Assessment of the Applicability of Large Language Models for Quantitative Stock Price Prediction

Frederic Voigt
Hamburg University of Applied Sciences
Hamburg, Germany
frederic.voigt@ventric.de

Kai von Luck
Hamburg University of Applied Sciences
Hamburg, Germany
kai.vonluck@haw-hamburg.de

Peer Stelldinger
Hamburg University of Applied Sciences
Hamburg, Germany
peer.stelldinger@haw-hamburg.de

ABSTRACT

In accordance with the findings presented in [34], this study examines the applicability of Machine Learning (ML) models and training strategies from the Natural Language Processing (NLP) domain in addressing time series problems, emphasizing the structural and operational aspects of these models and strategies. Recognizing the structural congruence within the data, we opt for Stock Price Prediction (SPP) as the designated domain to assess the transferability of NLP models and strategies. Building upon initial positive outcomes derived from quantitative SPP models in our ongoing research endeavors, we provide a rationale for exploring a range of additional methods and conducting subsequent research experiments. The outlined research aims to elucidate the efficacy of leveraging NLP models and techniques for addressing time series problems exemplified as SPP.

CCS CONCEPTS

• **Computing methodologies → Natural language processing**.

KEYWORDS

stock price prediction, quantitative analysis, stock embeddings, large language models, natural language processing, big data

ACM Reference Format:
Frederic Voigt, Kai von Luck, and Peer Stelldinger. 2024. Assessment of the Applicability of Large Language Models for Quantitative Stock Price Prediction. In *The PErvasive Technologies Related to Assistive Environments (PETRA) conference (PETRA '24), June 26–28, 2024, Crete, Greece.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3652037.3652047

1 INTRODUCTION

Few domains within the field of machine learning have undergone as extensive scrutiny as NLP. Large language models (LLMs), in particular, consistently demonstrate remarkable performance. Noteworthy is not only the substantial success of NLP in user applications but also the assimilation of its findings into various sub-disciplines of machine learning. This integration extends beyond the incorporation of individual models like Transformers to encompass the assimilation of comprehensive training strategies. A prevalent approach involves the pre-training of machine learning models on generalized tasks, followed by fine-tuning for specific tasks within their respective domains. A prime example is the widespread use of Transformers in image processing, as exemplified by the Vision Transformer (ViT) [13]. Similarly, the pervasive adoption of pre-trained image processing models is a common practice in the vision domain.

Engaging in systematic research on the integration of NLP techniques across various machine learning subdisciplines not only provides avenues for improving the performance of machine learning applications in the designated domain (e.g., in the vision domain) but also imparts valuable insights into models, training strategies, and application concepts [34].

The optimal target domain for adaptation is one that closely aligns with NLP tasks and text structures. Texts, inherently composed of sequences of words, word tokens, or individual characters, render many time series problems conducive to exploration and adaptation. A fundamental functionality inherent to LLMs involves forecasting future developments within a time series. This predictive capability, exemplified by predicting the continuation of a text or formulating responses to questions, relies on the analysis of the statistical characteristics of the preceding portion of the time series (i.e. the previously processed text).

Given this intrinsic property, we have chosen to focalize our attention on SPP and Stock Movement Prediction (SMP) as our targeted domain. The underlying concept of this research is to investigate the extent to which LLMs, proficient in generating new texts from input texts, can extrapolate and generate future stock data based on historic stock data. As delineated in [34], beyond the general compatibility, the structural characteristics of both NLP and stock data formats render them amenable to mutual conversion, thereby facilitating processing by identical models.

Beyond the direct transformation of stock price data into NLP-compatible inputs, there is also a promising avenue of research involving the direct utilization of textual inputs for stock price prediction. This expansive array of factors of text data includes information gleaned from newspaper articles, tweets, social media data, and business reports of individual companies. The incorporation of external factors to forecast stock prices is commonly known as "fundamental analysis"[35], contrasting with "quantitative analysis"[30], which seeks to predict future stock price movements based on historical data.

Both the models delineated in this discourse and the corresponding results are confined to quantitative methods. The principal

PETRA '24, June 26–28, 2024, Crete, Greece

ACM ISBN 979-8-4007-1760-4/24/06
https://doi.org/10.1145/3652037.3652047

justification for this limitation is grounded in the seamless adaptability of language models and NLP queries to quantitative methods. Strategies incorporating a second modality, particularly those involving fundamental analysis, can be perceived as adaptations of multimodal NLP approaches. Multimodal NLP models aspire to integrate NLP with another modality, exemplified by the fusion of vision (V+T). For future research endeavors, the exploration of models incorporating fundamental analysis is under consideration, reflecting an intention to broaden the scope beyond purely quantitative methods. As elucidated by Zou and Zhao et al. [44], the application of ML in the stock domain can be broadly categorized into four categories. In addition to SPP and SMP, there is also the option to focus on Portfolio Management or Trading Strategies. However, it is explicitly clarified that the latter two are not the primary focus of this research series, as they deviate from the time series character and the close association with NLP.

The underlying research question guiding this work is whether it is viable and meaningful to apply NLP techniques, particularly those derived from LLMs, to the domain of SPP. Preliminary evidence supporting the feasibility and significance of this approach is presented in the evaluation of models detailed in Section 4. While several researchers have employed Transformer models (commonly found in LLMs) for SPP, as documented in Section 2 and [34], to the best of our knowledge, no dedicated research focused on the systematic exploration of NLP adaptation in an entirely new problem domain exists, particularly within the domain of time series problems. Our decision to tackle this research aspect is bolstered by the absence of prior investigations specifically tailored to the context of SPP/SMP.

It is important to clarify that the primary objective of this research is not necessarily to formulate best performing SPP/SMP models; rather, it aims to scrutinize the transferability of NLP techniques in this unique context.

The main contributions of this work are as follows: We embark on a pioneering experiment aimed at adapting NLP strategies to tackle a different time series problem — stock price prediction. Our investigation yields promising initial results from quantitative models, suggesting the potential benefits of further exploring LLMs for this purpose. We demonstrate the applicability of adapted Word2Vec models, evaluate the performance of pre-trained stock transformers, and investigate transformers with a special focus on analyzing longer time series. Furthermore, we propose innovative approaches by directly adapting popular speech models for stock price prediction, treating stock trends as sentence-like structures.

2 RELATED WORK

The exploration of implementing SPP through machine learning ML methods has been extensively investigated in [34], particularly regarding models relevant to NLP techniques. Following this, a literature review conducted by Zou and Zhao et al. [44] has been published, offering a comprehensive overview of currently employed models and research approaches. The subsequent section builds upon the insights derived from this review. It is important to note that models already covered in [34] may not be explicitly listed in the following.

2.1 Fundamental Models

Several models, which adopt a fundamental approach, were not discussed in [34]. Even though this outline is constrained to quantitative approaches, these models are acknowledged for the sake of comprehensiveness. For instance, there are models trained using data derived from financial news, such as the TOPIX Finance Event Dictionary [3], BELT [12], the model developed by Li and Pan [23], or the model from Yang et al. [41]. Additionally, some models utilize data from social media, such as MFN from Wang et al. [37]. More unconventional approaches include HTML [40], which incorporates audio data from conference calls. A notable model in this realm is FinBERT [42], a pre-trained model specifically designed for financial applications. It leverages a vast text corpus comprising diverse financial texts to enhance its performance and understanding of financial language.

