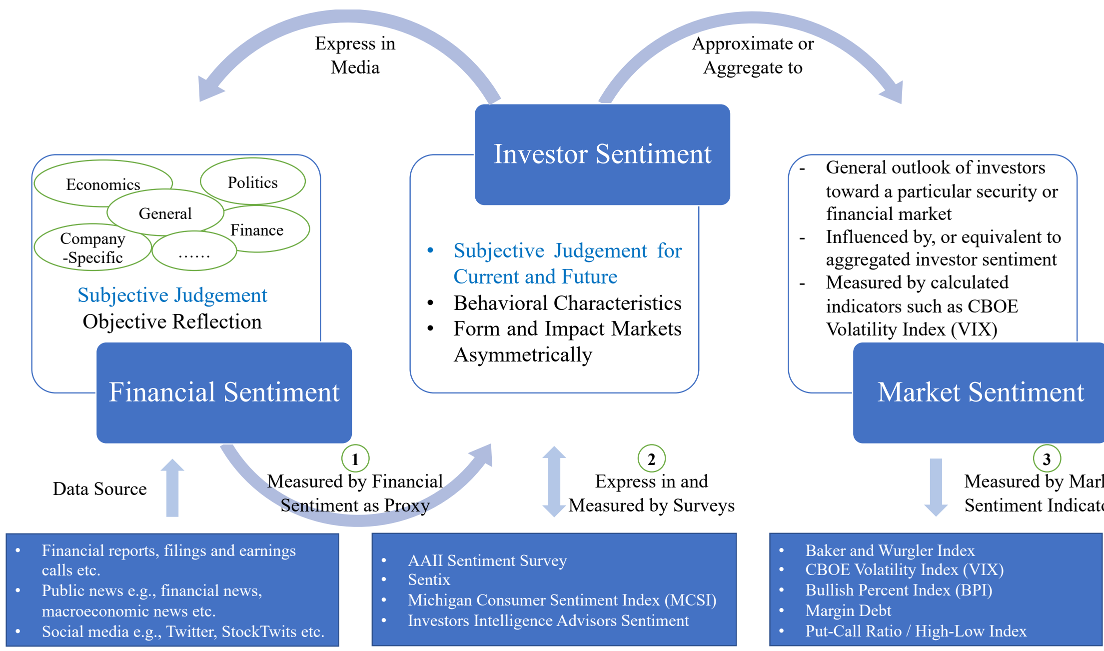

Fig. 1. Financial Sentiment, Investor Sentiment, and Market Sentiment.

2.1 Relationship among Different Sentiment Agents

The relationship among market sentiment, investor sentiment, and financial textual sentiment is illustrated in Figure 1. Firstly, investor sentiment, which indicates the degree of deviation of an asset value from its economic fundamentals, can be defined as investors' optimism or pessimism about future market activity [6] or as the way investors form beliefs [10]. Investor sentiment can be expressed and measured in two main forms including survey and financial texts [199]. The popular surveys include American Association of Individual Investors (AAII) Investor Sentiment Survey [1], Sentix Investor Confidence [2], or Investors Intelligence Sentiment Index [3]. The AAII Investor Sentiment Survey provides valuable insights into the perspectives of individual investors regarding the future direction of the market over a six-month period through a weekly survey in which investors can vote Bullish, Neutral, or Bearish. The Sentix Investor Confidence Index assesses the prospective economic outlook for the eurozone over a six-month period, and is derived from a comprehensive survey involving investors and analysts. A reading surpassing zero signifies a positive or optimistic outlook, while a reading below indicates a negative or pessimistic perspective. The Investors Intelligence Sentiment Index operates on contrarian principles and conducts surveys of more than one hundred independent market newsletters, evaluating the current stance of each author regarding the market, whether it be bullish, bearish, or indicating a correction. Investor sentiment can also be measured through textual data such as microblogs and analyst reports. This measure is derived from various communication platforms that exist on the internet and can be a proxy for investor sentiment or non-informational trading. In this sense, [3] summarized that some studies have adopted investor sentiment derived from social networks such as Twitter [196], StockTwits [140, 146] or Facebook [156]; message boards such as RagingBull.com [168], Yahoo! Finance [81], or Google searches [27].

Secondly, financial (textual) sentiment is measured by the degree of positivity or negativity in financial texts. Investor sentiment and financial textual sentiment are not independent but connect with each other. Investor sentiment could be measured by financial textual sentiment, especially with the increase of financial textual data today. This is because financial textual sentiment contains both subjective and objective information. The subjective information includes subjective judgment and analysis from investors and analysts, which is normally published on social media and self-media. The objective information, e.g., political and macroeconomics news, break news, and annual reports released by companies, is the objective reflection of conditions within the general environment, industries, markets, and firms [78]. The objective information is often leading in the sense that it is influential to investors' judgment, while the subjective information is often lagging because it is driven by investor sentiment and reflects investors' opinions. Thus, financial sentiment interacts with investor sentiment. From this perspective, financial sentiment can serve as a measure or proxy of investor sentiment, which subsequently influences trading strategies in the markets. Financial textual sentiment analysis differs from classic sentiment analysis in several key aspects. Firstly, it involves the frequent use of metaphorical expressions in financial communication, where figures of speech are employed to convey sentiments or describe market conditions. For example, "The market is riding a bull" is a common metaphor signifying a robust, upward market movement. Secondly, precision and brevity are of paramount importance in the financial world. Professionals use concise language to efficiently convey complex information. For instance, instead of stating "The company experienced a substantial increase in revenue and a corresponding improvement in profitability," a financial analyst might highlight, "The company posted robust revenue growth, driving higher profits.", which requires FSA to decode sentiments from concise sentence structures. Thirdly, the financial industry employs a unique set of terms and jargon with specific meanings. A thorough understanding of these terms is crucial for accurate interpretation and analysis of financial texts in FSA. For example, the "Price-to-Earnings (P/E) ratio" is a fundamental financial metric used to assess a company's valuation. A high P/E ratio may indicate that investors hold high expectations for future earnings. Furthermore, unlike classic sentiment analysis, which typically focuses on text alone, financial texts often integrate qualitative text with quantitative data, which requires FSA to not only understand the language used in financial texts but also to process

[1] https://www.aaii.com/sentimentsurvey
[2] https://www.sentix.de/
[3] https://www.investorsintelligence.com/x/us_advisors_sentiment.html

and analyze numerical information in conjunction with the textual context, to gain a comprehensive understanding of the sentiment. Lastly, FSA is often direction-dependent and the direction of events or changes holds critical importance in FSA. For instance, the word "profit" may carry both positive and negative sentiment depending on the direction. An increase in profit is generally regarded as positive, while a decrease is seen as negative. In practice, [181] concluded that there are six areas that cause FSA fail, i.e., irrealis mood (conditional mood, subjunctive mood, imperative mood), rhetoric (negative assertion, personification, sarcasm), dependent opinion, unspecified aspects, unrecognized words (entity, microtext, jargons), and external reference. Understanding these specificities is crucial for accurately analyzing financial sentiment, as they provide context and nuance to the language used in financial texts.

Further, market sentiment is the collective outlook of investors towards a specific financial market or security [160]. It is often used interchangeably with investor sentiment but fundamentally distinct as investors may hold varying viewpoints at different periods and markets. Market sentiment reflects the trading behavior of investors in a specific market, which is driven by investor sentiment or namely aggregated effect of investor sentiment [160]. It encapsulates the prevailing atmosphere or mood within the market, representing the crowd's psychological disposition, discernible through the trading activity and price fluctuations of the securities being exchanged. Broadly speaking, ascending prices signal an optimistic or bullish market sentiment, whereas descending prices signal a pessimistic or bearish market sentiment. The influence of investor sentiment on the market is asymmetric, which means the impact of investor sentiment on the market varies in different regimes of the market. Market sentiment can be measured by various proxy financial metrics, such as the degree of price movements and volatility computed from historical market data, and thus are backward-looking, lagging indicators. One of the most well-known market sentiment indicators is the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) [4], which measures expected market volatility based on real-time prices of the S&P 500 Index options over the next 30 days. The VIX tends to be higher when there is a greater level of fear and uncertainty in the market and lower in bull markets. Besides, the Equity Market Sentiment Index (EMSI) [9], High-Low Index, Bullish Percent Index (BPI) and the Baker and Wurgler Index [6, 7] are also popular indices for market sentiment or prevailing investor sentiment. Particularly, the Baker-Wurgler Index is generated from the first principal component of six proxies from market variables which are CEFD, dividend premium, equity issues, first-day return, IPO activity, and trading volume. [101] argued that the market sentiment of financial events does not equal to the semantic sentiment of financial news. The former represents the market reflections towards financial events in financial asset prices, while the latter represents the semantic understanding of the sentiment of news. [101] experimentally proved that using market sentiment representations can better predict stock price movements than using semantic sentiment representations.

2.2 FSA Research Scope

The scope of FSA studies can be broadly categorized into technique-driven and application-driven studies. The technique-driven FSA study, similar to other domain adoption of sentiment analysis, focuses on sentiment analysis of financial texts. However, the application-driven FSA study is unique to the finance domain, as financial sentiment can be used as a proxy of investor sentiment to make predictions in financial markets. Fundamentally, the objectives and scopes of FSA techniques and applications differ significantly. Nevertheless, an interactive relationship between techniques and applications also exists. For instance, the benchmark datasets employed for FSA technique studies usually require human annotation with sentiment polarities or intensity scores, which need to be sufficient, representative, and precise

[4]https://www.cboe.com/tradable_products/vix/

to train an unbiased model in different domains. On the other hand, the data sources adopted for FSA application studies are normally annotated by financial metrics, computed from the market data. They also require the data to be in a time series with a substantial amount of financial texts and representative periods to model the relationships between investor sentiment and financial metrics. As for the tasks, FSA techniques focus on the granularity which refers to the level of sentiment at which the sentiment is detected (e.g., targeted aspect-based sentiment analysis vs. sentence-level sentiment analysis). However, FSA applications explore various financial application scenarios, such as stock market movement prediction, financial risk prediction, portfolio management, FOREX market prediction, and cryptocurrency market prediction. The interactive relationship between FSA techniques and applications is twofold. First, the sentiment analysis lexicons and models developed through FSA techniques could be used for FSA applications by deriving the sentiment representation explicitly. Second, techniques such as feature engineering, text representation methods, algorithms, and evaluation metrics can also be considered for FSA applications. One point to highlight is that certain FSA applications such as portfolio management require more complex methods (e.g., reinforcement learning) and evaluations (e.g., trading simulation). Considering the difference between FSA techniques and applications in task definitions, datasets, methods, and their connections in learning algorithms, lexicon resources, and feature engineering methods, we believe an appropriate scope of an FSA survey should cover both domains.

2.3 Summary of Conceptualization

To summarize, it is a consensus that investor sentiment affects market dynamics and the measure of investor sentiment and quantification of its effect are critical to market prediction. Generally, investor sentiment can be measured by financial textual sentiment, sentiment surveys, and indices constructed from market data. Financial textual sentiment captures subjective judgment expressed by investors and also contains objective reflection, which drives investor sentiment. Market sentiment is the reflection of investor sentiment in investment behaviors. Investor sentiment is the psychological state of investors and it can be partially measured by financial textual sentiment and market sentiment, because of the interactions between investor sentiment and the other two aspects. Thus, FSA can be defined as a field of study that analyzes people's sentiment from financial texts, measures and quantifies investor sentiment from financial textual sentiment, and is finally grounded in the applications of market prediction and financial decision-making. We define the scope of this survey as covering both FSA techniques and FSA applications, motivated by the distinct yet interconnected nature of their datasets, methodologies, and targets.

3 LITERATURE REVIEW FRAMEWORK

The research in FSA can be categorized into two main streams. The first type focuses on the tasks in FSA, and its main objective is to study the techniques that are able to improve the performance of various FSA tasks such as paragraph and sentence-level sentiment analysis, (targeted) aspect-based sentiment analysis and development of financial lexicons and sentiment analysis models [1, 4, 5, 30, 40, 58, 66, 72, 73, 77, 88, 96, 97, 99, 109, 110, 113, 122, 133, 136, 139, 147, 152, 157, 160, 178]. The other group is application-driven or market-driven, which has received more attention in recent years where financial sentiment is treated as an intermediate output. The main objective is to use it for downstream applications, such as causality and correlation testing and financial forecasting [6, 13, 20, 31, 32, 34, 38, 43, 59, 63, 70, 75, 76, 79, 80, 85, 86, 95, 101, 104, 124, 128, 134, 145, 150, 151, 153, 154, 162, 167, 172–174, 177, 179, 180, 184, 185, 187, 190, 198, 200]. The sentiment can be represented in an explicit or implicit manner. The explicit representation refers to the generation of sentiment words, polarity, or intensity score [167], while the implicit representation of textual sentiment refers to the generation of feature embedding [95, 101]. Figure 2 illustrates our survey framework.

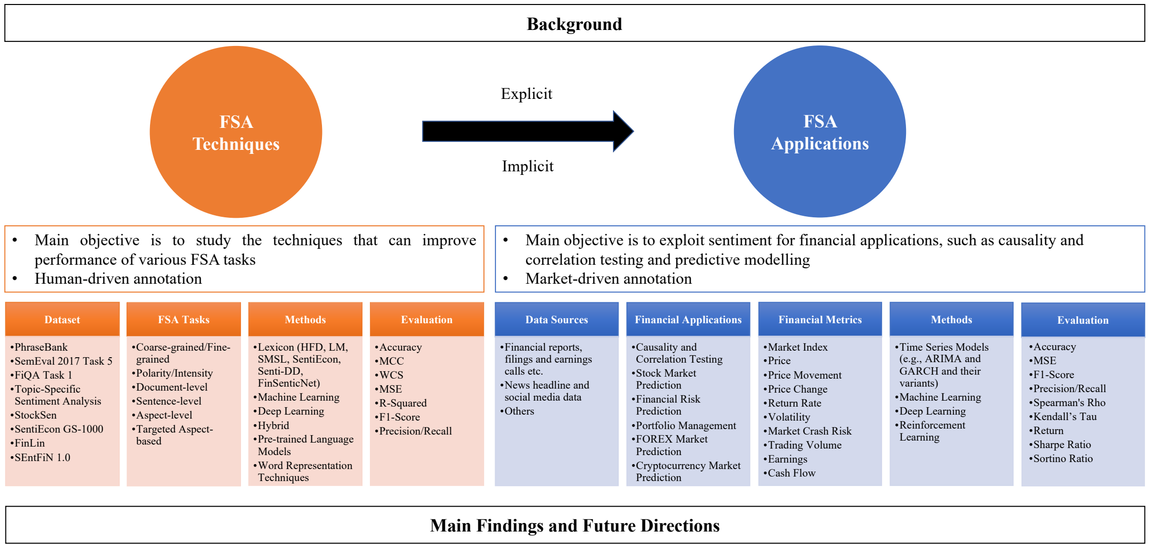

Fig. 2. FSA Review Framework.

4 FSA TECHNIQUES

4.1 Tasks

Sentiment analysis can be performed in a coarse-grained [170] or fine-grained manner. The fine-grained sentiment analysis can be studied from two perspectives: granularity and expression. Granularity refers to the level of sentiment at which the sentiment is detected and it includes document-level [121], paragraph-level [52], sentence-level [195] and aspect-level [141]. In the financial domain, the aspect-level approaches are known as Aspect-based FSA. To be more granular, the target element can be introduced where the sentiment detection is for that particular target. This is also known as stance detection defined by [159], or Targeted FSA. The task is to detect the text that is favorable or unfavorable to a specific given target. The most challenging but pragmatic task is called Targeted Aspect-based FSA (TABFSA), which aims to extract entities and aspects and detect their corresponding sentiment in financial texts. While most of the current FSA studies still adopt a sentiment polarity detection fashion (i.e., classification to positive or negative), sentiment can be also expressed by intensity score that is more consequential and nuanced for FSA compared to other sentiment analysis domains. Thus, intensity score-based FSA requires models in a regressive fashion.

4.2 Benchmark Datasets

The textual data used for FSA include email communications, social media posts (e.g., tweets), corporate reports, and daily news [18]. Financial corpora are labeled through either manual annotation [21, 35, 52, 57, 123, 181] or based on stock price [84]. Popular benchmark datasets have been summarized in Table 1. We summarize available FSA benchmark datasets not only to understand trends and templates in FSA annotation but also to facilitate researchers to find various public datasets that could be employed to evaluate their model performance and improve model generalization. Overall there is one document-level, four sentence-level, two target-level, and one targeted aspect-level dataset. The annotation is becoming more granular on target- and aspect-level. It is also observed that each public release of FSA benchmark

Benchmark Dataset	Number of Entries	Type of Annotation	Data Source
PhraseBank	Entries: 4,846 100% Agree (pos: 570, neu: 1391, neg: 303) 75% Agree (pos: 887, neu: 2146, neg: 420) 66% Agree (pos: 1,168, neu: 2,535, neg: 514) 50% Agree (pos: 1,363, neu: 2,879, neg: 604)	Sentence-level polarity (positive, neutral, negative)	News headline
SemEval 2017 Task 5	Entries: 2,836 Headline (pos: 653, neu: 38, neg: 451) Microblogs (pos: 1,086, neu: 27, neg: 581)	Targeted sentence-level sentiment score	News headline and microblogs
FiQA Task 1	Entries: 1,173 Headline (pos: 320, neu: 13, neg:165) Post (pos: 440, neu: 1, neg:234)	Targeted aspect-level sentiment score	News headline and posts
Topic-Specific Sentiment Analysis	Entries: 297*	Targeted document-level sentiment polarity (very negative, negative, slightly negative, neutral, slightly positive, positive, very positive)	News
StockSen	Entries: 55,171*	Sentence-level polarity (positive, neutral, negative)	StockTwits
SentiEcon GS-1000	Entries: 1,000*	Sentence-level polarity (positive, negative, none)	Business daily news
FinLin	Entries: 3,811*	Sentence-level sentiment score	Stocktwits, news articles, company reports and investor reports
SEntFiN	Entries: 10,753 pos: 5,074, neu: 5,517, neg: 3,814	Targeted sentence-level polarity (positive, neutral, negative)	News headline

Table 1. FSA Benchmark Datasets. * means annotation details is not publicly released.

dataset, particularly in open challenges such as SemEval 2017 Task 5 and FiQA Task 1 organized more recently, has promoted the research in FSA techniques. As highlighted by [131], the annotation is a challenging task as it is subject to human factors such as domain expertise as well as the annotator's private state and inference made into the text and earlier research also shows that the level of agreement on annotation varies on corpus annotated.

