Modeling the Trends Of Assets In The Financial Market With Machine Learning: A Systematic Literature Review

Randall Lionel Kharkrang[1], Giancarlo Succi[2], and Miras Tokanov[3]

[1]Innopolis University, Innopolis, Russia, l.kharkrang@innopolis.university
[2]Università di Bologna, Bologna, Italy, g.succi@unibo.it
[3]Innopolis University, Innopolis, Russia, m.tokanov@innopolis.university

Abstract

People have been trading for centuries, taking into account the possible change in the price of a particular product. With the development of machine learning, traders can rely on machines and their forecasting abilities. This study used a literature review method to analyze various studies to determine the type of analysis, the pre-processing methods used and the models most commonly used. This study gathered more than 200 papers from 1996 to 2022, chosen according to the criteria set by the standards of systematic literature reviews. The proposed work should assist researchers in structuring this emerging field and identifying the particular elements that require additional research.

1 Introduction

The stock market is a place where traders can come to buy and sell a company's stock. With the passage of time, an increasing number of individuals are becoming interested in the stock market. People's growing interest in this subject makes it more important for research.

The stock market is a volatile systems consisting of millions of data points per day, and people's livelihood are at stake when they invest in a stake. A human or team of humans, would not cognitively be able to handle all these inputs and changes in seconds notice, especially when it is their money. Automated Stock trading provides an alternative and efficient approach to stock trading and prediction that relies on mathematical precision and computing power. As far back as 1973, where Rosenberg and McKibben (1973) have used probabilistic models for predicting risk of returns from stocks. As the decades go by, and more computing power increased, more sophisticated techniques evolved for stock prediction/ asset evolution. We will address these advances in the context of Machine Learning.

According to Okoli (2015), a Systematic Literature Review (SLR) is "a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners". A SLR then offers the reader a vast and comprehensive analysis of the research conducted in the field, while also critically synthesizing research in ways that can be beneficial for a larger audience. A rigorous standalone literature review must be systematic in following a methodological approach, explicit in explaining the procedures by which it was conducted, comprehensive in its scope of including all relevant material, and, hence, reproducible by others who would follow the same approach in reviewing the topic. SLRs are important tools as they offer a baseline for scientific progress and for this reason are increasingly used by researchers Ozbayoglu, Gudelek, and Sezer (2020).

2 Related Work

Asset Evolution has been built on the foundation of CAPM developed by Shih et al. (2014) tracks the evolution of CAPM over the decades which eventually led us to our review of today.

We can divide the related works into 4 categories as follows:

- Studies that review and compare various Deep Learning and Machine Learning algorithms.
- Studies that apply specific techniques or a pipeline of techniques.
- Studies which delve into information retrieval of relevant financial technicalities.
- Studies that use State of the Art Deep Learning Techniques that resemble data representations that can be used to model financial assets

The field of finance has recently received a boom with regards to applications of Machine Learning and Deep Learning models (see Ozbayoglu, Gudelek, and Sezer, 2020).

Finance is an extremely diverse field ranging from Asset and Portfolio Management, Risk Assessment, Fraud Detection and many others.

In this SLR we want to explore various methods and ways by which Machine Learning has been used in Asset Evolution.

So far, we have found no relevant SLR that deals with modelling how a financial asset evolves over time, and taking into account textual data like news articles/tweets. However various papers and SLRs have attempted to study and compare various methods using machine learning to financial datasets as found in Zhang (2022), Feng, He, and Polson (2018), Thakkar and Chaudhari (2021). In contention to the methodologies involved, Support Vector Machines performed the best, whilst in the deep learning aspect, LSTMs and Attention based approaches saw optimal performance. As mentioned in Sharma et al. (2019), social media plays an important role in forecasting stock market returns. The main goals of this module are to use social media to access market sentiment and predict the behavior of a specific com-

pany's stock, classify the polarity of given text at the Document, sentence, or feature level and determine whether opinion in text is positive, negative, or neutral, and use machine learning algorithms to predict sentiment and find correlation between sentiment and stock prices. Islam et al. (2020) addresses forex currency prediction and conducted a review of the current state of the art, which showed that SVMs and Neural Networks were the dominant approaches. Other methodologies like Chaos Theory were also addressed by Islam et al. (2020). Mirete-Ferrer et al. (2022) presents a comprehensive survey of asset management in terms of price forecasting and value investing. SVR was concluded to be the highest performing model architecture. Neural networks showed excellent in-sample performance but mediocre out-of-sample performance. Broby (2022) discusses the use of predictive analytics in the field of finance as a whole, in terms of customer segmentation, Audit and Compliance Prediction, Stock Price, returns and volatility prediction etc. and also mentions the lack of using machine learning for unstructured data. Sharaf et al. (2022) talks in depth about the theoretical background and state of the art methodologies of recommender systems and their lack of usage in financial systems and finally addresses their strengths and weaknesses of the various types of recommender systems i.e., collaborative recommendation, content based etc. Saha, Gao, and Gerlach (2022) introduces a graph based formulation for stock movement prediction, describes the theoretical graph theory background, and uses graph neural networks.

3 Structure of the Systematic Literature Review

One of the main goals of this systematic literature review (SLR) is to summarise and categorize Machine Learning models and datasets used in asset allocation, asset evolution, asset management and other asset optimization methodologies that are reported in the literature. Besides providing us with a state-of-the-art summary of current research on this topic, this work will also contribute to highlighting issues that need to be solved as well as open gaps in the literature. This SLR will therefore be beneficial for researchers, developers, and various other types of IT practitioners and for anyone interested in software engineering, as it will allow them to:

a) Easily identify the best performing ML architectures in asset evolution;

b) Identify the most promising research directions for future work;

c) Develop new, more effective strategies for dealing with asset evolution from various sources

d) Integrate NLP into the decision pipeline

This work is organized as follows: Section 4 presents our methodology and outlines our research protocol; Section 5 discusses the main results of our work, while clustering and classifying them in meaningful ways; Section 6 offers a critical interpretation of our findings and specific answers to our research questions; Section 8 evaluates the limitations and various other shortcomings potentially affecting our SLR, while Section 9 summarizes what we achieved and points out future research plans/directions.

4 Protocol Development

4.1 Research Questions

One of the most important step in the production of an SLR involves the individuation of a research question (or of a series of research questions) that can be used to guide and inform its development. The research questions characterising this work are:

RQ$_1$ What are the most effective ML models and tools to predict the evolution of an asset in the financial market?

RQ$_2$ What are existing approaches for forecasting stock price trend based on news or tweets?

RQ$_3$ How accurately can machine learning models predict the stock price?

The motivation for **RQ1** comes from the requirement to search, gather, and explore the machine Learning models and deep learning models that effectively show good modelling and predictive performance on datasets revolving financial assets.

The motivation for **RQ2** was to explore state-of-the-art NLP methods that could process text from news sites and other sources of information and predict the impact on the price of goods (stock or token price) in the financial market.

The motivation for **RQ3** was to effectively used standard metrics to efectively, to understand how well these models performed

4.2 Literature Search

This section explains the research approach and methodology used to gather the studies included in our final reading log.

Search Keywords:

We extracted from the three research questions we formulated above a series of relevant

keywords that we believed might adequately characterise them. The keywords we selected for this study are: *a)* Asset Evolution, *b)* Asset pricing, *c)* Asset Management, *d)* Deep Learning, *e)* Machine Learning, *f)* Mathematical Finance,

Selected Databases:

We selected a series of databases, which we used to perform our initial searches. The databases we selected are: *a)* ScienceDirect, *b)* SpringerLink, *c)* SSRN, *d)* Google Scholar, *e)* Scopus, *f)* MDPI, *g)* IEEE

We recognize that the list of selected databases might not be the most comprehensive; however, we also note that these databases are among those that researchers most frequently use in their daily practice. In addition, these databases typically gather scholarly research from the most relevant sources in Computer Science and Finance.

Search Queries:

We then arranged the keywords we previously selected into more focused search queries by using Boolean operators. The list of search queries we developed follows below:

Q_1 (Deep Learning OR Machine Learning) AND Deep Learning AND Asset AND (evolution OR pricing

Q_2 (information retrieval) AND (Asset) AND (Evolution OR Pricing OR Forecasting)

Q_3 (Sentiment AND Stock price) AND (Prediction OR Forecasting)

Inclusion and Exclusion Criteria:

We applied a set of inclusion and exclusion criteria to determine which papers obtained from our preliminary searches should be included in the final reading log. In other words, the Inclusion and Exclusion criteria were used to ensure that only the most relevant sources were accepted in the SLR.

In the context of this SLR, the following criteria were used for inclusion:

Table 1: Database Papers Found.

Searched Databases	Q1	Q2	Q3
Google Scholar	306,000	199,000	622,000
SSRN	146	0	5
Scopus	35	72	72
ScienceDirect	11,451	11,438	9329
IEEE Xplore	5	7	135
Springer-Link	30669	26,212	26,948
MDPI	4	4	3
ACM	0	0	2,246
Arxiv	0	0	4

IC$_1$ The paper is published in a peer-reviewed journal or in a conference.

IC$_2$ The paper is related to finance/Machine learning.

IC$_3$ The paper uses Machine Learning And Statistics.

IC$_4$ The paper provides quantitative measurements or metrics involved and used by the finance sector

IC$_5$ The paper reports adopted preventive or corrective actions with respect to Asset Evolution/management

The exclusion criteria we used in this SLR are as follows:

EC$_1$ The paper is not written in English.

EC$_2$ The paper is an abstract, poster, summary, keynote, opinion piece, editorial (short paper) or it does not contain a significant amount of information.

EC$_3$ The paper is a duplicate or a secondary study.

EC$_4$ The paper can be considered as grey literature, e.g., a blog post, an unpublished manuscript or a graduate dissertation.

EC$_5$ The paper does not provide clear descriptions of methodologies used.

EC$_6$ The paper deals with asset evolution without suggesting appropriate metrics for predicting them.

EC$_7$ The paper does not satisfy IC1 or IC2.

EC$_8$ The paper does not fulfill at least one of IC3, IC4, and IC5.

Quality Assessment

To assess the quality of the studies included, we subsequently formulated a set of questions (or quality criteria):

QA$_1$ Does the study provide concrete examples of ML models and their performance?

QA$_2$ Is the dataset used large enough?

QA$_3$ Is the dataset from the financial domain?

QA$_4$ Does the study provide quantitative measurements of the metrics?

QA$_5$ Were the papers using state of the art NLP models and techniques and of relevance to financial data?

We evaluated the quality of each article in the final reading journal by assigning a weight to each of the quality criteria listed above. Specifically, we assigned 1 if the paper fully satisfied the question, 0.5 if it partially satisfied the question, and 0 otherwise.

Criteria for **QA$_1$** are the following:

Fully matched The study provides 2 or more examples of Machine learning models that outperform classical models in terms of performance

Partially matched The study includes not more than 2 examples of ML models that show slight/vast improvement over classical methods

Not matched The study contains only ML models that do not outperform classical methods using any metric of performance.

Criteria for QA_2 are the following:

Fully matched The dataset used was large (100000) data points

Partially matched The dataset was small (< 10000) data points

Not matched No data is provided.