2.2 Quantitative Models

The following section explores research efforts relevant to the quantitative models introduced in [34] and further discussed in this research outline.

Stock2Vec. The Stock2Vec model presented in [34] was designed with the motivation to embed the low dimensional and scalar inputs representing stock data into high dimensional vectors. The significance of high-dimensional embeddings for stock data is addressed explicitly, as seen in works such as by Daiya and Lin [7], and implicitly by models that employ high dimensional vector representations for companies or stock data, a characteristic shared by numerous models discussed in the following.

One example is the Adv-ALSTM proposed by Feng et al. [16], where a feature mapping layer is employed to create a high-dimensional latent space. Another model, the DTML model by Yoo et al. [43], utilizes Self-Attention based Encoders to generate stock embeddings from temporal price feature vectors, aiming to learn correlations. The DTRSI model by Nguyen and Yoon [25] focuses on creating feature vectors that provide data on companies related to the currently relevant one. Additionally, the CLVSA model by Wang et al. [38] employs an Encoder-Decoder architecture to extract temporal and local features. Also noteworthy is the HATR model by Wang, Heyuan, and Li [36], which employs graphs to express relationships between companies.

In addition, there are a number of publications that also attempt to represent companies as high-dimensional vectors without processing them directly in a model. This approach proves beneficial for capturing relationships between companies, understanding market correlations, and facilitating applications such as portfolio optimization. The objective might be for example to minimize risks by investing in stocks that demonstrate as little correlation as possible.

One such model is the Stock Embeddings model proposed by Dolphin, Smyth, and Ruihai [9]. In this model, an attempt is made to predict companies with a similar return ($= x_i^t - x_i^{t-1}$). The rationale is that stocks with a similar return value are likely exposed to similar fluctuations, indicating a level of similarity. Other stock embedding approaches by Dolphin, Smyth, and Ruihai include utilizing case base reasoning [10] or adopting a multimodal approach based on newspaper articles mentioning multiple companies [11].

The stock embedding model proposed by Du and Tanaka-Ishii [14] employs an Attention mechanism, utilizing key and value vectors extracted from financial news headlines alongside query vectors derived from quantitative data for SMP. The calculation of query vectors involves utilizing stock embeddings specific to each company. The explicit objective of the model is to train and refine these stock embeddings.

Sarmah et al.'s [31] approach involves the adaptation of Word2Vec training algorithms to generate embeddings for companies. The model uses random walks of a specified length extracted from a pruned graph network, treating them as sentences or "sentence-like structures". In this graph network, individual nodes represent companies, and the connections between them capture correlations in terms of return values.

The Asset Embeddings Model, crafted by Gabaix et al. [17], is notably distinctive as, to the best of our knowledge, the singular model that intermittently approaches certain facets of financial issues as challenges within the realm of NLP. The rationale behind this model aligns with the notion that "just as documents organize words in NLP [...] investors in financial markets organize assets"[17]. In this context, investors, including holdings, mutual funds, and exchange-traded funds (ETFs), are treated as sentences. The position of a company, determined by factors such as market capitalization or the company's proportion in the investor's portfolio, is likened to the position of a word in a sentence. In a Word2Vec adaptation, company embeddings are passed to the model, tasking it with identifying companies with a similar position in the investor's portfolio. Additionally, the model incorporates the BERT architecture, adapting Masked-Language Modeling to predict a masked company. Concatenated company embeddings are employed as input in this process. This unique approach sets the Asset Embeddings Model apart in its exploration of NLP techniques for the financial domain.

Pre-trained Transformer. Other quantitative models, not covered in [34], that rely on Transformers or at least employ an Attention mechanism include the previously mentioned CLVSA, which utilizes convolutional Long Short-Term Memory Neural Network (LSTM) [20] models in conjunction with Attention mechanisms. The HATR model incorporates a "Self-Attention layer and stacked gated causal convolution layers"[36] while the DTML model utilizes attentive LSTMs along with the aforementioned Self-Attention based Encoders. As for the DTRSI model, it stands out as, to the best of our knowledge, the sole model implementing any form of pre-training on quantitative data. This involves pre-training an LSTM-based model using feature vectors, followed by fine-tuning it for SMP in a subsequent step. Notably, the terms "fine-tuning" and "pre-training" in this context do not denote separate tasks, as in the envisioned models, but rather involve pre-training with a broad selection of companies and fine-tuning with a limited subset.

The Asset Embedding model employs not only Transformers but also entire NLP models like BERT, along with adapted pre-training tasks from the NLP field. In contrast to other Transformer-based models including those outlined in [34], the asset data is not fed into the Transformer as a concatenation of T market snapshots for $|C|$ companies ($X \in \mathbb{R}^{|C| \times T}$ or $X \in \mathbb{R}^{(|C| \cdot H) \times T}$ (for stacked Stock2Vec inputs with dimension H)) but is instead flattened. This means that the input is represented by the matrix $\hat{A} \in \mathbb{R}^{H \times |C|}$. This representation is defined as

$$\hat{A}[j, i] = E[f(i)][j] \tag{1}$$

with $E \in \mathbb{R}^{|C| \times H}$ as the embedding matrix of the company and $f(.)$ as the mapping for the company ranked within the context of the investor to the position in the embedding matrix. An approach using a similar representation, not yet presented in the previous work [34], is discussed in Section 6.2.

Recurrent Neural Networks. Prominent Recurrent Neural Network (RNN) models not previously discussed in [34] encompass the model by Rather, Agarwal and Sastry [29] and Ma et al.'s News2vec model [24], which incorporates LSTM-based fundamental analysis.

In [34], the significance of a model facilitating hierarchical decomposition for SPP has already been emphasized, suggesting the use of Clockwork RNNs [22] and Multiple Timescale RNNs [19] for this purpose. The Self-Attention-based HATR model employs stacked gated causal convolution layers, which can be interpreted as a form of hierarchical processing.

3 RESEARCH APPROACH

The goal of the presented research is to explore the applicability of NLP techniques within the SPP sector. Consequently, the ensuing chapter is structured to delve into a detailed examination of the financial domain from this perspective in Section 3.2. Subsequently, Section 3.3 provides a more in-depth explanation of the available data.

3.1 Problem Formulation

The primary objective of the conducted research is to effectuate a transformation of models or problems into a distinct domain. If the stock price of a company c_i at time t is denoted as x_i^t, then market snapshots X^t can encompass the prices of all companies $c_i \in C$ at a specific time t. Concatenating T market snapshots as $X \in \mathbb{R}^{|C| \times T}$ encapsulates the stock performance of all $c_i \in C$ over a time interval T. In many NLP applications, it is customary to represent a sentence as a concatenation of indexed word tokens $(v^1, v^2, ..., v^N)$ with $v_i \in V \subseteq \mathbb{N}$. These tokens are embedded into a dimension $\hat{H}$ with the assistance of an embedding matrix $\hat{E} \in \mathbb{R}^{|V| \times \hat{H}}$. The elucidated processing of stock data implies its compatibility with a broad spectrum of NLP models. The positions in the NLP input text correspond to the time dimension t in the market snapshot ($T \equiv N$), with the embedding dimension of the NLP text representing the respective company in the stock market ($|C| \equiv \hat{H}$). The approaches outlined in [34] for Pre-trained hierarchical models, unsupervised neuron models, and advanced Recurrent Neural Networks are considered as potential future models that necessitate inputs in the form of $X \in \mathbb{R}^{|C| \times T}$.