4.2.1 PhraseBank[106]. In 2014, [106] established a milestone dataset, i.e., PhraseBank, which includes 4,846 news annotated by 16 individuals who have adequate background knowledge in financial markets from an investor perspective. Based on the strength of agreement among annotators, it releases 4 reference datasets, namely 100%, 75%, 66%, and 50% agreement. In their study, [106] argues that the overall sentiment may be different from the prior sentiment polarity of individual words, and incorporating phrase-structure information and domain-specific use of language could improve the detection. [106] proposed Linearized Phrase Structure (LPS), which extends the quasi-compositional polarity-sequence framework proposed by [120] and is capable of accommodating key interactions between financial concepts and phrase-structure components. The LPS model first extracts the entities with semantic orientation, then projects the phrase structure, and finally performs the multi-label classification.

```
"sentence": "Pharmaceuticals group Orion Corp reported a fall in its third-quarter earnings that were hit by larger expenditures on R&D
and marketing .", "label": "negative"
```

4.2.2 SemEval 2017 Task 5. The SemEval 2017 Task 5 is for fine-grained FSA on news headlines and microblogs [25]. The training dataset includes 1,142 financial news headlines and 1,694 posts with their target entities and corresponding sentiment score as shown below. The test dataset consists of 491 financial news headlines and 794 posts. This task is to extract and detect the targets and their corresponding sentiment scores.

```
"id": 2, "company (target)": "Morrisons", "title": "Morrisons book second consecutive quarter of sales growth", "sentiment": 0.43
```

4.2.3 FiQA Task 1. The dataset is from FiQA Open Challenge Task 1 [103], which consists of 498 financial news headlines and 675 posts with their target entities, aspects, and corresponding sentiment score. The task is to extract and detect the targets, aspects, and their corresponding sentiment scores.


```
"sentence": "Royal Mail chairman Donald Brydon set to step down", "info": ["snippets": "['set to step down']", "target": "Royal Mail",
"sentiment_score": "-0.374", "aspects": "['Corporate/Appointment']"]
```


4.2.4 Topic-Specific Sentiment Analysis. [163] has created a benchmark dataset, which has 297 news documents extracted from the Thomson Reuters Newswire, for topic-specific sentiment analysis of economic texts. It has covered ten event topics that have significant financial impact such as Apple's iPad, the EuroZone crisis, GM's IPO, and the United-Continental merger. The 297 selected documents are equally distributed across all topics. A team of three experienced annotators is instructed to read and annotate the news documents as if they were investors in the company that was described in the topic statement [163]. The annotation uses a 7-point scale from very negative, negative, slightly negative, neutral, slightly positive, positive to very positive. The Kappa statistic, Intraclass Correlation, Robinson's A and Finn coefficient, and average percentage agreement are used to evaluate the degree of agreement between annotators and measure how reliable the annotation scheme is. This dataset is not publicly released but is available on a request basis.

4.2.5 StockSen [181]. The StockSen dataset consists of 55,171 financial tweets from StockTwits dated between 2019-06-06 and 2019-08-26. This dataset uses user annotations to investigate the common mistakes made by the lexicon, machine learning, and deep learning-based methods. It has shown that the same type of sentiment prediction models tend to have similar error patterns and identified six main error types that cause FSA fail. However, this dataset is not publicly released but is available on a request basis.

4.2.6 SentiEcon GS-1000 [122]. SentiEcon GS-1000 is a manually annotated gold standard dataset that contains 1,000 sentences extracted from the Esmeraldas Great Recession News Corpus. Two domain experts have classified each sentence as positive, negative, or none. The annotators were instructed to consider the information available in the sentences only for annotation. Annotation was carried out independently and a consensus was reached in differing cases.

4.2.7 FinLin [29]. The FinLin[5] corpus is released by Daudert in 2022 which aims to provide a novel and publicly available dataset for FSA to complement the current knowledge and foster research on FSA [29]. It contains a total of 3,811 texts including 3,204 stocktwits, 394 news articles, 127 company reports, and 86 investor reports. The corpus is annotated with a relevance score and a sentiment score in the range of $[0.0, 1.0]$ and $[-1.0, 1.0]$, respectively. Similarly, this dataset is not publicly released but is available on a request basis.


```
"type": "SW", "text": "$GM hot diggity dog", "created_at": "2018-09-19T13:12:15z", "entity": "GM", "id": 137731223, "sentiment": 0.7444,
"relevance": 0.0989, "annotations": ["sentiment": -0.0121, "spans": ["text_span": "hot"], "relevance": 0.6963, "sentiment": 0.7733, "spans":
["text_span": "hot diggity dog"], "relevance": 0.1049, "sentiment": 0, "spans": ["text_span": " hot diggity dog"], "relevance": 0.1159]
```


4.2.8 SEntFiN 1.0 [157]. In an effort to address the problem of scant benchmark dataset for fine-grained FSA, a challenging task that requires extensive human efforts for annotation, [157] released SEntFiN 1.0 and made publicly available to promote further research. SEntFiN is a human-annotated dataset that includes 10,753 news headlines with their entity and corresponding sentiment. It is common that multiple entities are present in a news headline with different

[5]https://github.com/TDaudert/FinLin

sentiment expressions and SEntFiN has 2,847 headlines that contain multiple entities, which may have conflicting sentiment [157].

"S No.": 1, "Title": "SpiceJet to issue 6.4 crore warrants to promoters", "Decisions": ['SpiceJet': 'neutral'], "Words": 8

4.2.9 Comparison of Benchmark Dataset. Table 1 shows that news data is the primary source for constructing benchmark datasets, followed by microblogs. News data is widely used for constructing sentiment analysis data in various fields, such as NewsMTSC for target-dependent sentiment classification on policy issues [64]. The quantity of entries varies across datasets, with those annotated by polarity typically featuring a higher number of labeled entries. For instance, in the case of the FiQA dataset, there are 498 entries derived from news and 675 from posts, presenting a potential challenge for model training and generalization. It is also worth noting that datasets originating from news sources adhere more to a formal English language structure. On the other hand, microblogs like tweets tend to feature more informal expressions and greater potential for aspect ambiguity, such as ticker names. This introduces additional complexities to FSA tasks. Furthermore, in terms of granularity, the fine-grained FSA dataset remains limited, with FiQA being the current preference for aspect-based FSA. Additionally, it is important to highlight that the majority of datasets annotate sentiment solely in terms of polarity, without capturing the intensity of sentiment.

4.3 Evaluation Metrics

4.3.1 Regression. The first group of metrics measures the closeness between the predicted value and ground truth in the context of a regression task. The popular metrics include Weighted Cosine Similarity (WCS), Mean Squared Error (MSE), and coefficient of determination or R-squared (R^2).

$$WCS = \frac{|P|}{|G|} \times \frac{\sum_{i=1}^{n}(G_i \times P_i)}{\sqrt{\sum_{i=1}^{n}(G_i^2)} \times \sqrt{\sum_{i=1}^{n}(P_i^2)}} \tag{1}$$

where P is the vector of scores predicted by the model and G is the vector of ground truth scores.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{2}$$

where y_i is the gold standard score and $\hat{y_i}$ is the score predicted by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})}{\sum_{i=1}^{n}(y_i - \bar{y_i})} \tag{3}$$

where y_i is the gold standard score and $\hat{y_i}$ is the score predicted by the model.

R^2 is a popular and intuitive measurement for how close the predictions fit the observations on a 0 to 1 scale. As a R^2 value closer to 1 signifies a good model, the model performance can be evaluated from R^2 even without comparing with other models. However, R^2 fails to determine whether the predictions are biased. MSE is the most common metric, which is often used in conjunction with R^2, for regression tasks. It places more weight on large errors by squaring to ensure that the trained model has no outlier predictions. The WCS metric is introduced by [25] to evaluate sentiment scores on a continuous scale between -1 and 1. It enables the comparison of the proximity between the ground truth vector and prediction vector, while not requiring exact correspondence between them for a given instance. The WCS is derived by weighting the cosine similarity with the proportion of scored instances aiming to reward models that attempt to predict all entries in the dataset.

4.3.2 Classification. The second group of metrics measures the categorical accuracy between predicted value and ground truth in the context of classification tasks. The popular metrics include Accuracy, Matthews Correlation Coefficient (MCC), and F1-Score.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{5}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

Here, Precision = TP/(TP + FP) and Recall = TP/(TP + FN). True Positive (TP) is the count of positive classes that are correctly predicted and True Negative (TN) is the count of negative classes that are correctly predicted. Similarly, False Positive (FP) is the count of positive classes that are incorrectly predicted and False Negative (FN) is the count of negative classes that are incorrectly predicted. Accuracy score can be calculated easily for both binary and multi-class classification. However, it cannot be considered a reliable measure when the data is imbalanced as it makes the classifier estimate over-optimistically on the majority class. F1-Score has addressed this issue and has been widely adopted in most application areas of machine learning. However, [23] argues that the F1-Score is independent of true negative, which is considered a conceptual flaw, and shows that MCC, which factors in TP, TN, FP, and FN together, can produce a score for evaluation of binary classification that is more informative and truthful than accuracy and F1-score.

4.4 Methods

4.4.1 Lexicon Approaches. The lexicon-based method is to detect the semantic orientation of the text based on the semantic orientation of the words in the text. Lexicon construction is a key element for sentiment analysis, which could be accomplished in manual, semi-automatic, or automatic manner [133]. The manual approach requires intensive efforts from creators with expert knowledge, which is slow but generally the accuracy is higher. On the other hand, the automatic approach is fast and scalable but often results in sacrifice in accuracy to some extent. The biggest advantage of the lexicon-based method is that no annotated dataset is required to perform FSA as it is unsupervised, which reduces the need for arduous manual annotation of the texts. Meanwhile, lexicons are useful to create features for supervised learning tasks. The challenge with lexicon-based approach is that it is time-consuming to build lexicons and also hard to generalize [160]. Also, it only can detect explicit sentiment and usually is less accurate than the learning-based method due to the constraint in coverage and quantification of sentiment intensity. More importantly, sentiment analysis is sensitive to domain [169] and generic domain-independent lexicons are often ineffective in FSA [133]. A general-purpose sentiment analysis lexicon may misclassify common words in financial texts [97]. For example, words like "liability" and "debt" are considered negative in general-purpose sentiment analysis, but are frequent and often neutral in the financial context. This makes it difficult to generalize the sentiment classifiers and underlines the need for finance domain-specific sentiment analysis [97].

The construction of lexicons in the financial domain is scant [133] as compared to general-purpose lexicons. In the context of FSA, there are six popular finance domain-specific lexicons as shown in Table 2, namely Henry's Financial Dictionary (HFD), Loughran and McDonald (LM) Word List, Stock Market Sentiment Lexicon (SMSL), SentiEcon, Senti-DD, and FinSenticNet. HFD, which includes 104 positive words and 85 negative words, is the first dictionary that was created specifically for the financial domain from earning press releases. It is used to measure the tone of earnings press releases which is an important element of the firm-investor communication process [66]. The weakness of HFD is its limited number of words which can result in low coverage. One prominent effort that advances the development of

Lexicon	Description	Approach	Source
HFD	104 positive words and 85 negative words	Manual, single word	Earnings press releases
LM	354 positive, 2,355 negative, 297 uncertainty, 904 litigious, 19 strong modal, 27 weak modal and 184 constraining words	Manual, single word	Text documents from the U.S. Securities and Exchange Commission
SMSL	20,550 entries: 10,534 positive, 10 neutral and 10,006 negative words	Automatic, single and multi-word	StockTwitts
SentiEcon	6,470 entries: 1,365 positive, 2,661 neutral and 2,444 negative words	Manual, single and multi-word	Business news articles from The Guardian and The New York Times
Senti-DD	2,573 entries: 1,597 positive and 976 negative direction-dependent words	Automatic, direction-dependent words e.g., (profit, up)	Phrasebank datasets
FinSenticNet	6,741 entries: 3,441 positive and 3,300 negative words and concepts	Manual, single and multi-word	Phrasebank, SemEval 2017 Task 5, FiQA Task 1 and SEntFiN datasets

Table 2. Financial Lexicons.

financial sentiment lexicon is the introduction of LM lexicon by Loughran and McDonald [97]. The authors manually examined the quality of GI (Harvard General Inquirer) and proposed a revised lexicon, which is specifically for financial texts. LM sentiment word list is created from the annual reports released by firms which includes 354 positive, 2,355 negative, 297 uncertainty, 904 litigious, 19 strong modal, 27 weak modal, and 184 constraining words [97]. It is the most commonly used lexicon we are aware of that is created for the financial domain. The limitation of HFD and LM is that they operate at the word level and fail to account for contextual nuances. [133] has proposed a novel and fast approach to built SMSL using labeled StockTwitts data and statistical measures. SMSL is created based on labeled tweets from StockTwits which is a microblog that is specialized in the stock market. This lexicon includes 20,550 words and phrases and shows competitive results in measuring investor sentiments [133]. Lastly, designed for sentiment analysis applications, SentiEcon is a comprehensive and large domain-specific computational lexicon for Finance and Economy. It consists of 6,470 entries of single and multi-word expressions, and each entry has tags denoting their semantic orientation and intensity [122]. One important note is that SentiEcon is designed to combine with a general domain lexicon as it only compiles entries whose domain-specific polarity is different from or not recorded in the general domain lexicon. FSA lexicons can be enriched and improved from the following aspects. First, the concept-level domain-specific lexicon can be added. For example, the concept of "making profit" is positive while "making loss" is a negative concept. Also, it is important to build context-aware and direction-dependent lexicon [136]. For example, the direction-dependent phrases "profit-up" and "loss-down" are positive while "profit-down" and "loss-up" have negative polarities. [136] has constructed context-aware sentiment lexicon using direction-dependent words, combined with LM to perform lexicon-based FSA, and achieved state-of-the-art performance on PhraseBank dataset. The most recent study by [42] has proposed FinSenticNet and conducted extensive experiments to assess the performance of various lexicons across benchmark datasets. The results indicate that the concept-level lexicon FinSenticNet outperforms both general-purpose and financial lexicons across evaluation datasets. Additionally, FinSenticNet surpasses FinBERT in SemEval 2017 Task 5 and FiQA Task 1, underscoring its effectiveness in accurately capturing and analyzing sentiment within the financial domain. Meanwhile, HFD shows promising results on both the PhraseBank and SEntiFin datasets, though it has fewer words. LM exhibits suboptimal performance mostly due to a very pronounced imbalance in the number of positive and negative words.

4.4.2 Machine Learning Approaches. Machine learning approaches make use of classification or regression algorithms to determine sentiment by constructing features. There are three important steps in machine learning approaches: feature engineering, feature selection, and algorithm selection. In terms of feature creation, it can be categorized into four types (1) linguistic features (e.g., n-grams, RF n-gram, verb, NER, and word cluster) (2) sentiment lexicon features (e.g., the proportion of positive and negative words, maximum, minimum and sum of sentiment score) (3) domain-specific features such as number (e.g., + number, - number, + number %, - number %), keyword + number (e.g., call + number %, put + number %), metadata (e.g., binary features such as source, user/official and entities/sentiment, value features and other features) and punctuation, and (4) word embeddings. In machine learning approaches, feature selection is equally important as noise features need to be identified and excluded. The popular feature selectors include Chi-squared, ANOVA, and mutual information [160]. In terms of algorithms, Support Vector Machines (SVM) is one of the most adopted algorithms, and other algorithms such as Bagging, Random Forest (RF), AdaBoost, Gradient Boosting (GB) and XGBoost (XGB) are also among the popular algorithms to be selected. For example, [73] has elaborately designed all above four types of features on SemEval 2017 Task 5 dataset, adopted a hill climbing algorithm to select the best features, and explored seven algorithms as follows: Bagging, RF, AdaBoost, GB, LASSO, Support Vector Regression (SVR) and XGB. An ensemble learning has been applied to different algorithms such as SVR + GB and SVR + XGB + AdaBoost + Bagging and achieved promising results. [30] established a strong baseline with a traditional feature engineering-based machine learning approach (MSE=0.0958) by treating aspect extraction as a classification task and sentiment detection as a regression task using Support Vector Classifier (SVC) and SVR respectively. The features generated include n-gram, tokenization, word replacement, and word embeddings using Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec. [5] has adopted linguistic features (uni-gram, bi-grams, and tri-grams) and semantic features (BabelNet synsets and semantic frames). The features are selected by the word-score correlation metric proposed by [39] and the sentiment regressor is trained using an SVR, which achieved an accuracy of 0.726 for microblogs and 0.655 for news headlines. [88] proposed an association rule mining-based Hierarchical Sentiment Classifier (HSC) which adopts the concept of financial and non-financial performance indicators, to classify financial texts into positive, neutral, or negative polarity on the PhraseBank dataset. [152] has used ontology information as a source of features in the SVM model on SemEval 2017 Task 5 dataset. The newly created domain ontology models various concepts from the financial domain and has four classes, which are sentiment, entity, property, and action. [147] has adopted a feature-light method that consists of an SVR with various kernels and word embedding vectors as features.

