The Predictive Power of Social Media Sentiment: Evidence from Cryptocurrencies and Stock Markets Using NLP and Stochastic ANNs

Giacomo di Tollo 1,*, Joseph Andria 2 and Gianni Filograsso 3

1 Dipartimento di Diritto, Economia, Management e Metodi Quantitativi, Università degli Studi del Sannio, 82100 Benevento, Italy
2 Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, 90100 Palermo, Italy; joseph.andria@unipa.it
3 Dipartimento di Economia, Università Ca’ Foscari di Venezia, 30121 Venezia, Italy; gianni.filograsso@unive.it

* Correspondence: giacomo.ditollo@unisannio.it; Tel.: +39-328-70-23-452

Abstract: Cryptocurrencies are nowadays seen as an investment opportunity, since they show some peculiar features, such as high volatility and diversification properties, that are triggering research interest into investigating their differences with traditional assets. In our paper, we address the problem of predictability of cryptocurrency and stock trends by using data from social online communities and platforms to assess their contribution in terms of predictive power. We extend recent developments in the field by exploiting a combination of stochastic neural networks (NNs), an extension of standard NNs, natural language processing (NLP) to extract sentiment from Twitter, and an external evolutionary algorithm for optimal parameter setting to predict the short-term trend direction. Our results point to good and robust accuracy over time and across different market regimes. Furthermore, we propose to exploit recent advances in sentiment analysis to reassess its role in financial forecasting; in this way, we contribute to the empirical literature by showing that predictions based on sentiment analysis are not found to be significantly different from predictions based on historical data. Nonetheless, compared to stock markets, we find that the accuracy of trend predictions with sentiment analysis is on average much higher for cryptocurrencies.

Keywords: cryptocurrencies; social media; sentiment analysis; prediction; machine learning

MSC: 91G60

1. Introduction

In the last few years, cryptocurrencies have drawn significant attention [1]. The introduction of a decentralized currency, i.e., a currency for which there is no central authority responsible for its economic value, is of paramount importance, as this feature poses serious problems for regulators, investors, and scholars alike [2]. Its primary function, unlike traditional currencies, is to actually cut off financial institutions and intermediaries, such as brokerages, exchanges, and banks, from transactions through a peer-to-peer blockchain technology. Furthermore, cryptocurrencies show other interesting features such as low transaction costs, safety of transactions and payments, and less control from central banks. As a consequence, all these characteristics have quickly made cryptocurrencies popular, especially among retail investors [3]. This rapid growth, however, has triggered high volume trading, a buying frenzy among retail investors, and pump-and-dump schemes, i.e., market manipulation fraud, which involves artificially raising the price of a cryptocurrency and then selling it at a higher price to other investors [4,5]. Therefore, high volatility and herding behaviour in the market, which are not coherent with the traditional features of either a currency or a long-term investment, are frequently observed in cryptocurrency markets. More generally, the growing importance and ballooning market capitalization...
of cryptocurrencies has sparked concern over potential and long-feared side effects and interplays between traditional asset classes and cryptocurrencies, especially with respect to an increase of systemic risk spillovers from the latter to the former [6].

Moreover, due to their unique features, studies have addressed the question of whether cryptocurrencies exhibit desirable properties against other asset classes (see Figure 1). Existing empirical findings agree that certain cryptocurrencies act as a good diversifier with respect to commodities and stocks [6,7], although a growing body of studies [8,9] point out that cryptocurrencies display poor safe haven and hedge properties [10].

For example, [11] showed that before the 2017 crash Bitcoin was a suitable hedge against energy commodities, while in the post-crash period this effect quickly faded away, leading to the conclusion that it is at most suitable as a diversifier. A hedge is generally denoted as an asset that is uncorrelated or negatively correlated with another asset or portfolio, while a diversifier is defined as as an asset that is positively but not perfectly correlated with another asset or portfolio [12].

![Figure 1. Daily returns of a selection of cryptocurrencies and market indexes.](image)

Consequently, a wide consensus has been reached over the speculative nature of cryptocurrencies [3,13–15], making them appealing for traders willing to reap profits from potential short-term sentiment-induced mispricings [16]; to date, the quest to predict cryptocurrency prices has become a mainstream issue thanks to the vast availability of free data across different market regimes. Despite being considered simplistic, perhaps the most popular and leading valuation model among Bitcoin proponents is the stock-to-flow (S2F) model, which was proposed by an anonymous institutional investor operating under the pseudonym Plan B in 2019. In essence, this model tries to predict Bitcoin prices using historical data on the ratio between the total supply of Bitcoin and the increase in supply; both of these quantities are known with high precision. However, the main flaw of the model lies in the misleading assumption that scarcity directly drives future Bitcoin prices, even the predictor is known without uncertainty. According to basic asset pricing theory, this is a piece of information that must be already fully discounted by the market, and as such fully reflected in the current prices, as argued, among others, by [17].

Thus, econometric and machine learning models for short-term price prediction of cryptocurrencies have blossomed in the last few years, with the goal of developing a rigorous and evidence-based framework for generating and appropriately assessing the quality of the forecasts. Various studies [18,19] have specifically proposed artificial intelligence methods for this purpose, while other contributions have suggested using them jointly with social media data analysis [20–22].

In our current contribution, we want to use a cutting-edge methodology to generate price forecasts for cryptocurrencies along with common stocks, the predictability of which is becoming one of the core research questions involving cryptocurrencies as a specific
financial asset. Indeed, there is mixed evidence in the literature about the profitability and value of market sentiment for prediction and trading purposes [23–28]. More generally, because there is broad evidence that machine learning methods are efficient for prediction tasks and are explicitly designed to approximate complex nonlinear dependencies, for asset pricing problems [29] we adopt an artificial neural network (ANN) model. ANNs are arguably among the most flexible approaches in machine learning and have been extensively used for prediction tasks in empirical contributions [19,30] (we refer to [31] for a technical discussion on the universal approximation theorem regarding feedforward networks).

