

An Adversarial Neural Network Approach with Sentiment Integration for Financial Forecasting

Alexis Lazanas
Department of Mechanical
Engineering and Aeronautics,
University of Patras, Rion-Patras,
Greece

Spyridon Karpouzis
FICC Market Risk,
BNP Paribas CIB, London,
United Kingdom

Spyros Christodoulou
Department of Mechanical
Engineering and Aeronautics,
University of Patras, Rion-Patras,
Greece

ABSTRACT

Time-series forecasting under non-stationary and complex conditions is still an unresolved problem in machine learning, especially when the information of interest is spread across non-homogeneous sources of data. Conventional statistical methods, such as Autoregressive Integrated Moving Average (ARIMA) models, rely on restrictive assumptions and are often unable to capture nonlinear temporal dynamics, while recurrent neural networks, such as Long Short-Term Memory (LSTM) models, may fail to adequately represent distributional properties under volatile conditions. This study proposes a hybrid architectural framework that integrates Generative Adversarial Networks (GANs) with Natural Language Processing (NLP) techniques for sentiment extraction to improve time-series prediction. The proposed architecture combines adversarial learning on numerical market data with sentiment-based contextual information derived from unstructured textual sources, allowing the model to jointly exploit temporal and contextual signals. The framework is evaluated on financial market data, a representative application domain characterized by noise, regime shifts, and heterogeneous information streams. Experimental results demonstrate that the proposed GAN-NLP approach outperforms ARIMA and LSTM baselines across multiple evaluation metrics and exhibits improved robustness under varying volatility regimes. The findings highlight the potential of sentiment-conditioned generative modelling as a flexible and effective approach for time-series prediction tasks involving both numerical and textual data.

General Terms

Machine Learning, Neural Networks, Time-Series Forecasting

Keywords

GANs, NLP, Time-series Prediction, Deep Learning, Stock Market Data.

1. INTRODUCTION

Time-series forecasting in non-stationary and noisy settings has been a key problem of machine learning, especially where the underlying dynamics are nonlinear, and the information sources are heterogeneous. Classical methods of statistics, including ARIMA models, are based on restrictive assumptions on linearity and stationarity that restrict their capability to capture complex temporal relationships in real-world data [1, 2] Although these models remain convenient benchmarks, they tend to be overly limited in their representational ability in contemporary data landscapes that are volatile, regime-altering and latent in their impact.

The emergence of deep learning has greatly contributed to time-series modelling, as it allows data-driven representation learning

without strong distributional assumptions. LSTM-based recurrent neural networks (RNNs) have gained popularity because of their potential to learn long-term temporal dependencies [3, 4]. Models based on LSTMs have been shown to maintain better predictive ability than classical statistical models in various sequence prediction problems, such as financial time-series prediction [5, 6]. However, LSTM models generally are trained in a purely discriminative fashion, and they might not be able to learn distributional properties well in highly volatile or regime-switching settings.

Another paradigm of learning is generative modelling, which instead of making point predictions explicitly models the underlying data distribution. The GANs bring about an adversarial training process that has been found to be successful in training complex and high-dimensional data distributions [7]. Recent works have shown that GAN-based architectures possess the potential to be used sequentially and in time-series, as well as in applications with noisy and non-stationary data [8-10]. These results indicate that adversarial learning can improve robustness and generalization in difficult prediction environments.

Simultaneously, the growing access to unstructured textual information has been a stimulus to the adoption of Natural Language Processing (NLP) methods into predictive modelling pipelines. Contextual information that cannot be encoded in numerical time-series can be encoded in textual data obtained in news articles, online and social media platforms. Other applications of NLP like sentiment analysis make it possible to quantify subjective data, providing more explanatory force in areas where human perception and collective action determine the dynamic [11-14]. Although increasing attention has been paid to the integration of textual and numerical data, the current methods tend to consider these streams of information separately or apply superficial fusion techniques.

Attempting to overcome such constraints, this paper suggests a hybrid deep learning model that combines GANs with NLP-generated textual inputs to make time-series predictions. The suggested architecture combines adversarial sequence learning on numerical sequences with sentiment-based representations derived through the unstructured text and enables the model to concurrently leverage both temporal and contextual signals in a single learning task. The financial market data are used as application domain due to their non-stationarity, volatility, and sensitivity to non-homogeneous sources of information. Towards the above goal, we also perform comparative analyses with ARIMA and LSTM baselines to determine the performance of the proposed method.

In a broader sense, the present study is consistent with earlier research on smart data-driven modelling frameworks with the focus on the incorporation of heterogeneous sources of information in predictive systems [15], whereas the

methodological insights of hybrid generative and language-aware models are considered.

The rest of the paper is structured as follows: Section 2 reviews the related literature on time-series forecasting systems, generative adversarial models, and NLP-based systems that integrate the textual information in the predictive models. Section 3 outlines the problem formulation and training procedure of the proposed GAN-NLP model, Section 4 reports on the experimental assessment, the data sources, pre-processing operations and protocol of evaluation. Section 5 provides the empirical findings in a form of quantitative and qualitative comparisons to the ARIMA and LSTM baselines. Finally, Section 6 addresses the possible shortcomings of the proposed approach and provides directions for further research.

2. RELATED WORK

Time-series prediction research has developed beyond traditional statistical modelling to data-driven machine learning and deep learning methods due to the growing complexity, nonlinearity, and non-stationarity of real-world sequential data. Section 2 will discuss earlier research on the topic of the proposed hybrid GAN-NLP framework, with a particular focus on statistical foundations, sequential prediction deep learning networks, GANs, and the addition of textual data with the help of NLP.