4.4.3 Deep Learning Approaches. Deep Learning models have achieved remarkable performance in FSA. It is able to construct complicated representations from textual data with a high level of abstraction [160]. The most popular deep learning algorithms used in FSA include Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) and their variants. For example, When target-aspect identification is jointly considered, [72] treated aspect extraction as a multi-class classification problem, as this task does not involve multiple aspects for one target, and adopted bidirectional Long Short-Term Memory (LSTM) to extract aspects using word embeddings such as GloVe, Google-News-Word2Vec, Godin, FastText, and Keras in-built embedding layer, while a multi-channel CNN is used for sentiment analysis task with enhanced vector combined from dependency tree, sentence word vector and snippet and target vector. The Bayesian optimization is used for hyper-parameters tuning to find out the most optimal parameters which achieves an F1-Score of 0.69 for aspect extraction and MSE of 0.112 for sentiment analysis. [139] has ensemble CNNs and RNNs with a voting strategy and a ridge regression for aspect and sentiment prediction. [99] has proposed FSA with Hierarchical Query-driven Attention (FISHQA) for financial polarity detection on the document level which outperforms benchmark models including SVM + Bag-of-Words (BoW), SVM + BoW TF-IDF, CNN-word [82], Bi-LSTM [61], LSTM-GRNN [164] and HAN [189] on a dataset which includes 7,648 documents annotated by three domain experts in the perspective that whether the corresponding bonds of the companies mentioned in the document will encounter the risk of default in the future. FISHQA has achieved an accuracy of 0.9446 and an F1-Score of 0.9449, followed by HAN, a model that adopts hierarchical networks with an attention mechanism of random initialization, having an accuracy of 0.9177 and an F1-Score of 0.9166.

4.4.4 Hybrid Approaches. Hybrid approaches refer to the ensemble of lexicon methods, machine learning models, and deep learning models and often yield superior performance. [1] has proposed a method by ensembling LSTM, CNN, Gated Recurrent Unit (GRU), and SVM. In terms of word embeddings for LSTM [67], CNN, and GRU, the word representation is learned through the stacked denoising auto-encoder network [171] by using Word2Vec [119] and GloVe [137]. In the SVR model, word TF-IDF, lexicon features and Vader sentiment [71] have been chosen as features. It has achieved a cosine similarity of 0.797 for microblogs data and 0.786 for news headlines on the SemEval 2017 Task 5. A hybrid of deep learning and lexicon-based technique that combines LSTM, CNN, Vector Averaging MLP and Feature Driven MLP (e.g., character n-grams (TF-IDF weighted counts of a continuous sequence of N characters), word n-grams (TF-IDF weighted counts of continuous sequence of N words), POS-tag, lexicons (e.g., positive count, negative count, net count, sum of positive score, sum of negative score, maximum of positive and negative scores)), which is proposed by [58] has achieved promising result (microblogs: Cosine=0.751, news headlines: Cosine=0.697) on the same dataset. The highest score (Cosine=0.745) for SemEval-2017 Task 5 Track 2 news headlines is reported by [109], which combines GloVe and DepecheMood to represent words and feed into CNN followed by global max-pooling, and the output is then concatenated with Vader sentiment score for two levels of drop and fully-connected layers. [77] combined the representation learned from CNN and Bidirectional GRU (Bi-GRU) with attention mechanism with manually engineered lexical, sentiment, and metadata features and obtained WCS scores of 0.723 and 0.744 for the Microblogs and the News Headline tracks in SemEval-2017 Task 5, respectively. Recently, MetaPro was proposed to improve FSA by understanding metaphors in financial text [113]. The linguistic intuition is that metaphors frequently appear in financial news, causing errors in sentiment analysis. MetaPro can paraphrase metaphors into their literal counterparts via data pre-processing, so that a sentiment classifier can achieve better performance in downstream applications. MetaPro consists of a multitask learning-based module for metaphor identification [112] and a WordNet[51]-based metaphor interpretation module [114]. A novel soft-parameter sharing method, termed Gated Bridging Mechanism (GBM), and a knowledge-enhanced masked word prediction technique are proposed. The average accuracy gain of three state-of-the-art sentiment analysis classifiers is 4.7% on the SemEval 2017 Task 5 news headline dataset.

4.4.5 Pre-Trained Language Models. The emergence of pre-trained language models and transfer learning has brought Natural Language Processing (NLP) research to a new era. This involves pre-training a neural network model on a large corpus of text, and the pre-trained models, such as Bidirectional Encoder Representations from Transformers (BERT) [36], are capable of capturing rich contextual information, enabling them to be adapted to various downstream tasks. Fine-tuning is the subsequent step in transfer learning which involves taking the pre-trained model and further training it on a task-specific dataset. This process fine-tunes the model's parameters to suit the specific requirements of the target application, thus enhancing its performance on the given task. Domain adaptation is a related concept that addresses the challenge of applying a pre-trained model to a specific domain or application for which it was not originally trained. This process involves adapting the model to perform effectively in the target domain, even if it differs significantly from the domain on which the model was pre-trained. In the finance domain, domain-specific transformer-based models have significantly enhanced the performance of various financial NLP tasks such as FinBERT [4, 96, 197] for FSA and FinBERT-MRC [197] for financial named entity recognition. In terms of the adoption of pre-trained language models in FSA, [186] reported a superior MSE of 0.08 using ULMFiT [68] on the FiQA Task 1. A more recent fine-tuned language model FinBERT [4], which is further pre-trained by TRC2-financial corpus[6], reported the best performance (MSE=0.07, R^2=0.55) on the FiQA Task 1. Similarly, [96] trained FinBERT using two general domain corpus including English Wikipedia and BooksCorpus, and three financial domain corpus including FinancialWeb, YahooFinance, and RedditFinanceQA. It has achieved the state-of-the-art performance on financial sentence boundary detection on the FinSBD English dataset with a mean score of 0.97 and FSA on the PhraseBank dataset with an accuracy of 0.94 and an F1-Score of 0.93. More recently, [178] proposed a Semantic and Syntactic Enhanced Neural Model (SSENM), which obtains input representation using BERT model and incorporates dependency graph and context words to supervise a target representation. This novel model captures semantic contextual information through a self-attentive mechanism. An edge-enhanced Graph Convolutional Network (E-GCN) is included to aggregate node-to-node features and a Manifold Mixup strategy is also developed to generate pseudo data in training to address the over-fitting problem potentially caused by limited data size. SSENM has significantly improved its performance with a WCS of 0.8441 for news headline and 0.8333 for microblogs in SemEval-2017 Task 5 and an MSE of 0.0717 in FiQA Task 1. Finally, it is important to highlight that while models primarily based on transformer encoder architectures, such as BERT, have significantly enhanced FSA, autoregressive decoder architectures like GPT (Generative Pre-trained Transformer) [142], have also demonstrated promise in FSA [50, 69, 194]. Bloomberg has introduced BloombergGPT [176], a Large Language Model in finance that outperforms similarly sized open models on financial NLP tasks including such as sentiment analysis, named entity recognition, news classification, question answering, etc.

4.4.6 Word Representation Techniques. Word representation is always an important component of sentiment analysis. This section aims to introduce both classic and state-of-the-art word representation techniques which can be used for FSA. In general, the word representation techniques could be grouped as classical models and representation learning models [126]. Classical models include categorical word representation (e.g., BoW and one hot encoding). This is the most straightforward method for text representation but they cannot capture positional and structural information as well as semantic relationships between words [72, 95]. Another classic model is weighted word representation which includes TF-IDF. Techniques such as TF-IDF can reduce the impact of common words but it is still built on the concept of the BOW model, which fails to capture the sequence of words in a document, semantics, and syntactical information of words [126]. To address the shortcomings of classic models, researchers study the methods to learn the distributed word representation in low-dimensional space. Overall representation learning can be categorized into contextual and non-contextual word representations. For non-contextual embeddings, the most popular model is Word2Vec, which is developed by [119] and can capture the semantics of words and manipulate the connectivity of words including sentimental similarity among words [160]. The problem with Word2vec is it only focuses on local context window knowledge but ignores global statistical information. Global Vectors (GloVe) [108] is thus presented. Meanwhile, FastText, which has decreased the training time but maintained the performance, is introduced by [12]. While the semantic and syntactic information of the text can be retained in non-contextual word representation models, there remains the problem of how the full context-specific representation can be kept. Therefore, contextual word representation techniques, such as generic context word representation (Context2Vec) [117], contextualized word vectors (CoVe) [116], embedding from Language Models (ELMo) [138], universal language model fine-tuning (ULMFiT) [68] and transformer-based pre-trained language models, are proposed to resolve this problem. It has been proven that transformer-based pre-trained language models are more efficient than LSTM or CNN models for language representation. Models such as GPT [143], XLNet [188], BERT [36] and Robustly optimized BERT approach (RoBERTa) [94] are from the transformer

[6]https://trec.nist.gov/data/reuters/reuters.html

family. GPT and XLNet are decoder-only architecture while BERT and RoBERTa are encoder-only architecture model. One of the most popular models for sentence embeddings is Sentence-BERT [144], which is a modification of the pre-trained BERT network which use siamese and triplet network structures, enabling it to effectively capture the semantic significance of sentence embeddings.

4.4.7 Summary of FSA Techniques. We summarize the characteristics of different technical trends in Tables 3. The evolution of FSA techniques has been marked by a progression from lexicon-based methods, conventional machine learning, and hybrid approaches to more advanced techniques involving deep learning and language models. It is observed that most of the research in FSA techniques has adopted the long-established PhraseBank, SemEval 2017 Task 5, and FiQA Task 1 as the benchmark datasets. First, the creation of a financial lexicon is still attracting researcher's attention although the lexicon approach is more often used in combination with learning-based methods in FSA. One trend in the development of financial lexicons is that it is shifting from single-word to multiple-word and direction-dependent expressions. This is particularly important in the finance domain as the sentiment of a financial term can be opposite for different directional words. Meanwhile, manual creation is still the main approach to building financial lexicons which requires intensive efforts from creators with expert knowledge but generally has higher accuracy. However, there is research [42, 133] which is pushing lexicon construction from manual to automatic approaches that enable us to address the slow process and low coverage issues in manual approach and build lexicons with increased speed and coverage. Second, in traditional machine learning methods, feature engineering is an important step and there are generally four types of features namely lexicon features, linguistic features, domain-specific features, and word embeddings. One type of feature that is less frequently investigated but is demonstrated effective in FSA is the domain-specific features such as numbers and keywords + numbers, especially when plenty of numbers are mentioned in the texts. In the financial context, for instance, a keyword of "revenue" followed by a positive symbol "+" and percentage can be a positive sign. Lastly, deep learning methods, represented by CNN and LSTM, and pre-trained language models are the mainstream techniques in FSA that have improved the performance significantly. Notably, the finance domain-specific BERT, namely FinBERT, is trained by using various data sources such as Reuters Corpora, Yahoo Finance, Raddit Finance, corporate reports, earnings call transcripts and analyst reports in different studies and pushed the boundary of research in FSA techniques. The most recent study by [41] has achieved state-of-the-art performance on SemEval 2017 Task 5 and FiQA Task 1 by incorporating multiple knowledge sources into the fine-tuning process of language models including RoBERTa.

5 FSA APPLICATIONS

FSA has been widely used in financial applications since the asymmetric and affective impact of news on market volatility has been discovered [45]. The application of FSA is mainly contextualized for two broader analytical purposes, i.e., hypothesis testing and predictive modeling in financial markets. Unlike the annotated benchmark datasets for FSA techniques, which aim to develop an accurate sentiment analysis model from a FSA technique standpoint, the data sources for FSA applications hold significant importance which offer supplementary information crucial for calibrating financial sentiment to their specific application scenario. We first review data sources for FSA applications and investigate how the techniques are adjusted for various application types.

Ref.	Dataset	Task	Method	Feature/Lexicon	Algorithm	Evaluation Metrics and Performance
[136]	PhraseBank	Sentence-level	Lexicon	LM and Senti-DD	Lexicon	Precision: 0.8238, Recall: 0.8128, F1-Score: 0.8105
[42]	PhraseBank	Sentence-level	Lexicon	FinSenticNet	Lexicon	Accuracy: 0.7619, F1-Score: 0.7216
[88]	PhraseBank	Sentence-level	Machine Learning	Performance indicator tags	Classification based on Multiple Association Rules (CMAR)	Precision: Pos: 0.83, Neg: 0.93, Neut: 0.86, Recall: Pos: 0.82, Neg: 0.93, Neut: 0.81, F1-Score: Pos: 0.83, Neg: 0.93, Neut: 0.83
[96]	PhraseBank	Sentence-level	Language Model	FinBERT	FinBERT	Accuracy: 0.94, F1-Score: 0.93
[73]	SemEval 2017 Task 5	Targeted, Sentence-level	Machine Learning	Linguistic, sentiment lexicon, domain-specific features and word embeddings	AdaBoost, Bagging, Random Forest, Gradient Boosting, LASSO, SVM and XGBoost	WCS (News: 0.7779, Microblogs: 0.7107)
[40]	SemEval 2017 Task 5	Targeted, Sentence-level	Machine Learning	Lexical features, semantic features	Linear regression with SGD, Lass with SGD, Ridge regression with SGD, SVR and RF	WCS (News: 0.655, Microblogs: 0.726)
[152]	SemEval 2017 Task 5	Targeted, Sentence-level	Machine Learning	Ontology based features	SVM	WCS (News and Microblogs: 0.7050)
[147]	SemEval 2017 Task 5	Targeted, Sentence-level	Machine Learning	Word embeddings	SVM	WCS (News: 0.733)
[1]	SemEval 2017 Task 5	Targeted, Sentence-level	Hybrid Approach	Lexicon features and word embeddings	SVM, GRU, LSTM, CNN	WCS (News: 0.786, Microblogs: 0.797)
[58]	SemEval 2017 Task 5	Targeted, Sentence-level	Hybrid Approach	Word embeddings using Word2Vec and GloVe, word n-grams, character n-grams, POS-tag, lexicons, pointwise mutual information	LSTM, CNN, Vector Averaging MLP and feature driven MLP	WCS (News: 0.697, Microblogs: 0.751)
[77]	SemEval 2017 Task 5	Targeted, Sentence-level	Hybrid Approach	Word embeddings, lexical, sentiment and metadata features	CNN, Bidirectional GRU (Bi-GRU)	WCS (News: 0.744)
[109]	SemEval 2017 Task 5	Targeted, Sentence-level	Hybrid Approach	GloVe and DepecheMood word embeddings, VADER lexicon	CNN, MLP	WCS (News: 0.745)
[113]	SemEval 2017 Task 5	Sentence-level	Hybrid Approach	RoBERTa	WordNet, RoBERTa, Transformer, masked word prediction, GBM	Avg. F1-Score: +0.040, Avg. accuracy: +0.047
[178]	SemEval 2017 Task 5	Targeted, Sentence-level	Language Model	BERT representation	Semantic and Syntactic Enhanced Neural Model (SSENM)	WCS (News: 0.8441, Microblogs: 0.8333)
[41]	SemEval 2017 Task 5	Targeted, Sentence-level	Language Model	RoBERTa	RoBERTa	WCS (News: 0.8483, Microblogs: 0.8122)
[30]	FiQA Task 1	Targeted, Aspect-level	Machine Learning	Linguistic and word embeddings	SVM	Aspect Extraction: F1-Score (News: 0.4240, Post: 0.5775), Sentiment Analysis: MSE (News: 0.1052, Microblogs: 0.1281)
[72]	FiQA Task 1	Targeted, Aspect-level	Deep Learning	GloVe, Google-News-Word2Vec, Godin, FastText, and Keras in-built embedding layer	CNN, LSTM	Aspect Extraction: F1-Score (0.69), Sentiment Analysis: MSE (0.112)
[139]	FiQA Task 1	Targeted, Aspect-level	Deep Learning	Pre-trained word embeddings using CNN	Ridge, Random Forest, CNN, GRU/Bi-GRU and Bi-LSTM	Aspect Extraction: F1-Score (0.6530), Sentiment Analysis: MSE (0.0926)
[178]	FiQA Task 1	Targeted, Aspect-level	Language Model	BERT representation	Semantic and Syntactic Enhanced Neural Model (SSENM)	Sentiment Analysis: MSE (0.0717), R^2 (0.4878)
[4]	FiQA Task 1	Targeted, Aspect-level	Language Model	FinBERT	FinBERT	Sentiment Analysis: MSE (0.07)
[41]	FiQA Task 1	Targeted, Aspect-level	Language Model	RoBERTa	RoBERTa	MSE: 0.0490
[157]	SEntFiN	Targeted, Sentence-level	Language Model	Pre-trained sentence representation	RoBERTa, FinBERT	Accuracy: 0.9429, F1-Score: 0.9327
[42]	SEntFiN	Targeted, Sentence-level	Lexicon	FinSenticNet	Lexicon	Accuracy: 0.5920, F1-Score: 0.5939
[99]	7648 documents (non-neg: 3681, neg: 3957)	Document-level	Deep Learning	Word embeddings	Hierarchical Query-driven Attention (FISHQA) using GRU	Accuracy: 0.9446, F1-Score: 0.9449
[198]	CCF BDCI, CCKS	Sentence-level	Language Model	RoBERTa	RoBERTa	Accuracy: 96.774%

Table 3. Performance of FSA.

5.1 Data Sources

The data sources used for FSA applications can be categorized into three types. The first type is the data that is officially released by the company which includes financial reports, filings, earnings calls, etc. The second type is survey sentiment indicators from sentiment surveys, which include AAII, UMSC, and Sentix. For example, AAII Investor Sentiment Survey is conducted weekly and UMSC is monthly [53]. Lastly, public news (e.g., financial news, macroeconomic news) and social media are another two important data resources for financial market prediction in recent years [185], which has grown rapidly over the last decade with the advancement of news and social media data processing technologies. There is strong evidence that social media and financial news influence market dynamics [90]. As the Internet and Web 2.0 phenomenon have grown, social media data from platforms such as StockTwits and Twitter have become increasingly important data sources [49]. For instance, [158] confirm that sentiment analysis of social media data is predictive of future stock market movement.