The contribution of this paper to the literature is twofold. First, from a methodology point of view, we propose to extend the stochastic neural network model [19] to new problem instances, including stock market data, while focusing on social media data analysis to predict the trend direction of markets (upwards or downwards). As for the social media data aspect, we use the Bidirectional Encoder Representations from Transformers (referred to as BERT) [32] to quantify the user experience regarding specific cryptocurrencies and traditional stocks through an analysis of Twitter data (see Section 3.2). Then, we set the values of two key hyperparameters of the network (i.e., the neural network topology and the learning rate) using an external evolutionary algorithm, the Relevance Estimation and Value Calibration (REVAC) [33], while considering the influence that the parameter setting has on the forecasting performance of the model [34]. To sum up, we aim to exploit recent advances in sentiment analysis, deep learning, and parameter tuning to reassess their role in generating accurate forecasts.

Second, we contribute to the empirical literature by evaluating the impact of short-term sentiment in trend fluctuations. We highlight that predictions with sentiment analysis are not found to be significantly different from forecasts simply based on historical data. Therefore, our results are not completely aligned with other studies (e.g., [19]) finding large and significant predictive power on the part of social media data, although the accuracy of trend prediction with sentiment analysis is found to be on average much higher for cryptocurrencies. Altogether, by reassessing the influence of sentiment analysis on trend forecasts and providing an updated tool for forecasting purposes based on recent developments in deep learning, this study provides trading implications for investors willing to gain an edge over buy-and-hold strategies.

The remainder of this paper is organized as follows. Section 2 reviews the related literature and state-of-the-art techniques for price prediction. Section 3 presents the data and methodology adopted to make predictions on the basis of sentiment analysis and deep learning. The results are presented in Section 4, along with a discussion of the implications for traders and investors. Finally, Section 5 concludes the paper with comments on possible future research directions.

2. Related Work

In this section, we provide a brief outline of the literature on price prediction based on statistical [35,36] and artificial intelligence-based models [30], with a particular focus on information recovered from sentiment analysis [37,38], which is nowadays considered particularly valuable for forecasting purposes. A comprehensive summary is reported in Table 1.

Statistical and econometric models, although outperformed by machine learning (ML) approaches for prediction purposes [35], are generally useful, as they provide a highly interpretable baseline. In this way, the distinctive characteristics of more sophisticated strategies that take into account potential nonlinearities in the data or allow for more complex modelling, e.g., based on large predictor sets including features, their interactions and their nonlinear transformations, can be emphasized. In [38], the authors compare and discuss a comprehensive summary of the previous studies in the field of cryptocurrencies price prediction from 2010 to 2020 and conclude that the latest contributions address the problem by focusing on ML models. These models have received increased attention in
study and analysis of cryptocurrencies, mainly thanks to their better achievements in terms of accuracy.

In [36], the authors focused on econometric modelling; they found and removed a seasonal component in the hourly Bitcoin data and subsequently generated closing price predictions using a simple AutoRegressive Integrated Moving Average (ARIMA) model. Relatively low accuracy was achieved, however, suggesting the possibility of further improvement by means of more sophisticated modelling.

Prediction with neural network models and sentiment analysis lies at the core of our contribution, as it has been largely shown that these are among the most powerful strategies for prediction purposes [30] despite being highly parametrized. A comprehensive test of econometric, Machine Learning (ML), and Deep Learning (DL) models was performed in [39] for cryptocurrency price prediction, including ARIMA, k-Nearest Neighbors (kNN), Support Vector Regression (SVR), Random Forests (RF), Long-Short Term Memory (LSTM) networks, Gated-Recurrence Units (GRU), LSTM-GRU networks (HYBRID), Temporal Convolution Network (TCN), and Temporal Fusion Transformer (TFT). The authors found that Recurrent Neural Networks (RNNs) with LSTM units outperformed other strategies, as [30,40] pointed out in the context of cryptocurrency price prediction. Furthermore, the same authors contributed to the literature by showing that DL strategies outperform ML and econometric approaches and that more complex and parameterized models tend to generate better forecasts in terms of accuracy, which is somewhat surprising and inconsistent with recent cutting-edge research in the field of ML-based equity asset pricing [29].

In a similar setting, ref. [41] examined the long-term performance of various ML-based strategies for the S&P500 out-of-sample directional movements and came to a similar conclusion, finding that a forecasting strategy based on a shallow LSTM network was the most effective method among those tested.

In [18], the authors tested both feed-forward artificial neural networks (ANNs) and more complex Long Short-term Memory (LSTM) networks to analyze the price dynamics of Bitcoin, Ethereum, and Ripple. Surprisingly, the LSTM did not significantly outperform ANNs in terms of accuracy, especially when the latter were fed a long-term history of returns as input, whereas the former model was capable of dealing with short predictive memory lengths more efficiently. The authors concluded that cryptocurrency markets are not even weakly efficient, proving that past returns contain valuable information and have predictive potential which can be exploited to make profits by trading accordingly.

A more involved model was deployed by [42], where a hybrid neural network based on a convolutional neural network (CNN) and an LSTM layer was proposed to forecast Bitcoin prices. In a nutshell, the CNN was used to extract influential features, which were then passed on to the LSTM layer for training and out-of-sample forecasting of the short-term price of Bitcoin. Moreover, the authors set up a model including a variety of diverse features, including transaction data, macroeconomic variables, investor attention, and technical indicators.