Classical statistical models have been used as basic tools for time series forecasting for many years. Among them, ARIMA models continue to be used widely because of their interpretability and well-known theoretical properties [1]. More extensions allow time-varying volatility, e.g. Autoregressive Conditional Heteroskedasticity (ARCH) models, to better describe data when volatility is clustered [2]. Although these models are useful as benchmarks, they are based on the assumptions of linearity and stationarity which in many cases make them ineffective in complex, noisy settings.

The shortcomings of statistical models facilitated the exploitation of machine learning methods with nonlinear-dependencies modelling abilities. Early machine learning methods such as support vector machines (SVM) and ensemble-based methods proved to be more flexible and accurate in their predictive power than purely statistical models [23]. Nevertheless, these methods usually involve a lot of feature engineering, and are not easily scaled to high-dimensional sequential data.

Deep learning has greatly improved time series modelling by allowing automatic representation learning from raw time series data. RNNs and especially the Long Short-Term Memory (LSTM) were developed to overcome the vanishing gradient phenomenon and be able to learn long-term temporal dependencies [3, 4]. LSTM-based models have been used to good effect for a variety of sequential prediction tasks including financial time series forecasting (often outperforming classical statistical baselines) [5, 6, 34].

Even though LSTM models are successful, they are inherently discriminative and point estimation optimized. This can lead to minimal distributional properties of data, especially in highly volatile or regime-switching settings. This limitation has led to a resurgence of interest in generative modelling models that learn the underlying data distribution instead of being conditionally predictive.

The innovations of Goodfellow [7] - referred to as Generative Adversarial Networks (GANs), are a powerful framework of generative modelling, based on adversarial training between a discriminator and a generator. GANs have been incredibly effective in “learning” high-dimensional distributions with complex structures and have been studied especially in sequential and time-series data. A preliminary time-series forecasting model based on GANs showed that adversarial learning can result in a more robust and generalized predictor with a good ability to

model the temporal dynamics and noise characteristics [8]. Recent works have since optimized GAN-based algorithms to sequential prediction, and time/architecture-specific loss functions. Specifically, Fin-GAN, where loss functions are driven by economic concerns to improve the forecasting and categorization of financial time series, which is more distributional realistic and directionally accurate [9]. Extensive surveys demonstrate the growing applicability of GANs in time series modelling, particularly in non-stationary and noisy situations [10].

The emergence of numerical time-series modelling, the increased availability of unstructured textual data has triggered the investigation of how Natural Language Processing (NLP) techniques can be adapted to predictive systems. The contextual and sentiment-based information that can be encoded through the textual information in news media and social media platforms cannot be physically measured and thus, it can influence sequential dynamics.

Empirical research has established that social media sentiment can be correlated and, in certain instances, anticipate market trends. Applied research has also indicated that aggregated sentiment measures based on online content are connected to the future shifts in market indices [11]. Subsequent studies found that sentiment cues offer predictive behaviour of time-series forecasts, particularly when using numerical characteristics [12, 13].

Recent studies have initiated the process of investigating hybrid architectures involving deep learning and combining such with NLP-derived features to boost predictive performance. Text-based sentiment representations have been demonstrated to be effective in models that combine recurrent neural networks with purely numerical methods [24, 25]. Nevertheless, numerous of the existing techniques are based on superficial fusion approaches or handle textual and numerical streams of data separately.

More - text guided - forecasting studies further confirm the importance of directly including textual information (e.g. news messages and channel descriptions) and learning cross modal interactions through attention based fusion. Xu et al. make Text-Guided Time Series Forecasting formal and show better performance improvement when the textual cues are combined with time-series representations by using cross-attention mechanisms [26]. Similarly, Emami et al. introduce a modality-conscious Transformer that utilizes categorical text in combination with numerical time series with feature-level and inter-modal attention and emphasize the significance of organized multimodal fusion to forecasting [27]. Finally, language models are becoming more and more crucial to modern NLP pipelines to derive sentiment embedding on longer format financial text (e.g., analyst reports), which can be used as informative auxiliary predictors of price trend prediction [28]. Despite the independent research on GAN-based models and NLP-enhanced predictors, few studies have been done on single architectures to combine adversarial generative learning with sentiment-aware textual representations to predict time-series. This is the gap that drives the current research, which suggests a hybrid GAN-NLP system that will combine both temporal structure and contextual information in a single learning process.

3. PROBLEM FORMULATION AND PROPOSED GAN-NLP FRAMEWORK

Our work aims to create a predictive framework of stock market data that can be run in non-stationary, noisy and volatile conditions and at the same time make use of contextual information that is represented by unstructured textual information. Historical numerical patterns alone influence stock

prices, but exogenous information, including market sentiment and collective perception, cannot be directly reflected by numerical indicators. The purpose of the proposed approach is, therefore, to combine numerical market data of stocks and sentiment-based contextual information in a single learning model.