Financial Metrics	Description
Market index	Predict market index such as S&P 500
Price	Predict the price e.g., stocks, FOREX, and cryptocurrency
Price movement	Predict price moving up or down
Price change or return	Predict price change or return
Price change rate or return rate	Predict price change rate or return rate
Trading volume	Predict trading volume
Volatility	Predict volatility
Change in target price	Predict change in analysts' price targets
Crash risk	Predict risk metrics such as Negative Coefficient of Skewness (NCSKEW) and Down-to-Up Volatility (DUVOL)
Earnings	Predict earnings
Cash flow	Predict cash flow

Table 4. Financial Metrics.

5.2 Financial Applications

One of the most attractive applications of FSA is to perform forecasting in financial markets [124, 183]. After selecting the data source, one or more sentiment measures are constructed through the aforementioned techniques, which could be in the form of explicit sentiment representation (e.g., sentiment polarity or score) or implicit sentiment representation (e.g., sentence embeddings). Financial sentiment can be utilized explicitly, for instance by analyzing financial texts with critical linguistic features such as content semantics [79] or investors' sentiment [104], for the sake of interpretability. Alternatively, it can be applied implicitly by directly encoding financial texts by neural networks and using the representations for learning downstream tasks [185] such as financial forecasting. Armed with the financial sentiment representation, we can investigate the hypotheses about how the sentiment interacts with the financial metrics and vice versa. The financial metrics are typically derived from either financial reports or financial market information. There are basically seven types of financial markets which include capital (e.g., stock and bond), commodity, money, derivatives, future, insurance, and foreign exchange markets. The most studied financial markets in FSA are the capital market and foreign exchange market. Table 4 has listed down the popular financial metrics defined in earlier studies. For example, stock market variables that can be predicted include stock price, stock price movement, return, volatility, and trading volume [134].

5.2.1 Causality and Correlation Testing. This area of study fundamentally focuses on the investigation of correlation and causality between sentiment from financial texts and market performance, which means the sentiment can be used to reflect the correlation and/or causality with other financial measures such as return, risk, and volatility [172]. For example, [173] conducted a correlation analysis between stock performance and sentiment from SeekingAlpha and StockTwits and concluded that while StockTwits messages and SeekingAlpha articles provide minimal correlation to stock performance in aggregate, a subset of experts contribute more valuable content with predictive power. It has also been demonstrated that public sentiment has a correlation with stock market movement and specifically good news has a positive impact on markets and increases optimism [104]. Using a Granger causality analysis and a Self-Organizing Fuzzy Neural Network to investigate the hypothesis that the mood states of the public can cause changes in the Dow Jones Industrial Average (DJIA) closing prices, [13] found that emotions from Twitter messages can be good predictors of market trends. The public mood states are measured by the OpinionFinder (i.e., positive vs. negative) and Google-Profile of Mood States which measures mood in six dimensions (i.e., alert, calm, happy, kind, sure and vital, kind) time series. Similarly, [124] has extracted sentiment attitude (e.g., positive vs negative) and sentiment emotion (e.g., joy, sadness, etc.) from financial news and tweets to perform Granger causality test before predicting stock market price movement and concluded that sentiment attitude does not seem to Granger-cause stock price changes while emotion does Granger-cause stock price changes on specific occasions. The addition of sentiment emotions has increased the model accuracy for certain stock price prediction. The sentiment attitude and stock price time series are verified to be stationary by analyzing the autocorrelation and partial autocorrelation and performing the augmented Dickey-Fuller or the Ljung-Box t-statistic tests. [158] developed and applied an active learning approach to perform sentiment analysis of tweet streams in the equity market. The Granger causality test demonstrates that sentiment derived from stock-related tweets can serve as indicators of stock price movements a few days in advance. The results are improved further by adopting the SVM classifier to categorize tweets into three sentiment polarities (i.e., positive, negative, and neutral) instead of two polarities only (i.e., positive and negative) [158]. [31] find that Twitter activity is related to market participants' interest in and attention to 8-K filings, which in turn affect stock price and volume reactions to 8-K filings. The results also show that positive abnormal sentiment is not significantly associated with stock price reactions but is significantly negatively associated with stock volume reactions. [62] examined the correlation between four distinct investor emotions (fear, gloom, joy, stress) and S&P 500 index returns using the Threshold Generalized Auto-regressive Conditional Heteroskedasticity (TGARCH) model and discovered that fear emotion has a significant and lasting impact on conditional volatility and market returns. It is also found that the abnormal returns associated to emotion experience rapid reversal within 5 days. [26] investigates the effect of investor sentiment on stock price crash risk, which is measured by the Negative Coefficient of Skewness (NCSKEW) and the Down-to-Up Volatility (DUVOL) of the weekly stock return, by testing their main hypothesis by performing OLS analysis, and concludes that the risk of collapse is impacted by investor sentiment.

5.2.2 Stock Market Prediction. The stock market prediction is a challenging task due to its inherently noisy and volatile characteristics. Most earlier research in market prediction uses historical stock trading data, the technical indicators of stock trading, and macroeconomic variables as input data. The inclusion of financial textual data and application of NLP techniques in financial forecasting is an emerging research field [101, 124, 182, 183]. Traditional financial news providers such as Bloomberg and Thomson Reuters, have pioneered to provide commercial FSA service [124, 134]. Today, many investment banks and fund managers are exploiting financial sentiment to make better predictions of the financial market. Financial institutions such as Two Sigma and D. E. Shaw have included financial sentiment signals, in addition to traditional structured transaction data, to improve their machine learning model for algorithmic trading [124]. Practical traders agree that any results above 50% are value-added to their day-to-day trading [128]. Conventionally there are two major schools of thought in stock market analysis: fundamental analysis and technical analysis [128]. Fundamental analysis is to evaluate stocks from their intrinsic value perspective from economy, and industry conditions to the financial strength of individual companies. Financial indicators such as earnings, expenses, assets, and liabilities are part of fundamental analysis. Technical analysis attempts to identify opportunities from statistical trends in the movements of stock's price volume. Popular technical indicators include Simple Moving Average (SMA), Exponential Moving Average (EMA), and Moving Average Convergence/Divergence(MACD). Natural language-based financial forecasting techniques, as suggested by [19], could be classified as technical analysis. Essentially, the intrinsic value does not change with the sentiment and indicators that measure market sentiment, such as the High-Low Index and Bullish Percent Index which are important indicators in technical analysis.

Investor sentiment refers to the degree to which investors' beliefs about future firm valuation deviates from fundamental information and existing studies show that investor sentiment has a significant impact on stock prices and market participants' activities [118]. The studies in stock market prediction include stock index, stock price, stock

Ref.	Data Source	Period	Text Representation	Markets	Method	Task	Performance
[38]	Thomson Reuters and Bloomberg	Oct-2006 to Nov-2013	Event Embedding	US	CNN	Stock price prediction, S&P500 index prediction	Accuracy: 64.21%, MCC: 0.40 for index and Accuracy: 65.48%, MCC: 0.41 for individual stock
[95]	Thomson Reuters	Oct-2011 to Jul-2017	Feature extraction from news title using CNN and from event tuple using TransE model	US	SVM, LSTM	Stock price movement prediction	Accuracy: 55.44%, F1-Score: 0.7133
[32]	News and comments from Engadget	1-Jan-2006 to 15-Aug-2008	Sentiment (objective, subjective, negative, positive) using SentiWordNet.	US	SVM with Multiple Kernel Learning	Stock price change rate prediction	Sharp (MAE: 0.2368, MAPE: 1.3501, RMSE: 0.3025), Panasonic (MAE: 0.2673, MAPE: 1.3178, RMSE: 0.3435), Sony (MAE: 0.7001, MAPE: 1.4630, RMSE: 0.8865)
[179]	Thomson Reuters	Jan-2007 to Aug-2012	Frames, BOW, and part-of-speech specific DAL (FWD) features and SemTree data representations	US	SVM	Stock price movement prediction, Stock price change prediction	MCC: ConsumerStaples (0.1550 for change and 0.1147 for movement), Information Technology (0.1017 for change and 0.0801 for movement), Telecommunication Services (0.3049 for change and 0.0770 for movement)
[185]	Twitter	1-Jan-2014 to 1-Jan-2016	Learnt Embeddings	US	GRU, VAEs	Stock price movement prediction	Accuracy: 0.5823, MCC: 0.080796
[79]	Earning calls	1-Jan-2010 to 31-Dec-2017	Market, semantic (Doc2Vec, BoW) and pragmatic features	US	Ridge Regression, Logistic Regression, LSTM, Ensemble	Change in target price	Regression (MSE: 0.00137, R^2: 0.1718), Classification (Accuracy: 0.482, F1-Score: 0.475)
[134]	Twitter	22-Dec-2012 to 29-Oct-2015	Sentiment indicators based on SMSL lexicon which combines with AAII, II, UMSC, Sentix	US	MR, NN, SVM, RF, Ensemble	Return, trading volume and volatility	Daily return (Lowest NMAE: NDQ, 7.58), Daily trading volume (Lowest NMAE: DJIA, 5.84), Daily volatility (Lowest NMAE: DJIA, 2.79)
[200]	EDT Dataset	PRNewswire: 1-Mar-2020 to 30-Apr-2021, Businesswire: 16-Aug-2020 to 6-May-2021	High-level event detector incorporates entire article's representation and Low-level model detects events, which results to discover events at article-level	US	Multi-class Classification with MLM Loss	Stock price movement prediction	Trade at End Strategy:- Average Return: 1.74%, Exceed return of $84443 (844%) in 1-day trading and Trade at Best Strategy:- Average Return: 9.11%
[76]	Stocktwits	4-Mar-2013 to 28-Feb-2018	CNN as base model of sentiment index	US	Empirical Mode Decomposition (EMD) with LSTM based model	Stock price prediction	Accuracy: 70.56%, MAPE: 1.65%, MAE: 2.39
[177]	EastMoney.com	1-Jul-2017 to 30-Apr-2020	Skip-gram	China	CNN and LSTM	Stock price prediction	MAE: 2.38, MSE: 7.27, RMSE: 2.69
[101]	Hundsun Electronics	2-January-2018 to 18-June-2021	Market-driven sentiment representations	China	Pre-training, BiLSTM and GCN	Stock price movement prediction	Accuracy: 0.6726, MCC: 0.3452
[63]	Twitter	14-June-2017 to 30-Aug-2017	Twitter Sentiment Score (TSS)	UK	Linear Regression	Stock price prediction	Accuracy: 67.22%
[175]	Twitter	Jan-2017 to Nov-2017	Daily tweet-level embeddings	US	Cross-modal attention based Hybrid Recurrent Neural Network(CH-RNN)	Stock price movement prediction	Accuracy: 59.15%
[161]	Twitter	14-June-2017 to 30-Aug-2017	Tweet embeddings by self-supervised learning	US	Attentive LSTM	Stock price movement prediction	Accuracy (BIGDATA22: 54.81%, ACL18: 58.72%, CIKM18: 55.86%), MCC (BIGDATA22: 0.0952, ACL18: 0.2065, CIKM18: 0.0899)
[149]	Twitter	1-Jan-2014 to 1-Jan-2016	Universal Sentence Encoders (USE)	US	Graph Attention Network	Stock price movement prediction	F1-Score: 0.605, Accuracy: 0.608, MCC: 0.195

Table 5. Stock Market Prediction

price movement, return rate, and volatility using time series models (e.g., ARIMA and GARCH), machine learning, deep learning, and reinforcement learning approaches. [38] has proposed a novel neural tensor network to learn event embeddings, and a deep CNN to model the combined effects of long-term and short-term events for event-driven stock price movement prediction on the S&P 500 index and its individual stocks. Accuracy and MCC are used to measure the model performance and a simulation is performed to evaluate the profitability of the proposed model, which has demonstrated that the deep learning method is effective in event-driven market prediction. [95] has extracted features from news title via CNN and from event tuple (an event can be defined as a tuple (Agent, Predicate, Object) e.g., Apple sues Samsung where A is Apple, P is sue and O is Samsung.) using knowledge graph embedding (i.e., TransE model) and combined with daily trading and technical analysis data. This approach is evaluated using an SVM model as a machine learning method and an LSTM model as a deep learning method to predict stock price movement. The best-performed model is achieved through joint learning of event tuples and text, which has solved the text sparsity problem in feature extraction. [32] has combined technical indicators with sentiment information to predict future prices using regression models and demonstrated that combining technical indicators and sentiment indicators has produced better prediction than using one of them only. The sentiment information is used explicitly where sentiment (i.e., subjective, objective, negative, positive) for each news and comment is obtained using SentiWordNet, and the count of positive, negative, and objective texts for the target company is used as sentiment features. [179] proposed FWD features (Frame, bag-of-Words, part-of-speech DAL score) and SemTree data representations, and adopted SVM to predict stock price movement (polarity and change). [185] proposed StockNet, a deep generative model, which consists of Market Information Encoder, Variational Movement Decoder, and Attentive Temporal Auxiliary, for stock market prediction based on binary movement where a rise in stock price is denoted by one and a fall is by zero. A two-year Twitter data is selected which targets 88 stocks (8 stocks from the Conglomerates industry and the top 10 stocks in terms of capital size in Basic Materials, Consumer Goods, Financial, Healthcare, Industrial Goods and Technology, Utilities, and Services). The proposed model is evaluated by Accuracy and MCC and achieved state-of-the-art performance on a new stock movement prediction dataset, which is also made publicly available[7]. Wu et al. [175] presented novel cross-modal attention based hybrid recurrent neural network (CH-RNN), which is inspired by the DA-RNN model and also released the social text-driven stock prediction dataset built by aggregating stock prices from Yahoo Finance alongside relevant social media discourse, primarily from Twitter. Earnings calls is another data source used for financial forecasting. Earlier research has shown that more information is disclosed in earnings calls [54] than company filings alone and they influence investor sentiment in the short term [14]. The most recent studies using earnings calls include [79], which used 12,285 earnings calls for the S&P 500 companies. It studies the pragmatic correlation with analysts' pre-call judgment and predicts the changes in analysts' post-call forecast as a regression problem and a 3-class classification task by using Ridge Regression and Logistic Regression respectively. The models are evaluated by MSE and R^2 for the regression model and F1-Score and accuracy for the classification model. The results demonstrate that earnings calls are moderately predictive of changes in analysts' target prices. Earning calls has shown predictive power of investment sentiment in the short term, increasing absolute returns [24].

Multiple studies have investigated the integration of company relationships into the prediction of stock market movements. Notably, [22] incorporated company relationships using Graph Convolutional Neural Networks, while [149] proposed a deep attentive learning approach for predicting stock movements based on information from social media texts and company correlations. Recently, [192] introduced Data-axis Transformer with Multi-Level contexts (DTML) for stock movement prediction. DTML leverages temporal and global market context to learn dynamic correlations and outperforms existing approaches, resulting in a substantial annualized return in investment simulations. Another initiative to advance stock market movement prediction involves self-supervised learning from sparsely noisy tweets [161] and a newly created dataset for stock market forecasting is also made publicly available [8]. The proposed SLOT method employs self-supervised learning to generate shared embeddings for stocks and tweets, enabling accurate predictions for less popular stocks. Moreover, it exploits multi-level relationships between stocks inferred from tweets, thereby bolstering its robustness. The work of [101] came up with two important hypotheses: (1) market sentiment does not equal semantic sentiment; (2) the stock price of a target company is also impacted by its related company. Thus, they proposed a multi-source aggregated classifier for stock price movement prediction. They first pre-trained a market-driven sentiment classifier to generate sentiment representations for news. Then, they proposed a classifier to predict the stock price movement for a target company, which aggregates the quantitative indicators and news sentiment of the target company and the news sentiment of its related companies, respectively. This model achieved 67.26% accuracy and 34.52% MCC averaged over 6 blue chip stocks in the Chinese market. The backtesting is also conducted to show the improvements over strong baselines in the Sharpe ratio.

While many studies have combined textual sentiment with fundamental and technical indicators, another type of research is to combine various sources of sentiment with different frequencies for stock market prediction. Principal Component Analysis (PCA) and Kalman Filter (KF) are two popular methods to combine sentiment indicators with

[7] https://github.com/yumoxu/stocknet-dataset
[8] https://github.com/deeptrade-public/slot

different frequencies. [134] applied the KF procedure which is able to aggregate various sources of sentiment with distinct frequencies (e.g., daily, weekly, monthly) and generate a more representative and less noisy latent variable as a newly created daily sentiment indicator. Specifically, the daily sentiment indicator from Twitter and weekly and monthly sentiment indicators from surveys (e.g., AAII, II, UMSC, Sentix) are extracted using the KF procedure. Multiple Regression, Neural Networks, SVM, RF, and Ensemble Averaging are used to perform predictions. Also, [134] highlighted that the common issues with evaluation in stock market prediction, which include that either out-of-sample data is not used to evaluate the model performance or the test data size is too small, are addressed in the study. Meanwhile, it is limited to utilize statistical tests to evaluate the accuracy of prediction and this is resolved in [134] by applying Diebold-Mariano test in addition to the evaluation criteria of Normalized Mean Absolute Error (NMAE). Unlike earlier studies which used textual features to predict market movements, [200] proposed event-driven trading strategies to predict stock market movements by detecting corporate events, which is considered as the driving force of market movements, from news articles. The bi-level event detection model is trained with masked-language model (MLM) loss. The authors employed two trading strategies to perform experimentation on the EDT dataset[9]. The first strategy, Trade at End Strategy keeps the transactions already started on hold for k days and closes on the last day. This strategy gave an estimated 1.74% average return and exceeded the return with 844% in 1-day trading. The second strategy, Trade at Best Strategy completes the transaction within k trading days, at the best price. This method resulted in an estimate of 9.11% average returns, that exceeded all the sentiment-based models.