As far as sentiment analysis is concerned, Ref. [43] analyzed tweets about Bitcoin to assess whether they conveyed positive or negative sentiments, then used them as input for a Recursive Neural Networks (RNNs). In [38], the authors evaluated the predictive power of sentiment and explored statistical and deep-learning methods to predict the future price of Bitcoin by contributing an analysis of financial and sentiment features extracted from economic and crowdsourced data. A novel perspective was investigated by [19]; in their proposed modelization, a stochastic neural network model was introduced to perform cryptocurrency price predictions with a broad range of features. They found that social sentiment data plays a key role in forecasting. More precisely, 23 features were retrieved from three different main sources. First, market-based data were used, such as the number of transactions, intra-day lows and highs, market capitalization, and volume. Second, crypto-specific and blockchain-based features were taken into account, namely, the mining difficulty and profitability, hashrate, transaction fees, and confirmation time. Finally, the influence of sentiment on prices was factored in by including the volume of
tweets and Google Trends data. In [44], the authors found that sentiment analysis based on the Valence-Aware Dictionary and sEntiment Reasoner tool is an invaluable predictor in cryptocurrency markets. A comprehensive study with different cryptocurrencies and an ensemble method combining cutting edge ML algorithms showed that sentiment data and Google Trends were especially effective at forecasting the short-term fluctuations of cryptocurrencies.

The practical value of sentiment data was assessed in [10]. The authors proposed capturing cryptocurrency market sentiment by creating an ad hoc crypto-specific sentiment dictionary based on posts on a popular Chinese social media platform. Trend direction forecasting based on sentiment data and historical market prices shows promising results in terms of accuracy and recall compared to previous studies. Similar conclusions have been reached for Dash price prediction by [45]; alternatively, ref. [46] used random forests to forecast Bitcoin prices and documented that news feeds and tweets had little predictive power. In similar fashion, ref. [47] constructed an hourly sentiment index by extracting and classifying signals from Twitter to predict the price fluctuations of a small-cap alternative cryptocurrency. Forecasts based on the Extreme Gradient Boosting Regression Tree Model were found to be particularly accurate, supporting the view that sentiment analysis provides additional value to predictions.

Moreover, several efforts have been made to combine neural networks with text and data mining approaches; Ref. [48] created a machine learning model based on the price of Bitcoin, Google trends data, and custom related features. To this end, the authors compared a neural network with LSTM layers, a Gradient Boosting Regression Model, and an XGBoost model; according to their results, the first approach, based on deep learning, was the most efficient for prediction purposes. A combination of Twitter and Google Trends data was used as input in a simple multivariate linear regression model by [20], which proved to be effective at generating a signal for the price direction. In [21], the authors employed headline-based and tweet-based predictions, which were modelled by means of logistic regression, linear support vector machine, and naive Bayes models for a classification task, that is, predicting price increases and decreases. The authors did not find a robustly outperforming classifier across different cryptocurrencies, although their baseline logistic regression model performed relatively well across different datasets.

Similarly, Ref. [37] compared Neural Network (NN), support Vector Machine (SVM), and Random Forest (RF) models while using market and Twitter data as input features, showing that the sentiment itself is effective enough to generate high quality forecasts without controlling for market features, at least for a subgroup of cryptocurrencies. Their results are consistent with [48], where the authors showed that neural networks outperform other families of models.