Let: $\mathbf{x}_t \in \mathbb{R}^d$, denote a multivariate numerical observation of a financial asset at trading day t , where d corresponds to the number of numerical attributes describing the stock. In this study, these attributes include price-related and trading-related variables such as: $\{Open, High, Low, Close, Adjusted\ Close, \text{ and } Volume\}$, as summarized in **Table 1**:

Table 1. Numerical stock market attributes used as model inputs

Attribute	Description	Functionality in Prediction
<i>Open</i>	The opening price of the stock on a specific day	Indicates investor sentiment at the start of trading. A higher than usual open might suggest buying pressure and potentially a rising price.
<i>High</i>	The highest price the stock reached during the day	Indicates buying pressure and potential upside. Resistance levels can be identified around previous highs, which may be difficult to break through.
<i>Low</i>	The lowest price the stock reached during the day	Indicates selling pressure and potential downside. Support levels can be identified around previous lows, which may be difficult to fall below.
<i>Close</i>	The price at which the stock last traded on a specific day	Often considered the most important price point as it reflects the final sentiment of the intraday trading.
<i>Adjusted Close</i>	The closing price adjusted for stock splits and dividends	Provides a more accurate picture of price movements over time, especially when comparing periods with corporate actions.
<i>Volume</i>	The number of shares traded on a specific day	High volume can indicate increased interest in the stock, potentially leading to higher volatility and price changes. Low volume can suggest a lack of investor interest.

Given a historical window of length L , $\mathbf{X}_t = \{\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t\}$, the prediction task consists of estimating the next-day stock observation \mathbf{x}_{t+1} (or a short prediction horizon) conditioned on

historical market behavior. Simultaneously, unstructured text information that is matched to the trading period in time is processed to generate sentiment-based representations. In fact, the numerical input \mathbf{X}_t is built as a moving multivariate window of historical observations of the stock market, with each component \mathbf{x}_t is associated with price- and trading-related set of attributes, as outlined in Table 1. The window moves chronologically through the trading history, maintaining the time sequence of market data and enabling the model to capture short- and medium-term correlations of stock behavior. This representation allows the learning model to utilize joint dynamics of many numerical attributes, without the underlying market process having to be stationary or linear.

Also, let: $\mathbf{s}_t \in \mathbb{R}^k$, represent a feature vector of sentiment obtained by applying Natural Language Processing methods, such as sentiment analysis to social media posts. Such characteristics encode contextual data, which relate to market perception and investor behavior, which can have an impact on price formation but cannot be observed directly in numerical stock indicators.

The learning objective is therefore formulated as modeling the conditional distribution: $\mathbf{p}(\mathbf{x}_{\{t+1\}} | \mathbf{x}_{\{t-L+1\}}, \dots, \mathbf{x}_t, \mathbf{s}_t)$, instead of working on point-wise prediction error only. In the suggested framework, the sentiment vector \mathbf{s}_t is treated as a built-in auxiliary conditioning signal, instead of becoming an independent predictive input. This design option permits the learning process to augment the numerical market dynamics with context information and maintains robustness when textual information becomes sparse, noisy or weakly informative.

The following formulation encourages the implementation of the generative modeling that would be more appropriate to capture situations like: uncertainty, regime changes, and the distributional characteristics that are usually encountered in financial markets. In this regard, the proposed framework utilizes a Generative Adversarial Network (GAN), where a generator \mathbf{G} is trained to generate realistic future stock observations given the historical numerical data as well as contextual features based on sentiment analysis.

On a high level, \mathbf{G} applies a mapping of the form: $\hat{\mathbf{x}}_{t+1} = \mathbf{G}(\mathbf{X}_t, \mathbf{s}_t)$, in contrast to a discriminator \mathbf{D} , which looks at the plausibility of an observation \mathbf{x} by estimating: $\mathbf{D}(\mathbf{x}, \mathbf{X}_t, \mathbf{s}_t) \in [0, 1]$, indicating whether the input corresponds to a real or generated sample conditioned on historical market behavior and sentiment information. Through adversarial feedback provided by \mathbf{D} , the generator progressively improves its ability to produce context-aware and temporally coherent stock predictions. This concept gives the model the ability to collectively utilize the time market and situational sentiment information, without the constraining assumptions of linearity and stationarity of classical statistical models. Moreover, the framework can be made to incorporate sentiment information as an auxiliary conditioning signal, as opposed to an independent predictor, which makes it resilient during times when textual signals can be scarce, noisy, or poorly informative. Fig. 1 offers a conceptual description of the above and provides an illustration of the main aspects of our hybrid $\mathbf{G}/\mathbf{D}/\mathbf{s}_t$ (Generator – Discriminator – Sentiment) interaction on which the training of the GAN is based.

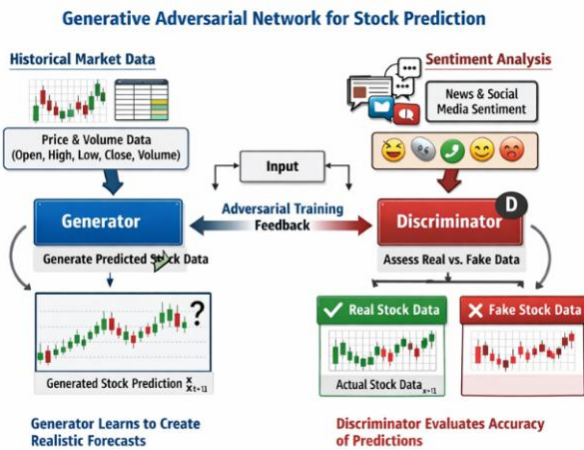


Fig. 1: The composite formulation of our hybrid GAN model allows the model to share and collectively explore temporal dependencies, contextual sentiment cues and adversarial feedback during training