Criteria for QA_3 are the following:

Fully matched The dataset used was from the finance domain and a gold standard

Partially matched The dataset used was not from finance but was a gold standard and had some resemblance to financial data

Not matched The dataset used was not from fiance and no resemblance to financial data

Criteria for QA_4 are as follows:

Fully matched The study used quantitative measurements

Not matched No data is provided

Criteria for QA_5 are the following:

Fully matched The paper uses state of the art NLP models and is of relevance to financial data

Partially matched The study with either uses NLP techniques or uses financial dataset but not both

Not matched The study does not use any financial dataset nor NLP techniques

Thus, the total score assigned to each paper could range from 0 to 5. The results of this quality assessment procedure can be found on 3. This process of quality assessment was performed on the 205 papers selected for inclusion in the reading log and it was instrumental

to determine, evaluate, and further assess, as noted above, the overall quality of each of the papers we included in this study. As the tables below demonstrate, the vast majority of the papers included in our study were of high quality, as expected. Hence, we can infer that the results we gathered from their analysis are scientifically sound.

However, the main goal of this Systematic Literature Review was not to simply evaluate the papers per se, rather, its aim was to collect and structure, in a meaningful way, information for future use.

Table 2: Paper selection procedure

Source	Initial Selection	Removed papers								Selected Papers		
		EC1	EC2	EC3	EC4	EC5	EC6	EC7	EC8	RQ1	RQ2	RQ3
MDPI	11	0	0	0	0	0	0	0	7	4	0	0
Google Scholar	851,000	305,000	50	200,000	51,900	10	250,000	3954	40,000	85	1	-
SSRN	151	0	0	25	0	0	15	100	0	10	0	-
Scopus	179	0	0	30	0	0	31	45	70	3	0	0
Science Direct	22,889	0	2	80	0	800	0	10,000	12000	2	5	0
IEEE Xplore	147	0	0	10	0	1	0	0	0	50	86	-
Springer Link	83,829	0	0	600	424	400	2000	400	80,000	5	0	0
ACM	2246	0	500	79	437	219	15	400	592	0	4	-
Arxiv	4	0	0	0	0	0	0	0	0	0	0	0

Table 3: Papers' Ranking

Quality Level	Score Range	Count	Percentage
Poor	[0-1.5]	20	9%
Good	[2-4]	100	49%
Excellent	[4.5-5]	85	42%
Total	—	205	100%

We note that the vast majority of the papers we included in our final reading log ($\sim 91\%$) were of good or excellent quality. Hence, we can infer that the data on which we based our SLR are scientifically sound. We also note that the hypen "-" in Table 2 indicates that they also satisfy that particular research question as well as the other one. For example, in Google Scholar, we note that 85 and 1 paper exclusively address RQ1 and RQ2 respectively, these
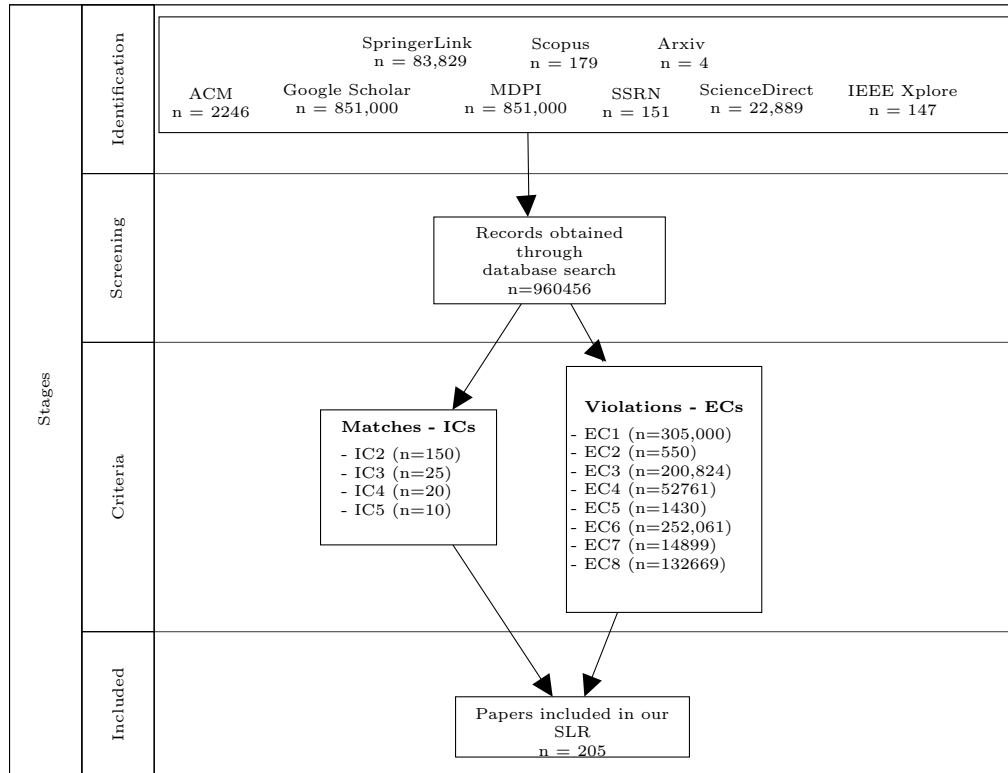
Figure 1: Prisma Flow Chart Diagram

papers collectively address RQ3 together. Figure 1 for Prisma template was obtained from Bugayenko et al. (2022)

5 Results

This section summarises the results obtained from reading and extracting the data from the papers we included in the final reading log.

5.1 Preliminary Clustering

As noted above, we performed this SLR on some of the most popular and trusted databases available to researchers.

We note that we attributed papers to single databases (even if some papers could be found across different databases), based on the chronological order of the searches performed. We also notice that two databases (Google Scholar and IEEE Explore) indexed the most (41 + 41% ∼ 82%) of the papers included in our reading log. Searches performed on MDPI and ACM rendered the least (∼ 1.9% each) number of relevant papers amounting to only of the total papers included in the log, while Arxiv accumulated zero number of papers.

Thus, in Figure 2 above and Table 2 above, we only represented papers, which we attributed to a single repository. The attribution was subjective in character and yet followed the following rule. The paper was attributed to one database (e.g. Google Scholar) rather than to the other (e.g., Scopus) based on the chronological order of the searches we performed. We think that while this is perhaps not an ideal practice, it does not constitute a significant problem for this work.

We also present the geographical distribution of the studies we included in our log (based on first author's affiliation). Figure 3 shows that the studies we selected have been published in at least 13 countries. The Other represents countries with publications less than or equal to 3, which includes Japan, Bangladesh, Greece, etc. This ensures a sufficient degree of cultural diversity for our reading log.

We also display a distribution of publications by year from 1996-2022 Figure 5. We can clearly see a rise in use of machine learning methods in asset allocation/forecasting, including
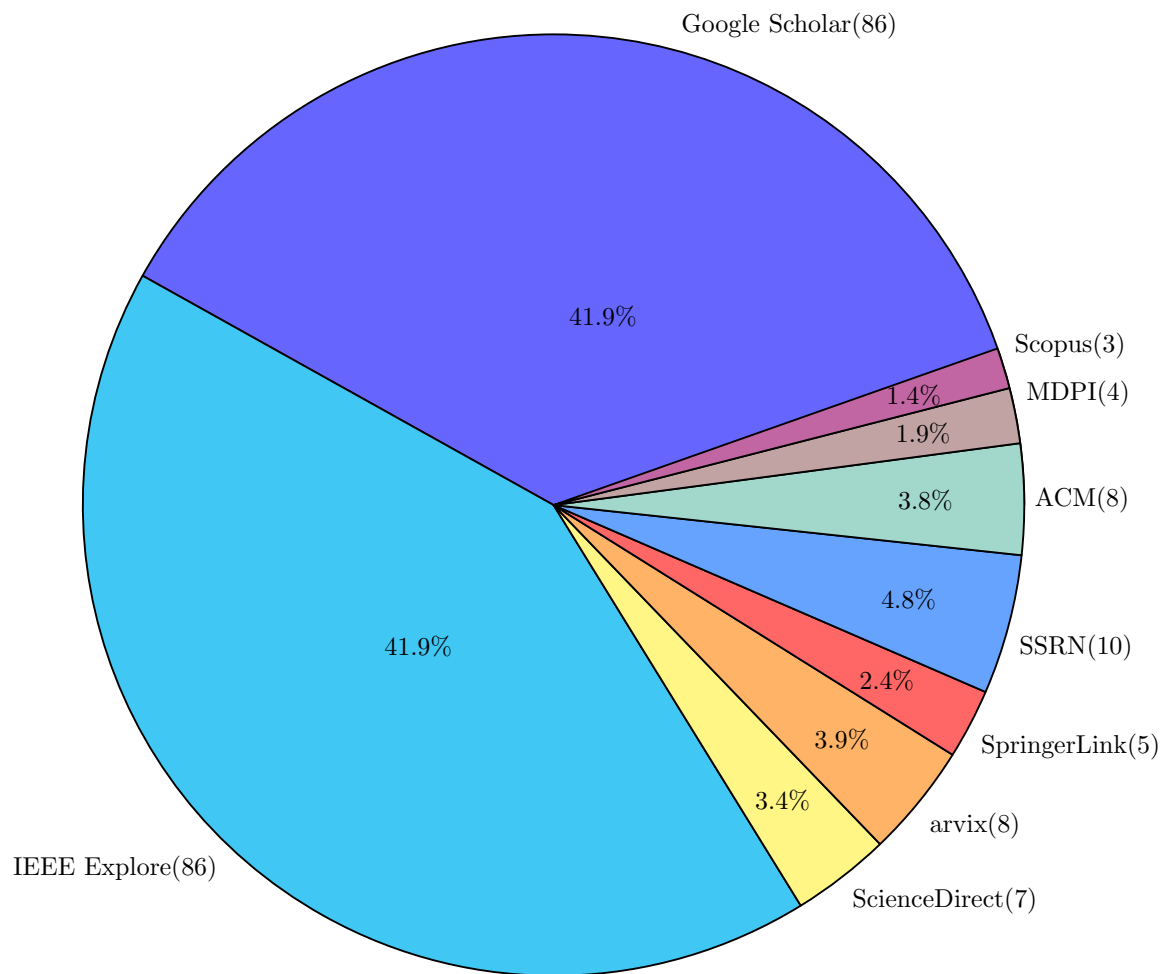
Figure 2: Distribution of studies by Database based on search queries

the use of NLP techniques. We have also presented the same data in, Table 5 where for simplicity the distribution is shown over a period of 5-years.

We also, visually display the distribution of Journals that contained collected papers in Figure 6. Others include journals such as Journal of Econometrics, Hindawi, Financial Letters etc. All of which are ≤ 3.