**Table 1. Summary of the literature on statistics and artificial-intelligence-based price prediction.**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Type of Prediction</th>
<th>Performance Measures</th>
<th>Sets of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albariqi and Winarko (2020) [49]</td>
<td>2-days to 60-days prices</td>
<td>Accuracy: 81.3%, precision: 81%, recall: 94.7%</td>
<td>1300 observations (August 2010-October 2017)</td>
</tr>
<tr>
<td>Atsalakis, Atsalaki, Pasious and Zopounidis (2019) [50]</td>
<td>Price movements</td>
<td>RMSE: 0.376, MSE: 0.0014, MAE: 0.0307</td>
<td>2201 daily closing prices from September 2011 to October 2017</td>
</tr>
<tr>
<td>Charandabi and Kamyar (2021), [30]</td>
<td>Actual price, short-term prediction</td>
<td>Accuracy: 50%</td>
<td>2 years observations</td>
</tr>
<tr>
<td>Derbentsev, Datserko, Stepanenko and Bezko (2019) [51]</td>
<td>5 to 30 days price movement</td>
<td>RMSE: 0.04-0.08</td>
<td></td>
</tr>
<tr>
<td>Hitam, Ismail and Saeed (2019) [52]</td>
<td>Cryptocurrency daily prices</td>
<td>Accuracy: 78.9%</td>
<td>OHLC (open/high/low/closing) daily prices from 2013 to 2018</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Type of Prediction</th>
<th>Performance Measures</th>
<th>Sets of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khedr, Arif El-Bannany, Alhashmi and Sreedharan (2021) [35]</td>
<td>Survey of previous contributions from 2010 to 2020</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Li and Dai (2021) [42]</td>
<td>3 days ahead price prediction</td>
<td>Precision: 64%; Recall 81%; F1 69%</td>
<td>Bitcoin historical prices, macroeconomic indicators, and investor attention. Data from December 2016 to August 2018. 5 days and 6 months historical series.</td>
</tr>
<tr>
<td>Mahboubbeh and Heidari (2020) [53]</td>
<td>5 days ahead forecasting</td>
<td>average MAPE: 1.14%</td>
<td>Daily prices and 26 additional features, gathered from Blockchain Info.</td>
</tr>
<tr>
<td>Madan, Saluja, Shau-rya, Zhao (2015) [54]</td>
<td>Price movements</td>
<td>Accuracy: 98.7% for daily data; 8% to 50% for high frequency data (10 s and 10 min timeframes)</td>
<td>Data from September 2014 to December 2020. Data from January 2015 to December 2017 (2585 positive, 1669 negative and 3200 irrelevant tweets). Hourly-based analysis from April 2013 to July 2017.</td>
</tr>
<tr>
<td>Nayak (2022) [55]</td>
<td>Daily, weekly, monthly closing prices</td>
<td>MAPE: 0.031%; MSE: 0.01893; UT: 0.052; ARV 0.016.</td>
<td>Data from January 2015 to December 2020. Data from January 2015 to December 2017 (2585 positive, 1669 negative and 3200 irrelevant tweets). Hourly-based analysis from April 2013 to July 2017.</td>
</tr>
<tr>
<td>Pant, Neupane, Poudel, Pokhrel, and Lama (2018) [40]</td>
<td>Next day’s price</td>
<td>Accuracy for sentiment classification 81.39% and 77.62% for overall RNN</td>
<td>Daily closing price, starting from April 2013 to February 2019.</td>
</tr>
<tr>
<td>Pratama, Nugroho and Sukiyono (2020) [56]</td>
<td>Next day closing price</td>
<td>MAPE: 1.883% for hybrid method between backpropagation and genetic (GABPNN)</td>
<td>BTC prices from October 2013 to April 2017 (1278 observations), OHLC prices and volumes.</td>
</tr>
<tr>
<td>Radiyo, Munajat and Budi (2017) [57]</td>
<td>Next day weighted value</td>
<td>ARIMAX-MSE: 0.0003; RNN-MSE: 0.0014</td>
<td>Data from April 2017 to October 2019, BTC volumes, weighted prices, sentiment and Tweets volumes. 1-min spaced BTC data from January 2012 to March 2021, OHLC prices, volumes, chosen currencies, weighted Bitcoin prices, Tweets by Elon Musk about cryptocurrencies from 2009 until 2021. 80-days data, hourly and daily granularity. The dataset contains OHLC prices, transaction volumes, and social data retrieved from Twitter.</td>
</tr>
<tr>
<td>Valencia, Gómez-Espinoza and Valdés-Aguirre (2019) [37]</td>
<td>Daily closing price, price movements</td>
<td>RMSE: 0.097; R² 95.2% on ETH</td>
<td>Daily closing prices from July 2017 to July 2020.</td>
</tr>
</tbody>
</table>

3. Data and Methodology

In this section, we describe the building blocks of our methodology that we use to predict the price movements of stocks and cryptocurrencies in Section 4. First, we introduce the concepts of data retrieval and sentiment analysis in the framework of financial stock trends in Section 3.1. Then, we discuss how to convert these concepts into numerical aggregates via the BERT and roBERTa models in Section 3.2. These aggregates, along with the historical prices, are used as inputs to the neural network outlined in Section 3.3. The data at hand are described in Section 3.4, and because Neural Networks require preliminary operations over data before their use, we outline the preprocessing operations in Section 3.5, while the procedure used to partition the dataset into a training and test set is outlined in Section 3.6.

3.1. Data Retrieval, Sentiment Analysis, and Financial Stock Trends

The process of collecting information from raw data by means of retrieving, selecting, cleaning, transforming, and modeling techniques is known as Data Science. Nowadays, a huge amount of information lies on social media platforms such as Facebook or Twitter, where people post their opinion, mood, or sentiment not only exclusively related to their private lives but also to commercial products, experiential events, brand services, and the latest worldwide breaking news (i.e., wars, pandemics, migration, climate change, natural disasters, sport) [60–64]. Organizations, governments, and companies are increasingly tasked with selecting, extracting, and mining large amounts of user information that flows on the web and social networks.
In this view, social networks are not only virtual communities where people meet, they represent effective merchandising means by which companies bring their products and services directly into their targeted clients’ houses. This huge amount of structured or unstructured digitized data available from various sorts of sources, i.e., Big Data, is the oil of the new economy. For example, the use of Big Data for marketing strategies to attain a superior customer experience and improve sales revenues has become a must-have core skill for every company to attract target consumers and secure a dominant position. To this end, companies have undergone a significant transformational process, resulting in turning their businesses into data-driven organizations. Big Data are a precious source of information by which an effective Sentiment Analysis study can be conducted. Sentiment Analysis, in a few words, by inferring people’s sentiments and opinions from social media contents, aims at revealing how well a brand is perceived by the marketplace. This information can help companies and organizations to better understand their clients and consequently develop strategies to ensure their satisfaction and retention. The process of detecting and extracting sentiment or opinion from online text has become essential for almost every kind of business, such as insurance and financial services, media and healthcare companies, education, leisure and hospitality, e-commerce, and sport.

In [65], data were collected from Twitter on people’s opinions on movies, products, books, etc., in order to analyze their sentiments using Support Vector Machine (SVM) and Particle Swarm Optimization (PSO), while [66] used four different machine learning algorithms: Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM), for classification of human sentiment. In [67], a Hierarchical Twitter Sentiment Model (HTSM) was proposed to analyze people’s sentiments and feelings about an HTC product (smartphones) and the Coronavirus topic. In [68], a hybrid Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) model was proposed to deal with the sentiment analysis problem and was tested on the IMDB and Amazon movie review datasets.