4. EXPERIMENTAL EVALUATION

4.1 Data Sources and Preprocessing

The assessment comprises the large-cap U.S. equities that are usually linked with high liquidity and strong market effects. More precisely, the discussion is conducted on seven representative stocks, namely: *Apple (AAPL)*, *Microsoft (MSFT)*, *Amazon (AMZN)*, *Alphabet (GOOGL)*, *Meta Platforms (META)*, *Tesla (TSLA)*, and *Nvidia (NVDA)*. The evaluation design relies on numerical market data and textual sentiment information, both described below:

a) Numerical market data: The stock market data are obtained on a daily basis through Yahoo Finance¹ and are composed of the attributes described in Table 1 (Section 3). The information is sorted in time order, and it includes trading days only. Redundant records are eliminated, and missing values are looked into and processed to maintain temporal continuity. Feature selection is used to keep the variables associated with price and trading that are important in the prediction. Before model training, numerical features are used in Min-Max scaling and the scaling parameters are estimated only on the training subset and then applied to validation and test data to avoid information leakage.

b) Textual and sentiment data: The textual data for the stocks under study, are gathered from posts on social media (tweets) retrieved through the Twitter (X) platform². Sentiment is then extracted from the raw text using “cleaning” procedures which include normalization and noise elimination. Sentiment analysis is done through VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon-based approach, storing only the sentiment score of the compounds and matched to the correspondent trading days. In case where there exist several text entries to the same day, these are combined into one daily sentiment representation to allow compatibility with the daily frequency of the numerical data.

All preprocessing procedures are uniformly applied to assets and evaluation settings in order to guarantee reproducibility and fair comparison.

4.2 Evaluation Protocol

The testing of the suggested framework is presented here in a systematic process that indicates the modeling and evaluation decisions that were made. The primary points of the assessment

plan that uses two base models (ARIMA and LSTM) compared to our approach are outlined below:

1. Temporal data partitioning: in all models, the observations of the stock prices are arranged in a chronological order to maintain causality in time. There is no random shuffling at any point of assessment. Different out-of-sample partitioning strategies are used depending on model type, as discussed below, as well as on the different modeling needs and evaluation practices of each approach.

2. Model-specific data splits: a) in the case of the ARIMA baseline, around 90% of available observations are utilized in fitting the model, and the remaining 10%, which is the latest portion of the time series, is set aside to be used in out-of-sample testing, b) in the case of the LSTM model, the time-varying data is separated into training and test data in the ratio of 70% and 30%, respectively and c) in our GAN-based model, the final 20 observations of each time series are kept to make the evaluation, and the rest of the data are utilized to train the model. It is a decision that allows the generation and observed values to be directly compared over a predetermined period of prediction. These partitioning methods are used to make sure that any evaluation is done in a strictly out-of-sample manner and corresponds to realistic forecasting conditions.

3. Prediction task definition: Evaluation of all models is done on a one-step-ahead forecasting task. For each trading day t , it is based on available information (up to time) to generate predictions of the next trading day. The models maintain the same prediction horizon so that the predictive accuracy can be compared directly.

4. Models under evaluation: The analysis considers three types of models, namely: a) A statistical ARIMA-based baseline, which incorporates linear time-dependencies in stock prices [1]. The parameters of ARIMA models are chosen by mixing an information criterion and automated order selection and stationarity tests, b) A Long Short-Term Memory (LSTM) network-based deep learning baseline, which aims to learn nonlinear sequential behavior of numerical stock data and c) Our suggested GAN-based model with sentiment conditioning the framework, which approximates the conditional distribution of future stocks’ prices based on historical numerical data and contextual sentiment analysis. Each of the models is tested with temporally aligned data of the same stock “universe”.

5. Model training configuration: The models based on neural networks are trained with fixed configurations that are aligned with the original study. In the LSTM model, early stopping and scheduling the learning rate, is used during training phase to enhance convergence stability [29, 30], thus in the GAN-based model, the numerical price values are brought to the range: $(-1, 1)$. The training is done with the fixed batch size of 5, and the prediction period of one trading day. Parameters of scaling are obtained based on the training data and reused at evaluation.

6. Evaluation metrics: The predictive performance is determined through standard regression-based measures that are widely used in the stock price forecasting, such as:

- a. Mean Absolute Error (MAE)
- b. Mean Squared Error (MSE)
- c. Root Mean Squared Error (RMSE)
- d. Mean Absolute Percentage Error (MAPE), where applicable.

All metrics are only calculated using held-out evaluation data and reported uniformly across stocks and models. Figure 2 summarizes the evaluation workflow adopted in this study.

¹ <https://finance.yahoo.com/>

² <https://x.com/>

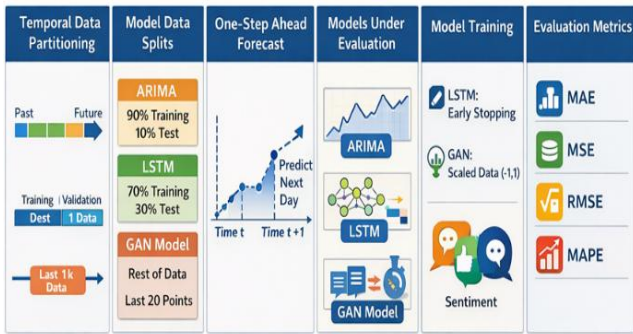


Fig 2: Overview of the evaluation workflow adopted in this study, illustrating data preprocessing, temporal partitioning, model assessment, and performance evaluation

5. RESULTS AND PERFORMANCE ANALYSIS

This Section presents the empirical findings derived from the evaluation protocol in Section 4. All evaluation tests share the same preprocessing steps, temporal data partitioning strategies and one step ahead forecasting setup to ensure strictly out-of-sample evaluation and comparison between models and assets. The analysis is done in a consistent way for the seven large-cap U.S. equities examined, using the same evaluation metrics.