The expected return could be investigated from the time series or cross-section perspective. A time series perspective is how average returns change over time while a cross-section perspective is how average returns change across different stocks or portfolios. While the majority of the studies investigate the effects of investor sentiment from a time series perspective, [6] pioneered and demonstrated investor sentiment, broadly defined, affects the cross-section of stock returns significantly via a set of empirical results. When estimation of sentiment is high, stocks that are unattractive to arbitrageurs and meanwhile attractive to optimists and speculators, tend to generate relatively low subsequent returns [6]. This includes distressed stocks, extreme growth stocks, high volatility stocks, non-dividend-paying stocks, small stocks, unprofitable stocks, and younger stocks.

5.2.3 Financial Risk Prediction. One essential indicator of instability and risk is financial volatility, which is a popular metric used in financial forecasting. Volatility is commonly defined as the standard deviation of a stock's returns over a pre-defined period of time. The volatility of return is defined as follows [172]:

$$v_{[s,s+\tau]} = \sqrt{\frac{\sum_{t=s}^{s+\tau}(r_t - \bar{r})^2}{\tau}} \tag{7}$$

where r_t is the return and $\bar{r}$ is mean of return.

$$r_t = P_t / P_{t-1} - 1 \tag{8}$$

where P_t is the adjusted close price.

There are studies in risk prediction using various data sources such as financial reports [83, 130, 166, 172], financial news [130, 166], message boards [129] and earning calls [174]. [172] investigates the significance of sentiment words on financial risk, by formulating a regression task to predict future real-value risk given sentiment and a ranking task to rank the risk levels, using financial reports from 1996 to 2006 i.e., Section 7: management's discussion and analysis of

[9]https://github.com/Zhihan1996/TradeTheEvent/tree/main/data

Ref.	Data Source	Period	Text Representation	Markets	Method	Task	Performance
[134]	Twitter	22-Dec-2012 to 29-Oct-2015	Sentiment indicators based on SMSL lexicon which combines with AAII, II, UMSC, Sentix	US	MR, NN, SVM, RF, Ensemble	Return, trading volume and volatility	Daily return (Lowest NMAE: NDQ, 7.58), Daily trading volume (Lowest NMAE: DJIA, 5.84), Daily volatility (Lowest NMAE: DJIA, 2.79)
[145]	10-K reports of companies from the U.S. SEC	2006 to 2015	Feature vector generated by the weights of lexicons from LM and the word weighting scheme includes TC, TF, TF-IDF and BM25	US	GARCH, SVM	Volatility	MSE: 0.111, R^2: 0.527
[174]	Quarterly earning calls from the US stock market	2006 to 2013	Uni-grams, bi-grams, named entities, part-of-speech tags and probabilistic frame-semantic features	US	Linear Regression, Linear SVM, Gaussian SVM, Gaussian Copula Models	Volatility	Spearman (Pre-2009: 0.425, 2009: 0.422, Post-2009: 0.375), Kendall: (Pre-2009: 0.315, 2009: 0.310, Post-2009: 0.282)
[184]	StockTwits	14-Aug-2017 to 22-Aug-2018	Sentiment polarity score computed by augmented sentic computing	US	SAVING, GARCH, EGARCH, TARCH, GJR, GP-vol, VRNN, NSVM, LSTM, s+LSTM	Volatility	Negative Log-Likelihood (NLL): -3.0642
[172]	10-K Form, an annual report required by the Securities and Exchange Commission (SEC)	1996 to 2006	LM lexicon is chosen with six types of sentiment words i.e., positive, negative, uncertainty, legal context, strong and weak confidence. The BoW model is adopted and TF-IDF and LOG1P are selected as word features to represent the 10-K reports	US	SVR	Volatility	Regression (MSE: 0.14894), Ranking (Kendall's Tau: 0.60458, Spearman's Rho: 0.63403)
[34]	RavenPack, Twitter, Thomson Reuters	2019	Latent Dirichlet Allocation (LDA) model to extract feature vectors	UK	Logistic Regression	Volatility	Accuracy (Headlines: 65%, Tweets: 65%, Stories: 67%), F1-Score (Headlines: 64%, Tweets: 64%, Stories: 63%)

Table 6. Financial Risk Prediction.

financial conditions and results of operations (MD&A), which contains the most important forward-looking statements about the companies, in the 10-K Form, an annual report required by U.S. SEC, provides a comprehensive overview of the company's business and financial conditions, and includes audited financial statements. The LM lexicon is chosen with six types of sentiment words (i.e., positive, negative, uncertainty, legal context, strong and weak confidence). The BoW model is adopted and TF-IDF and LOG1P are selected as word features to represent the 10-K reports. As compared to original texts, the LM lexicon has reduced the word dimension significantly from hundreds of thousands to less than two thousand. The regression task is evaluated by MSE and the ranking task is by Spearman's Rho and Kendall's Tau. The experimental results demonstrate that the models trained on sentiment words generate relatively better performance than models trained on origin texts, which attests to the importance of the financial sentiment lexicons on financial risk prediction. Additionally, the trained models also suggest that there is a strong correlation between financial sentiment words and financial risk. [174] examined the correlation between earnings calls and financial risk, which is defined as the volatility of stock prices in the next week, by using uni-grams, bi-grams, part-of-speech tags, named entities and probabilistic frame-semantic features to build Gaussian copula model which is evaluated by the Spearman's correlation and Kendall's tau. There are three datasets of transcribed quarterly earnings calls from the U.S. stock market during the Great Recession period, where the pre-2009 dataset includes earnings calls from 2006 to 2008 when the economic downturn began, the 2009 dataset consists of earnings calls from 2009 when the financial crisis spread globally and the post-2009 dataset contains earnings calls from 2010 to 2013 when the global economy recovers. It has demonstrated that the quarterly earnings calls can be used to predict the volatility of stocks in the limited future. One of the most recent studies done by [145], which performed volatility forecasting using textual features from financial disclosures (i.e., TC, TF, TF-IDF, BM25) with dimension reduction technique of PCA, as well as market features (e.g., current volatility and sector) from factual market information using SVM with Radial Basis Function (RBF) kernel. The proposed approach to sentiment analysis significantly outperforms current state-of-the-art methods and shows the information in the 10-K reports is valuable for volatility prediction. The model performance is evaluated by MSE and R^2 metrics. [184] proposed the sentiment-aware volatility forecasting (SAVING) model, which combines symbolic and sub-symbolic AI approaches by integrating grounded knowledge into neural networks, which incorporates market sentiment to predict stock return fluctuation. The proposed model outperforms not only pure statistical models such as GARCH and its variants, which are commonly used econometric time series models for volatility prediction, and Gaussian-process volatility model, but also the latest autoregressive deep neural network architectures e.g., neural stochastic volatility model and variational recurrent neural network. [34] proposed a market volatility classifier based on Latent Dirichlet Allocation (LDA) topic modeling. The paper suggests a strong negative correlation between positive tweets and next-day volatility by observing the relationship between financial news, tweets, and FTSE100. The study also indicated the dependence of the model's accuracy on a number of topics chosen.

5.2.4 Portfolio Management. One type of research that was less focused earlier [104] but emerged recently is to exploit the opinions posted by investors to invest in stock markets properly by optimizing portfolios. Existing approaches largely treat market prediction as a classification, regression, or ranking task but are not optimized for making profitable investment decisions. However, the decision-making and trading strategies should be incorporated and also improve practical applicability [104]. To address this challenge, researchers started to investigate the adoption of financial sentiment for portfolio management. [86] applied a semi-supervised learning method to stock microblogs and proposed Follow-the-Loser online portfolio selection strategy. The proposed approach includes a user model, an emotion classifier using MLP, and a portfolio selection strategy. The concept of online portfolio selection is to apply online learning in the machine learning literature to the portfolio selection problem, which aims to maximize the cumulative returns over sequential multiple periods of time [86]. This is different from offline or batch portfolio selection (e.g., Markowtitz's Mean-Variance Theory), which balances the expected return and risk focusing on a single period of time. [167] firstly predicts the quality of opinions followed by investment recommendations. The quality estimation for investment opinions adopted features associated with the author, content, and stocks in discussion, and opinions with the highest predicted qualities are selected as high-quality opinions. When generating portfolios, a score is generated for each stock by aggregating the sentiment about stocks in the opinions weighted by the predicted qualities of opinions. Experiments are conducted on a real-world dataset and demonstrate the effectiveness of the proposed strategy in recommending high-quality opinions and profitable portfolios. [187] applied the Gaussian inverse reinforcement learning method in which the market dynamics is modeled as a Markov decision process and investor sentiment is regarded as a series of actions taken at different market states. The S&P500 index return, which is measured at 15 minutes internal, is used as the market return and the sentiment from Thomson Reuters news is used as the proxy of investor sentiment toward the general U.S. market. It is often that markets do not react to noisy signals, which largely exist in investor sentiment signals. The investor sentiment reward-based trading system is designed to filter out noisy signals and extract only effective signals that generate either negative or positive market responses. The annualized performance including the mean return, volatility, Sharpe ratio, and Sterling ratio are adopted to measure the model performance. [150] has proposed Policy for Return Optimization using FInancial news and online Text (PROFIT), which formulates the stock prediction into a reinforcement learning problem and leverages financial news and tweets to model stock-affecting signals and optimize trading decisions to increase profitability. The trading performance is evaluated by Sharpe ratio, Sortino ratio, cumulative return, and maximum drawdown and compared with baselines spanning different formulations including regression, classification, and ranking. The proposed PROFIT system has outperformed benchmark systems significantly. [190] has proposed a novel and generic state-augmented RL framework called State Augmented Reinforcement Learning (SARL), which can integrate heterogeneous data sources into standard RL training pipelines for learning portfolio management strategies. The portfolio performance is measured by portfolio value and Sharpe ratio. The Bitcoin [74] and HighTech [37] datasets are used for the experiments which have achieved significantly better portfolio value and Sharpe ratio than baseline models including equal weight and Deep Portfolio Management [74]. [43] proposed stock embedding, a vector representation of stocks in a financial market, which uses a neural network framework consisting of text feature distiller and price movement classifier to acquire such vectors from stock price history and news articles. The stock embedding could be applied to other financial tasks such as portfolio optimization other than price prediction. The proposed method has outperformed baseline models in both price movement classification and portfolio optimization tasks. [104] has included public mood from online news and social media and selected RF, Multi-Layer Perceptron, and LSTM to take a historical series of lagged data and public mood and generate the optimal portfolio allocation automatically. The proposed methodology outperforms the equal-weighted portfolio allocation strategy consistently and shows that it is always beneficial to the model performance to include the financial sentiment. [85] proposed a sentiment-aware deep reinforcement model for portfolio allocation on a daily basis. The sentiment polarity is added using Valence Aware Dictionary and sEntiment Reasoner (VADER) [71]. The trading performance is evaluated by Sharpe ratio and annualised return which shows it is more robust than benchmarks across Sharpe ratio and annualised returns. [20] presented a sentiment-based RF model to generate a portfolio of Chinese stocks that can achieve higher returns. The proposed method shows the importance of choosing suitable methods for stock characteristics and stock selection methods in a highly volatile market. This method generated higher returns than the Shanghai Stock Index. [151] proposed a modified LSTM, namely time-aware LSTM (t-LSTM) that learns the time-aware representations to produce a ranked list of predicted stock return ratios based on expected profit. The model captures the relevant market trends by using hierarchy-based temporal attention for ranking stocks. In intra-day situations, the model outperformed the SOTA methods by over 8%, and by 10% in risk-adjusted returns.

5.2.5 FOREX Market Prediction. There has been a significant focus on forecasting stock market trends, while the foreign exchange market has received comparatively less attention in predictive efforts. Earlier studies had investigated the connection between macroeconomic fundamentals and exchange rates in the short run using Flexible Fourier Form regression method using absolute returns as a measure of volatility [90]. The impact of macroeconomic announcements, which is collected from Bloomberg World Economic Calendar (e.g., GDP, interest rates, and consumer confidence indexes), on USD/EUR exchange rate volatility is estimated. The observations are divided into 5-minute intervals, totaling 288 in 24 hours, from 28-Oct-2003 to 20-Jan-2004 and the results suggest that macroeconomic news significantly increases the volatility of exchange rates immediately after the announcement. It also shows that the degree of significance varies by news category and country. Also, [46] concludes that macroeconomic news arrivals affect currency markets over time. The average news effects correspond to the direct channel for price impact which is absorbed immediately, but total news effects are not reflected quickly. [75] presented Forex-foreteller (FF), which utilizes news articles to forecast the movement of foreign currency markets. FF combines language models, topic clustering, and sentiment analysis to identify relevant news articles which are used together with historical stock index and currency exchange values to build a linear regression model to perform forecasting and generate warning messages. [128] proposed a novel approach that adopted TF-IDF weighted features scaled by sentiment sum score using SentiWordNet to predict intra-day directional-movements of currency-pair using SVM and demonstrated the existence of a promising predictive relationship between foreign exchange market and financial news. [153] proposed a FOREX market prediction system that performs sentiment analysis of news headlines by exploiting word sense disambiguation and predicts the directional movement of EUR/USD exchange rate and improved prediction accuracy. [154] presented a novel approach that includes news story events in the economy calendar to predict intra-day directional movement of currency pairs using SVM, RF, and XGBoost algorithms and achieved promising results. [180] investigates the efficacy of high-frequency news

Ref.	Data Source	Period	Text Representation	Markets	Method	Task	Performance
[86]	Yahoo! JAPAN textream	1-Jan-2013 to 1-Jan-2016	Emotion polarity	Japan	MLP, reliability and cumulative reliability score for each individual using sentiment polarity, Follow-the-Winner and Follow-the-Loser strategy	Portfolio selection	Cumulative return: 1.151
[167]	StockTwits	2014	Sentiment polarity provided in tweets	US	Linear regression for opinion quality score	Portfolio selection	Cumulative return: 30%
[187]	Thomson Reuters	2-Jan-2008 to 31-Dec-2015	News sentiment from Thomson Reuters News	US	Gaussian inverse reinforcement learning	Portfolio selection	Return: 17.39%, Sharpe ratio: 0.85, Sterling Ratio: 0.76 (about 3 times all the market benchmarks)
[150]	Twitter (English), financial news headlines (Chinese) aggregated by Wind.com.cn from major financial website like Sina and Hexun for China and Hong Kong	US: Jan-2014 to Dec-2015, China and Hong Kong: Jan-2015 to Dec-2015	BERT Embeddings	US, China, Hong Kong	Deep Deterministic Policy Gradient (DDPG) framework	Portfolio selection	Sharpe ratio (US: 1.03, China and Hong Kong: 1.29), Sortino Ratio (US: 1.87, China and Hong Kong: 1.99), Cumulative Return (US: 29.64, China and Kong Kong: 40.88), Maximum Drawdown(US: 5.01, China and Hong Kong: 6.78)
[190]	Bitcoin of 10 different cryptocurrencies, HighTech daily closing asset price and financial news	Bitcoin: 30-Jun-2015 to 30-Jun-2017, HighTech: 20-Oct-2006 to 20-Nov-2013	Auto Phrase, Glove, Word2Vec and FastText Embeddings	US	State augmented reinforcement learning	Portfolio selection	Portfolio value is improved by 140.9% and 15.7% for Bitcoin and HighTech respectively as compared to the state-of-the-art RL algorithm DPM, Sharpe ratio (Bitcoin: 14.78, HighTech: 7.73)
[43]	Wall Street Journal (WSJ), Reuters & Bloomberg (R&B)	WSJ: Jan-2000 to Dec-2015, R&B: Oct-2006 to Nov-2013	Stock Embeddings	US	Deep learning	Price movement prediction, portfolio selection	Price movement prediction (WSJ: Accuracy is 0.601 for 16 years, 0.550 for 3 years and 0.525 for 1 year, R&B: Accuracy is 0.688 for 7 years, 0.675 for 3 years and 0.512 for 1 year), Portfolio selection (An average of the realised annual gains increased to 17.2% and 35.5% for WSJ and R&B respectively)
[104]	Financial and sentiment data for 15 different stocks	24-Jan-2012 to 2-Jun-2017	The number of positive, negative, and neutral comments, a measure of change in positive and negative comments compared with the previous days (change) and a measure of positive and neutral versus negative reviews (sentiment score) for each day and each stock	US	EW, LSTM, MLP, RF	Portfolio selection	Wealth of the portfolio (LSTM + Sentiment for five portfolios: 2.23, 2.72, 2.30, 2.81 and 1.65)
[151]	US: Twitter, China and Hong Kong: Wind.com.cn	US: Jan-2014 to Dec-2015, China and Hong Kong: Jan-2015 to Dec-2015	Time-aware representations of news and tweets	US, China, Hong Kong	Time-aware LSTM (t-LSTM)	Portfolio selection	US S&P 500: Return ratio: 1.34, Sharpe ratio: 0.96, China and Hong Kong: Return ratio: 1.44, Sharpe ratio:1.19
[85]	Wharton Research Data Services	1-Jan-2001 to 2-Oct-2018	Sentiment polarity using VADER	US	Deep reinforcement learning	Portfolio selection	Sharpe ratio: 2.07, Annualized return: 22%
[20]	RESSET database	1-Jan-2016 to 31-Jul-2018	Lexical Model to give confidence index	China	DT, LR, SVM, RF	Portfolio selection	Accuracy: 79.6%, Holding Period yield: 5.41
[59]	Bloomberg	Jan-1990 to Dec-2018	3D standardized features	US	RF, LSTM	Portfolio selection	Daily return (LSTM: 0.64%, RF: 0.54%)
[162]	Sina Guba and Eastmoney Guba, RESSET	2008 to 2018	Gubalex - stock sentiment lexicon	China	Probabilistic Neural Network (PNN)	Price movement prediction, portfolio selection	Accuracy: 86.3%

Table 7. Portfolio Management.