Table 4: Database Advantage and Limitations

Database	Advantages	Limitations
Google Scholar	This database is helpful in finding open access articles. It provides papers' citations data, which is a useful indicator of an article's popularity within a given scientific community.	Some papers (those not peer-reviewed) can be problematic, a few more are behind a paywall
Scopus,ScienceDirect and SSRN	Scientific databases with effective and precise search tools and analytic statistics	Most of its journals as well as search features are behind paywall
IEEE Xplore	Published papers are all peer-reviewed by leading experts in the field	Paywall
SpringerLink	High Quality peer-reviewed papers, with reasonably priced access	Paywall
MDPI	Peer-Reivewed and fast reviewd	Not as reputable as other journals

Table 5: Paper Distribution Around 5-year Period

| Year Range	Count	Percentage
1996-2004	4	1.9%
2005-2010	11	5.2%
2010-2015	20	9.6%
2016-2020	90	43.2%
2021-2022	80	38.4%
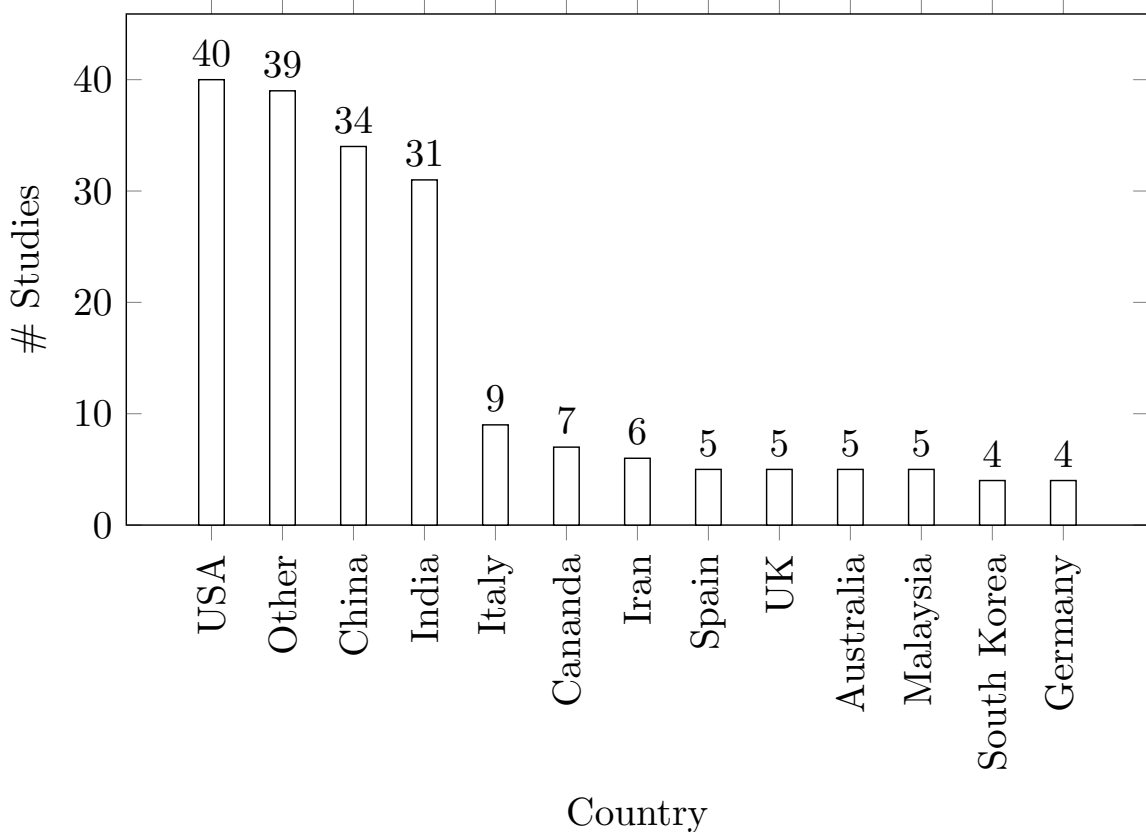
Figure 3: Distribution of studies by country of origin, based on first author's affiliation

5.2 Further Clustering

We present further clustering based on the following parameters a) Algorthims, b) Datasets, and c) Metrics.

In addition, we also clustered the datasets used in the papers included in the final reading log. It is worth noting that a single dataset was found in multiple articles and use of certain algorithms were prevalent in many articles. Table 6 lists the algorithms used for training models and the statistics of their usage. We observed that LSTMs and SVMs were the most frequently mentioned algorithms. We concluded that it can probably be considered as the most widely used in the field.
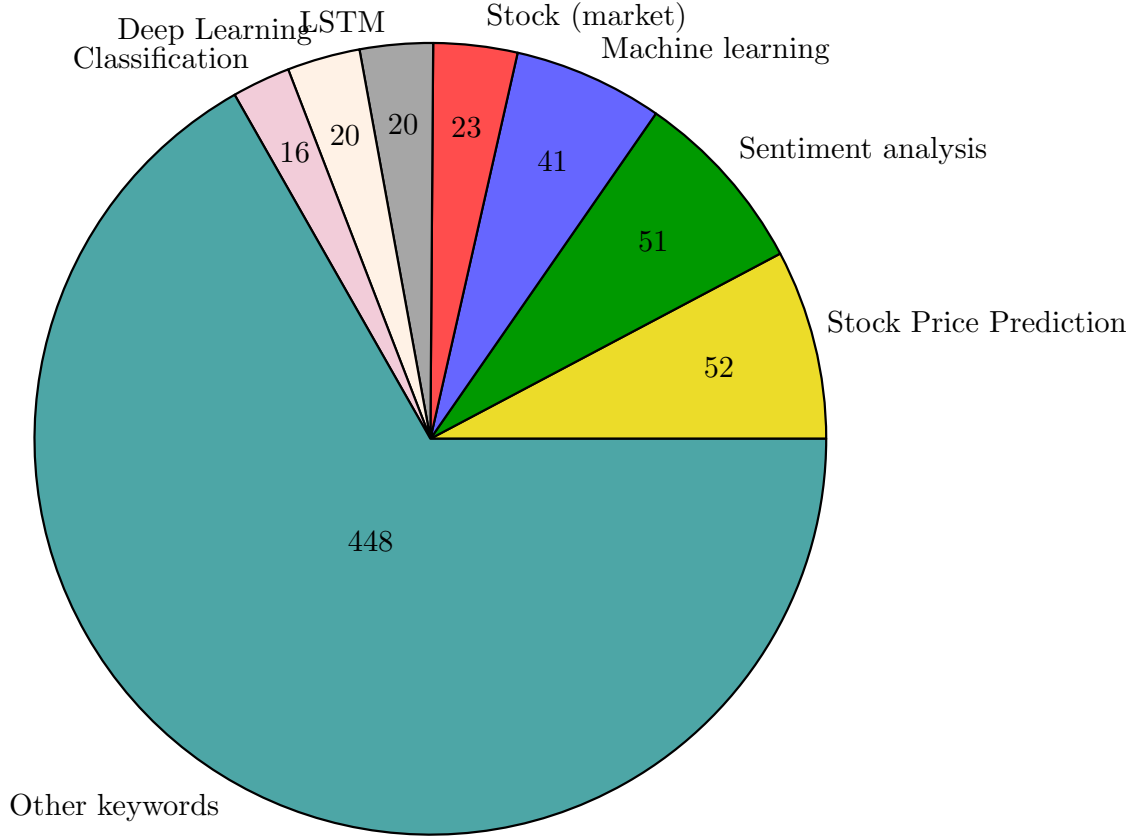
Figure 4: Distribution of keywords

"Other" Algorthims mentioned in Table 6 include k-Means Clustering, Transformer architectures, P-Trees, XGBoost etc. All of these algorithms used were less ≤ 2 in number. We also note a discrepancy in the number of algorithms used compared to total number of papers. This is due to the fact that some algorithms appeared in more than one papers, or used 2 or more algorithms in a pipeline as their solution.

Furthermore, Table 7 below displays a list of metrics used for evaluating a model.

Others here include Average absolute difference rate, Beta Value, Weighted Recall etc. All of these metrics are of count ≤ 2
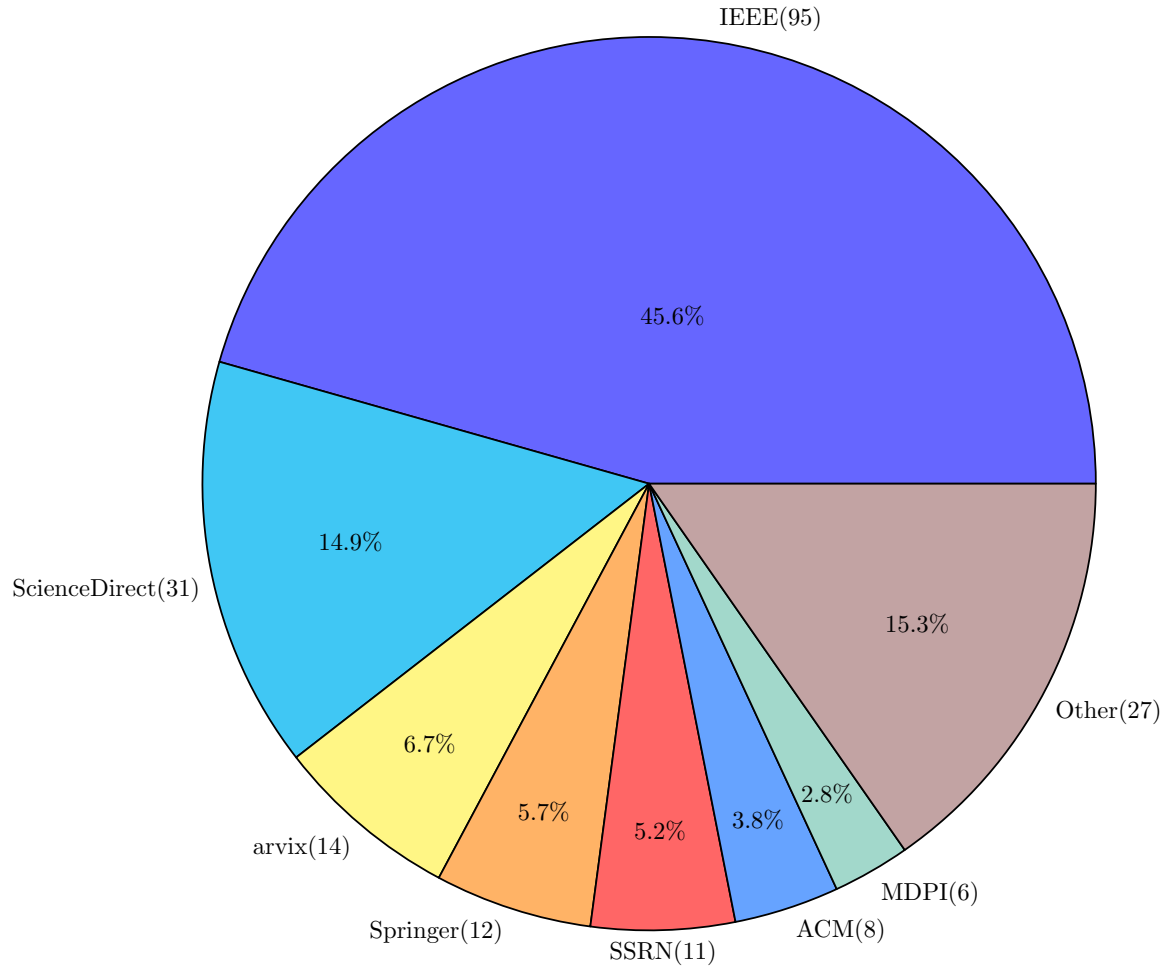
Finally, we present in Table 8, list of datasets used and also count for each.
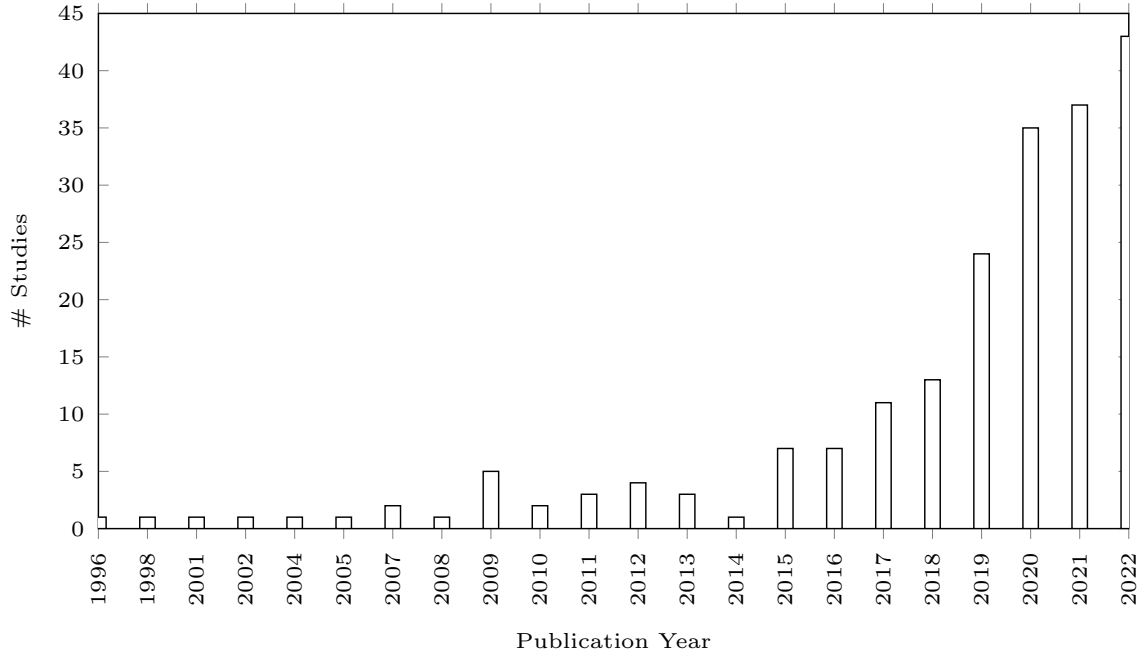
Figure 5: Distribution Of Studies By Year