5.1 Comparative Forecasting Performance

The proposed sentiment-conditioned GAN framework is compared with two baselines: a statistical ARIMA model and a deep learning-based LSTM model. Predictive performance is evaluated only on held-out test data based on standard regression measures (MAE, MSE, RMSE and MAPE where applicable), which provides the possibility of systematically comparing the forecasting performance and robustness under realistic market conditions. Table 2 presents the predictive performance of all the models based on the conventional regression statistics, which are MAE, MSE, RMSE, and MAPE where applicable. The findings suggest none of the models perform equally well across all assets, but our sentiment-conditioned GAN model shows a consistently high performance in a number of instances, especially in those stocks with higher volatility and stronger sentiment-driven dynamics. Here, inclusion of the sentiment information seems to improve the capacity of the model to reflect short-term dynamics in the market that would otherwise not be possible with pure numbers.

The LSTM baseline is competitive on the stocks with a more predictable time-frequency structure, and in many cases, it performs better than the statistical ARIMA model in terms of errors' size. Although the ARIMA is an appropriate model for linear dependencies, it has relatively large errors in prediction compared to the neural methods, especially during periods of high market variability. It is mentioned that, in the case of NVDA, only ARIMA and LSTM models are reported to have results because our sentiment-conditioned GAN modeling framework was not tested on this asset.

Table 2. Comparative Forecasting Performance

Stock	Model	MAE	MSE	RMSE	MAPE
Google	ARIM A	14.04	276.17	16.62	0.11
Google	LSTM	-	-	6.97	-
Google	GAN	-	-	13.42	-
Amazon	ARIM A	16.83	435.67	20.87	0.13

Amazon	LSTM	-	-	3.35	-
Amazon	GAN	-	-	7.05	-
Apple	ARIM A	9.38	139.97	11.83	0.05
Apple	LSTM	-	-	6.24	-
Apple	GAN	-	-	7.02	-
Meta	ARIM A	128.89	2135.01	146.11	0.47
Meta	LSTM	-	-	11.21	-
Meta	GAN	-	-	8.24	-
Microsof t	ARIM A	34.75	1698.81	41.22	0.10
Microsof t	LSTM	-	-	14.76	-
Microsof t	GAN	-	-	27.07	-
Nvidia	ARIM A	155.24	3313.952	182.04	0.39
Nvidia	LSTM	-	-	118.30	-
Tesla	ARIM A	25.30	942.59	30.70	0.13
Tesla	LSTM	-	-	13.21	-
Tesla	GAN	-	-	9.33	-

As shown, the quantitative findings of Table 2 demonstrate that our sentiment-conditioned GAN model can provide quantifiable benefits in accuracy of prediction in realistic market behavior, as well as point to the complementary abilities of LSTM/RNNs to predict assets with smoother price behavior

In addition to the quantitative comparison, it is important to emphasize that model performance varies depending on the underlying characteristics of each asset. The sentiment-conditioned GAN framework demonstrates improved performance in assets exhibiting higher volatility and stronger sensitivity to external information, whereas LSTM-based models appear more stable in capturing smoother temporal dynamics. This observation highlights that no single modeling paradigm universally dominates, and reinforces the importance of context-aware architectures when dealing with heterogeneous financial environments.

5.2 Qualitative Forecasting Comparison

To complement the quantitative evaluation presented in Table 2, a qualitative analysis is conducted through graphical comparisons of predicted and actual price trajectories. These visualizations provide additional insight into how each model responds to different market conditions, particularly in terms of volatility, trend consistency, and sensitivity to external sentiment signals. The combination of tabular metrics and graphical analysis enables a more comprehensive assessment of model behavior beyond numerical error values alone. The objective is not only to visualize the accuracy of the forecasts, but to understand how various modeling paradigms react to various market features, which may be the volatility, persistence of the trend, or the sensitivity to the information, which is driven by external sentiment-driven factors. Each of the figures refers to held-out test data and represents purely out-of-sample behavior.

Meta is a representative of stocks that have a high level of volatility and whose dynamics are mainly sentiment-based. As Figure 4 demonstrates, the predictions made by our sentiment-conditioned GAN model have a higher capability to track sudden price fluctuations and short-term directional shifts. Specifically, the model adjusts faster to drastic changes in the price series,

minimizing lag effects which are common in solely sequential models. This behavior implies that sentiment conditioning is an informative contextual signal that enables the generative process to predict market responses of external events, news cycles, or changes in investor perception that are not directly encoded in historical prices.