3.2 Aims and Objectives in the Financial Domain

The research endeavors to unveil insights into whether the incorporation of NLP methods, specifically using LLMs, can significantly enhance the comprehension and predictive capabilities related to stock price movements. The objective is not solely to position these adapted NLP models as definitive guides in SPP or to assert their superiority; instead, the emphasis lies on conducting a comprehensive exploration to assess the feasibility of integrating these models and techniques into the SPP domain. Additionally, it seeks to address the question of whether an improved understanding of how NLP models operate can be derived from this integration, thereby contributing to advancements in the understanding and application of NLP in the financial forecasting landscape.

As SPP serves as an auxiliary problem in this context, it is prudent not to confine the investigation to a specific (profitable) class of investment types, markets, countries, exchanges, or time periods. The intention is to embrace a broad and inclusive approach, recognizing the diverse landscape of financial variables and conditions, thereby ensuring a more holistic evaluation in the varied realms of quantitative finance.

Consequently, the application of both predictive and generative models in SPP is warranted, as these model types are prevalent in NLP and therefore merit further exploration within the financial domain. A nuanced evaluation of the predictive capabilities of machine learning models across various market dynamics is ensured by this multifaceted approach, facilitating a more robust and comprehensive analysis of their effectiveness in different financial environments. In addition to the LLM usage emphasized in this overview, other NLP models and techniques are also under scrutiny, as detailed in [34].

In traditional SPP research, a notable challenge revolves around addressing unexpected events (Black Swan events [33]). While it is feasible to hedge against such Black Swan events, as exemplified in portfolio optimization or risk minimization practices [9], it is crucial to underscore that certain unexpected occurrences are, by definition, not predictable, especially within the domain of quantitative methods. The inherent unpredictability of such events can significantly impact stock markets, thereby influencing the accuracy of predictive models. As a result, this research recognizes the intrinsic limitation of not directly accounting for Black Swan events. Instead, the focus is on exploring how the market evolves post the occurrence of unexpected events, with the objective of predicting the subsequent effects and trajectories of stocks.

3.3 Data

In order to ensure least possible restrictiveness in the target domain, it is imperative to conduct an extensive series of sub-experiments. Leveraging our research access to the Alpha Vantage[1] database provides us with the means to enhance our investigations through the utilization of diverse datasets.

Models spanning various time intervals, ranging from interday data to minute-level resolution, are being developed. Our research is confined to data post January 1st, 2000, a limitation imposed by data accessibility constraints.

Furthermore, our analysis extends across diverse asset classes, including equities, commodities, indices, funds, currencies, and cryptocurrencies—all of which are accessible as historical data in our database. Depending on the year, we have successfully amassed data from over 7,000 to more than 10,000 distinct assets.

[1] https://www.alphavantage.co/

The research is strategically designed to encompass a broad spectrum of markets, ranging from smaller local exchanges to specific industry sectors, and extending to major indices such as the S&P 500. Additional information, such as industry classification, geographical location, annual turnover for each company, and more, is readily available and can be integrated into future models. In nearly all instances of stock market data access, stock prices are not provided directly but rather as ψ distinct features representing various aspects of prices within a given time interval. These features typically include the highest, lowest, opening, and closing prices for the interval, along with the associated trading volume. Furthermore, the incorporation of technical indicators—metrics derived from quantitative data—adds another layer of interest to our research, and this is made possible through our data access capabilities.

In addition to the option of independently gathering fundamental analysis data, a practice common in numerous other SPP projects, we have access to a comprehensive repository of fundamental data using the Alpha Vantage database. Nonetheless, at present, these data sets are not pertinent to the forthcoming exploration of multimodal NLP approaches.

4 QUANTITATIVE MODELS

The models following quantitative approach are briefly outlined below. The purpose of this overview is not to delve into their intricacies or assess their performance in comparison to other research endeavors. Instead, its sole intention is to underscore that the transference of NLP techniques to SPP holds promise and merits further exploration as a valuable research avenue.

4.1 Stock2Vec

In contrast to the methods discussed in Section 2.2, the present research endeavors to directly predict regression data instead of classifying companies. Our primary objective is not solely to attain a commendable intrinsic evaluation, as proposed by [31], characterized by representation quality independent of a specific task. The embedding models enumerated in Section 2.2 are primarily designed either to form clusters based on specific sectors or to articulate market correlations through their high-dimensional vector representations. Our overarching objective is to enhance extrinsic performance by leveraging embeddings for downstream tasks. Therefore the development of a more application-oriented implementation, grounded in the translation of the NLP problem into the SPP domain, is targeted.

In the realm of NLP, there are typically two explicit approaches for training embeddings. The skip-gram (SG) approach involves providing word token embeddings to a model to predict the surrounding context of size $2 \cdot s$, whereas continuous bag of words (CBOW) models utilize word context to predict the central word.

Following the mapping described in Section 3.1 ($T \equiv N$ and $|C| \equiv \hat{H}$) the embeddings are trained by the SG und CBOS (continuous bag of stocks) task.

The aim is to train an embedding matrix $E \in \mathbb{R}^{|C| \times H}$. A company c_i is therefore represented by an embedding vector $e_i \in \mathbb{R}^H$. To embed the information about the current price of a stock, the vectors are shifted in the vector space which is defined as

$$e_i^t = x_i^t \cdot E[i] \tag{2}$$

where x_i^t is an extracted price feature (e.g. the closing price) or the whole feature vector. The rest of this section is based on a single price feature for presentation reasons.

For the SG task

$$F_{\text{X-SG}}(x_i^t) = e_i^t \cdot W_{\text{X-SG}}^T , \tag{3}$$

with $W_{\text{X-SG}} \in \mathbb{R}^{(2 \cdot s) \times H}$ is calculated. The resulting $2 \cdot s$ dimensional vector is compared to the past and future stock prices $(x_i^{t-s}, x_i^{t-(s-1)}, ..., x_i^{t-1}, x_i^{t+1}, ..., x_i^{t+(s-1)}, x_i^{t+s})$. If whole price features are used ($X \in \mathbb{R}^{(|C| \cdot \psi) \times T}$), the calculation can be performed for each price feature and the results stacked as representing vectors. This approach will be further addressed in future publications.

For the CBOS task

$$F_{\text{X-CBOS}}(\{x_i^{t-s}, ..., x_i^{t+s}\} \setminus \{x_i^t\}) = \left(\sum_{j \in \{x_i^{t-s}, ..., x_i^{t+s}\} \setminus \{x_i^t\}} (E[g(j)] \cdot j) \right) \cdot W_{\text{X-CBOS}}^T \tag{4}$$

(with $W_{\text{X-CBOS}} \in \mathbb{R}^{1 \times H}$ and $g(x_i^t) = i$) is calculated trying to predict x_i^t.

An alternative approach involves capitalizing on the high-dimensional nature of stock price information data. This dimensionality is characterized by the fact that, in a market snapshot X^t, the stock prices of $|C|$ companies are concurrently represented. Consequently, the scalar x_i^t can be employed to predict the price of another company $c_{j \neq i}$, enabling contextualization of the embedding vector with respect to the market X^t, rather than solely for the individual time series element x_i. Approaches wherein the prices of other companies are predicted are denoted with "C", while methods aiming to predict the price for the same c_i are marked with "X".