Table 6: Algorithms Used

Algorithm	Count	Percentage
SVM	43	20%
LSTM	43	20%
Naive Bayes	22	10.5%
ARIMA	17	8%
Random Forest	12	5.7%
ANN	11	5.2%
Linear Regression	11	5.2%
kNN	10	4.8%
Logistic Regression	9	4.3%
RNN	7	3.3%
CNN	6	2.8%
Decision Trees	6	2.8%
GRU	6	2.8%
GAN	3	1.4%
Others	102	49%

Figure 6: Distribution of studies By Journals

The Others included datasets like Nifty, Shanghai Stock Exchange, etc. Others include datasets that are ≤ 3 in total frequency of papers.

Table 7: Metrics Used

Metrics	Count	Percentage
Accuracy	60	28.8%
R-Squared	39	18.75%
RMSE	32	15.3%
MSE	15	7.2%
MAE	15	7.2%
Precision	15	7.2%
Recall	14	6.7%
MAPE	10	4.8%
F-Score	8	4.3%
ROC-AUC	4	1.9%
MCC	3	1.4%
Correlation	3	1.4%
Sharpe Ratio	3	1.4%
Others	18	8.6%

6 Discussion

Table 8: Datsets Used In Studies

Dataset	Count	Percentage
Twitter	25	12%
S&P	21	10%
Yahoo Finance	12	5.7%
NASDAQ	11	5.2%
Reuters Website	5	2.4%
NSE	5	2.4%
NYSE	5	2.4%
Google Trend Index	4	1.9%
Others	139	66.8%

6.1 Datasets Used

Twitter, S&P and Yahoo Finance are the most predominant datasets that are used by the researchers in our SLR as shown in Table 8. The reason for this could be due to the fact that these are publicly available datasets, and are are typically huge in terms of number of sample points.

NASDAQ, Reuters constituted the remaining dominant datasets used by researchers. Twitter was the dominant source for Stock prediction using News articles (see Dong et al., 2020; Bollen, Mao, and Zeng, 2011; Sharma et al., 2019).

Some Articles used ML models for predicting sales performance using Blogs like in Liu et al. (2007). Other articles, collected and built their own custom collection of stocks, but did not mention the source as in Song et al. (2020)

News sites, as well as encyclopedia entries, are commonly used for text classification tasks because they offer a broad set of topics and (or) an abundance of texts.

6.2 Stock Preprocessing/Text Preprocessing

This section is devoted in commenting the on the preprocessing steps that were used before applying any Machine Learning models.

In terms of Stock pre-processing, some authors used predictor variables directly from publicly available datasets like S&P 500 (see Feng, He, and Polson, 2018), while some drastically change their data representation, converting them into graph structures, that takes correlation of asset returns to build connectivity between assets (see Son and Lee, 2022; Uddin, Tao, and Yu, 2021), while some converted the stocks into embeddings that are part of graph neural network pipelines (see Wu et al., 2019a). As found in Wu et al. (2019a), embeddings of stocks and news articles, returned best performance in terms of Sharpe Ratio and Profit&Loss. Shynkevich et al. (2015) divided pre-processing into three steps: feature extraction, selection and representation. Author mentioned that the Bag-of-Words approach is used for feature

extraction, and all articles are processed in the following way. Following that, words with capital letters are converted to lowercase, words with two or fewer characters are filtered out, and stop words are deleted. Finally, Porter's stemming algorithm is applied to each word in order to extract stems in order to treat different forms of the same word equally.

6.3 Algorithms

As observed from Table 6, SVM is the most popular used algorithm in our collection of studies. SVM works well for linearly separable data, however, if data is non-linear, SVM will employ kernel tricks to project the data into higher dimension , and try to separate the data in that dimension. SVM however, does not perform well with noisy data set, and does not work well with large datasets, that are prevalent in the finance industry.

Other approaches for forecasting Stock prices, was Random Forest as shown in Sadorsky (2021). A 20-day forecast horizon, tree bagging and random forests methods produce accuracy rates of between 85% and 90% while logit models produce accuracy rates of between 55% and 60%. Random Forests are highly interpretable, and according to this study, showed more promising results.

Enke and Thawornwong (2005) was one of the first instance in our collection of using Neural networks(NN), as well as using General Regression Neural Networks(GRNN). They made an importatnt point that eventhough GRNN provides better performance measurement, they did not generate as much profit compared to the vanilla NN. The NN showed better prediction of signs on highly volatile periods, hence paramters should be trained to make robust predicitons where higher returns are available.

The Other most popular algorthim that was used LSTM. The LSTM model was either used as a standalone architecture (see Koudjonou and Rout, 2020; Shen et al., 2021), or it was used with combination of other models like CNN-BiLSTM as in Wang et al. (2021) and combining traditional Time Series methods like ARIMA with LSTM (see Hua, 2020)

Other methodologies include Reinforcement Learning in stock trading using Deep Q networks (see Dang, 2019), as well as using and ensemble of Deep Q-learning agents as in Carta et al. (2021). Others used other policy optimization procedures like PPO, A2C as in Durall (2022)

Graph based approaches have also been in the rise, as it is seems natural to consider stocks as nodes of graphs as in Son and Lee (2022); Fazli et al. (2021)

Explainable AI is one of the other factors that should taken into account when building a robust and interpretable Neural Network, as we are allocating clients' resources and financial assets. Babaei, Giudici, and Raffinetti (2022) uses Shapely values of their complex ML model, can contribute to the shift in Z-score of their portfolio. Shapely Values and LIME were also used in Jakubowski et al. (2022) for asset failure prediction. LIME and Shapely values are quite easy to understand but fail to scale for large feature space. Other researchers opted to integrate fuzzy logic into system as in Chen, Cheng, and Tsai (2014); Xie, Rajan, and Chai (2021)

Further other methodologies applied heuristic based optimization approaches like Genetic Algorthim (see Thakkar and Chaudhari, 2022; Chen, Zhong, and Chen, 2020), fruitfly Optimization (see Tian, 2020)

Sharma et al. (2019) used in his work Generalized additive models (GAM). GAMs combine the concept of additive models with the standard generalized linear models. By estimating nonparametric functions of predictor variables that are connected to the dependent variable via some link function, generalized additive models aim to maximize prediction quality of a dependent variable from many distributions. The GAM model has the advantage of easily decomposing and incorporating new components, such as identifying and incorporating a new source of seasonality.

Algorithms like AdaBoost and Naive Bayes were used in the article Haider and Mehrotra (2011). To correctly identify the weights of the parameters, AdaBoost calls a weak classifier repeatedly in a series of rounds and The Naives Bayes classifier constructs the model using

estimator classes. The precision values of numerical estimators are determined by analyzing the training data.

To conclude, there are a number of ways and approaches to time series stock/ asset prediction/forecasting. Nowadays, NN based approaches can be considered a state-of-the-art approach being the one that is most widely used by researchers. The main advantage of NN is that in most of the cases it outperforms other approaches. Its major limitation is that it requires a lot of trials to succeed because none knows exactly, at the time of writing at least, the reason for this superior performance. This means that one needs to try a lot of different combinations of hyperparameters and different preprocessing techniques in order for the model to achieve optimal performance.

6.4   Quality Assessment Metrics

Fortunately, most papers in our SLR, use accuracy in combination with other representative metrics. For example, F-1 Score, Recall and Precision was used with accuracy. F-1 score eliminates the main limitation of accuracy and leads to more precise and reliable results.

For asset prediction/forecasting, R^2 was used. It was used in Zhang (2022); Chen, Pelger, and Zhu (2019); Uddin and Yu (2020); Son and Lee (2022); Wu et al. (2019a); Li and Mei (2020); Huang et al. (2021) and others. R^2 has a lot of issues, R-squared does not measure goodness of fit. It can be arbitrarily low when the model is completely correct, it can be arbitrarily close to 1 even if the model is completely wrong, to name a few. Thankfully, most of the researchers also compared it with other metrics like adjusted R^2.

Another important metric seen in quantitative financial literature, including our study, is Sharpe Ratio. It was used in Son and Lee (2022) for example. Sharpe ratio is simple formula , but it relies too much on standard deviation and treatment of volatility as the same.

Most popular metrics among stock price time series prediction papers is Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). For example, all these metrics have been used in these articles: Dong et al. (2020), Ji, Wang, and

Yan (2021), Guo (2020), Yu and Wu (2019), Yu, Wang, and Zhang (2017) and Ray, Ganguli, and Chakrabarti (2021). Mean Squared Error (MSE) and Root Mean Square Error (RMSE) penalize large prediction errors in comparison to Mean Absolute Error (MAE). However, because it uses the same units as the dependent variable, RMSE is more widely used than MSE to compare the performance of the regression model to other random models. In comparison to a non-differentiable function like MAE, MSE is a differentiable function that makes it easier to perform mathematical operations. As a result, despite being more difficult to interpret than MAE, RMSE is often used as the default metric for calculating Loss Function in many models.

Based on the statistical analysis performed above, it seems reasonable and appropriate to summarize the information obtained and to conduct a critical review of our research questions, which is what we will do next.

7   Critical Review of Research Questions

7.1   Analysis of RQ1

From The papers collected, there is a strong representation of LSTMs being used for Time Series prediction. But as mentioned in Feng, He, and Polson (2018), LSTMs seem to be overly optimistic and they predict less market crashes. A big problem with LSTMs is that they also require lots more training time, especially on large datasets, and tend to overfit due to their complex and expressive structure.

Cong et al. (2022) applied an interpretable tree based method, called P-Tree to generate a stochastic discount factor model and test assets for cross-sectional asset pricing. Unlike other tree algorithms, P-Trees accommodate imbalanced panels of asset returns and grow under the no-arbitrage condition. The model is quite simple, requires little training time, and deals with non-linearity, with the added benefit of being highly interpretable.

Son and Lee (2022) establishes a graph based approach to asset pricing.To reflect the connectedness between asset returns for risk exposure estimation, they used a graph convolutional

26

network (GCN), which takes into account the graph structure between firms, rather than a simple neural network. They also presented a way to construct a graph of firms that fits our goal, asset pricing, using the correlation of returns. Second, They proposed a forward stagewise modeling architecture that sequentially adds latent factors to inherit factors from the latent factor model of the previous step, such as an observable factor model or PCA-based latent factor model. The model achieved state of the art performance and beats every other benchmark model, the model also achieved the highest Sharpe ratio of the tangency factor portfolio.