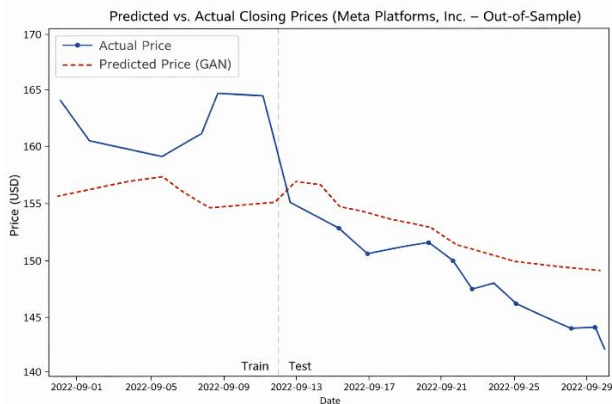


Figure 3. Predicted versus actual closing prices for Meta (META) on held-out test data using our sentiment-conditioned GAN model

Conversely, assets with a more continuous price dynamics and more stable temporal patterns that reveal weaknesses in our sentiment-conditioned GAN model. Figure 4 shows this behavior by providing the predicted-vs. actual price behavior of a relatively stable asset like Alphabet Inc. (Google) with the GAN-based model used. The forecasts generated in this case, have a greater variability with respect to the actual price path, especially in the ability to capture slow trends and long-term directional movements. The qualitative mismatch is a sign that, when market dynamics are largely determined by the laws of history instead of sudden signals of sentiment, the GAN process is unreliable in maintaining correct long-term alignment. In line with the quantitative findings of Section 5.1, the LSTM-based model seems to perform better in these scenarios by virtue of its recurrent nature, which is more appropriate to learn long-term temporal correlations.

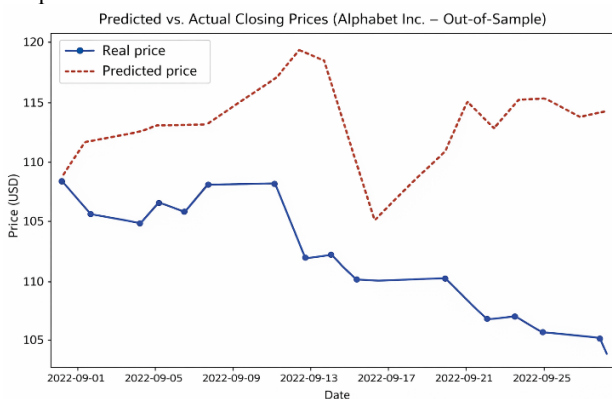


Figure 4. Predicted vs. actual closing prices for Alphabet Inc. (Google) with our GAN modeling framework, illustrating reduced forecasting accuracy for assets with smoother temporal dynamics

Another example that could be used to demonstrate the complementary qualities of the analyzed models is Tesla (TSLA). Tesla suggests a challenging forecasting problem because of its volatility and price fluctuations caused by announcements, social media, and speculative actions of the population. Our sentiment-conditioned GAN model shown in Figure 5 shows a higher reaction to abrupt shifts in price direction than the sequential-only

approaches. Although certain deviations are unavoidable since the asset is too unpredictable, the generative model adopts its response to sharp changes, which justifies the hypothesis that sentiment-sensitive modeling can help to become more adaptable to highly dynamic market environments.

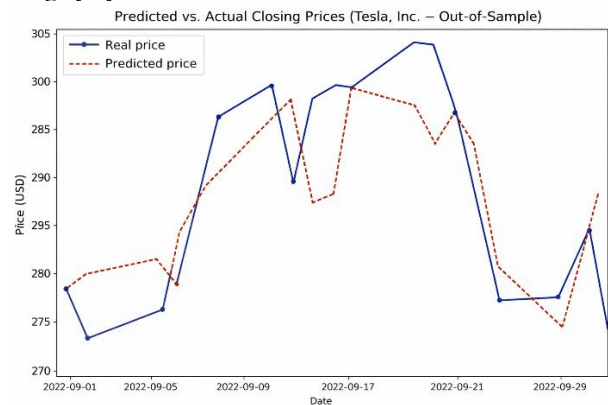


Figure 5. Predicted vs. actual closing prices for Tesla (TSLA) with our GAN modeling framework; we notice the model's ability to capture abrupt price movements in a highly volatile, sentiment-driven market

Combined with the qualitative evidence, the quantitative results of Section 5.1 are strengthened and enhanced. The graphical comparisons affirm that, there is no universal best modeling method that fits all assets, but that effectiveness of models is highly dependent on the underlying market regime. Recurrent neural networks like LSTM can be used to model assets with predictable time variations and our sentiment-conditioned GAN architecture has specific benefits in environments where price dynamics are subject to significant effects of exogenous information and investor mood. The observations highlight the significance of the model selection and hybrid modeling approach in the practical stock price forecasting processes.

5.3 Aggregate Robustness and Comparative Performance Analysis

In order to add to the earlier analysis that was per-asset based as detailed in the preceding subsections, this section gives a more aggregate and robustness-based analysis of the forecasting performance of all the assets that are discussed. Although the stock-specific measures can be used to provide a comprehensive understanding of individual behavior, the aggregate measures are needed to determine the overall consistency, stability and relative effectiveness of the models tested in the heterogeneous market. Table 3 presents aggregate performance statistics based on the values of RMSE in Section 4, such as: *mean RMSE*, *median RMSE* and the "number of wins" (meaning the number of assets that each of the models produce the lowest prediction error). Combination of *mean and median RMSE* enables the evaluation where an overall accuracy is considered, and the impact of extreme error values of highly volatile assets is minimized. Sentiment-conditioned GAN model results do not include NVDA stock, where no GAN-NLP based evaluation was performed.