In [69], the authors introduced a novel context-aware deep learning-driven Persian sentiment analysis approach which they validated on the Persian movie review and hotel review datasets. In [70], a recommendation approach was presented integrating sentiment analysis and genre-based similarity in collaborative filtering methods. Experiments were conducted on music and movie datasets containing information about item reviews and genres. In [71], the authors introduced different volumetric, sentiment, and social network approaches to predict elections in three Asian countries, i.e., Malaysia, India, and Pakistan, from Twitter posts. For the sentiment-based price prediction, we refer the reader to the discussion in Section 2 and to references therein (e.g., [10,19,38,44–47]).

3.2. BERT and roBERTa Models

It has been shown that language model pretraining is effective in improving many natural language processing tasks [72,73]. The feature-based and fine-tuning strategies used, respectively, task-specific architectures and task-specific parameters, apply pretrained language representations to downstream level tasks. However, by using unidirectional language models, both techniques significantly limit the effective gains from applying a pretrained representation, as the choice of architectures that can be used during pretraining is heavily restricted.

In this contribution, we rely on the BERT model, which stands for Bidirectional Encoder Representations from Transformers [32]. BERT attenuates the above-mentioned drawback of unidirectionality by using a “masked language model” (MLM) pretraining objective. In detail, the MLM objective, by enabling the representation to merge both the left and right context, allows pretraining of a deep bidirectional Transformer. In addition to MLM, BERT accounts for Next Sentence Prediction task (NSP) by allowing for sentence embedding tasks. The latter aspect is a very relevant one, as many important downstream tasks, such as Question Answering (QA) and Natural Language Inference (NLI), require
information about the linkages between two sentences, which is not directly captured by language modeling.

From an operational point of view, we used BERTweet, a robustly optimized BERT model (roBERTa) trained on English tweets [74]. Specifically, roBERTa overcomes the main drawback of the BERT model, i.e., the application of a static masking strategy, by adopting a dynamic masking, that is to say, generating a masking pattern every time a sequence is passed to the model. The latter aspect is particularly crucial when dealing with large datasets. Moreover, the auxiliary NSP task is removed and larger batch sizes (up to 32K sequence) as well as a larger byte-level Byte-Pair Encoding (BPE) vocabulary are considered, leading to a better end-task accuracy. The roBERTa model outputs labels for each post published by the user. In more detail, it returns the labels NEG, NEU, and POS, and the model computes, how negative, neutral, and positive, respectively, a sentence is.

3.3. Neural Networks

Most existing empirical studies in the field of financial forecasting rely on well-established neural network models, including shallow and deep MLPs, networks with LSTM layers, which are particularly suitable for prediction based on time series data, and sometimes on CNNs, although their effectiveness in finance is yet to be fully understood and documented [75,76]. As we discuss in Section 2, hybrid neural network models are feasible, although the underlying economic motivation for this choice of architecture and for its use for financial forecasting is not fully clear.

In this contribution, we rely instead on a model that is deeply rooted in financial theory. The main aim of the proposed approach is to incorporate stochasticity in the neural network modelization; therefore, we follow [19] to investigate the predictability of trends in cryptocurrency and stock markets. The proposed strategy can accommodate stochasticity into both MLP and LSTM layers by incorporating randomness into the activation function, and this can be basically achieved by inducing randomness in the features to mimic the randomness of financial markets. To formalize the concept, which is outlined in Figure 2, we introduce the following formulation of a stochastic layer:

\[ s_t = h_t + \gamma \xi_t \cdot f(h_t, s_{t-1}) \quad 0 < \gamma < 1 \quad (1) \]

where \( h_t \) is the activation at time step \( t \), \( \gamma \) is a perturbation factor controlling for the amount of randomness, \( \xi_t \) denotes a generic multivariate random variable at time \( t \) drawn from an i.i.d. process, and \( f(\cdot) \) denotes a family of functional forms by means of which the activation at time \( t \) is related to the activation at the previous time step. Following [19], we denote the stochastic transformation of the layer with \( s_t \).

A parsimonious and financially sound representation of the reaction function \( f(h_t, s_{t-1}) \) is simply given by the difference between the current and the previous stochastic activation, that is:

\[ f(h_t, s_{t-1}) = h_t - s_{t-1} \quad (2) \]

This implies that the process is memoryless and independent of previous stochastic activations.
Figure 2. Flowchart of the forecasting algorithm: the model integrates stochastic neural networks with a REVAC parameter tuning method. Data are extracted from historical prices and sentiment analysis is performed with the roBERTa model.
Moreover, the importance of stochastic weights decays exponentially over time due to the fact that $0 < \gamma < 1$. This property can be shown by recursively substituting the value of $s_{t-1}$ in Equation (1):

$$s_t = (1 + \gamma \xi_t) h_t - \gamma \xi_t s_{t-1} = (1 + \gamma \xi_t) h_t - \gamma \xi_t [(1 + \gamma \xi_{t-1}) h_{t-1} - \gamma \xi_{t-1} s_{t-2}] = \ldots$$

$$= (1 + \gamma \xi_t) h_t - \gamma \xi_t (1 + \gamma \xi_{t-1}) h_{t-1} + \gamma^2 \xi_t \xi_{t-1} s_{t-2} = \ldots$$

$$= \ldots = (1 + \gamma \xi_t) h_t + \sum_{i=1}^{t-1} (-\gamma)^{t-i} (1 + \gamma \xi_i) h_i \prod_{j=i+1}^{t} \xi_j$$

(3)