The C-CBOS approach is expressed as

$$F_{\text{C-CBOS}}(\{x_1^t, x_2^t, ..., x_{|C|}^t\} \setminus \{x_i^t\}) = \left(\sum_{j \in \{x_1^t, x_2^t, ..., x_{|C|}^t\} \setminus \{x_i^t\}} (E[g(j)] \cdot j) \right) \cdot W_{\text{C-CBOS}}^T \tag{5}$$

with $W_{\text{C-CBOS}} \in \mathbb{R}^{1 \times H}$ using x_i^t as a prediction target. The C-SG adaption is defined as

$$F_{\text{C-SG}}(x_i^t) = e_i^t \cdot W_{\text{C-SG}}^T , \tag{6}$$

with $W_{\text{C-SG}} \in \mathbb{R}^{(|C|-1) \times H}$ trying to predict $\{x_1^t, x_2^t, ..., x_{|C|-1}^t, x_{|C|}^t\} \setminus \{x_i^t\}$.

By amalgamating predictions (in the CBOS models) or the associated loss functions, it becomes possible to create composite models (SG + CBOS and C+X). We have conducted a parameter grid search for all the implementations mentioned so far. This search was tested on three distinct features: the daily closing price (a common feature in the SPP domain), the daily trading volume, and a composite of all ψ features combined. The best results of this grid search are presented in Table 1. Specific details regarding the implementation will be made available in subsequent publications.

Model	CBOS	SG	SG + CBOS
Closing Price	1.638	6.312	4.897
Volume	0.583	3.701	2.364
All Features	0.014	0.051	0.058

Table 1: Denormalized mean-squared error (MSE) loss calculated on the test set for various Stock2Vec implementations. Due to normalization across all features in the "All Features" model, the loss values might appear lower. The models were trained on interday stock prices of S&P 500 companies since January 1st, 2000.

The Stock2Vec models can be employed for SPP /SMP within other quantitative models by establishing a mapping $x_i^t \leftarrow e_i^t$.

4.2 Stock Transformer

Pre-trained language models, especially those based on Transformers, currently hold a position of significant interest within the realms of ML and NLP. However, the popularity of pre-trained models extends beyond these domains and also encompasses areas such as computer vision. As indicated in [34], these are now being transferred for stock price predictions.

Long Range Transformers. Handling lengthy inputs presents challenges especially for Transformer models, given that classical Transformers exhibit a time complexity of $O(n^2 \cdot k)$ and a space complexity of $O(n^2 + n \cdot k)$ with respect to the input length n and the model dimension k [26]. Particularly for SPP, we assume that incorporating long trends is pivotal for achieving accurate predictions.

In addition to alternative methods, such as global and local Attention exemplified in models like Longformer [1] or Sparse-Attention [5], as well as various other approaches anticipated for future use, there are strategies involving the segmentation of input sequences and the integration of past input for the ongoing calculation.

Three Transformer-based models have been devised, utilizing a contextual framework to assimilate preceding inputs into the model. These models draw inspiration from classical recurrent neural networks, including Elman RNNs [15], Jordan RNNs [39], or LSTMs, in terms of how the input chunks are processed. The Jordan Network-based model is comparable to the Transformer-XL model [6], the Elman RNN-based model shares similarities with the Recurrent Memory Transformer model [2], and the LSTM-based model draws parallels to the Block-Recurrent Transformers [21].

The multi-layer Transformer, denoted as $F_T(.)$, takes the current chunk of X, represented as $\mathcal{X}^\kappa \in \mathbb{R}^{|C| \times \theta_\kappa}$, and the context $M^\kappa \in \mathbb{R}^{|C| \times \theta_M}$ at iteration step κ, where θ_κ denotes the chunk size and θ_M the context length. Additionally, for future examinations, it is possible to utilize (stacked) interval feature vectors ($\mathcal{X} \in \mathbb{R}^{(|C| \cdot \psi) \times \theta_\kappa}$) or (stacked and weighted) Stock2Vec embeddings ($\mathcal{X} \in \mathbb{R}^{(|C| \cdot H) \times \theta_\kappa}$).

The two tensors $\mathcal{X}^\kappa$, and M^κ can be either concatenated and processed as $F_T(M^\kappa \oplus \mathcal{X}^\kappa)$ (Merged-Attention) or the context can be employed for the keys and values of the Transformer Attention Layers (Cross-Attention).

The model is initially designed as a pure Encoder architecture primarily suitable for a predictive approach. However, it is intended to

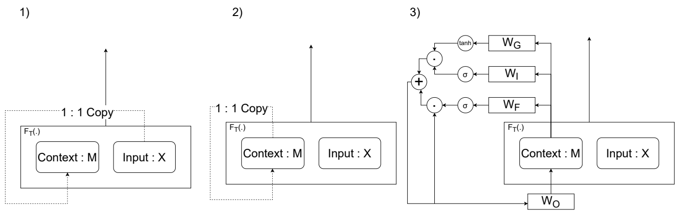

Figure 1: Sketch of the Transformer models. 1) Jordan-Network based model, 2) Elman RNN based model, 3) LSTM based model.

be employed for further investigations on Decoders for generative predictions.

The context for the current iteration κ is defined based on the model in consideration. For the Jordan-based model, it is articulated as $M^{\kappa+1} = F_T(M^\kappa, \mathcal{X}^\kappa)$ for Cross-Attention or $M^{\kappa+1}[i,j] = F_T(M^\kappa, \mathcal{X}^\kappa)[i,j]$ with $1 \le i \le |C|$ and $\theta_M + 1 \le j \le \theta_M + \theta_\kappa$ for Merged-Attention. In the case of the Elman RNN-based model, the context is defined as $M^{\kappa+1}[i,j] = F_T(M^\kappa, \mathcal{X}^\kappa)[i,j]$ with $1 \le j \le \theta_M$ and $1 \le i \le |C|$ [2]. Regarding the LSTM-based model, context related calculations are done both before and after processing through $F_T(.)$. Before entering $F_T(.)$, the context is calculated as $M^\kappa = \dot{M}^{\kappa-1} \cdot W_O^T + b_O$, and subsequently as

$$f^\kappa = \sigma(W_F^T \cdot \ddot{M}^\kappa + b_F)$$
$$i^\kappa = \sigma(W_I^T \cdot \ddot{M}^\kappa + b_I)$$
$$g^\kappa = \tanh(W_G^T \cdot \ddot{M}^\kappa + b_G)$$
$$\dot{M}^\kappa = f^\kappa \odot \dot{M}^{\kappa-1} + g^\kappa \odot i^\kappa$$

with $\ddot{M}^\kappa[i,j] = F_T(.)[i,j]$ with $1 \le i \le |C|$ and $1 \le j \le \theta_M$ for Merged-Attention or $\ddot{M}^\kappa = F_T(.)$ for Cross-Attention.

A visualization of all models is depicted in Figure 1 and a pseudocode for the algorithm can be seen in Algorithm 1. On an SPP task displayed in Section 4.2 ind Table 3, we show that the use of longer inputs seemingly has a positive effect on performance.