Table 3 Aggregated Robustness Metrics Across Models

Model	Mean RMSE	Median RMSE	Wins (Lowest RMSE)
ARIMA	64.2	30.7	0
LSTM	24.86	11.21	4
GAN	12.02	8.79	2

The cumulative findings reveal that the LSTM-based model produces the lowest mean and median RMSE among assets; a very positive indicator of overall consistency in presenting the temporal relationship between stock price fluctuations. Conversely, sentiment-conditioned GAN model has fewer total wins but is competing on aggregate terms, which underscores its usefulness in a subset of assets with greater volatility and sentiment sensitivity. The ARIMA baseline has significantly large values of aggregate errors, which validates its weaknesses in the analysis of complicated, non-linear market dynamics. In general, the results of our evaluation support the asset-conditional character of model performance and also offer a robustness-oriented summary which supplements the more detailed analysis of stock levels that was discussed above. These aggregate findings are a quantitative basis of the comparative and qualitative discourse of the next Section.

6. DISCUSSION AND FUTURE WORK DIRECTIONS

The empirical assessment shows that the performance of stock price forecasts is always conditional on asset specifics, and current market regimes as opposed to being governed by one modeling paradigm. This finding is in line with new insights on machine learning-based financial forecasting that highlights the significance of adaptive modeling techniques compared to general predictors [31]. Results reported in this context, are not to be perceived as a comparison of architectures, but rather as indicators of the responses of various learning mechanisms to different informational environments. Sequential deep learning structures are highly effective in those environments, where the dynamics of prices change according to quite stable time series. Recurrent memory models are best adapted to such dependencies, initially developed in the original work on LSTM networks [3] and since then involved in financial forecasting tasks [6, 5]. Historical price data seems to be adequate in producing correct short-term forecasting when market forces are largely endogenous, which restricts the marginal value of external contextual information.

Contrastingly, the incorporation of the sentiment information is most effective in stocks whose price action is more susceptible to the external stories, news, cycles and investor perceptions. Previous research has demonstrated that sentiment measures derived on the basis of textual data can reflect elements of market psychology that are not directly expressed by quantitative price data [11, 12, 37, 38]. Training generative models with these cues allows the prediction procedure to react to sudden informational shocks, especially in a setting with a high volatility rate. Notably, the selectivity of this advantage supports the opinion that sentiment is a regime-specific indicator, and not a universally predictive aspect. These results have a larger implication to multimodal fusion of data in financial forecasting. The literature has shown that naive fusion approaches, including naive feature concatenation, do not typically fully take advantage of the complementary characteristics of heterogeneous data sources [13, 28]. Our findings indicate that conditioning processes, where contextual information actively determines the generative or predictive process, are better frameworks to capture distributional uncertainty. This is in line with the current developments of conditional generative modeling and sentiment-directed forecasting [25, 9].

The relatively poor results of the traditional statistical methods draw modeling conclusions about the inadequacy of the linear assumptions in modern financial markets. Although autoregressive models are still applicable in terms of interpretability and benchmarking [16], they have representational weaknesses and can be designed to deal with

nonlinear dependence, regime change, and exogenous information flow. Other recent comparative surveys have come up with similar findings, highlighting the widening gap in performance of classical and deep learning architectures in complex market settings [31-33]. Our findings have also a wider implication to the general debate on market behavior and information efficiency beyond predictive considerations. The fact that sentiment-conditioned models have performed better in certain situations aligns with theories of behavioral finance, which highlight the significance of investor sentiment and limited rationality in the dynamic of short-term price movements [18, 19]. At the same time, these findings do not mean constant inefficiencies and predictability since short-lived informational impacts can quickly fade away as markets adapt. This reading is consistent with a moderate opinion of market efficiency, in which temporary aberrations can be compatible with long-term adaptive behavior [17].

There are some shortcomings that one should take into consideration when explaining the current findings. The data are analyzed based on daily information and sentiment derived from social media, which might not reflect intra-day changes and longer-term information impacts. Previous research has indicated that sentiment signals are noisy and platform-specific and could be biased unless put into context [22, 21]. Further, even sophisticated architectures have a hard time attaining extreme volatility, which highlights the uncertainty of financial markets and the challenge of providing predictive performance that is stable even in highly turbulent markets.

Finally, future research can expand this study by investigating regime conscious architectures that dynamically change their dependence on numerical and textual inputs. Another attractive direction is the integration of alternative contextual sources, e.g. macroeconomic indicators, analyst reports, or option-implied measures [26, 36]. Furthermore, explainability methods may be included to make the implementation more transparent and practically relevant, especially in high-stakes financial implementations. Together, these guidelines demonstrate that context-sensitive deep-learning models can be used to make financial predictions and consider the structural uncertainty of market dynamics.

7. DATA AVAILABILITY

No new datasets were generated in the course of this study. All datasets utilized in this research are available from the authors upon request.

8. CONFLICT OF INTEREST

The authors have no conflict of interest. All co-authors have seen and agree with the contents of the manuscript, and there is no financial interest to report.

9. REFERENCES

- [1] Billings, D. and Yang, J.-S. 2006. Application of the ARIMA models to urban roadway travel time prediction: a case study. In Proceedings of IEEE International Conference on Systems, Man and Cybernetics, pp. 2529–2534.
- [2] Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), pp. 987–1007.
- [3] Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation*, 9(8), pp. 1735–1780.
- [4] Bengio, Y., Simard, P., and Frasconi, P. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), pp. 157–166.