Ref.	Data Source	Period	Text Representation	Markets	Method	Task	Performance
[75]	Bloomberg	Apr-2010 to Mar-2013	LM lexicon to identify relevant keywords. AFINN dictionary to measure general emotions	FOREX	Linear Regression	Predict the change in currency value and generate warning messages	Precision (Argentina: 0.18, Brazil: 0.28, Chile: 0.33, Colombia: 0.25), Recall (Argentina: 0.60, Brazil: 0.63, Chile: 1, Colombia: 1)
[128]	MarketWatch.com, Google RSS reader API	2008 to 2011	TF-IDF weighted features scaled by sentiment sum score using SentiWordNet	FOREX	SVM	FOREX price movement prediction	Accuracy: 0.8333, Precision (Pos: 0.6667, Neg: 0.8889), Recall (Pos: 0.6667, Neg: 0.8889)
[153]	MarketWatch.com	2008 to 2012	WSD-Sentiment + TF-IDF, SentiWordNet	FOREX, EUR/USD	SVM	FOREX price movement prediction	Accuracy: 0.5926, Precision: 0.5710, Recall: 0.5735
[154]	Forexfactory.com, First Word FX News	1-Oct-2015 to 31-Oct-2017	TF, TF-IDF, LM, and FX dictionaries	FOREX, EUR/USD, GBP/USD, USD/CHF, USD/JPY	SVM, RF, XGB	FOREX price movement prediction	Accuracy (EUR/USD: 0.638, GBP/USD: 0.663, USD/CHF: 0.631, and USD/JPY: 0.641)
[180]	Dow Jones Newswire	1-Jan-2016 to 30-Jun-2018	Sentiment generated from FinBERT	FOREX, EUR/USD	SVM	FOREX price movement prediction	Accuracy: 0.503, F1-Score: 0.538

Table 8. FOREX Market Prediction.

sentiment, which is represented by a 4-dimensional time series extracted by a FinBERT-based model, for FOREX market prediction without including other semantic features. Experimental results show that their model outperforms benchmark approaches for sentiment analysis and conclude that news sentiment alone may have predictive power, though it is relatively weak, for FOREX price movements.

Ref.	Data Source	Period	Text Representation	Markets	Method	Task	Performance
[80]	Google Trends, Twitter	Dec-2018 to May-2019	Google trend rate, positive sentiment rate, negative sentiment rate	Cryptocurrency	Hidden Markov Model	Cryptocurrency market movement prediction	Accuracy: 54%, AUC: 52%
[70]	Sina-Weibo	2021	Cryptocurrency sentiment dictionary	Cryptocurrency	Autoregression, LSTM	Cryptocurrency market movement prediction	Precision: 87%, Recall: 92.5%
[148]	Twitter	28-May-2021 to 25-Sep-2021	Sentiment score from fine-tuned FinBERT	Cryptocurrency	LSTM	Financial sentiment, BTC volume correlation	Sentiment: Accuracy: 0.8352, F1-Score: 0.8515, BTC Volume: Pearson's R: 0.1584
[132]	Twitter	01-Sep-2021 to 01-Nov-2021	Sentiment score from VADER	Cryptocurrency	Granger causality test, Vector Autoregression	Cryptocurrency price	MAPE: 0.0038
[87]	Twitter	4-Jun-2018 to 4-Aug-2018	VADER, LM lexicon and cryptocurrency lexicon	Cryptocurrency	Granger causality test	Cryptocurrency price return	Significant predictive power ($p < 0.05$) on Bitcoin, Bitcoin Cash and Litecoin
[155]	Twitter	4-Sep-2014 to 31-Aug-2018	Volume of tweets	Cryptocurrency	Granger causality test, Vector Autoregression	Volume, return and realized volatility	Number of previous day tweets are significant drivers of Bitcoin volume and realized volatility, but not returns.
[55]	Twitter	24-Jun-2019 to 12-Aug-2019	LM lexicon, stock market opinion lexicon	Cryptocurrency	Regression	Return and volatility	Bitcoin prices are partially predicted by momentum on social media sentiment

Table 9. Cryptocurrency Market Prediction

5.2.6 Cryptocurrency Market Prediction. Cryptocurrencies have experienced remarkable value growth, surpassing the most substantial historical bubbles of the past three centuries [2]. More studies have been conducted to understand the dynamics of cryptocurrency market behavior over time since recently, with a primary focus on on causality and correlation tests [87, 132, 155]. The research demonstrated that investor sentiment holds significant nonlinear predictive power for the returns of major cryptocurrencies [125]. [80] proposed a hidden Markov Model to construct a transition matrix from Markov Chains on positive and negative sentiment, trading volume, and closing price to predict the upward or downward market trend. This study also observed that the market tends to respond more to positive sentiments in a bearish market and responds more to negative sentiments in a bullish market. The study however only used Bitcoin for the study and other cryptocurrencies were not included in the dataset. [70] worked on analyzing the sentiment of Chinese social media and its effects on the cryptocurrency market. The study proposed LSTM-based RNN model to predict the cryptocurrency price and it has achieved better precision and recall than baseline auto regressive models.

5.2.7 Explainable FSA Applications. The notion of explainability holds paramount importance in FSA applications where decisions can have significant consequences [16]. [191] has classified the explanation procedure into textual, visual, by example, simplification and feature relevance. The majority of studies in FSA applications emphasize explainability through visual, feature relevance, and simplification techniques. Specifically, in visual explanation, [33] employs knowledge graphs to establish visual connections among event entities extracted from stock news articles. This method provides users with a graphical representation of the relationship between features and their corresponding predictions. Another approach, as presented in [89], involves conducting deconvolution on the penultimate layer preceding the output to generate a visual attentive map. This approach, named CLEAR (Class Enhanced Attentive Response), produces a graphical representation indicating the timeframe during which the stock-picking agent exhibits the highest degree of attention, along with a separate plot corresponding to the sentiment class of the stock. Regarding feature relevance explainability, [17] adopts various configurations of a permutation importance technique to prune less significant technical indicators. Subsequently, decision tree techniques are implemented for stock market forecasting. The proposed method was compared with LIME and exhibited greater reliability. In a similar vein, [135] employs aspect-based sentiment analysis to examine the correlation between stock price movement and the most pertinent aspects identified in tweets. The polarity of each aspect is derived using a SenticNet-based graph convolutional network (GCN) [91]. This approach mirrors the feature relevance technique, with its focus on discerning top-contributing aspects along with their associated polarity values. Notably, this work emphasizes the interplay between financial variables rather than making direct financial predictions. This information serves as a foundation for further analysis, leveraging the relationship between the price movement of individual stocks and the sentiment associated with popular terms detected in tweets. In the simplification procedure, [8] integrates sentiment analysis of text with technical analysis of historical stock prices to train a random forest stock forecasting model, which is further explained through LIME. Furthermore, [60] implements LIME in conjunction with LSTM-CNN, accurately identifying pivotal words that align with the target sentiment. In reference to [193], text explanations are generated utilizing the advanced natural language generation transformer decoder, GPT-2 [143], with the added consideration of incorporating specific keywords within the generated text. The presented methodology, known as soft-constrained dynamic beam allocation (SC-DBA), involves the extraction of keywords associated with different tiers of anticipated market volatility. This extraction process is facilitated by a distinct network designed for analyzing harvested news titles. The evaluation of the quantitative performance is based on assessing both the fluency and practical relevance of the generated explanation.

5.2.8 Summary of FSA Applications. We summarize different FSA applications in Tables 5, 6, 7, 8 and 9 by types of FSA applications. The most frequently used data sources in FSA applications are microblogs and news, followed by annual reports filed by companies. The financial metrics predicted include stock index, stock price, stock price movement, stock price change, return rate, volatility, stock market crash risk, and FOREX rate change. As for methods for predictive modeling, similar to many other NLP tasks, deep learning has received more attention than traditional machine learning (e.g., SVM) and time series modeling (e.g., GARCH) in recent years with more promising performance contributing to its capability to perform high-level abstraction from complex data, though is less explainable than feature-based machine learning. The variants of Recurrent Neural Networks (e.g., Attentive LSTM, DA-RNN), and advancements in Graph Convolution Networks (e.g., Graph Attention Networks) have demonstrated remarkable state-of-the-art performance. Meanwhile, Reinforcement Learning has started to apply to portfolio optimization tasks and we have observed more work adopting it for intelligent asset allocation recently. In terms of markets, the US stock market is the most studied market and the tasks are moving beyond market movement prediction to risk prediction and portfolio management. Meanwhile, there has been a notable increase in research aimed at forecasting both the FOREX and cryptocurrency markets using textual data. Particularly, the exploration of the cryptocurrency market is still in its infancy, primarily centered on causality and correlation tests. Early investigations reveal that the cryptocurrency market exhibits reduced predictability when compared to the stock market, a trait attributed to the heightened dynamism inherent in cryptocurrency markets. As for model performance, practical traders agree that any results above 50% accuracy are value-added to their day-to-day trading [128] and literature reviewed in our survey has demonstrated that effectiveness. The evaluation metrics adopted in FSA techniques also apply to FSA applications depending on whether it is a regression (e.g., MSE) or classification task (e.g., F1-Score). Additionally, the trading simulation results are consistently adopted as a measurement of performance for portfolio management systems and popular metrics include accumulated return, average percentage gain per transaction, and Sharpe ratio.

6 MAIN FINDINGS

6.1 FSA Scopes and Sentiment Types

To address our first research question on what is the scope of FSA in today's context, and what is the relationship among FSA, investor sentiment, and market sentiment, we surveyed the recent publications. FSA studies have evolved with the increase of financial textual data over the years particularly in public news and social media. In today's context, the scope of FSA research has been extended to the field of study that not only analyzes people's sentiments from financial texts but also investigates the predictability of financial textual sentiment on the financial market. The objectives and scopes of FSA techniques are fundamentally different from FSA applications but there is also an interactive relationship between them. While FSA techniques aim to study the techniques that can improve the performance of various FSA tasks (e.g., targeted aspect-based FSA) with human-driven annotation, the main objective of FSA applications is to exploit sentiment for financial applications, such as causality and correlation testing and financial forecasting with market-driven annotation computed from real-world market data. Here, financial sentiment serves as a proxy of investor sentiment, which affects the market dynamics. There is a complex connection exists in FSA, investor sentiment, and market sentiment. Market sentiment is the aggregated effect of investor sentiment and a reflection of investor sentiment in their investment behaviors. Investor sentiment can be partially measured by financial textual sentiment, sentiment surveys, and market sentiment. This has established the theoretical foundation of using multiple data sources such as financial texts, sentiment survey, and market data to perform financial forecasting.

6.2 Trends in FSA Techniques

Our second research question is: What trends are emerging from the latest tasks, benchmark datasets, and methods in the newest FSA technique studies? We observed that the more benchmark datasets (e.g., SEntFiN in 2022) are annotated with substantial entries for fine-grained FSA tasks. Meanwhile, the creation of financial lexicons is extending from conventional word-level to concept-level and direction-dependent expressions. As for methods, the feature engineering process has factored in domain-specific features such as numbers in finance. The deep learning and hybrid approach which ensembles lexicon, machine learning, and deep learning methods has shown promising model performance. Moreover, the pre-trained language presentation models such as BERT are able to capture general language representation from large-scale corpora but lack domain-specific knowledge [93]. To improve the domain application of pre-trained language models, researchers have attempted to train domain-specific pre-trained language models such as FinBERT [4, 96] but it requires large domain-specific corpus (e.g., news, corporate reports, earnings call transcripts and analyst reports) and substantial computing resources. This has pushed the boundary of research in FSA techniques and improved the model performance significantly. Finally, autoregressive decoder architectures such as GPT have shown potential in FSA. Bloomberg has unveiled BloombergGPT, a Large Language Model specialized in finance, which surpasses similar sized open models in financial NLP tasks, including FSA.

6.3 Trends in FSA Applications

Our third research question is: What data sources, tasks, methods, and financial markets can be used in FSA application-focused domains? FSA applications have gained increasing attention than FSA techniques in recent years largely attributed to the increase of various textual data sources and technologies. The main application of FSA is in causality and correlation testing and predictive modeling in financial markets, or named natural language-based financial forecasting which is brought up by [183]. While causality and correlation testing were focused on by earlier studies, the adoption of financial textual data using NLP techniques to extract sentiment in financial forecasting is an emerging research field. We have identified six financial forecasting tasks including stock market movement prediction, stock market risk prediction, portfolio management, FOREX market prediction, and cryptocurrency market prediction. Financial sentiment, which serves as a proxy for investor sentiment or non-informational trading, has demonstrated its effectiveness in financial forecasting, especially in the stock market. Sentiment from the three main sources (i.e., corporation-released, media-expressed, and internet-posted texts) has been found to have important effects on market movement. Particularly, negative sentiment has proved to be the strongest influence. Specifically, media-expressed texts (e.g., financial news) are the most commonly used data source followed by internet-posted texts (e.g., Twitter and StockTwits) and corporation-released texts (e.g., annual reports). The media-expressed texts have been applied to all six financial forecasting tasks while corporation-released texts are more used for stock market risk prediction. A more recent trend is to combine different sources of sentiment for financial forecasting. Currently, the application of FSA in financial markets focuses on the stock market, FOREX market, and cryptocurrency market, where the former market is the most commonly studied and the latter two markets are emerging. Many other financial markets such as commodities, money, derivatives, future, and insurance are not explored yet. It is also observed that FSA applications are shifting from market predictions to trading strategies such as portfolio management to improve practical applicability. In terms of methods, deep learning and reinforcement learning with financial market prediction is regarded as one of the most charming topics. In every instance, the importance of explainability is paramount in FSA applications due to the profound consequences that decisions can entail.

6.4 FSA and Financial Forecasting

Our fourth research question is: How financial sentiment is involved in financial forecasting and the focus is on FSA techniques or applications? We found that financial sentiment can be represented in an explicit or implicit manner for financial forecasting, with the latter (implicit) being more frequently adopted. The explicit use of financial sentiment is to derive sentiment polarity, intensity, or sentiment lexicons and use it for predictive modeling while the implicit use is to generate the features and text representation such as word or event embeddings. The approaches to generate features and word representations in FSA applications are similar to word representation techniques, summarized under FSA techniques in Section 4.4.6. The implicit use of sentiment is more popular than explicit use and also can achieve better performance, largely attributed to the fact that this way carries more complete information than direct sentiment extraction. Notably, the FSA model demonstrates potential in the filtration of textual data, allowing for the retention of more pertinent information crucial for financial forecasting. Empirical studies also indicate that embeddings generated from models trained on market sentiment exhibit superior performance in financial forecasting compared to those trained on semantic sentiment. A thorough investigation into the interplay between financial sentiment models and their influence on the performance of financial applications is a crucial area for further exploration. Meanwhile, the algorithms and evaluation metrics adopted in FSA techniques can also be applied to financial forecasting except for complex application tasks such as portfolio management which requires reinforcement learning and trading simulations in many studies. Our view is that the application of financial sentiment in predictive modeling also can be regarded as an FSA task that uses market-driven annotation, as compared to the human-driven annotation in the research in FSA techniques.

7 CHALLENGES AND FUTURE DIRECTIONS

7.1 FSA Techniques

Firstly, as elaborated by [106], [4] and [96], there is a lack of high-quality and large-scale open-source finance domain-specific annotations for FSA. The main challenge is that the creation of FSA benchmark datasets is usually expensive and requires expert knowledge [96, 106]. The research in fine-grained FSA has gained more attention after the release of the SemEval 2017 Task 5 and FiQA Task 1 datasets. In terms of data annotation, the nested target annotation schema, which is proposed by [100] and goes beyond the traditional target, aspect, and sentiment annotation, could open a new space for FSA. Secondly, lexical resources are limited and scattered. Since finance is a highly professional domain, general-purpose sentiment lexicons usually fail to take into account the domain-specific connotations and the heavy reference to prior knowledge. For example, words like “liability” and “debt” are considered negative in general-purpose sentiment analysis, but are frequent and have neutral meanings in the financial context. This makes it difficult to generalize the sentiment classifiers and underlines the need for finance domain-specific sentiment analysis [97]. Further, sentiment intensity scores are more consequential and nuanced for FSA compared to other domains. Whereas most of the current FSA studies still adopt a polarity detection fashion (i.e., classification to positive or negative). Further, it is important to improve the capability of generalization for BERT-based models. The successful application of FinBERT in FSA tasks is largely dependent on the corpora used to pre-train the language model. Presently, financial news, annual reports, earning call transcripts, and analyst reports are adopted by variant FinBERT but microblog corpora have not been explored. Next, the incorporation of knowledge and adoption of GCN are demonstrated to be useful in sentiment analysis but few of the earlier research have attempted to incorporate knowledge in FSA, which could be a promising direction for research in FSA techniques. Finally, incorporating linguistic intuitions in deep learning models [56, 65, 115] is another direction to improve the rationality of model design, because many deep learning-based methods focused on algorithm novelties and ignored linguistic intuitions. The future study could focus on improving the six areas that cause FSA fail, i.e., irrealis mood (conditional mood, subjunctive mood, imperative mood), rhetoric (negative assertion, personification, sarcasm), dependent opinion, unspecified aspects, unrecognized words (entity, microtext, jargons), and external reference. Future research can explore the incorporation of multimodal data into FSA, such as text, images, audio, and video, to gain a more comprehensive understanding of financial sentiment. This could involve analyzing earnings call transcripts, financial news articles, social media posts, and even multimedia content. Also, the contextual FSA which is to develop methods that can understand and analyze the context in which financial sentiments are expressed will be crucial. This could involve sentiment disambiguation, where sentiment is understood in relation to specific events, companies, or market conditions.