Pre-Trained Stock Transformer. Adapting two prevalent pre-training tasks from the NLP domain, we integrate them into the SPP and SMP domain. Masked Language Modeling (MLM), where a word in a sentence is masked, and the model predicts it, transforms into Masked Price Modeling (MPM). Here, a single x_i^t is masked, and the model predicts it. The Next Sentence Prediction (NSP) task, involving predicting whether a pair of sentences are consecutive, adapts into Trend-Matching (TM). In TM, two trends of stocks are separated by a separation token "[SEP]", and the model predicts whether they follow each other. Additional tasks, such as time step masking (masking a complete market snapshot X^t), company masking (masking of $\mathcal{X}[i,j]$ with $1 \le j \le \theta_\kappa$ for one c_i) , and complete section masking ($X[i,j]$ with $-s \le j \le s$ and $-l \le i \le l$), are also developed. In addition to masking tasks, the pre-training model can also undertake the task of directly predicting future stock prices, specifically forecasting $X^{t+\epsilon}$, denoted as Trend Prediction

[2]For the Elman RNN-based model, there is no Cross-Attention version, as it would be equivalent to the Jordan-based model.

Algorithm 1 Stock Transformer Algorithm

$M \leftarrow$ random_init$(|C|, \psi, \theta_M)$
for $i = 1$ to int$(T \div \theta_\kappa)$ **do**
  $\mathcal{X} \leftarrow X[:, i \times \theta_\kappa : (i+1) \times \theta_\kappa]$
  if merged_attention **then**
    $out \leftarrow$ MHSA(cat$(M, \mathcal{X})$, cat$(M, \mathcal{X})$, cat$(M, \mathcal{X})$)
    if jordan **then**
      $M \leftarrow out[:, \theta_M + 1 :]$
    else if elman **then**
      $M \leftarrow out[:, : \theta_M]$
    else if lstm **then**
      $M \leftarrow$ lstm_model$(out[:, : \theta_M])$
    end if
  else if cross_attention **then**
    $out \leftarrow$ MHSA$(M, M, \mathcal{X})$
    if jordan **then**
      $M \leftarrow out$
    else if lstm **then**
      $M \leftarrow$ lstm_model(out)
    end if
  end if
end for

Model	Jordan-based	RNN-based	LSTM-based
MPM	1.305e−4	2.071e−4	1.141e−4
TP ($\epsilon = 1$)	0.721	0.684	2.616
TP ($\epsilon = 50$)	0.721	0.680	2.615
TP ($\epsilon = 500$)	0284.79	1878.80	0017.00
TM	0.637	0.645	0.645

Table 2: Best test set results of the pre-training tasks per model. MPM and TP are reported as MSE, denormalized with closing price. For TM, the results are presented in terms of Cross Entropy.

(TP). These tasks are self-supervised, promoting better generalizability, a promising research direction mentioned by [44], and demonstrated by techniques like those presented by Hendrycks et al. [18]. All tasks can also be executed as SMP tasks or distribution modeling. Figure 2 provides a visualization of the pre-training tasks, including MPM, TM, Timestep Masking, Company Masking, and TP. For precise implementation details, function structures, and comprehensive examination results, we refer interested readers to upcoming publications.

Initial results from a smaller hyperparameter gridsearch are outlined in Table 2, where models are trained on S&P 500 data with one-minute resolution from January 1st, 2000, to August 2023. It is worth noting the challenge in predicting prices further into the future and the difficulty of the TM task underlined by the high entropy. The displayed results serve as initial insights, with acknowledgment that not all models might have been trained on the same (model specific optimal) parameters, rendering the results illustrative rather than fully representative.

Fine-Tuned Stock Transformer. Similar to NLP, pre-trained generalized Transformers can be effectively employed for valuable

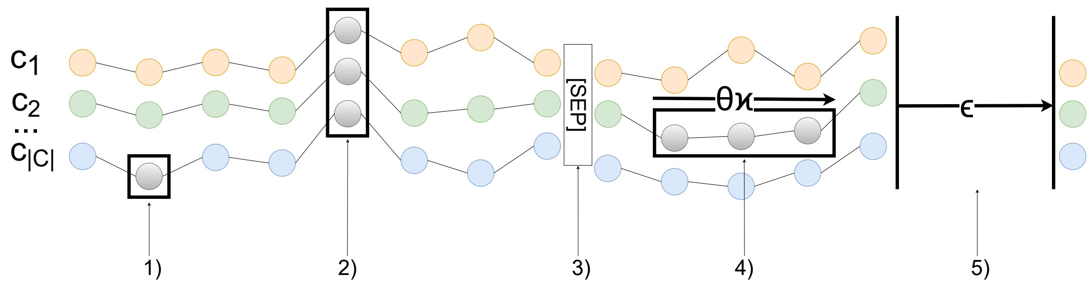

Figure 2: Visualization of the various pre-training tasks. 1) MPM, 2) Timestep Masking, 3) TM, 4) Company Masking, 5) TP.

Model	Jordan-based	RNN-based	LSTM-based
Scratch	0.863	0.549	2.342
Pre-Trained	0.733	0.530	2.111
Small T	0.899	0.599	2.566

Table 3: MSE test set results of Stock Transformers on one-minute resolution S&P 500 data for the $\epsilon = 1$-TP task on closing prices. "Scratch" denotes training the Transformer from scratch, while "Pre-Trained" signifies initializing the model with weights from a prior run. "Small T" means that the parameters $T = 128$ with $\theta_\kappa = 128$ were used, in contrast to $T = 512$ with $\theta_\kappa = 128$ for the other two runs. Values are denormalized based on the closing price.

tasks during fine-tuning in the financial domain. The initial fine-tuning task involves TP with $\epsilon = 1$. Subsequent experiments will explore additional fine-tuning tasks such as SMP and generative (Decoder-based) predictions of entire price trends. The preliminary fine-tuned results are presented in Table 3, demonstrating that, at least for these specific instances, pre-training yields a positive impact on performance compared to non-pre-trained results. As highlighted in Section 4.2, the exhibited results provide initial insights, acknowledging the potential divergence in (model specific optimal) training parameters among models and, consequently, their indicative rather than fully comprehensive nature. In Table 3 we also show that a larger period under consideration (in this case $T = 512$ with $\theta_\kappa = 128$) leads to better results than a smaller one ($T = 128$ with $\theta_\kappa = 128$) as already mentioned in Section 4.2. Initial evaluation utilizes MSE loss, with plans to expand assessment criteria in the future to include other metrics, such as investment simulations [14][38].

5 SUMMARY

One valuable insight derived from our exploration is the apparent suitability of models inspired by LLMs, as detailed in Section 4, for operating effectively in the SPP domain. Building upon this observation, further research will delve into the adaptability of NLP models.

Given the promising outcomes achieved by Transformer models, closely related to LLMs, it becomes advantageous to consider the direct utilization of language models, particularly LLMs, in future investigations. This application extends beyond the $X \in \mathbb{R}^{|C| \times T}$ format outlined in Section 4, encompassing a more detailed format expounded upon in Section 6.2.

Both the emphasis on Transformer models capable of processing substantial data volumes (a key focus in NLP research) and the utilization of pre-trained models demonstrate a positive impact on performance. Consequently, an exploration of NLP techniques is warranted, evaluating their potential applicability.