Finally, let us briefly describe the forward propagation step in the stochastic version of a standard MLP. Denoting with $x_i$ a generic preprocessed vector of input data at time $t$ of size $n$, where $n$ is the number of features, consider a user-defined model architecture with $k = 1, \ldots, l$ layers and a vector of weights $w^k$. The main element of novelty in the forward pass step is the presence of a stochastic module attached to each layer of the network, by which a desired degree of randomness is introduced in the model, where the amount of randomness depends on the hyperparameter $\gamma$. Therefore, assuming a linear activation for the input layer as in the standard MLP case, i.e., $z_t = b_t + w^1 s_t$, a stochastic module is introduced for each $k = 1, \ldots, l$, where we characterize the hidden layer with $h_t = F(z^k_t)$ and where $F$ includes but is not limited to ReLU or logistic activation functions. Finally, for each $k$, we draw from an i.i.d. collection of multivariate random variables at time $t$, $\xi^k_t$, where the dimension of $\xi^k_t$ is equal to the number of features, and finally write the stochastic layer as follows:

$$s_t = h_t + \gamma \xi^k_t \odot f(h_t, s_{t-1})$$

(4)

where $\odot$ denotes the Hadamard product. Therefore, $f(\cdot)$ takes into account the activation at time $t$ and the stochastic activation at the previous time step to capture the random propagation of the features over time, where the degree of randomness is regulated through the perturbation factor $\gamma$ and the randomly generated vector $\xi^k_t$. Finally, note that the standard MLP is a special case of its stochastic version, with $\gamma = 0$.

In our contribution, the presented stochastic neural network model is trained to predict whether the trend direction of stock market and cryptocurrency prices will move upwards or downwards, where the trend direction labels are denoted with $l_t$ and the labels are respectively equal to $+1$ or $-1$ depending on the direction. Our neural networks have been implemented in Python using the Keras library. From an operational point of view, experiments have been run on a laptop equipped with 16 GB RAM and an Intel Core Intel 2.80 GHz CPU.

3.4. Data

For our data, we tackled the problem with different sets of data consisting of all stocks belonging to different stock exchange indices from France (CAC40: 40 stocks), Germany (DAX: 40 stocks), the United Kingdom (FTSE: 100 stocks), the United States (Nasdaq, 100 stocks), and Japan (Nikkei, 225 stocks). For each index component, we collected the daily closing price of three different periods: from 1 January 2006 to 1 January 2007 (i.e., a pre-crisis scenario); from 1 January 2009 to 1 January 2010 (i.e., a post crisis scenario); and from 1 January 2022 to 1 January 2023 (the current scenario). In a nutshell, we used observations for three distinct scenarios about 505 assets.

Furthermore, we analyse several of the most widespread cryptocurrencies (BTC, ETH, BNB, XRP, DOGE, ADA, MATIC, DAI, TRX, DOT, LTC, UNI, for a total of 12 cryptocurrencies), but only in the current scenario, as historical prices for the pre-crisis and post-crisis scenarios were not available. Henceforth, we follow the definition of closing price for cryptocurrencies provided by coinmarketcap.com and commonly used in the practice and in the literature: because cryptocurrency exchanges support 24/7 trading, with listed assets operational 100% of the time, the closing price is set equal to the latest available price for
any calendar day. From an operational point of view, data have been retrieved with a script designed and implemented in Python that runs on the *Yahoo Finance* website.

### 3.5. Data Preprocessing

Before moving on to the results obtained from our modelization, a description of the actual procedure for handling and processing the data is crucial in order to understand its key patterns, to remove potential anomalies in the data, and more generally to adjust the values of the features in the dataset to a common scale. In this contribution, we follow the approach described in [77–79]:

- **Removal and replacement:** remove the input features including more than 30% of missing or anomalous values, such as not available (N/A) data due to delisting or other corporate operations; then, replace missing values, if any, with a rule of thumb, such as the average value of the variable over time or by propagating the last available observation forward to the next available;
- **Normalization:** the normalization step is necessary, especially for neural networks, to facilitate the convergence of the training algorithm towards a global optimum and thereby obtain stable parameters for the model. Hence, let $x_t$ be the value before normalization of a generic input feature at time $t$, and let $\bar{x}_t$ be its normalized value. The relationship between the two can be stated as follows:

$$\bar{x}_t = \log_u (x_t + 1)$$  \hspace{1cm} (5)

where $u = x_{\text{max}} + 1$, such that the condition $\bar{x}_t \in [0, 1]$ is satisfied. For problems in which the features might potentially contain negative values, ref. [77] proposed a different formulation:

$$\bar{x}_t = \log_u \left( |\min(0, x_{\text{min}})| + x_t + 1 \right)$$  \hspace{1cm} (6)

In our case, because we do not have negative values, we used Equation (5) in what follows. We chose to use the interval [0, 1] for each input node. As [77] have argued, a natural way of normalizing data is the min-max linear transformation to the range [0, 1]; however, such a formulation is excessively sensitive to outliers due to linearity. Thus, if we used the min–max formula we would lose a substantial amount of information, as it would set a non-negligible number of values close to one or zero right away. For this reason, according to other related works that used similar transformations [80], we opt for the logarithmic transformation of the data.

For the network parameters, there is a broad range of literature stressing the importance of parameter settings in the framework of neural network training. Every possible parameter setting schema has its own pros and cons. Thus, we have chosen to set the most important neural network parameters (i.e., the neural network topology and the learning rate) using an external parameter setting procedure; we resorted to REVAC (Relevance Estimation and Value Calibration), which is based on evolutionary computation, and used its variation operators to estimate parameter distributions. For each iteration, REVAC generates a new set of parameter values and evaluates them according to a user-defined criterion (see [34] for more information). Of course, other parameter setting (and control) methods can be used, which is left for further work.

### 3.6. Training and Test Set

In order to perform the learning phase and test the network’s performance, the data at hand were partitioned into a *training set* and *test set*, with the first used to train the network. In order to avoid overfitting, learning is generally stopped depending on the performance on the test set, which is used to assess algorithm performance.