7.2 FSA Applications

One challenge with FSA applications is the lack of publicly released textual data sources which are in a time series with a substantial amount of financial texts and representative periods to model the relationships between investor sentiment and financial markets. Meanwhile, the attention of the FSA application is moving beyond the stock market to other financial markets and shifting from stock market prediction to trading strategies such as portfolio management to improve practical applicability. The adoption of reinforcement learning could open a new avenue for portfolio management but it still remains relatively less explored. Further, earlier studies focus on news sentiment, the emotions derived from news could also influence investors' behaviors in financial markets [44]. Lastly, the financial market is driven by different types of news which include macroeconomic factors, geopolitics, and company-specific factors [170], which means the sentiment can be derived from different perspectives such as macroeconomics, microstructure factors, event-oriented, and company-specific [99]. From this perspective, the aspect-based FSA could be adopted to extract aspect-level features to forecast future market performance [186] which could also improve the model interpretability and explainability. The notion of explainability holds paramount importance in FSA applications where decisions can entail significant consequences. Methods for generating human-understandable explanations for model predictions will become a focus. With the rise of cryptocurrencies and emerging markets, there will likely be a growing interest in sentiment analysis tailored to these specific assets and markets. FSA can continue to play a crucial role in assessing risk and managing portfolios. One potential area of future research is to develop models that can provide more accurate risk assessments and aid in making more informed investment decisions. Lastly, future research could delve deeper into the intersection of behavioral economics and FSA, exploring the psychological factors that influence sentiment and decision-making in financial markets.

8 CONCLUSION

This survey conducted a comprehensive review of FSA research from both technique and application perspectives including their interactive relationship. The scope of FSA research has been redefined and the relationship among FSA, investor sentiment, and market sentiment. Specifically, through the review of FSA techniques, we have included the latest benchmark datasets, and the methods, which include lexicon, machine learning, deep learning and hybrid approaches, pre-trained language models, and word representation techniques used for the FSA study. As for FSA applications, we summarized that the main FSA application in financial markets is hypothesis testing and predictive modeling. Predictive modeling has received more attention in recent years, particularly in the stock market and FOREX market. In terms of tasks in predictive modeling, the application in the stock market has moved beyond traditional market movement prediction to market risk prediction and portfolio management. Market prediction typically is treated as a classification or regression problem, however, the decision-making and trading strategies are incorporated into portfolio management which has improved practical applicability. The study in the FOREX market is less than the stock market but has become an emerging field. In terms of methods for predictive modeling, machine learning, deep learning, and reinforcement learning have become the three mainstream approaches.

ACKNOWLEDGMENTS

The authors would like to thank Sunisth Kumar, who has helped collecting and reviewing some papers about FSA applications under the NTU-India Connect Programme.

REFERENCES

[1] Md Shad Akhtar, Abhishek Kumar, Deepanway Ghosal, Asif Ekbal, and Pushpak Bhattacharyya. 2017. A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis. In *Proceedings of EMNLP*. 540–546.

[2] Khamis Hamed Al-Yahyaee, Mobeen Ur Rehman, Walid Mensi, and Idries Mohammad Wanas Al-Jarrah. 2019. Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches. *The North American Journal of Economics and Finance* 49 (2019), 47–56.

[3] M Ángeles López-Cabarcos, Ada M Pérez-Pico, Maria Luisa López Perez, et al. 2020. Investor sentiment in the theoretical field of behavioural finance. *Economic research-Ekonomska istraživanja* 33, 1 (2020), 2101–2228.

[4] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).

[5] Mattia Atzeni, Amna Dridi, and Diego Reforgiato Recupero. 2017. Fine-grained sentiment analysis on financial microblogs and news headlines. In *Semantic Web Evaluation Challenge*. Springer, 124–128.

[6] Malcolm Baker and Jeffrey Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61, 4 (2006), 1645–1680.

[7] Malcolm Baker and Jeffrey Wurgler. 2007. Investor sentiment in the stock market. *Journal of economic perspectives* 21, 2 (2007), 129–152.

[8] Harit Bandi, Suyash Joshi, Siddhant Bhagat, and Dayanand Ambawade. 2021. Integrated technical and sentiment analysis tool for market index movement prediction, comprehensible using xai. In *2021 ICCICT*. IEEE, 1–8.

[9] Arindam Bandopadhyaya and Anne Leah Jones. 2006. Measuring investor sentiment in equity markets. *Journal of Asset Management* 7, 3 (2006), 208–215.

[10] Nicholas Barberis, Andrei Shleifer, and Robert Vishny. 1998. A model of investor sentiment. *Journal of Financial Economics* 49, 3 (1998), 307–343.

[11] Anna Blajer-Golebiewska, Dagmara Wach, and Maciej Kos. 2018. Financial risk information avoidance. *Economic Research-Ekonomska Istraživanja* 31, 1 (2018), 521–536.

[12] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

[13] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1 (2011), 1–8.

[14] Robert M Bowen, Angela K Davis, and Dawn A Matsumoto. 2002. Do conference calls affect analysts' forecasts? *The Accounting Review* 77, 2 (2002), 285–316.

[15] Erik Cambria, Qian Liu, Sergio Decherchi, Frank Xing, , and Kenneth Kwok. 2022. SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis. In *LREC*. 3829–3839.

[16] Erik Cambria, Rui Mao, Melvin Chen, Zhaoxia Wang, and Seng-Beng Ho. 2023. Seven Pillars for the Future of Artificial Intelligence. *IEEE Intelligent Systems* 38, 6 (2023), 62–69.

[17] Salvatore Carta, Alessandro Sebastian Podda, Diego Reforgiato Recupero, and Maria Madalina Stanciu. 2021. Explainable AI for financial forecasting. In *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 51–69.

[18] Samuel WK Chan and Mickey WC Chong. 2017. Sentiment analysis in financial texts. *Decision Support Systems* 94 (2017), 53–64.

[19] Jaebin Jay Chang. 2020. Natural Language Processing as a Predictive Feature in Financial Forecasting. (2020).

[20] Mingqin Chen, Zhenhua Zhang, Jiawen Shen, Zhijian Deng, Jiaxing He, and Shiting Huang. 2020. A Quantitative Investment Model Based on Random Forest and Sentiment Analysis. *Journal of Physics: Conference Series* 1575, 1 (2020), 012083. https://doi.org/10.1088/1742-6596/1575/1/012083

[21] Siyi Chen and Frank Xing. 2023. Understanding Emojis for Financial Sentiment Analysis. In *Proceedings of the 44th International Conference on Information Systems (ICIS)*. 1–16.

[22] Yingmei Chen, Zhongyu Wei, and Xuanjing Huang. 2018. Incorporating corporation relationship via graph convolutional neural networks for stock price prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1655–1658.

[23] Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* 21, 1 (2020), 1–13.

[24] Lauren Cohen, Dong Lou, and Christopher J Malloy. 2020. Casting conference calls. *Management Science* 66, 11 (2020), 5015–5039.

[25] Keith Cortis, André Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. 2017. Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. Association for Computational Linguistics (ACL).

[26] Huijie Cui and Yanan Zhang. 2020. Does investor sentiment affect stock price crash risk? *Applied Economics Letters* 27, 7 (2020), 564–568.

[27] Zhi Da, Joseph Engelberg, and Pengjie Gao. 2011. In search of attention. *The Journal of Finance* 66, 5 (2011), 1461–1499.

[28] Sanjiv R. Das and Mike Y. Chen. 2007. Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science* 53 (2007), 1375–1388. Issue 9.

[29] Tobias Daudert. 2022. A multi-source entity-level sentiment corpus for the financial domain: the FinLin corpus. *Language Resources and Evaluation* 56, 1 (2022), 333–356.

[30] Dayan de França Costa and Nadia Felix Felipe da Silva. 2018. INF-UFG at FiQA 2018 Task 1: predicting sentiments and aspects on financial tweets and news headlines. In *Companion Proceedings of the The Web Conference 2018*. 1967–1971.

[31] Roger S Debreceny, Asheq Rahman, and Tawei Wang. 2021. Is User-Generated Twittersphere Activity Associated with Stock Market Reactions to 8-K Filings? *Journal of Information Systems* 35, 2 (2021), 195–217.

[32] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, and Akito Sakurai. 2011. Combining technical analysis with sentiment analysis for stock price prediction. In *IEEE DASC*. 800–807.

[33] Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z Pan, and Huajun Chen. 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of The 2019 World Wide Web Conference*. 678–685.

[34] Justina Deveikyte, Hélyette Geman, Carlo Piccari, and Alessandro Provetti. 2020. A Sentiment Analysis Approach to the Prediction of Market Volatility. *ArXiv* abs/2012.05906 (2020).

[35] Ann Devitt and Khurshid Ahmad. 2007. Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of ACL*. 984–991.

[36] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*. 4171–4186.

[37] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2014. Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of EMNLP*. 1415–1425.

[38] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *IJCNN*.

[39] Adam Drake, Eric Ringger, and Dan Ventura. 2008. Sentiment regression: Using real-valued scores to summarize overall document sentiment. In *IEEE ICSC*. 152–157.

[40] Amna Dridi, Mattia Atzeni, and Diego Reforgiato Recupero. 2019. FineNews: fine-grained semantic sentiment analysis on financial microblogs and news. *International Journal of Machine Learning and Cybernetics* (2019), 1–9.

[41] Kelvin Du, Frank Xing, and Erik Cambria. 2023. Incorporating Multiple Knowledge Sources for Targeted Aspect-based Financial Sentiment Analysis. *ACM Transactions on Management Information Systems* 14, 3 (2023), 1–24. https://doi.org/10.1145/3580480

[42] Kelvin Du, Frank Xing, Rui Mao, and Erik Cambria. 2023. FinSenticNet: A Concept-Level Lexicon for Financial Sentiment Analysis. In *IEEE SSCI*. 109–114.

[43] Xin Du and Kumiko Tanaka-Ishii. 2020. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In *Proceedings of ACL*. 3353–3363.

[44] Darren Duxbury, Tommy Gärling, Amelie Gamble, and Vian Klass. 2020. How emotions influence behavior in financial markets: a conceptual analysis and emotion-based account of buy-sell preferences. *The European Journal of Finance* 26, 14 (2020), 1417–1438.

[45] Robert F Engle and Victor K Ng. 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance* 48, 5 (1993), 1749–1778.

[46] Martin DD Evans and Richard K Lyons. 2005. Do currency markets absorb news quickly? *Journal of International Money and Finance* 24, 2 (2005), 197–217.

[47] Eugene F Fama. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25, 2 (1970), 383–417.

[48] Eugene F Fama. 1991. Efficient capital markets: II. *The journal of finance* 46, 5 (1991), 1575–1617.

[49] Weiguo Fan and Michael D Gordon. 2014. The power of social media analytics. *Commun. ACM* 57, 6 (2014), 74–81.

[50] Georgios Fatouros, John Soldatos, Kalliopi Kouroumali, Georgios Makridis, and Dimosthenis Kyriazis. 2023. Transforming Sentiment Analysis in the Financial Domain with ChatGPT. *Machine Learning with Applications* 14 (2023), 100508.

[51] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database.* Bradford Books.

[52] Paul Ferguson, Neil O'Hare, Michael Davy, Adam Bermingham, Paraic Sheridan, Cathal Gurrin, and Alan F Smeaton. 2009. Exploring the use of paragraph-level annotations for sentiment analysis of financial blogs. (2009).

[53] Kenneth L Fisher and Meir Statman. 2000. Investor sentiment and stock returns. *Financial Analysts Journal* 56, 2 (2000), 16–23.

[54] Richard Frankel, Marilyn Johnson, and Douglas J Skinner. 1999. An empirical examination of conference calls as a voluntary disclosure medium. *Journal of Accounting Research* 37, 1 (1999), 133–150.

[55] Xiang Gao, Weige Huang, and Hua Wang. 2021. Financial Twitter Sentiment on Bitcoin Return and High-Frequency Volatility. *Virtual Economics* 4, 1 (2021), 7–18.

[56] Mengshi Ge, Rui Mao, and Erik Cambria. 2022. Explainable Metaphor Identification Inspired by Conceptual Metaphor Theory. In *Proceedings of AAAI*. 10681–10689.

[57] Manoochehr Ghiassi, James Skinner, and David Zimbra. 2013. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications* 40, 16 (2013), 6266–6282.

[58] Deepanway Ghosal, Shobhit Bhatnagar, Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya. 2017. IITP at SemEval-2017 task 5: an ensemble of deep learning and feature based models for financial sentiment analysis. In *Proceedings of SemEval-2017*. 899–903.

[59] Pushpendu Ghosh, Ariel Neufeld, and Jajati Keshari Sahoo. 2022. Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters* 46 (2022), 102280.

[60] Shilpa Gite, Hrituja Khatavkar, Ketan Kotecha, Shilpi Srivastava, Priyam Maheshwari, and Neerav Pandey. 2021. Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science* 7 (2021), e340.

[61] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks* 18, 5-6 (2005), 602–610.

[62] John Griffith, Mohammad Najand, and Jiancheng Shen. 2020. Emotions in the stock market. *Journal of Behavioral Finance* 21, 1 (2020), 42–56.

[63] Xinyi Guo and Jinfeng Li. 2019. A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. *SNAMS*, 472–477.

[64] Felix Hamborg, Karsten Donnay, and Bela Gipp. 2021. Towards target-dependent sentiment classification in news articles. In *Diversity, Divergence, Dialogue: 16th International Conference, iConference 2021, Proceedings, Part II 16*. Springer, 156–166.

[65] Sooji Han, Rui Mao, and Erik Cambria. 2022. Hierarchical Attention Network for Explainable Depression Detection on Twitter Aided by Metaphor Concept Mappings. In *Proceedings of COLING*. 94–104.

[66] Elaine Henry. 2008. Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973)* 45, 4 (2008), 363–407.

[67] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[68] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In *ACL*. 328–339.

[69] Nan Hu, Peng Liang, and Xu Yang. 2023. Whetting All Your Appetites for Financial Tasks with One Meal from GPT? A Comparison of GPT, FinBERT, and Dictionaries in Evaluating Sentiment Analysis. SSRN: https://ssrn.com/abstract=4426455.

[70] Xin Huang, Wenbin Zhang, Yiyi Huang, Xuejiao Tang, Mingli Zhang, Jayachander Surbiryala, Vasileios Iosifidis, Zhen Liu, and Ji Zhang. 2021. LSTM Based Sentiment Analysis for Cryptocurrency Prediction. In *DASFAA*.

[71] Clayton Hutto and Eric Gilbert. 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of AAAI*, Vol. 8.

[72] Hitkul Jangid, Shivangi Singhal, Rajiv Ratn Shah, and Roger Zimmermann. 2018. Aspect-based financial sentiment analysis using deep learning. In *Companion Proceedings of the The Web Conference 2018*. 1961–1966.

[73] Mengxiao Jiang, Man Lan, and Yuanbin Wu. 2017. Ecnu at semeval-2017 task 5: An ensemble of regression algorithms with effective features for fine-grained sentiment analysis in financial domain. In *Proceedings of SemEval-2017*. 888–893.

[74] Zhengyao Jiang, Dixing Xu, and Jinjun Liang. 2017. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059* (2017).

[75] Fang Jin, Nathan Self, Parang Saraf, Patrick Butler, Wei Wang, and Naren Ramakrishnan. 2013. Forex-foreteller: Currency trend modeling using news articles. In *Proceedings of ACM SIGKDD*. 1470–1473.

[76] Zhigang Jin, Yang Yang, and Yuhong Liu. 2020. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications* 32 (07 2020).

[77] Sudipta Kar, Suraj Maharjan, and Thamar Solorio. 2017. RiTUAL-UH at SemEval-2017 Task 5: Sentiment Analysis on Financial Data Using Neural Networks. In *Proceedings of SemEval-2017*. 877–882.

[78] Colm Kearney and Sha Liu. 2014. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis* 33 (2014), 171–185.

[79] Katherine Keith and Amanda Stent. 2019. Modeling Financial Analysts' Decision Making via the Pragmatics and Semantics of Earnings Calls. In *Proceedings of ACL*. 493–503.

[80] Kwansoo Kim, Sang-Yong Lee, and Saïd Assar. 2021. The dynamics of cryptocurrency market behavior: sentiment analysis using Markov chains. *Industrial Management and Data Systems* ahead-of-print (11 2021).

[81] Soon-Ho Kim and Dongcheol Kim. 2014. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization* 107 (2014), 708–729.

[82] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of EMNLP*. 1746–1751.

[83] Shimon Kogan, Dimitry Levin, Bryan R Routledge, Jacob S Sagi, and Noah A Smith. 2009. Predicting risk from financial reports with regression. In *Proceedings of NAACL-HLT*. 272–280.

[84] Moshe Koppel and Itai Shtrimberg. 2006. Good news or bad news? Let the market decide. In *Computing Attitude and Affect in Text: Theory and Applications*. Springer, 297–301.

[85] Prahlad Koratamaddi, Karan Wadhwani, Mridul Gupta, and Sriram G. Sanjeevi. 2021. Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal* (2021).

[86] Shinta Koyano and Kazushi Ikeda. 2017. Online portfolio selection based on the posts of winners and losers in stock microblogs. In *IEEE SSCI*.

[87] Olivier Kraaijeveld and Johannes De Smedt. 2020. The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money* 65 (2020), 101188.

[88] Srikumar Krishnamoorthy. 2018. Sentiment analysis of financial news articles using performance indicators. *Knowledge and Information Systems* 56, 2 (2018), 373–394.

[89] Devinder Kumar, Graham W Taylor, and Alexander Wong. 2017. Opening the black box of financial ai with clear-trade: A class-enhanced attentive response approach for explaining and visualizing deep learning-driven stock market prediction. *arXiv preprint arXiv:1709.01574* (2017).

[90] Helinä Laakkonen. 2004. The impact of macroeconomic news on exchange rate volatility. *Bank of Finland discussion paper* 24 (2004).

[91] Bin Liang, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. 2022. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems* 235 (2022), 107643.

[92] Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5, 1 (2012), 1–167.

[93] Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-BERT: Enabling language representation with knowledge graph. In *Proceedings of AAAI*, Vol. 34. 2901–2908.

[94] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs.CL]

[95] Yang Liu, Qingguo Zeng, Huanrui Yang, and Adrian Carrio. 2018. Stock price movement prediction from financial news with deep learning and knowledge graph embedding. In *Pacific Rim Knowledge Acquisition Workshop*. Springer, 102–113.

[96] Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2020. FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining.. In *IJCAI*. 4513–4519.