The Stock2Vec approach distinguishes itself from those presented in Section 2.2 in other publications. Unlike approaches directly targeting relationship representation or clustering, Stock2Vec focuses on embeddings designed to enhance usability in SPP. Particularly noteworthy is its promising performance, especially when predicting based on x_i^t to $x_{j \neq i}^t$.

6 FUTURE RESEARCH OUTLINE

Having demonstrated that the integration of NLP models into SPP is a potentially promising approach, this section delineates our methodology for future research. We aim to assess the efficacy of the adapted techniques within the financial domain and evaluate their performance comprehensively.

6.1 Refinement of the Quantitative Models

The models outlined in Section 4 will undergo further optimization in the future, involving a continued hyperparameter search and performance comparison against prominent SPP models. Ongoing tasks, such as SMP or distribution modeling, will also be targeted. Future iterations will explore the introduction of additional information, incorporating trainable embeddings and supplementary embeddings contingent on sector, industry, or country-specific factors. Plans include incorporating insights from the performance of entire markets, as exemplified in [25], to enhance predictions for individual company stock prices. The evaluation and expansion of additional pre-training tasks will remain a focus for future developments.

Stock2Vec models, although not directly utilized for SPP thus far, will be trained and evaluated in conjunction with Transformers. Their significance extends to the adapted language models detailed in Section 6.2. In forthcoming investigations, we will delve into the potential adoption of SMP in lieu of SPP. Additionally, the indexing and categorization of individual regression values, aiming for closer alignment with NLP, as well as exploring the prediction of value distributions, will be subjects of further discussion. A comprehensive intrinsic evaluation and exploration of practical applications in downstream tasks will form integral components of our future research agenda.

Addressing Black Swan events, as discussed in Section 3.2, can be improved without relying on multimodal data. For instance, the detection of extreme fluctuations (volatility) could trigger a partial reset of learned parameters when existing rules may no longer be applicable. Introducing a recognition mechanism as an emergency switch to express uncertainty about future predictions due to unforeseen events or to modify the learning process between such events, given their potential outlier status, is also under consideration.

From a technical standpoint, the consideration of numerous markets, companies, and time periods, as mentioned in Section 3.3, should be incorporated. Furthermore, exploration into other NLP models, beyond the scope of LLMs as presented in [34], is planned for further research.

6.2 Adapted Speech Models

In conventional NLP models, a tuple of word vectors $(v^1, v^2, \dots, v^T)$ is typically embedded, resulting in the concatenation of embeddings denoted as $\left[\hat{e}^1 \quad \hat{e}^2 \quad \dots \quad \hat{e}^T\right]$, which is then input into an NLP speech model $F_{\text{NLP}}(.)$.

This approach differs from the Transformer structures introduced in Section 4.2, particularly in the context that in NLP models, each point along the first dimension (temporal dimension/sentence position dimension) corresponds to one single information point (i.e. a word token). However, in the Stock Transformers presented in Section 4.2, there are $|C|$ information points (each representing a company's stock price), which becomes problematic, especially when employing Stock2Vec models. In NLP models, embeddings are concatenated, whereas in the Stock Transformer, they would need to be stacked to preserve the information delivered across the time axis t. This departure contradicts the processing mechanism of NLP models, where embeddings are concatenated, presenting a practical challenge in implementation as the input dimension expands to $(|C| \cdot H) \times T$, potentially reaching limits swiftly with a sufficiently large H.

Therefore, a new representation is highly desirable, aiming to bridge the gap with language models. This new representation should allow for effective utilization of well-established language models such as GPT-2 [28] or BERT [8] in a manner aligned with their design principles.

Additionally, this approach introduces an intriguing option regarding the extensibility of the model. Companies can enter or exit the stock market, especially over prolonged periods which means that C is expandable. This dynamic nature is evident, for instance, when start-ups go public or established companies face bankruptcy. For representations akin to those used in Section 4.2, particularly models not utilizing a Stock2Vec embedding, this poses a challenge due to their lack of extensibility. In such cases, H is directly derived from $|C|$ (or $|C| \cdot \psi$), and each c_i assumes a fixed position in the model.

For the approach adapting speech models, X^t is transformed into the structure of a sentence within an NLP problem by utilizing a flattened concatenation of market snapshots as "sentences". A flattened market snapshot $a \in \mathbb{R}^{H\times|C|}$ is defined as

$$a^t[j,i] = e_i^t[j] \,. \tag{7}$$

This representation, akin to the one used by Gabaix et al. [17], is novel in its application.

To represent the entire input X, the input $A \in \mathbb{R}^{H\times((T+1)\cdot|C|)}$ is defined as

$$A = \left[a^1 \quad [\text{PUNC}] \quad a^2 \quad [\text{PUNC}] \quad \dots \quad a^T\right] \,. \tag{8}$$

Here, the trainable $[\text{PUNC}] \in \mathbb{R}^{H\times 1}$ is employed to separate the market snapshots. In the context of most NLP models, a simple punctuation symbol such as " . " would typically be utilized to delimit sentences. The novel representation of X as A can now be incorporated not only into conventional predictive language models such as BERT but also into generative language models like GPT-2. The intriguing aspect of this prospect lies in training a (generative) language model to predict

$$B = \left[a^{T+1} \quad [\text{PUNC}] \quad a^{T+2} \quad [\text{PUNC}] \quad \dots \quad a^{T+\mathbb{T}}\right] \,. \tag{9}$$

This corresponds to training a generative text model with a target text. An alternative approach involves granting the model the flexibility to predict only what it is certain of, rather than compelling it to articulate stock prices (/movements) about which it lacks confidence. Or to put it more formally, $\hat{B}$ could be defined as

$$\hat{B} = \left[\hat{a}^{T+1} \dots \quad \hat{a}^{T+\mathbb{T}}\right] \tag{10}$$

with $\hat{a}^{T+1} \notin \mathbb{R}^{H\times|C|}$ e.g. $\hat{a}^{T+1} = \left[e_1^{T+1}, e_3^{T+1} \dots e_{|C|}^{T+1}\right]$. Two layers can be used to transform c_i and x_i^t from the output embeddings.

The approach described in this section is visualized together with the Stock Transformer from Section 4.2 and the classical language models in Figure 3 and a pseudo-code can be seen in Algorithm 2. Future experiments will also attempt to tokenize the float number inputs to have an NLP-like classification problem.

Algorithm 2 Stoks2Sentence

$sentence \leftarrow$ empty list
for t in T **do**
  $x_t \leftarrow X[:, t]$
  $x_t \leftarrow x_t^\top$ ▷ Transpose
  $indices \leftarrow \text{range}(1, |C|)$
  $embeddings \leftarrow E[indices]$ ▷ Stock2Vec Embeddings
  $x_{scaled} \leftarrow embeddings \cdot x_t$
  $sentence \leftarrow sentence + x_{scaled}$
  $sentence \leftarrow sentence + [\text{PUNC}]$
end for

6.3 Fundamental Models

In addition to the discussed quantitative models, the fundamental analysis approach to stock data holds particular promise. From an NLP perspective, this can be characterized as a multimodal approach. Within the SPP domain, textual data sources such as financial websites, annual reports, or social media entries would be leveraged, as elaborated in various models in [34] or Section 2.1.