For our experiments, we used 252 daily closing prices (observations) for each scenario and for each stock belonging to each stock exchange index. Please note that our approach requires a series of historical prices to predict the upward/downward trend on the day after
the last observation. We have decided to partition data to ensure 70 percent observations for the training set and 30 percent for the test set. We used 176 daily observations for predicting the upward/downward movement of the stock market.

For each stock, for \( i \in 1 \ldots 53 \) we trained the network by providing as input the daily closing prices of days \( d \in 1 \ldots 176 + i - 1 \) and the observed upward/downward movement registered by observing the closing price at time \( 176 + i \). Then, we used the trained model to perform 23 predictions: for each stock, for \( i \in 1 \ldots 23 \), we provided as input the daily closing prices of days \( d \in 53 + i \ldots 176 + 53 + i - 1 \) and the observed upward/downward movement registered by observing the closing price at time \( 176 + 53 + i \).

In a nutshell, our neural networks had 176 input neurons for experiments that used only historical data as input. For the experiments that used historical data and roBERTa output as input indicators, we instead had 179 input neurons, as we used roBERTa to provide information about positivity, neutrality, and negativity (i.e., three neurons) on the tweets mentioning the specific stock (or cryptocurrency) posted on the day corresponding to the last observed historical price.

4. Results

In this section, we outline the results of our neural network approach: Section 4.1 discusses the experiments performed in the first and second scenarios using historical data as input values for stocks, while Section 4.2 describes the results of the experiments performed in the third scenario using historical values and outputs from roBERTa as input values for both stocks and cryptocurrencies.

4.1. Results with Historical Stock Market Data

In these experiments, we used historical daily closing prices as the sole inputs of the network, as data about tweets for the time periods 2006–2007 and 2009–2010 were not available. To assess the quality of our predictions, we rely on a conventional accuracy metric: first of all, for classification purposes, recall that there are basically four types of outcomes for a prediction, i.e., two when the prediction is right (predict true with true realization or predict false with false realization) and two when the prediction is wrong (predict true with false realization and the other way round). Denoting with \( N \) the sample size, with \( y_i \) and \( \hat{y}_i \) respectively the actual and the estimated outcome and with 1 an indicator function, we formally define the above-mentioned rates as follows:

\[
TP = N^{-1} \sum_{i=1}^{N} 1_{y_i=\hat{y}_i=1} \quad \text{True positive}
\]

\[
TN = N^{-1} \sum_{i=1}^{N} 1_{y_i=\hat{y}_i=0} \quad \text{True negative}
\]

\[
FP = N^{-1} \sum_{i=1}^{N} 1_{\hat{y}_i=1,y_i=0} \quad \text{False positive}
\]

\[
FN = N^{-1} \sum_{i=1}^{N} 1_{\hat{y}_i=0,y_i=1} \quad \text{False negative}
\]

From the first two baseline rates, it is possible to recover the accuracy metric, i.e., the percentage of correct forecasts:

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN}
\]

Table 2 reports the main statistics of accuracy on the test set of experiments performed on stocks belonging to the four different stock exchange indices introduced in Section 3.4. We remark that we cannot perform experiments on cryptos, since historical prices are not available for these two periods. Our neural network approach offers a satisfactory performance with respect to the state of the art, as shown in Table 1.
Table 2. Accuracy of stock market trend prediction under different scenarios. The tests were performed with stochastic neural networks + REVAC using exclusively historical data as input to forecast the future trend.

<table>
<thead>
<tr>
<th>Index</th>
<th>CAC40</th>
<th>DAX</th>
<th>FTSE 100</th>
<th>Nasdaq</th>
<th>Nikkei</th>
<th>Crypto</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 January 2006–1 January 2007</td>
<td>Min 0.6888; Mean 0.9629; Md 0.822; Std 0.0595</td>
<td>Max 0.7608; Mean 0.844; Md 0.451</td>
<td>Max 0.6592; Mean 0.829; Md 0.0571</td>
<td>Max 0.6888; Mean 0.033; Md 0.822; Std 0.0584</td>
<td>Max 0.5832; Mean 0.9125; Md 0.8538; Std 0.1942</td>
<td>N/A</td>
</tr>
<tr>
<td>1 January 2009–1 January 2010</td>
<td>Min 0.6304; Mean 0.9347; Md 0.822; Std 0.0651</td>
<td>Max 0.7555; Mean 0.844; Md 0.0485</td>
<td>Max 0.6666; Mean 0.933; Md 0.844; Std 0.069</td>
<td>Max 0.6666; Mean 0.955; Md 0.822; Std 0.0691</td>
<td>Max 0.6172; Mean 0.693; Md 0.823; Std 0.1874</td>
<td>N/A</td>
</tr>
</tbody>
</table>

4.2. Results with Stocks and Twitter Historical Data

In this set of experiments we use the degree of positivity, neutrality, and negativity provided by roBERTa over the tweets of the last day, on the time period 2022–2023 as network’s input, along with historical daily closing prices. Table 3 reports the main statistics of the accuracy metric (8) on the test set of experiments performed on both stocks belonging to the four different stock markets and cryptocurrencies prices presented in Section 3.4. By visual inspection, the results seem to be better than those reported in Table 2. However, a Wilcoxon pairwise test run on the accuracy values of experiments performed on the third scenario with either historical prices or the combination of prices+roBERTa, does not allow us to reject the null hypothesis that the results of the two experiments are drawn from the same distribution, although the latter are definitely better in absolute terms. Nonetheless, we can state that the prediction exercise on cryptos is definitely better than those obtained on stock market data, suggesting two possible explanations.