7.2 Analysis of RQ2

Most successful and applicable for stock price time series prediction models is LSTM and ARIMA.

Social media content contains a wealth of important information about the stock. The stock financial index variables can only represent the development trend of the stock price, but investor sentiment can describe the potential trend of the stock price, which is usually overlooked in traditional prediction methods. Ji, Wang, and Yan (2021) used deep learning technology to extract text features that can represent investor sentiment and greatly improve prediction performance by using an LSTM-based model.

As found in Mehta, Malhar, and Shankarmani (2021), the ARIMA model performs best when forecasting short-term results, while the LSTM model performs better when forecasting long-term results. Because the user must run the model daily for each stock in the stock market, we need short term prediction to be more accurate. For this scenario, we do not need long term prediction to be more accurate, so the ARIMA model provides the best RMSE score.

The information presented in the collected dataset is in raw form. To remove attribute bias in model development, the data should be cleaned for empty observations and scaled. Each neural network approach employed in Al Aradi and Hewahi (2020), had a unique treatment and selection criterion as in other works.

7.3 Analysis of RQ3

Haq et al. (2021) described an accuracy amount of 59.44 for their STOCKNET + MFSS method, outperforming various methods in caparison, like RAND and ARIMA models.

Li and Mei (2020) displayed a supreme performance of neural based methods. Showing a result of $2.9(10^-5)$ R^2 value, in comparison to statistical methods.

Qiu et al. (2020) displayed, an RNN based model, returned a MSE of 0.2592, displaying better peroformance.

Shen et al. (2021) used an LSTM network for effective multinational trade forecasting. It had smaller values of root mean square error (RMSE) and mean absolute percentage error (MAPE) than did time-series models and economic structural models. On the export forecast, RMSE improved by 17.048%

As we can see, over the years, even on completely different datasets and data representation, Neural Network models, showed improved performance over the years.

7.4 Synoptic Summary

Table 9 presents as summary of our results to our research questions tackled in our study.

We identified two main approaches that are used for asset evolution. The first one being, the standard model of using Time Series stocks, and predict their prices in a given window as shown in Dhafer et al. (2021). The other being that they mine text/new articles as displayed in Gidófalvi (2001) and Dhafer et al. (2021), which used BP Neural Networks based on public opinion. Typically for the second method, old ML methods like Naive Bayes are being used, or small parameter sized neural network models.

From the methods mentioned in Section 6.3, we have identified a few promising approaches. One being mentioned in Sarmah et al. (2022), which uses Learnt embedded Representation of Stock Correlation Matrix. We can further extract textual information uses State of the Art Models like FinBert and further embed such information in our graph structure.

8 Limitations, Threats to Validity and Review Assessment

The analysis of the potential weaknesses of a research work plays a fundamental role in encouraging readers to think more deeply about the topic as well as in appreciating possible future research directions. In this section, we thus adopt a critical stance on our findings. More specifically, we reflect on a few issues that may have influenced the objectivity, transparency, reproducibility, and academic soundness of our research.

Table 9: Synoptic summary Of Our Results

RQs	Summary
RQ1	A lot of methods are used in asset pricing/forecasting. The most dominant model is LSTM, followed by SVM, Random Forest, GRU, and other NN based approahces like attention Graph Neural Networks
RQ2	Most popular models for time series stock price forecasting is Long short-term memory (LSTM) and Autoregressive integrated moving average (ARIMA). The ARIMA model performs best when forecasting short-term results, while the LSTM model performs better when forecasting long-term results.
RQ3	Researchers used a list of metrics, with R^2 and MSE being the most common for most regression/time series prediction tasks, and Accuracy being the most coming for classification

8.1 Limitations

We start by analyzing circumstances and conditions that may limit the relevance and significance of our work. In this context, we might consider two aspects as potentially detrimental to our research:

First, one may see as problematic the fact that we searched only six databases. However, we note that the databases we searched are the most frequently used by researchers in computer science. In addition, one of such databases (Google Scholar) aggregates most of the contents available on other indexes. Thus, it is safe to assume that despite being run on a limited number of databases, our searches were quite comprehensive (more on this in Section 8.2 below) Second, one may object that one of our inclusion criteria (English requirement) limited our research capability. Cross-cultural issues have become especially important in modern science . We

are aware of these issues and recognize them as vitally important for a fair characterisation of scientific findings. However, given that most publications on the topic are written in English, and since leading journals in the field only accept submission in this language, we believe the restriction we adopted is neither unusual nor unjustified. In any case, the geographical distribution of our studies (Figure 3) demonstrates that there is sufficient cultural variability among the studies we included in our reading log.

8.2 Threats to Validity

Next, we address a series of possible biases that could cast doubt on the legitimacy or general validity of our work.. These are: *a)* retrieval bias, *b)* selector bias, *c)* expectation bias, and *d)* within-study bias..

Retrieval bias: may occur when the articles selected for inclusion are inadequate or when the searches performed by the authors are partial or incomplete. To avoid this bias, we conducted comprehensive searches on several different databases (six) and did not put any restriction on year of publication. In addition, we only selected high-quality articles, as shown in subsection analysis above in Section 4.2 and obviously included only works that matched our research topic.

Selector bias: occurs when the studies selected for inclusion in the reading log are affected by the authors' previous knowledge of the topic. In brief, when the findings are collected subjectively. To avoid the occurrence of this bias, we performed all the searches collaboratively. In other words, each member of the team was directly involved in them. In addition, to safeguard objectivity, team members were also asked to cross-check the activities of other members.

Expectation bias: can be considered as an extension of the *Selector bias* and typically occurs when data extraction and synthesis is severely constrained by the authors' conscious or unconscious expectations about the results. Our research team consists of academics, students, and industry practitioners. Furthermore, it comprises specialists in machine learning,

engineers, and AI scientists, thus it draws from various backgrounds and relies on multiple scientific approaches. Given the multidisciplinary character of our research team and its interdisciplinary stance, we believe we managed to reduce the occurrence of this bias to the minimum.

Within-study bias: develops as a result of a deficiency of data extraction templates. For this purpose, we designed a data extraction template (on Google sheet) and each team member actively participated in filling it.

9 Conclusion

The aim of this Systematic Literature Review was to investigate the usage of Machine Learning in Asset Evolution/Forecasting. This study attempted to answer the following research questions **RQ1**: What are the most effective ML models to predict evolution of an asset in the financial market? **RQ2**: What the existing approaches for forecasting stock price trend based on news/tweets or textual data? and **RQ3**: How accurately can machine learning models predict the stock price?

With regards to **RQ1**, we found a high representation of studies using LSTMs either as base models or standalone models, and showed to be highly effective. Concerning **RQ2**, we found LSTM and SVMs dominate the modelling part, which is usually preceded by embeddings. With respect to **RQ3**, we can state that effectiveness in identification is dependent on the data and on the algorithms used. Shen et al. (2021) showed a RMSE improvement of 17.048%, Qiu et al. (2020) returned an MSE of 0.2592.

Our results may also suggest that stock price prediction could be heavily improved using textual data and the appropriate embeddings. In any case, further work is to corroborate any of these speculations. We nevertheless hope that this systematic literature review will broaden interest in using hybrid sources for financial asset evolution using Machine Learning and will provide new grounds for more detailed explorations into these extraordinarily rich and fascinating set of phenomena.

References

Aasi, B., S. A. Imtiaz, H. A. Qadeer, M. Singarajah, and R. Kashef. 2021. Stock price prediction using a multivariate multistep lstm: A sentiment and public engagement analysis model. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. doi:10.1109/IEMTRONICS52119.2021.9422526.

Akyildirim, E., D. K. Nguyen, A. Sensoy, and M. Šikić. 2021. Forecasting high-frequency excess stock returns via data analytics and machine learning. *European Financial Management* .

Al Aradi, M., and N. Hewahi. 2020. Prediction of stock price and direction using neural networks: Datasets hybrid modeling approach. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*. doi: 10.1109/ICDABI51230.2020.9325697.

Alamsyah, A., M. T. Arasyi, and B. Rikumahu. 2018. Supporting investment decision using socio-economic issues exploration and stock price prediction. In *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*. doi:10.1109/SAIN.2018.8673343.

Alostad, H., and H. Davulcu. 2015. Directional prediction of stock prices using breaking news on twitter. In *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*. doi:10.1109/WI-IAT.2015.82.

Alsulmi, M. 2022. From ranking search results to managing investment portfolios: Exploring rank-based approaches for portfolio stock selection. *Electronics* 11:4019–.

Armano, G., M. Marchesi, and A. Murru. 2005. A hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences* 170:3–33. Computational Intelligence in Economics and Finance.

Asraf Roslan, M. A., and M. H. Fazalul Rahiman. 2018. Stock prediction using sentiment analysis in twitter for day trader. In *2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC)*. doi:10.1109/ICSGRC.2018.8657614.

Attanasio, G., L. Cagliero, P. Garza, and E. Baralis. 2019. Combining news sentiment and technical analysis to predict stock trend reversal. In *2019 International Conference on Data Mining Workshops (ICDMW)*. doi:10.1109/ICDMW.2019.00079.

Babaei, G., P. Giudici, and E. Raffinetti. 2022. Explainable artificial intelligence for crypto asset allocation. *Finance Research Letters* 47:102941–.

Barak, S., A. Arjmand, and S. Ortobelli. 2017. Fusion of multiple diverse predictors in stock market. *Information Fusion* 36:90–102.

Barra, S., S. M. Carta, A. Corriga, A. S. Podda, and D. R. Recupero. 2020. Deep learning and time series-to-image encoding for financial forecasting. *IEEE/CAA Journal of Automatica Sinica* 7:683–92.

Batra, R., and S. M. Daudpota. 2018. Integrating stocktwits with sentiment analysis for better prediction of stock price movement. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. doi:10.1109/ICOMET.2018.8346382.

Bollen, J., H. Mao, and X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2:1–8.

Broby, D. 2022. The use of predictive analytics in finance. *The Journal of Finance and Data Science* .

Bugayenko, Y., K. Daniakin, M. Farina, R. Ikramov, F. Jolha, Z. Kholmatova, A. Kruglov, W. Pedrycz, and G. Succi. 2022. Metrics for recommending corrective and preventive actions (capas) in software development projects: a systematic literature review. doi:10.21227/1zxb-q808.

Cakra, Y. E., and B. Distiawan Trisedya. 2015. Stock price prediction using linear regression based on sentiment analysis. In *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. doi:10.1109/ICACSIS.2015.7415179.

Carta, S., D. R. Recupero, R. Saia, and M. M. Stanciu. 2020. A general approach for risk controlled trading based on machine learning and statistical arbitrage. In *International Conference on Machine Learning, Optimization, and Data Science*, 489–503. Springer.

Carta, S. M., S. Consoli, A. S. Podda, D. R. Recupero, and M. M. Stanciu. 2021. Ensembling and dynamic asset selection for risk-controlled statistical arbitrage. *IEEE Access* 9:29942–59.

Chang, P.-C., C.-H. Liu, J.-L. Lin, C.-Y. Fan, and C. S. Ng. 2009. A neural network with a case based dynamic window for stock trading prediction. *Expert Systems with Applications* 36:6889–98.

Chao, C.-H., I.-H. Ting, T.-H. Tsai, and M.-C. Chen. 2019. Opinion mining and the visualization of stock selection in quantitative trading. In *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)*. doi:10.1109/TAAI48200. 2019.8959875.

Chen, B., J. Zhong, and Y. Chen. 2020. A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from shanghai stock exchange. *Expert Systems with Applications* 145:113104–.