While the further development and specification of this approach remain open for future investigation, several research directions

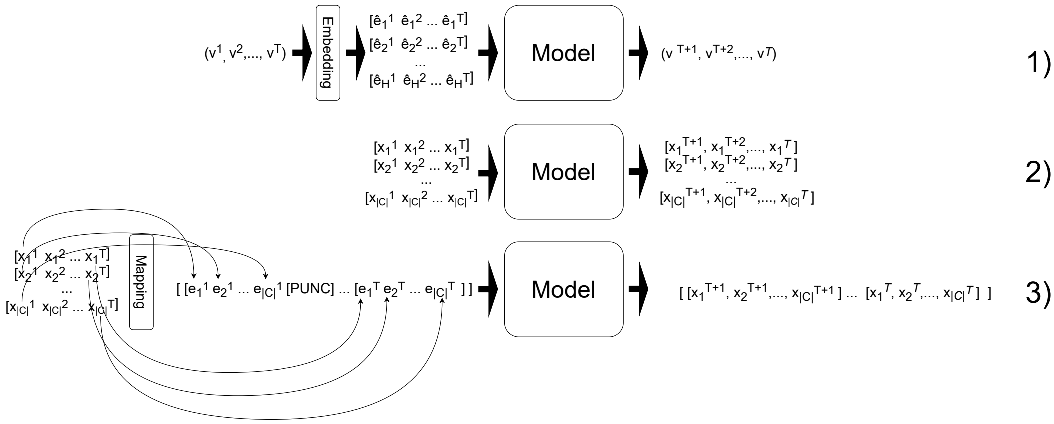

Figure 3: Outline of the processing approaches: 1) shows classical NLP models, 2) shows the approaches of most of the previous research in the SPP area and those presented in Section 4.2 and 3) the new approach to adapt the language models from Section 6.2.

are delineated. The refinement of Stock2Vec models, especially if proven beneficial for model performance, could be advanced further using textual data, as demonstrated in [11] or [14]. Models from the V+T domain, such as ViT [13], UNITER [4], or CLIP [27], can be adapted to extend the Transformer models presented in Section 4.2 to encompass additional modalities, including textual data.

Given the extensive body of literature exploring fundamental analysis approaches, this domain emerges as a highly promising avenue for future research and experimentation, particularly in the context of SPP/SMP performance. Notably, this perspective suggests a potential utility and explanatory support for models when confronted with unforeseen events. In essence, this aspect expands upon the assumption of market normalcy discussed in Section 3. Furthermore, the area proves intriguing and pertinent when considering the adaptability of multimodal NLP models, which serves as a focal point in the underlying research outline.

7 CONCLUSION

So far the exploration of SPP/SMP has predominantly adopted three different yet not necessarily distinct perspectives. Firstly, fundamental analysis, although not immediately pertinent in this context. Conversely, there exist quantitative Deep Learning models and other quantitative approaches such as statistical time series models like ARIMA and GARCH [32] (the latter often excluded from the classical ML/Deep Learning paradigm). The advantage of employing Deep Learning techniques lies in the capacity of networks to autonomously acquire multiple facets, requiring minimal assumptions and background knowledge. Consequently, numerous publications referenced in Section 2 have already underscored that utilizing ML for quantitative analysis is a promising avenue of research. We have demonstrated that LLM inspired architectures and NLP techniques can serve as wellsprings of inspiration for ML methods within the SPP domain, particularly under quantitative frameworks. Subsequent research is imperative to ascertain the extent of the efficacy of these techniques and to explore other potentially beneficial avenues, especially in relation to fundamental models. Within this paper, we have provided a preliminary framework for prospective investigations and offered a rationale for our conviction that research in this direction is promising.

ACKNOWLEDGMENTS

The data on which the quantitative models presented in this paper were trained was collected courtesy of research access kindly provided by Alpha Vantage[1].

REFERENCES

[1] Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. *CoRR* abs/2004.05150 (2020). arXiv:2004.05150 https://arxiv.org/abs/2004.05150

[2] Aydar Bulatov, Yuri Kuratov, and Mikhail S. Burtsev. 2023. Scaling Transformer to 1M tokens and beyond with RMT. *ArXiv* abs/2304.11062 (2023). https://api.semanticscholar.org/CorpusID:258291566

[3] Deli Chen, Yanyan Zou, Keiko Harimoto, Ruihan Bao, Xuancheng Ren, and Xu Sun. 2019. Incorporating Fine-grained Events in Stock Movement Prediction. In *Proceedings of the Second Workshop on Economics and Natural Language Processing*, Udo Hahn, Véronique Hoste, and Zhu Zhang (Eds.). Association for Computational Linguistics, Hong Kong, 31–40. https://doi.org/10.18653/v1/D19-5105

[4] Yen-Chun Chen, Linjie Li, Licheng Yu, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, Yu Cheng, and Jingjing Liu. 2019. UNITER: Learning UNiversal Image-TExt Representations. *CoRR* abs/1909.11740 (2019). arXiv:1909.11740 http://arxiv.org/abs/1909.11740

[5] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. 2019. Generating Long Sequences with Sparse Transformers. *CoRR* abs/1904.10509 (2019). arXiv:1904.10509 http://arxiv.org/abs/1904.10509

[6] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G. Carbonell, Quoc Viet Le, and Ruslan Salakhutdinov. 2019. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 2978–2988. https://doi.org/10.18653/V1/P19-1285

[7] Divyanshu Daiya and Che Lin. 2021. Stock Movement Prediction and Portfolio Management via Multimodal Learning with Transformer. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 3305–3309. https://doi.org/10.1109/ICASSP39728.2021.9414893

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/V1/N19-1423

[9] Rian Dolphin, Barry Smyth, and Ruihai Dong. 2022. Stock Embeddings: Learning Distributed Representations for Financial Assets. arXiv:2202.08968 [q-fin.ST]

[10] Rian Dolphin, Barry Smyth, and Ruihai Dong. 2023. Industry Classification Using a Novel Financial Time-Series Case Representation. arXiv:2305.00245 [cs.LG]

[11] Rian Dolphin, Barry Smyth, and Ruihai Dong. 2023. A Machine Learning Approach to Industry Classification in Financial Markets. In *Artificial Intelligence and Cognitive Science*, Luca Longo and Ruairi O'Reilly (Eds.). Springer Nature Switzerland, Cham, 81–94.

[12] Yingzhe Dong, Da Yan, Abdullateef Ibrahim Almudaifer, Sibo Yan, Zhe Jiang, and Yang Zhou. 2020. BELT: A Pipeline for Stock Price Prediction Using News. In *2020 IEEE International Conference on Big Data (Big Data)*. 1137–1146. https://doi.org/10.1109/BigData50022.2020.9378345

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy

[14] 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 the 58th Annual Meeting of the Association for Computational Linguistics*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (Eds.). Association for Computational Linguistics, Online, 3353–3363. https://doi.org/10.18653/v1/2020.acl-main.307

[15] J. L. Elman. 1990. Finding Structure In Time. *Cognitive Science* 14 (1990), 179–211.

[16] Fuli Feng, Huimin Chen, Xiangnan He, Ji Ding, Maosong Sun, and Tat-Seng Chua. 2018. Enhancing Stock Movement Prediction with Adversarial Training. In *International Joint Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:173991126

[17] Xavier Gabaix, Ralph Koijen, and Motohiro Yogo. 2023. Asset Embeddings. *SSRN Electronic Journal* (01 2023). https://doi.org/10.2139/ssrn.4507511 Available at SSRN: https://ssrn.com/abstract=4507511 or http://dx.doi.org/10.2139/ssrn.4507511 . Preliminary and incomplete as stated by the authors..