[97] Tim Loughran and Bill McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66, 1 (2011), 35–65.

[98] Brian M Lucey and Michael Dowling. 2005. The role of feelings in investor decision-making. *Journal of Economic Surveys* 19, 2 (2005), 211–237.

[99] Ling Luo, Xiang Ao, Feiyang Pan, Jin Wang, Tong Zhao, Ningzi Yu, and Qing He. 2018. Beyond Polarity: Interpretable Financial Sentiment Analysis with Hierarchical Query-driven Attention.. In *IJCAI*. 4244–4250.

[100] Yun Luo, Hongjie Cai, Linyi Yang, Yanxia Qin, Rui Xia, and Yue Zhang. 2022. Challenges for Open-domain Targeted Sentiment Analysis. *arXiv preprint arXiv:2204.06893* (2022).

[101] Yu Ma, Rui Mao, Qika Lin, Peng Wu, and Erik Cambria. 2023. Multi-source Aggregated Classification for Stock Price Movement Prediction. *Information Fusion* 91 (2023), 515–528.

[102] Yu Ma, Rui Mao, Qika Lin, Peng Wu, and Erik Cambria. 2024. Quantitative Stock Portfolio Optimization by Multi-task Learning Risk and Return. *Information Fusion* 104 (2024), 102165.

[103] Macedo Maia, Siegfried Handschuh, André Freitas, Brian Davis, Ross McDermott, Manel Zarrouk, and Alexandra Balahur. 2018. WWW'18 open challenge: financial opinion mining and question answering. In *Companion Proceedings of the The Web Conference 2018*. 1941–1942.

[104] Lorenzo Malandri, Frank Z Xing, Carlotta Orsenigo, Carlo Vercellis, and Erik Cambria. 2018. Public mood–driven asset allocation: The importance of financial sentiment in portfolio management. *Cognitive Computation* 10, 6 (2018), 1167–1176.

[105] Burton G Malkiel. 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives* 17, 1 (2003), 59–82.

[106] Pekka Malo, Ankur Sinha, Pekka Korhonen, Jyrki Wallenius, and Pyry Takala. 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology* 65, 4 (2014), 782–796.

[107] Xiliu Man, Tong Luo, and Jianwu Lin. 2019. Financial sentiment analysis (FSA): A survey. In *IEEE ICPS*. 617–622.

[108] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of ACL: system demonstrations*. 55–60.

[109] Youness Mansar, Lorenzo Gatti, Sira Ferradans, Marco Guerini, and Jacopo Staiano. 2017. Fortia-FBK at SemEval-2017 Task 5: Bullish or Bearish? Inferring Sentiment towards Brands from Financial News Headlines. In *Proceedings of (SemEval-2017)*. 817–822.

[110] Huina Mao, Pengjie Gao, Yongxiang Wang, and Johan Bollen. 2014. Automatic construction of financial semantic orientation lexicon from large-scale Chinese news corpus. *Institut Louis Bachelier* 20, 2 (2014), 1–18.

[111] Rui Mao, Kelvin Du, Yu Ma, Luyao Zhu, and Erik Cambria. 2023. Discovering the Cognition behind Language: Financial Metaphor Analysis with MetaPro. In *IEEE ICDM*. IEEE, 1211–1216.

[112] Rui Mao and Xiao Li. 2021. Bridging Towers of Multitask Learning with a Gating Mechanism for Aspect-based Sentiment Analysis and Sequential Metaphor Identification. In *Proceedings of AAAI*. 13534–13542.

[113] Rui Mao, Xiao Li, Kai He, Mengshi Ge, and Erik Cambria. 2023. MetaPro Online: A Computational Metaphor Processing Online System. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Vol. 3. 127–135.

[114] Rui Mao, Chenghua Lin, and Frank Guerin. 2018. Word Embedding and WordNet Based Metaphor Identification and Interpretation. In *Proceedings of ACL*, Vol. 1. 1222–1231.

[115] Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-End Sequential Metaphor Identification Inspired by Linguistic Theories. In *Proceedings of ACL*. 3888–3898.

[116] Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. 2017. Learned in translation: Contextualized word vectors. *NeurIPS* 30 (2017).

[117] Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional LSTM. In *Proceedings of SIGNLL*. 51–61.

[118] G Mujtaba Mian and Srinivasan Sankaraguruswamy. 2012. Investor sentiment and stock market response to earnings news. *The Accounting Review* 87, 4 (2012), 1357–1384.

[119] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *NeurIPS* 26 (2013).

[120] Karo Moilanen, Stephen Pulman, and Yue Zhang. 2010. Packed feelings and ordered sentiments: Sentiment parsing with quasi-compositional polarity sequencing and compression. In *Proc. WASSA Workshop at ECAI*. 36–43.

[121] Rodrigo Moraes, João Francisco Valiati, and Wilson P GaviãO Neto. 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications* 40, 2 (2013), 621–633.

[122] Antonio Moreno-Ortiz, Javier Fernández-Cruz, and Chantal Pérez Chantal Hernández. 2020. Design and evaluation of SentiEcon: A fine-grained economic/financial sentiment lexicon from a corpus of business news. In *Proceedings of LREC*. 5065–5072.

[123] Mohamed M Mostafa. 2013. More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications* 40, 10 (2013), 4241–4251.

[124] Andrius Mudinas, Dell Zhang, and Mark Levene. 2019. Market trend prediction using sentiment analysis: lessons learned and paths forward. *arXiv preprint arXiv:1903.05440* (2019).

[125] Muhammad Abubakr Naeem, Imen Mbarki, and Syed Jawad Hussain Shahzad. 2021. Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears. *International Review of Economics & Finance* 73 (2021), 496–514.

[126] Usman Naseem, Imran Razzak, Shah Khalid Khan, and Mukesh Prasad. 2021. A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. *Transactions on Asian and Low-Resource Language Information Processing* 20, 5 (2021), 1–35.

[127] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41, 16 (2014), 7653–7670.

[128] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2015. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications* 42, 1 (2015), 306–324.

[129] Thien Hai Nguyen and Kiyoaki Shirai. 2015. Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of ACL-IJCNLP*. 1354–1364.

[130] Clemens Nopp and Allan Hanbury. 2015. Detecting risks in the banking system by sentiment analysis. In *Proceedings of EMNLP*. 591–600.

[131] Neil O'Hare, Michael Davy, Adam Bermingham, Paul Ferguson, Páraic Sheridan, Cathal Gurrin, and Alan F Smeaton. 2009. Topic-dependent sentiment analysis of financial blogs. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. 9–16.

[132] Sotirios Oikonomopoulos, Katerina Tzafilkou, Dimitrios Karapiperis, and Vassilios Verykios. 2022. Cryptocurrency Price Prediction using Social Media Sentiment Analysis. In *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 1–8.

[133] Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2016. Stock market sentiment lexicon acquisition using microblogging data and statistical measures. *Decision Support Systems* 85 (2016), 62–73.

[134] Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2017. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications* 73 (2017), 125–144.

[135] Keane Ong, Wihan van der Heever, Ranjan Satapathy, Gianmarco Mengaldo, and Erik Cambria. 2023. FinXABSA: Explainable Finance through Aspect-Based Sentiment Analysis. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. 773–782.

[136] Jihye Park, Hye Jin Lee, and Sungzoon Cho. 2021. Automatic Construction of Context-Aware Sentiment Lexicon in the Financial Domain Using Direction-Dependent Words. *arXiv preprint arXiv:2106.05723* (2021).

[137] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of EMNLP*. 1532–1543.

[138] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of NAACL-HLT*.

[139] Guangyuan Piao and John G Breslin. 2018. Financial aspect and sentiment predictions with deep neural networks: an ensemble approach. In *Companion Proceedings of the The Web Conference 2018*. 1973–1977.

[140] Juan Piñeiro-Chousa, Marcos Vizcaíno-González, and Ada María Pérez-Pico. 2017. Influence of social media over the stock market. *Psychology & Marketing* 34, 1 (2017), 101–108.

[141] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, and al. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *International Workshop on Semantic Evaluation*. 19–30.

[142] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).

[143] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[144] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Conference on Empirical Methods in Natural Language Processing*.

[145] Navid Rekabsaz, Mihai Lupu, Artem Baklanov, Allan Hanbury, Alexander Dür, and Linda Anderson. 2017. Volatility prediction using financial disclosures sentiments with word embedding-based IR models. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (2017).

[146] Thomas Renault. 2017. Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance* 84 (2017), 25–40.

[147] Leon Rotim, Martin Tutek, and Jan Šnajder. 2017. TakeLab at SemEval-2017 Task 5: Linear aggregation of word embeddings for fine-grained sentiment analysis of financial news. In *Proceedings of SemEval-2017*. 866–871.

[148] Jayit Saha, Smit Patel, Frank Xing, and Erik Cambria. 2022. Does Social Media Sentiment Predict Bitcoin Trading Volume?. In *Proceedings of the 43rd International Conference on Information Systems (ICIS)*. 1–9.

[149] Ramit Sawhney, Shivam Agarwal, Arnav Wadhwa, and Rajiv Shah. 2020. Deep attentive learning for stock movement prediction from social media text and company correlations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 8415–8426.

[150] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Shah. 2021. Quantitative Day Trading from Natural Language using Reinforcement Learning. In *Proceedings of NAACL-HLT*. 4018–4030.

[151] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. 2021. FAST: Financial News and Tweet Based Time Aware Network for Stock Trading. In *Proceedings of EACL*. 2164–2175.

[152] Kim Schouten, Flavius Frasincar, and Franciska de Jong. 2017. Commit at semeval-2017 task 5: Ontology-based method for sentiment analysis of financial headlines. In *Proceedings of SemEval-2017*. 883–887.

[153] Saeed Seifollahi and Mehdi Shajari. 2019. Word sense disambiguation application in sentiment analysis of news headlines: an applied approach to FOREX market prediction. *Journal of Intelligent Information Systems* 52, 1 (2019), 57–83.

[154] Hamed Naderi Semiromi, Stefan Lessmann, and Wiebke Peters. 2020. News will tell: Forecasting foreign exchange rates based on news story events in the economy calendar. *The North American Journal of Economics and Finance* 52 (2020), 101181.

[155] Dehua Shen, Andrew Urquhart, and Pengfei Wang. 2019. Does twitter predict Bitcoin? *Economics letters* 174 (2019), 118–122.

[156] Antonios Siganos, Evangelos Vagenas-Nanos, and Patrick Verwijmeren. 2017. Divergence of sentiment and stock market trading. *Journal of Banking & Finance* 78 (2017), 130–141.

[157] Ankur Sinha, Satishwar Kedas, Rishu Kumar, and Pekka Malo. 2022. SEntFiN 1.0: Entity-aware sentiment analysis for financial news. *Journal of the Association for Information Science and Technology* (2022).

[158] Jasmina Smailović, Miha Grčar, Nada Lavrač, and Martin Žnidaršič. 2014. Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences* 285 (2014), 181–203.

[159] Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of *SEM*. 159–169.

[160] Sahar Sohangir, Dingding Wang, Anna Pomeranets, and Taghi M Khoshgoftaar. 2018. Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data* 5, 1 (2018), 1–25.

[161] Yejun Soun, Jaemin Yoo, Minyong Cho, Jihyeong Jeon, and U Kang. 2022. Accurate Stock Movement Prediction with Self-supervised Learning from Sparse Noisy Tweets. In *IEEE International Conference on Big Data*. 1691–1700.

[162] Yunchuan Sun, Mengting Fang, and Xinyu Wang. 2018. A novel stock recommendation system using Guba sentiment analysis. *Personal and Ubiquitous Computing* 22 (2018), 575–587.

[163] Pyry Takala, Pekka Malo, Ankur Sinha, and Oskar Ahlgren. 2014. Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. In *LREC*, Vol. 2014. Citeseer, 2152–2157.

[164] Duyu Tang, Bing Qin, and Ting Liu. 2015. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of EMNLP*. 1422–1432.

[165] Richard M Tong. 2001. An operational system for detecting and tracking opinions in on-line discussion. In *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification*, Vol. 1.

[166] Ming-Feng Tsai and Chuan-Ju Wang. 2014. Financial keyword expansion via continuous word vector representations. In *Proceedings of EMNLP*. 1453–1458.

[167] Wenting Tu, David W Cheung, Nikos Mamoulis, Min Yang, and Ziyu Lu. 2016. Investment recommendation using investor opinions in social media. In *Proceedings of ACM SIGIR*. 881–884.

[168] Robert Tumarkin and Robert F Whitelaw. 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal* 57, 3 (2001), 41–51.

[169] Peter D Turney and Michael L Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* 21, 4 (2003), 315–346.

[170] Marjan van de Kauter, Diane Breesch, and Véronique Hoste. 2015. Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications* 42, 11 (2015), 4999–5010.

[171] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research* 11, 12 (2010).

[172] Chuan-Ju Wang, Ming-Feng Tsai, Tse Liu, and Chin-Ting Chang. 2013. Financial sentiment analysis for risk prediction. In *Proceedings of IJCNLP*. 802–808.

[173] Gang Wang, Tianyi Wang, Bolun Wang, Divya Sambasivan, Zengbin Zhang, Haitao Zheng, and Ben Y Zhao. 2015. Crowds on wall street: Extracting value from collaborative investing platforms. In *Proceedings of ACM CSCW*. 17–30.

[174] William Yang Wang and Zhenhao Hua. 2014. A semiparametric Gaussian copula regression model for predicting financial risks from earnings calls. In *Proceedings of ACL*. 1155–1165.

[175] Huizhe Wu, Wei Zhang, Weiwei Shen, and Jun Wang. 2018. Hybrid deep sequential modeling for social text-driven stock prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*. 1627–1630.

[176] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564* (2023).

[177] Shengting Wu, Yuling Liu, Ziran Zou, and Tien-Hsiung Weng. 2022. S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. *Connection Science* 34, 1 (2022), 44–62.

[178] Chunli Xiang, Junchi Zhang, Fei Li, Hao Fei, and Donghong Ji. 2022. A semantic and syntactic enhanced neural model for financial sentiment analysis. *Information Processing & Management* 59, 4 (2022), 102943.

[179] Boyi Xie, Rebecca Passonneau, Leon Wu, and Germán G Creamer. 2013. Semantic frames to predict stock price movement. In *Proceedings of ACL*. 873–883.

[180] Frank Xing, Duc Hong Hoang, and Dinh-Vinh Vo. 2020. High-frequency news sentiment and its application to forex market prediction. In *Proceedings of HICSS*.

[181] Frank Xing, Lorenzo Malandri, Yue Zhang, and Erik Cambria. 2020. Financial sentiment analysis: an investigation into common mistakes and silver bullets. In *Proceedings of COLING*. 978–987.

[182] Frank Z Xing, Erik Cambria, and Roy E Welsch. 2018. Intelligent asset allocation via market sentiment views. *IEEE Computational Intelligence Magazine* 13, 4 (2018), 25–34.

[183] Frank Z Xing, Erik Cambria, and Roy E Welsch. 2018. Natural language based financial forecasting: a survey. *Artificial Intelligence Review* 50, 1 (2018), 49–73.

[184] Frank Z Xing, Erik Cambria, and Yue Zhang. 2019. Sentiment-aware volatility forecasting. *Knowledge-Based Systems* 176 (2019), 68–76.

[185] Yumo Xu and Shay B Cohen. 2018. Stock movement prediction from tweets and historical prices. In *Proceedings of ACL*. 1970–1979.

[186] Steve Yang, Jason Rosenfeld, and Jacques Makutonin. 2018. Financial aspect-based sentiment analysis using deep representations. *arXiv preprint arXiv:1808.07931* (2018).

[187] Steve Y Yang, Yangyang Yu, and Saud Almahdi. 2018. An investor sentiment reward-based trading system using Gaussian inverse reinforcement learning algorithm. *Expert Systems with Applications* 114 (2018), 388–401.

[188] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *NeurIPS*. 5754–5764.

[189] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of NAACL-HLT*. 1480–1489.

[190] Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. 2020. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In *Proceedings of AAAI*, Vol. 34. 1112–1119.

[191] Wei Jie Yeo, Wihan van der Heever, Rui Mao, Erik Cambria, Ranjan Satapathy, and Gianmarco Mengaldo. 2023. A comprehensive review on financial explainable AI. *arXiv preprint arXiv:2309.11960* (2023).

[192] Jaemin Yoo, Yejun Soun, Yong-chan Park, and U Kang. 2021. Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2037–2045.

[193] Jie Yuan and Zhu Zhang. 2020. Connecting the dots: forecasting and explaining short-term market volatility. In *Proceedings of ACM ICAIF*. 1–8.

[194] Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. 2023. Instruct-FinGPT: Financial Sentiment Analysis by Instruction Tuning of General-Purpose Large Language Models. *arXiv preprint arXiv:2306.12659* (2023).

[195] Pu Zhang and Zhongshi He. 2015. Using data-driven feature enrichment of text representation and ensemble technique for sentence-level polarity classification. *Journal of Information Science* 41, 4 (2015), 531–549.

[196] Wei Zhang, Xiao Li, Dehua Shen, and Andrea Teglio. 2016. Daily happiness and stock returns: Some international evidence. *Physica A: Statistical Mechanics and its Applications* 460 (2016), 201–209.

[197] Yuzhe Zhang and Hong Zhang. 2023. FinBERT–MRC: Financial Named Entity Recognition Using BERT Under the Machine Reading Comprehension Paradigm. *Neural Processing Letters* (2023), 1–21.

[198] Lin Zhao, Lin Li, and Xinhao Zheng. 2021. A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts. *2021 IEEE CSCWD* (2021), 1233–1238.

[199] Guofu Zhou. 2018. Measuring Investor Sentiment. *Annual Review of Financial Economics* 10, 1 (2018), 239–259.

[200] Zhihan Zhou, Liqian Ma, and Han Liu. 2021. Trade the Event: Corporate Events Detection for News-Based Event-Driven Trading. In *Findings of ACL-IJCNLP*. 2114–2124.