Chen, L., M. Pelger, and J. Zhu. 2019. Deep learning in asset pricing. doi:10.48550/ARXIV. 1904.00745.

———. 2020. Internet appendix for deep learning in asset pricing. *Available at SSRN 3600206* .

Chen, Y.-S., C.-H. Cheng, and W.-L. Tsai. 2014. Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting. *Applied intelligence* 41:327–47.

Cheng, S.-H. 2010. Forecasting the change of intraday stock price by using text mining news of stock. In *2010 International Conference on Machine Learning and Cybernetics*. doi: 10.1109/ICMLC.2010.5580879.

Chua, C., M. Milosavljevic, and J. R. Curran. 2009. A sentiment detection engine for internet stock message boards. In *ALTA*.

Coelho, J., D. D'almeida, S. Coyne, N. Gilkerson, K. Mills, and P. Madiraju. 2019. Social media and forecasting stock price change. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*. doi:10.1109/COMPSAC.2019.10206.

Cong, L. W., G. Feng, J. He, and X. He. 2022. Asset pricing with panel tree under global split criteria. *Available at SSRN 3949463* .

Consoli, S., L. T. Pezzoli, and E. Tosetti. 2022. Neural forecasting of the italian sovereign bond market with economic news. *arXiv preprint arXiv:2203.07071* .

Coyne, S., P. Madiraju, and J. Coelho. 2017. Forecasting stock prices using social media analysis. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/ CyberSciTech)*. doi:10.1109/DASC-PICom-DataCom-CyberSciTec.2017.169.

Cuomo, S., F. Gatta, F. Giampaolo, C. Iorio, and F. Piccialli. 2022. An unsupervised learning framework for marketneutral portfolio. *Expert Systems with Applications* 192.

Dang, Q.-V. 2019. Reinforcement learning in stock trading. In *International conference on computer science, applied mathematics and applications*, 311–22. Springer.

Darapaneni, N., A. R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar, P. Bhagat, U. Aiyer, and Y. Agarwal. 2022. Stock price prediction using sentiment analysis and deep learning for indian markets. doi:10.48550/ARXIV.2204.05783.

Dash, R. 2020. Performance analysis of an evolutionary recurrent legendre polynomial neural network in application to forex prediction. *Journal of King Saud University-Computer and Information Sciences* 32:1000–11.

Day, M.-Y., and C.-C. Lee. 2016. Deep learning for financial sentiment analysis on finance news providers. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. doi:10.1109/ASONAM.2016.7752381.

Deng, Y., Q. Xie, and Y. Wang. 2018. Research on investor sentiment and stock market prediction based on weibo text. In *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, 1–5. doi:10.1109/SPAC46244.2018.8965607.

Devi, K., V. Bhaskaran, and G. Kumar. 2015. w optimized svm for stock market prediction. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. doi:10.1109/ICIIECS.2015.7192906.

Devitt, A., and K. Ahmad. 2007. Sentiment polarity identification in financial news: A cohesion-based approach.

Dhafer, A. H., F. M. Nor, W. Hashim, N. R. Shah, K. F. B. Khairi, and G. Alkawsi. 2021. A narx neural network model to predict one-day ahead movement of the stock market index of malaysia. In *2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 1–7. IEEE.

Dong, Y., D. Yan, A. I. Almudaifer, S. Yan, Z. Jiang, and Y. Zhou. 2020. Belt: A pipeline for stock price prediction using news. In *2020 IEEE International Conference on Big Data (Big Data)*. doi:10.1109/BigData50022.2020.9378345.

Duong, D., T. Nguyen, and M. Dang. 2016. Stock market prediction using financial news articles on ho chi minh stock exchange. New York, NY, USA: Association for Computing Machinery. doi:10.1145/2857546.2857619.

Durall, R. 2022. Asset allocation: From markowitz to deep reinforcement learning. doi:10.48550/ARXIV.2208.07158.

Enke, D., and S. Thawornwong. 2005. The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications* 29:927–40.

Fan, H. 2022. The digital asset value and currency supervision under deep learning and blockchain technology. *Journal of Computational and Applied Mathematics* 407:114061–.

Fang, M., and S. Taylor. 2021. A machine learning based asset pricing factor model comparison on anomaly portfolios. *Economics Letters* 204:109919–.

Fataliyev, K., A. Chivukula, M. Prasad, and W. Liu. 2021. Stock market analysis with text data: A review. doi:10.48550/ARXIV.2106.12985.

Fazli, M., P. Alian, A. Owfi, and E. Loghmani. 2021. Rps: Portfolio asset selection using graph based representation learning. doi:10.48550/ARXIV.2111.15634.

Felizardo, L. K., F. C. L. Paiva, A. H. R. Costa, and E. Del-Moral-Hernandez. 2022. Reinforcement learning applied to trading systems: A survey. *arXiv preprint arXiv:2212.06064*
.

Feng, G., J. He, and N. G. Polson. 2018. Deep learning for predicting asset returns.

Gao, Y., L. Zhou, Y. Zhang, C. Xing, Y. Sun, and X. Zhu. 2010. Sentiment classification for stock news. In *5th International Conference on Pervasive Computing and Applications*, 99–104. doi:10.1109/ICPCA.2010.5704082.

Gidófalvi, G. 2001. Using news articles to predict stock price movements. In *Semantic Scholar*.

Gu, S., B. Kelly, and D. Xiu. 2021. Autoencoder asset pricing models. *Journal of Econometrics* 222:429–50. Annals Issue:Financial Econometrics in the Age of the Digital Economy.

GUMUS, A., and C. O. SAKAR. 2021. Stock market prediction by combining stock price information and sentiment analysis. *International Journal of Advances in Engineering and Pure Sciences* 33:18–27.

Guo, Y. 2020. Stock price prediction based on lstm neural network: the effectiveness of news sentiment analysis. In *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*. doi:10.1109/ICEMME51517.2020.00206.

Guo, Y., and X. Chen. 2022. Forecasting the mid-price movements with high-frequency lob: A dual-stage temporal attention-based deep learning architecture. *Arabian Journal for Science and Engineering* 1–22.

Göçken, M., M. Özçalıcı, A. Boru, and A. T. Dosdoğru. 2016. Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Systems with Applications* 44:320–31.

Hagenau, M., M. Liebmann, M. Hedwig, and D. Neumann. 2012. Automated news reading: Stock price prediction based on financial news using context-specific features. In *2012 45th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2012.129.

Hagenau, M., M. Liebmann, and D. Neumann. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems* 55:685–97.

Haider, S. A., and R. Mehrotra. 2011. Corporate news classification and valence prediction: A supervised approach. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, 175–181.

Hajirahimi, Z., and M. Khashei. 2022. Hybridization of hybrid structures for time series forecasting: a review. *Artificial Intelligence Review* 1–61.

Haq, A. U., A. Zeb, Z. Lei, and D. Zhang. 2021. Forecasting daily stock trend using multi-filter feature selection and deep learning. *Expert Systems with Applications* 168:114444–.

He, H., S. Gao, T. Jin, S. Sato, and X. Zhang. 2021. A seasonal-trend decomposition-based dendritic neuron model for financial time series prediction. *Applied Soft Computing* 108:107488–.

Henrique, B. M., V. A. Sobreiro, and H. Kimura. 2019. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications* 124:226–51.

Herrmann, H., and B. Masawi. 2022. Three and a half decades of artificial intelligence in banking, financial services, and insurance: A systematic evolutionary review. *Strategic Change* .

Hua, Y. 2020. Bitcoin price prediction using arima and lstm. In *E3S Web of Conferences*, vol. 218, 01050. EDP Sciences.

Huang, C.-L., and C.-Y. Tsai. 2009. A hybrid sofm-svr with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications* 36:1529–39.

Huang, S.-C., C.-C. Chiou, J.-T. Chiang, and C.-F. Wu. 2020. A novel intelligent option price forecasting and trading system by multiple kernel adaptive filters. *Journal of Computational and Applied Mathematics* 369:112560–.

Huang, Y., Y. Gao, Y. Gan, and M. Ye. 2021. A new financial data forecasting model using genetic algorithm and long short-term memory network. *Neurocomputing* 425:207–18.

Islam, M. S., E. Hossain, A. Rahman, M. S. Hossain, and K. Andersson. 2020. A review on recent advancements in forex currency prediction. *Algorithms* 13:186–.

Islam, T., and M. I. Jabiullah. 2022. Empirical asset pricing via deep learning algorithms. In R. Asokan, D. P. Ruiz, Z. A. Baig, and S. Piramuthu, eds., *Smart Data Intelligence*, 109–19. Singapore: Springer Nature Singapore. ISBN 978-981-19-3311-0.

Ismail, M. S., M. S. Md Noorani, M. Ismail, F. Abdul Razak, and M. A. Alias. 2020. Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from kuala lumpur stock exchange. *Applied Soft Computing* 93:106422–.

Itkin, A. 2019. Deep learning calibration of option pricing models: some pitfalls and solutions. doi:10.48550/ARXIV.1906.03507.

Jakubowski, J., P. Stanisz, S. Bobek, and G. J. Nalepa. 2022. Performance of explainable ai methods in asset failure prediction. In D. Groen, C. de Mulatier, M. Paszynski, V. V.

Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, eds., *Computational Science – ICCS 2022*, 472–85. Cham: Springer International Publishing. ISBN 978-3-031-08760-8.

Jariwala, G., H. Agarwal, and V. Jadhav. 2020. Sentimental analysis of news headlines for stock market. In *2020 IEEE International Conference for Innovation in Technology (INOCON)*. doi:10.1109/INOCON50539.2020.9298333.

Jessica, and R. S. Oetama. 2019. Sentiment analysis on official news accounts of twitter media in predicting facebook stock. In *2019 5th International Conference on New Media Studies (CONMEDIA)*. doi:10.1109/CONMEDIA46929.2019.8981836.

Ji, X., J. Wang, and Z. Yan. 2021. A stock price prediction method based on deep learning technology. *International Journal of Crowd Science* 5:55–72.

Jomthanachai, S., W. P. Wong, and K. W. Khaw. 2022. An application of machine learning regression to feature selection: a study of logistics performance and economic attribute. *Neural Computing and Applications* 1–25.

Kalra, S., and J. S. Prasad. 2019. Efficacy of news sentiment for stock market prediction. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. doi:10.1109/COMITCon.2019.8862265.

Kalyani, J., P. H. N. Bharathi, and P. R. Jyothi. 2016. Stock trend prediction using news sentiment analysis. doi:10.48550/ARXIV.1607.01958.

Kanavos, A., G. Vonitsanos, A. Mohasseb, and P. Mylonas. 2020. An entropy-based evaluation for sentiment analysis of stock market prices using twitter data. In *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*. doi:10.1109/SMAP49528.2020.9248440.

Karaca, Y., Y.-D. Zhang, and K. Muhammad. 2020. Characterizing complexity and self-similarity based on fractal and entropy analyses for stock market forecast modelling. *Expert Systems with Applications* 144:113098–.

42

Kaushal, A., and P. Chaudhary. 2017. News and events aware stock price forecasting technique. In *2017 International Conference on Big Data, IoT and Data Science (BID)*. doi:10.1109/BID.2017.8336565.

Kaya, M. I. Y., and M. E. Karsligil. 2010. Stock price prediction using financial news articles. In *2010 2nd IEEE International Conference on Information and Financial Engineering*. doi:10.1109/ICIFE.2010.5609404.

Kesavan, M., J. Karthiraman, R. T. Ebenezer, and S. Adhithyan. 2020. Stock market prediction with historical time series data and sentimental analysis of social media data. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. doi:10.1109/ICICCS48265.2020.9121121.