[18] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. 2019. Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/a2b15837edac15df90721968986f7f8e-Paper.pdf

[19] Wataru Hinoshita, Hiroaki Arie, Jun Tani, Hiroshi G. Okuno, and Tetsuya Ogata. 2011. Emergence of hierarchical structure mirroring linguistic composition in a recurrent neural network. *Neural Networks* 24, 4 (2011), 311–320. https://doi.org/10.1016/j.neunet.2010.12.006

[20] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-term Memory. *Neural computation* 9 (12 1997), 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

[21] DeLesley Hutchins, Imanol Schlag, Yuhuai Wu, Ethan Dyer, and Behnam Neyshabur. 2022. Block-Recurrent Transformers. In *Advances in Neural Information Processing Systems*, Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (Eds.). https://openreview.net/forum?id=uloenYmLCAo

[22] Jan Koutnik, Klaus Greff, Faustino Gomez, and Juergen Schmidhuber. 2014. A Clockwork RNN. In *Proceedings of the 31st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 32)*, Eric P. Xing and Tony Jebara (Eds.). PMLR, Bejing, China, 1863–1871. https://proceedings.mlr.press/v32/koutnik14.html

[23] Yang Li and Yi Pan. 2022. A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics* 13, 2 (01 Mar 2022), 139–149. https://doi.org/10.1007/s41060-021-00279-9

[24] Ye Ma, Lu Zong, Yikang Yang, and Jionglong Su. 2019. News2vec: News Network Embedding with Subnode Information. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Hong Kong, China, 4843–4852. https://doi.org/10.18653/v1/D19-1490

[25] Thi-Thu Nguyen and Seokhoon Yoon. 2019. A Novel Approach to Short-Term Stock Price Movement Prediction using Transfer Learning. *Applied Sciences* 9, 22 (2019). https://doi.org/10.3390/app9224745

[26] Bo Peng, Eric Alcaide, Quentin Anthony, Alon Albalak, Samuel Arcadinho, Stella Biderman, Huanqi Cao, Xin Cheng, Michael Chung, Leon Derczynski, Xingjian Du, Matteo Grella, Kranthi Gv, Xuzheng He, Haowen Hou, Przemyslaw Kazienko, Jan Kocon, Jiaming Kong, Bartłomiej Koptyra, Hayden Lau, Jiaju Lin, Krishna Sri Ipsit Mantri, Ferdinand Mom, Atsushi Saito, Guangyu Song, Xiangru Tang, Johan Wind, Stanisław Woźniak, Zhenyuan Zhang, Qinghua Zhou, Jian Zhu, and Rui-Jie Zhu. 2023. RWKV: Reinventing RNNs for the Transformer Era. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 14048–14077. https://doi.org/10.18653/v1/2023.findings-emnlp.936

[27] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning*. https://api.semanticscholar.org/CorpusID:231591445

[28] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. https://api.semanticscholar.org/CorpusID:160025533

[29] Akhter Mohiuddin Rather, Arun Agarwal, and V.N. Sastry. 2015. Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications* 42, 6 (2015), 3234–3241. https://doi.org/10.1016/j.eswa.2014.12.003

[30] Jerald Pinto David Runkle Richard DeFusco, Dennis McLeavey. [n. d.]. *Quantitative Investment Analysis*. . John Wiley and Sons. (Cited on pages 1 and 3)..

[31] Bhaskarjit Sarmah, Nayana Nair, Dhagash Mehta, and Stefano Pasquali. 2022. Learning Embedded Representation of the Stock Correlation Matrix using Graph Machine Learning. arXiv:2207.07183 [q-fin.CP]

[32] Granville Tunnicliffe Wilson. 2016. Time Series Analysis: Forecasting and Control,5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *Journal of Time Series Analysis* 37 (03 2016), n/a–n/a. https://doi.org/10.1111/jtsa.12194

[33] Katarzyna Tworek, Agnieszka Bieńkowska, Liliana Hawrysz, and Jolanta Maj. 2023. The Model of Organizational Performance Based on Employees' Dynamic Capabilities–Verification During Crisis Caused by Black Swan Event. *IEEE Access* 11 (2023), 45039–45055. https://doi.org/10.1109/ACCESS.2023.3273608

[34] Frederic Voigt. 2023. Adapting Natural Language Processing Strategies for Stock Price Prediction. DC@KI2023: Proceedings of Doctoral Consortium at KI 2023. , 20–29 pages. https://doi.org/10.18420/ki2023-dc-03

[35] Ahmed. S. Wafi, Hassan Hassan, and Adel Mabrouk. 2015. Fundamental Analysis Models in Financial Markets – Review Study. *Procedia Economics and Finance* 30 (2015), 939–947. https://doi.org/10.1016/S2212-5671(15)01344-1 IISES 3rd and 4th Economics and Finance Conference.

[36] Heyuan Wang, Shun Li, Tengjiao Wang, and Jiayi Zheng. 2021. Hierarchical Adaptive Temporal-Relational Modeling for Stock Trend Prediction. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3691–3698. https://doi.org/10.24963/ijcai.2021/508 Main Track.

[37] Heyuan Wang, Tengjiao Wang, and Yi Li. 2020. Incorporating Expert-Based Investment Opinion Signals in Stock Prediction: A Deep Learning Framework. In *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:214262181

[38] Jia Wang, Tong Sun, Benyuan Liu, Yu Cao, and Hongwei Zhu. 2019. CLVSA: A Convolutional LSTM Based Variational Sequence-to-Sequence Model with Attention for Predicting Trends of Financial Markets. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-2019)*. International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2019/514

[39] Antoni Wysocki and Maciej Ławryńczuk. 2015. Jordan neural network for modelling and predictive control of dynamic systems. In *2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*. 145–150. https://doi.org/10.1109/MMAR.2015.7283862

[40] Linyi Yang, James Ng, Barry Smyth, and Ruihai Dong. 2020. HTML: Hierarchical Transformer-based Multi-task Learning for Volatility Prediction. https://doi.org/10.1145/3366423.3380128

[41] Linyi Yang, Zheng Zhang, Su Xiong, Lirui Wei, James Ng, Lina Xu, and Ruihai Dong. 2018. Explainable Text-Driven Neural Network for Stock Prediction. In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*. 441–445. https://doi.org/10.1109/CCIS.2018.8691233

[42] Yi Yang, Mark Christopher Siy Uy, and Allen Huang. 2020. FinBERT: A Pretrained Language Model for Financial Communications. *CoRR* abs/2006.08097 (2020). arXiv:2006.08097 https://arxiv.org/abs/2006.08097

[43] 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* (Virtual Event, Singapore) *(KDD '21)*. Association for Computing Machinery, New York, NY, USA, 2037–2045. https://doi.org/10.1145/3447548.3467297

[44] Jinan Zou, Qingying Zhao, Yang Jiao, Haiyao Cao, Yanxi Liu, Qingsen Yan, Ehsan Abbasnejad, Lingqiao Liu, and Javen Qinfeng Shi. 2023. Stock Market Prediction via Deep Learning Techniques: A Survey. arXiv:2212.12717 [q-fin.GN]