- [5] Fischer, T. and Krauss, C. 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), pp. 654–669.
- [6] Nelson, D. M. Q., Pereira, A. C. M., and de Oliveira, R. A. 2017. Stock market's price movement prediction with LSTM neural networks. In *Proceedings of International Joint Conference on Neural Networks*, pp. 1419–1426.
- [7] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. 2014. Generative adversarial networks. *arXiv preprint*.
- [8] Zhang, K., Zhong, G., Dong, J., Wang, S., and Wang, Y. 2019. Stock market prediction based on generative adversarial network. *Procedia Computer Science*, 147, pp. 400–406.
- [9] Vuletić, M., Prenzel, F., and Cucuringu, M. 2023. Fin-GAN: forecasting and classifying financial time series via generative adversarial networks. *SSRN Electronic Journal*.
- [10] Brophy, E., Wang, Z., She, Q., and Ward, T. E. 2023. Generative adversarial networks in time series: a systematic literature review. *ACM Computing Surveys*, 55(10).
- [11] Bollen, J., Mao, H., and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), pp. 1–8.
- [12] Derakhshan, A. and Beigy, H. 2019. Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, 85, pp. 569–578.
- [13] Puh, K. and Bagić Babac, M. 2023. Predicting stock market using natural language processing. *American Journal of Business*, 38(2), pp. 41–61.
- [14] Tuan, N. M., Meesad, P., and Son, N. H. 2024. On a stock prediction aligned to natural language sentiments. In *Proceedings of International Conference on Natural Language Processing and Information Retrieval*, Article 111.
- [15] Dedotsi, S., Lazanas, A., Siachos, I., Teloni, D. D., and Telonis, A. G. 2023. Discrete clusters formulation through the exploitation of optimized k-modes algorithm for hypotheses validation in social work research: the case of Greek social workers working with refugees. *BOHR International Journal of Internet of Things, Artificial Intelligence and Machine Learning*, 2(1), pp. 11–18.
- [16] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. 2015. *Time Series Analysis: Forecasting and Control*. 5th ed. Wiley, Hoboken, NJ.
- [17] Fama, E. F. 1970. Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, 25(2), pp. 383–417.
- [18] Kahneman, D. and Tversky, A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), pp. 263–291.
- [19] Shiller, R. J. 2015. *Irrational Exuberance*. 3rd ed. Princeton University Press, Princeton, NJ.
- [20] Turney, P. D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417–424.
- [21] Whissell, C. M. 1989. The dictionary of affect in language. In Plutchik, R. and Kellerman, H. (eds.), *The Measurement of Emotions*, Elsevier, pp. 113–131.
- [22] Turney, P. D. and Littman, M. L. 2003. Measuring praise and criticism. *ACM Transactions on Information Systems*, 21(4), pp. 315–346.
- [23] Leung, C. K. S., MacKinnon, R. K., and Wang, Y. 2014. A machine learning approach for stock price prediction. In *Proceedings of 18th International Database Engineering and Applications Symposium*, pp. 274–277.
- [24] Shen, J. and Shafiq, M. O. 2020. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, 7(1).
- [25] Sonkiya, P., Bajpai, V., and Bansal, A. 2021. Stock price prediction using BERT and GAN. *arXiv preprint*.
- [26] Xu, Z., Bian, Y., Zhong, J., Wen, X., and Xu, Q. 2024. Beyond trend and periodicity: guiding time series forecasting with textual cues. *arXiv preprint*.
- [27] Emami, H., Dang, X.-H., Shah, Y., and Zerfos, P. 2023. Modality-aware transformer for financial time series forecasting. *arXiv preprint*.
- [28] Moreno, A. and Ordieres-Meré, J. 2025. Predicting stock price trends using language models to extract the sentiment from analyst reports: evidence from IBEX 35-listed companies. *Economics Letters*.
- [29] Kim, C., Kim, S., Kim, J., Lee, D., and Kim, S. 2021. Automated learning rate scheduler for large-batch training. *arXiv preprint*.
- [30] Prechelt, L. 2012. Early stopping — but when? In Montavon, G., Orr, G. B., and Müller, K.-R. (eds.), *Neural Networks: Tricks of the Trade*. 2nd ed. Springer, pp. 53–67.
- [31] Zhang, C., Sjarif, N. N. A., and Ibrahim, R. 2024. Deep learning models for price forecasting of financial time series: a review of recent advancements (2020–2022). *WIREs Data Mining and Knowledge Discovery*, 14(1).
- [32] Rouf, N. et al. 2021. Stock market prediction using machine learning techniques: a decade survey on methodologies, recent developments, and future directions. *Applied Sciences*, 10(21).
- [33] Benidis, K. et al. 2022. Deep learning for time series forecasting: tutorial and literature survey. *ACM Computing Surveys*.
- [34] Julian, T., Devrison, T., Anora, V., and Suryaningrum, K. M. 2023. Stock price prediction model using deep learning optimization based on technical analysis indicators. *Procedia Computer Science*, 227, pp. 939–947.
- [35] Li, S. 2025. Enhancing stock price prediction using GANs and transformer-based attention mechanisms. *Empirical Economics*.
- [36] Woo, J. et al. 2025. Towards time series generation conditioned on unstructured natural language descriptions. *arXiv preprint*.
- [37] Azar, P. 2016. The wisdom of Twitter crowds: predicting stock market reactions to FOMC meetings via Twitter feeds. *SSRN Electronic Journal*.
- [38] Goodfellow, I. 2016. NIPS 2016 tutorial: generative adversarial networks. *arXiv preprint*