Khairi, T. W. A., R. M. Zaki, and W. A. Mahmood. 2019. Stock price prediction using technical, fundamental and news based approach. In *2019 2nd Scientific Conference of Computer Sciences (SCCS)*. doi:10.1109/SCCS.2019.8852599.

Khedmati, M., and P. Azin. 2020. An online portfolio selection algorithm using clustering approaches and considering transaction costs. *Expert Systems with Applications* 159:113546–.

Kim, J., J. Seo, M. Lee, and J. Seok. 2019. Stock price prediction through the sentimental analysis of news articles. In *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*. doi:10.1109/ICUFN.2019.8806182.

Kim, J.-M., D. H. Kim, and H. Jung. 2021. Applications of machine learning for corporate bond yield spread forecasting. *The North American Journal of Economics and Finance* 58:101540–.

Kogan, S., D. Levin, B. R. Routledge, J. S. Sagi, and N. A. Smith. 2009. Predicting risk from financial reports with regression. 272–80. Association for Computational Linguistics.

Kohzadi, N., M. S. Boyd, B. Kermanshahi, and I. Kaastra. 1996. A comparison of artificial neural network and time series models for forecasting commodity prices .

Koudjonou, K. M., and M. Rout. 2020. A stateless deep learning framework to predict net asset value. *Neural Computing and Applications* 32:1–19.

Koutlis, C., S. Papadopoulos, M. Schinas, and I. Kompatsiaris. 2020. Lavarnet: Neural network modeling of causal variable relationships for multivariate time series forecasting. *Applied Soft Computing* 96:106685–.

Kumar, D. 2020. Stock forecasting using natural language and recurrent network. In *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*. doi:10.1109/ICETCE48199.2020.9091732.

Kumar, S., R. Gupta, P. Kumar, and N. Aggarwal. 2021. A survey on artificial neural network based stock price prediction using various methods. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. doi:10.1109/ICICCS51141.2021. 9432329.

Lauren, P., and P. Watta. 2019. A conversational user interface for stock analysis. In *2019 IEEE International Conference on Big Data (Big Data)*. doi:10.1109/BigData47090.2019.9005635.

Lee, H., M. Surdeanu, B. MacCartney, and D. Jurafsky. 2014. On the importance of text analysis for stock price prediction. In *LREC*.

Li, C., and K. Liu. 2022. Path signature-based phase space reconstruction for stock trend prediction. *International Journal of Data Science and Analytics* 1–12.

Li, H., R. Jia, and X. Wan. 2022. Time series classification based on complex network. *Expert Systems with Applications* 194:116502–.

Li, J., H. Bu, and J. Wu. 2017. Sentiment-aware stock market prediction: A deep learning method. In *2017 International Conference on Service Systems and Service Management*. doi:10.1109/ICSSSM.2017.7996306.

Li, R., D. Fu, and Z. Zheng. 2017. An analysis of the correlation between internet public opinion and stock market. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*. doi:10.1109/ICISCE.2017.41.

Li, W., and F. Mei. 2020. Asset returns in deep learning methods: An empirical analysis on sse 50 and csi 300. *Research in International Business and Finance* 54:101291–.

Li Im, T., P. Wai San, C. Kim On, R. Alfred, and P. Anthony. 2013. Analysing market sentiment in financial news using lexical approach. In *2013 IEEE Conference on Open Systems (ICOS)*. doi:10.1109/ICOS.2013.6735064.

liang Chen, T., and F. yu Chen. 2016. An intelligent pattern recognition model for supporting investment decisions in stock market. *Information Sciences* 346-347:261–74.

Lien Minh, D., A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon. 2018. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access* .

Lima Paiva, F. C., L. K. Felizardo, R. A. d. C. Bianchi, and A. H. R. Costa. 2021. Intelligent trading systems: a sentiment-aware reinforcement learning approach. In *Proceedings of the Second ACM International Conference on AI in Finance*, 1–9.

Lin, Y., S. Liu, H. Yang, and H. Wu. 2021. Stock trend prediction using candlestick charting and ensemble machine learning techniques with a novelty feature engineering scheme. *IEEE Access* 9:101433–46.

Liu, T., and Z. Yu. 2022. The analysis of financial market risk based on machine learning and particle swarm optimization algorithm. *EURASIP Journal on Wireless Communications and Networking* 2022:1–17.

Liu, Y., X. Huang, A. An, and X. Yu. 2007. Arsa: A sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 607–614. Association for Computing Machinery. ISBN 9781595935977. doi:10.1145/1277741.1277845.

Liu, Y., and L. Zhang. 2020. Cryptocurrency valuation and machine learning. *SSRN Electronic Journal* .

Lopez-Pujalte, C., M. J. Tena-Mateos, and A. Muñoz-Cañavate. 2022. A technology watch/competitive intelligence–based decision-support system optimised with genetic algorithms. *Journal of Information Science* .

Ly, T. H., and K. Nguyen. 2020. Do words matter: Predicting ipo performance from prospectus sentiment. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*. doi:10.1109/ICSC.2020.00061.

Malawana, M., and R. M. K. T. Rathnayaka. 2020. The public sentiment analysis within big data distributed system for stock market prediction– a case study on colombo stock exchange. In *2020 5th International Conference on Information Technology Research (ICITR)*. doi: 10.1109/ICITR51448.2020.9310871.

Mankar, T., T. Hotchandani, M. Madhwani, A. Chidrawar, and C. Lifna. 2018. Stock market prediction based on social sentiments using machine learning. In *2018 International Conference on Smart City and Emerging Technology (ICSCET)*. doi:10.1109/ICSCET.2018. 8537242.

Manujakshi, B., M. G. Kabadi, and N. Naik. 2022. A hybrid stock price prediction model based on pre and deep neural network. *Data* 7:51–.

Mehta, Y., A. Malhar, and R. Shankarmani. 2021. Stock price prediction using machine learning and sentiment analysis. In *2021 2nd International Conference for Emerging Technology (INCET)*. doi:10.1109/INCET51464.2021.9456376.

Mendoza-Urdiales, R. A., J. A. Núñez-Mora, R. J. Santillán-Salgado, and H. Valencia-Herrera. 2022. Twitter sentiment analysis and influence on stock performance using transfer entropy and egarch methods. *Entropy* .

Merello, S., A. Picasso Ratto, Y. Ma, O. Luca, and E. Cambria. 2018. Investigating timing and impact of news on the stock market. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW).* doi:10.1109/ICDMW.2018.00191.

Mirete-Ferrer, P. M., A. Garcia-Garcia, J. S. Baixauli-Soler, and M. A. Prats. 2022. A review on machine learning for asset management. *Risks* 10.

Mishev, K., A. Gjorgjevikj, I. Vodenska, L. Chitkushev, W. Souma, and D. Trajanov. 2019. Forecasting corporate revenue by using deep-learning methodologies. In *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO).* doi: 10.1109/ICCAIRO47923.2019.00026.

Mittermayer, M.-A. 2004. Forecasting intraday stock price trends with text mining techniques. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the.* doi:10.1109/HICSS.2004.1265201.

Mohan, S., S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu. 2019. Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService).* doi:10.1109/ BigDataService.2019.00035.

Mustapa, F. H., and M. T. Ismail. 2019. Modelling and forecasting s&p 500 stock prices using hybrid arima-garch model. In *Journal of Physics: Conference Series*, vol. 1366, 012130. IOP Publishing.

Nagar, A., and M. Hahsler. 2012. Using text and data mining techniques to extract stock market sentiment from live news streams.

Nguyen, T. H., K. Shirai, and J. Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* .

Nishitha, S. N. T., S. Bano, G. G. Reddy, P. Arja, and G. L. Niharika. 2020. Stock price prognosticator using machine learning techniques. In *2020 4th International Conference on*

Electronics, Communication and Aerospace Technology (ICECA). doi:10.1109/ICECA49313. 2020.9297644.

Nivetha, R. Y., and C. Dhaya. 2017. Developing a prediction model for stock analysis. In *2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*. doi:10.1109/ICTACC.2017.11.

Okoli, C. 2015. A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems* 37.

Orte, F., J. M. M. McWilliams, M. J. Sánchez, and P. Solana. 2022. A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction .

Ozbayoglu, A. M., M. U. Gudelek, and O. B. Sezer. 2020. Deep learning for financial applications: A survey. *Applied Soft Computing* 93:106384–.

Padhi, D. K., N. Padhy, A. K. Bhoi, J. Shafi, and S. H. Yesuf. 2022. An intelligent fusion model with portfolio selection and machine learning for stock market prediction. *Computational Intelligence and Neuroscience* 2022.

Pagolu, V. S., K. N. Reddy, G. Panda, and B. Majhi. 2016. Sentiment analysis of twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. doi:10.1109/SCOPES. 2016.7955659.

Pang, B., and L. Lee. 2008. Opinion mining and sentiment analysis .

Parida, N., D. Mishra, K. Das, N. K. Rout, and G. Panda. 2021. On deep ensemble cnn–sae based novel agro-market price forecasting. *Evolutionary Intelligence* 14:851–62.

Peng, D. 2019. Analysis of investor sentiment and stock market volatility trend based on big data strategy. In *2019 International Conference on Robots & Intelligent System (ICRIS)*. doi:10.1109/ICRIS.2019.00077.

Petrozziello, A., L. Troiano, A. Serra, I. Jordanov, G. Storti, R. Tagliaferri, and M. La Rocca. 2022. Deep learning for volatility forecasting in asset management. *Soft Computing* 26:8553–74.

Peykani, P., F. S. Seyed Esmaeili, M. Mirmozaffari, A. Jabbarzadeh, and M. Khamechian. 2022. Input/output variables selection in data envelopment analysis: A shannon entropy approach. *Machine Learning and Knowledge Extraction* 4:688–99.

Pinelis, M., and D. Ruppert. 2022. Machine learning portfolio allocation. *The Journal of Finance and Data Science* 8:35–54.

Pooja, B. L., S. Kanakaraddi, and M. M. Raikar. 2018. Sentiment based stock market prediction. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*. doi:10.1109/CTEMS.2018.8769159.

Qi, Y., W. Yu, and Y. Deng. 2021. Stock prediction under covid-19 based on lstm. In *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. doi:10.1109/IPEC51340.2021.9421323.

Qin, Y., D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell. 2017. A dual-stage attention-based recurrent neural network for time series prediction .

Qiu, H., and F. Liu. 2019. Candlestick analysis in forecasting u.s. stock market: Are they informative and effective. In *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*. doi:10.1109/ICBDA.2019.8713248.

Qiu, Y., H.-Y. Yang, S. Lu, and W. Chen. 2020. A novel hybrid model based on recurrent neural networks for stock market timing. *Soft Computing* 24:15273–90.

Ray, P., B. Ganguli, and A. Chakrabarti. 2021. A hybrid approach of bayesian structural time series with lstm to identify the influence of news sentiment on short-term forecasting of stock price. *IEEE Transactions on Computational Social Systems* .

Reddy, N. N., E. Naresh, and V. Kumar B.P. 2020. Predicting stock price using sentimental analysis through twitter data. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. doi:10.1109/CONECCT50063. 2020.9198494.

Riabykh, A., D. Surzhko, M. Konovalikhin, and S. Koltcov. 2022. Sttm: an efficient approach to estimating news impact on stock movement direction. *PeerJ Computer Science* 8:e1156–.

Rødseth, H., R. J. Eleftheriadis, Z. Li, and J. Li. 2019. Smart maintenance in asset management–application with deep learning. In *International Workshop of Advanced Manufacturing and Automation*, 608–15. Springer.

Rosenberg, B., and W. McKibben. 1973. The prediction of systematic and specific risk in common stocks. *Journal of Financial and Quantitative Analysis* 8:317–333–.

Ruan, J., W. Wu, and J. Luo. 2020. Stock price prediction under anomalous circumstances. In *2020 IEEE International Conference on Big Data (Big Data)*. doi:10.1109/BigData50022. 2020.9378030.

Rubesam, A. 2022. Machine learning portfolios with equal risk contributions: Evidence from the brazilian market. *Emerging Markets Review* 51:100891–.

Sadeghi, A., A. Daneshvar, and M. M. Zaj. 2021. Combined ensemble multi-class svm and fuzzy nsga-ii for trend forecasting and trading in forex markets. *Expert Systems with Applications* 185:115566–.

Sadorsky, P. 2021. A random forests approach to predicting clean energy stock prices. *Journal of Risk and Financial Management* 14.

Saha, S., J. Gao, and R. Gerlach. 2021. Stock ranking prediction using list-wise approach and node embedding technique. *IEEE Access* 9:88981–96.

———. 2022. A survey of the application of graph-based approaches in stock market analysis and prediction. *International Journal of Data Science and Analytics* 1–15.

Sangsavate, S., S. Tanthanongsakkun, and S. Sinthupinyo. 2019. Stock market sentiment classification from fintech news. In *2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE)*. doi:10.1109/ICTKE47035.2019.8966841.

Sarmah, B., N. Nair, D. Mehta, and S. Pasquali. 2022. Learning embedded representation of the stock correlation matrix using graph machine learning. *arXiv preprint arXiv:2207.07183* .

Sarode, S., H. G. Tolani, P. Kak, and C. S. Lifna. 2019. Stock price prediction using machine learning techniques. In *2019 International Conference on Intelligent Sustainable Systems (ICISS)*. doi:10.1109/ISS1.2019.8907958.

Senhora, F. V., H. Chi, Y. Zhang, L. Mirabella, T. L. E. Tang, and G. H. Paulino. 2022. Machine learning for topology optimization: Physics-based learning through an independent training strategy. *Computer Methods in Applied Mechanics and Engineering* 398:115116–.

Shah, D., H. Isah, and F. Zulkernine. 2018. Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data (Big Data)*. doi: 10.1109/BigData.2018.8621884.

Sharaf, M., E. E.-D. Hemdan, A. El-Sayed, and N. A. El-Bahnasawy. 2022. A survey on recommendation systems for financial services. *Multimedia Tools and Applications* 81:16761–81.

Sharma, D. K., H. Hota, K. Brown, and R. Handa. 2022. Integration of genetic algorithm with artificial neural network for stock market forecasting. *International Journal of System Assurance Engineering and Management* 13:828–41.

Sharma, V., R. Khemnar, R. Kumari, and B. R. Mohan. 2019. Time series with sentiment analysis for stock price prediction. In *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*. doi:10.1109/ICCT46177.2019.8969060.

Shen, M.-L., C.-F. Lee, H.-H. Liu, P.-Y. Chang, and C.-H. Yang. 2021. Effective multinational trade forecasting using lstm recurrent neural network. *Expert Systems with Applications* 182:115199–.

Shih, Y.-C., S.-S. Chen, C.-F. Lee, and P.-J. Chen. 2014. The evolution of capital asset pricing models. *Review of Quantitative Finance and Accounting* 42:415–48.

Shynkevich, Y., T. McGinnity, S. Coleman, and A. Belatreche. 2015. Stock price prediction based on stock-specific and sub-industry-specific news articles. In *2015 International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2015.7280517.

Son, B., and J. Lee. 2022. Graph-based multi-factor asset pricing model. *Finance Research Letters* 44:102032–.

Song, C., W. Chen, L. Fu, and A. Arshad. 2020. Improving stock price prediction based on investing sentiments. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. doi:10.1109/ICCC51575.2020.9345303.

Sonkiya, P., V. Bajpai, and A. Bansal. 2021. Stock price prediction using bert and gan. doi:10.48550/ARXIV.2107.09055.

Sönksen, J. 2022. *Machine learning for asset pricing*, 337–66. Cham: Springer International Publishing. ISBN 978-3-031-15149-1. doi:10.1007/978-3-031-15149-1_10.

Suman, N., P. K. Gupta, and P. Sharma. 2017. Analysis of stock price flow based on social media sentiments. In *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*. doi:10.1109/ICNGCIS.2017.34.

Sun, T., J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang. 2017. Predicting stock price returns using microblog sentiment for chinese stock market. In *2017 3rd International Conference on Big Data Computing and Communications (BIGCOM)*. doi:10.1109/BIGCOM.2017.59.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* .

Tang, X., C. Yang, and J. Zhou. 2009. Stock price forecasting by combining news mining and time series analysis. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. doi:10.1109/WI-IAT.2009.48.

Thakkar, A., and K. Chaudhari. 2020. Crest: cross-reference to exchange-based stock trend prediction using long short-term memory. *Procedia Computer Science* 167:616–25.

———. 2021. A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications* 177:114800–.

———. 2022. Information fusion-based genetic algorithm with long short-term memory for stock price and trend prediction. *Applied Soft Computing* 128:109428–.

Thu, H. G. T., T. N. Thanh, and T. L. Quy. 2021. A neighborhood deep neural network model using sliding window for stock price prediction. In *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*. doi:10.1109/BigComp51126.2021.00022.

Tian, Z. 2020. Echo state network based on improved fruit fly optimization algorithm for chaotic time series prediction. *Journal of Ambient Intelligence and Humanized Computing* 1–20.

Tiwari, S., A. Bharadwaj, and S. Gupta. 2017. Stock price prediction using data analytics. In *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*. doi:10.1109/ICAC3.2017.8318783.

Tsai, M.-F., and C.-J. Wang. 2012. Post-modern portfolio theory for information retrieval. *Procedia Computer Science* .

Turney, P. D., and M. L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv preprint cs/0212012* .

Uddin, A., X. Tao, and D. Yu. 2021. Attention based dynamic graph learning framework for asset pricing. In *Proceedings of the 30th ACM International Conference on Information*

& *Knowledge Management*, CIKM '21, 1844–1853. New York, NY, USA: Association for Computing Machinery. ISBN 9781450384469. doi:10.1145/3459637.3482413.

Uddin, A., and D. Yu. 2020. Latent factor model for asset pricing. *Journal of Behavioral and Experimental Finance* 27:100353–.

Vadlamudi, S. 2017. Stock market prediction using machine learning: A systematic literature review. *American Journal of Trade and Policy* .

Wang, H., J. Wang, L. Cao, Y. Li, Q. Sun, and J. Wang. 2021. A stock closing price prediction model based on cnn-bislstm. *Complexity* 2021.

Wang, L., L. Zhao, et al. 2022. Digital economy meets artificial intelligence: Forecasting economic conditions based on big data analytics. *Mobile Information Systems* 2022.

Wang, Y., and Y. Wang. 2016. Using social media mining technology to assist in price prediction of stock market. In *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. doi:10.1109/ICBDA.2016.7509794.

Wang, Z., S.-B. Ho, and Z. Lin. 2018. Stock market prediction analysis by incorporating social and news opinion and sentiment. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. doi:10.1109/ICDMW.2018.00195.

Wei, G., W. Zhang, and L. Zhou. 2015. Stock trends prediction combining the public opinion analysis. In *2015 International Conference on Logistics, Informatics and Service Sciences (LISS)*. doi:10.1109/LISS.2015.7369692.

Wu, D., X. Wang, and S. Wu. 2022. Jointly modeling transfer learning of industrial chain information and deep learning for stock prediction. *Expert Systems with Applications* 191:116257– .

Wu, J.-L., C.-S. Yang, K.-H. Liu, and M.-T. Huang. 2019a. A deep learning model for dimensional valencearousal intensity prediction in stock market. In *2019 IEEE 10th International*

Conference on Awareness Science and Technology (iCAST). doi:10.1109/ICAwST.2019.8923244.

Wu, Q., Z. Zhang, A. Pizzoferrato, M. Cucuringu, and Z. Liu. 2019b. A deep learning framework for pricing financial instruments. *CoRR* .

Wuthrich, B., V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang. 1998. Daily stock market forecast from textual web data. doi:10.1109/ICSMC.1998.725072.

Xie, B., R. Passonneau, L. Wu, and G. Creamer. 2013. Semantic frames to predict stock price movement.

Xie, C., D. Rajan, and Q. Chai. 2021. An interpretable neural fuzzy hammerstein-wiener network for stock price prediction. *Information Sciences* 577:324–35.

Xu, Y., W. Lin, and Y. Hu. 2020. Stock trend prediction using historical data and financial online news. In *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*. doi:10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00227.

Yadav, A., C. K. Jha, and A. Sharan. 2020. Optimizing lstm for time series prediction in indian stock market. *Procedia Computer Science* International Conference on Computational Intelligence and Data Science.

Yang, L., Z. Zhang, S. Xiong, L. Wei, J. Ng, L. Xu, and R. Dong. 2018. Explainable text-driven neural network for stock prediction. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*. doi:10.1109/CCIS.2018.8691233.

Yin, L., N. Zhang, L. He, and W. Fang. 2016. A study of relationship between investor sentiment and stock price based on text mining. In *2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*. doi:10.1109/IIKI.2016.49.

Yu, M.-H., and J.-L. Wu. 2019. Ceam: A novel approach using cycle embeddings with attention mechanism for stock price prediction. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)*. doi:10.1109/BIGCOMP.2019.8679218.

Yu, Y., S. Wang, and L. Zhang. 2017. Stock price forecasting based on bp neural network model of network public opinion. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. doi:10.1109/ICIVC.2017.7984716.

Yun, H., G. Sim, and J. Seok. 2019. Stock prices prediction using the title of newspaper articles with korean natural language processing. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. doi:10.1109/ICAIIC.2019.8668996.

Yun, K. K., S. W. Yoon, and D. Won. 2021. Prediction of stock price direction using a hybrid ga-xgboost algorithm with a three-stage feature engineering process. *Expert Systems with Applications* 186:115716–.

Yıldırım, S., D. Jothimani, C. Kavaklıoğlu, and A. Başar. 2018. Classification of "hot news" for financial forecast using nlp techniques. In *2018 IEEE International Conference on Big Data (Big Data)*. doi:10.1109/BigData.2018.8621903.

Zeng, X., J. Cai, C. Liang, and C. Yuan. 2022. A hybrid model integrating long short-term memory with adaptive genetic algorithm based on individual ranking for stock index prediction. *Plos one* 17:e0272637–.

Zhang, C. 2022. Asset pricing and deep learning.

Zhang, M., J. Yang, M. Wan, X. Zhang, and J. Zhou. 2022. Predicting long-term stock movements with fused textual features of chinese research reports. *Expert Systems with Applications* 210:118312–.

Zhang, W., X. Gong, C. Wang, and X. Ye. 2021. Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting* 40:1479–500.

Zhang, W., C. Li, Y. Ye, W. Li, and E. W. Ngai. 2015. Dynamic business network analysis for correlated stock price movement prediction. *IEEE Intelligent Systems* .

Zhang, Z., S. Zohren, and S. Roberts. 2020. Deep learning for portfolio optimization. *The Journal of Financial Data Science* 2:8–20.

Zhu, L., H. Wu, and M. T. Wells. 2021. A News-based Machine Learning Model for Adaptive Asset Pricing. doi:10.48550/ARXIV.2106.07103.

Zolotareva, E. 2021. Applying convolutional neural networks for stock market trends identification. In *International Conference on Artificial Intelligence and Soft Computing*, 269–82. Springer.