First, our findings imply that social medias do not add particular value to the trend forecasting of these instruments, since the overall effect is not statistically significant. Second, our results also suggest that unregulated and less liquid markets, such as cryptocurrency exchanges, can be nonetheless influenced by short-term sentiment more than stock markets, causing more volatility and mispricings, ultimately paving the way to profitable trading opportunities.

Table 3. Accuracy of stock markets and cryptocurrencies trend prediction under the third scenario. The tests were performed with stochastic neural networks + REVAC using historical data and the degree of positivity, neutrality, and negativity provided by the roBERTa model as inputs to forecast the future trend (Historical + roBERTa model); historical data only (Historical only model).

<table>
<thead>
<tr>
<th>Index</th>
<th>CAC40</th>
<th>DAX</th>
<th>FTSE 100</th>
<th>Nasdaq</th>
<th>Nikkei</th>
<th>Crypto</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 January 2022–1 January 2023</td>
<td>Min 0.6895; Max 0.933; Mean 0.894; Md 0.822; Std 0.0543</td>
<td>Min 0.5914; Max 0.974; Mean 0.868; Md 0.822; Std 0.0814</td>
<td>Min 0.6818; Max 0.984; Mean 0.829; Md 0.818; Std 0.0596</td>
<td>Min 0.7538; Max 0.975; Mean 0.826; Md 0.822; Std 0.056</td>
<td>Min 0.7719; Max 0.983; Mean 0.851; Md 0.841; Std 0.1142</td>
<td>Min 0.6919; Max 0.990; Mean 0.874; Md 0.835; Std 0.0901</td>
</tr>
</tbody>
</table>

In order to test the goodness of the predictions obtained with a baseline model, we compare our results with those obtained by a multilayer perceptron network with the hyperparameter setting proposed by [49]. For the sake of readability, we report here the settings proposed by the authors, with \( N \) denoting the input nodes and \( m \) the output nodes:

- Learning rate: 0.01;
- Learning algorithm: Adam;
• Batch Size: 128;
• Hidden Layer 1 Node Size: $\sqrt{(m + 2)N + 2}\sqrt{\frac{N}{m+2}}$;
• Hidden Layer 2 Node Size: $m\sqrt{N}\frac{N}{(m+2)}$;
• Dropout: 0.5;

Moreover, we keep using the two sets of input features discussed in the present section. Hence, for simplicity, we depart from [49] in this respect and we set up a forecasting model with historical prices and another one with historical prices and sentiment analysis based on the roBERTa approach. Please note that a comparison between the results based on a MLP architecture in Table 4 and the accuracy reported in Table 3 provides a confirmation that our proposed modelization performs regularly better. We therefore assume that the benefits, both in terms of average and of volatility of accuracy, can be essentially boiled down to parameter tuning and to a tailored model for financial time series data.

Table 4. Accuracy of stock markets and cryptocurrencies trend prediction under the third scenario. Results with obtained with the multilayer perceptron network with the hyperparameter setting proposed by [49].

<table>
<thead>
<tr>
<th>Index</th>
<th>CAC40</th>
<th>DAX</th>
<th>FTSE 100</th>
<th>Nasdaq</th>
<th>Nikkei</th>
<th>Crypto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Mean</td>
<td>Min</td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>1 January 2022–1 January 2023</td>
<td>0.4910</td>
<td>0.7513</td>
<td>0.6812</td>
<td>0.7192</td>
<td>0.1281</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.4152</td>
<td>0.6315</td>
<td>0.6814</td>
<td>0.7201</td>
<td>0.2613</td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusions

The aim of this paper is to investigate the predictive power of short-term social media sentiment with respect to both stock and cryptocurrency markets, which have price discovery processes that are well known [81] for being substantially different. Other stylized facts about stocks, cryptocurrencies, and their interplays have been widely documented in the literature [11].

Therefore, we explore the role played by social media sentiment in forecasting the trends of both markets by proposing a toolkit for forecasting in which recent developments in deep learning, sentiment analysis, and hyperparameter tuning are assembled in a unique model. A simple analysis of the accuracy of out-of-sample predictions shows that cryptocurrencies are far more predictable and that short-term sentiment has a substantial influence on model accuracy, especially as compared with data from stock markets. Nonetheless, we find mixed evidence with respect to the predictability of trends when using sentiment analysis. For each dataset in the last scenario, neither a pairwise comparison based on the Wilcoxon test between naive forecasts using historical prices nor predictions including sentiment analysis based on the roBERTa model were found to be statistically different.

Therefore, our findings do not fully support results from previous studies pointing to high and significant predictability of trends according to machine learning and sentiment analysis-based models. Our reassessment of forecasts relies on a robust and updated modelization which is tailored to the peculiar nature of financial time series, i.e., a recently proposed stochastic neural network, coupled with an evolutionary approach controlling for the optimality of two key hyperparameters.

However, it is important to acknowledge limitations that should be considered when interpreting our results. Although several assets and cryptocurrencies have been taken into account, the sample size of our study is relatively small with regard to the time horizon,
which may limit the generalizability of our findings. Although in our contribution we consider different scenarios to control for different market regimes, further work should focus on longer time series and assess how sentiment data-based predictions potentially improve returns, risk, Sharpe Ratios, and accuracy indicators over time. Ideally, further details on the evolution of the predictive power of sentiment should be provided and evaluated across further dimensions. For instance, a breakdown of the results should evaluate whether there are industry-specific patterns and provide a financial interpretation of the predictive power fluctuations of sentiment data over time while allowing for comparisons between different models. Furthermore, our approach relies on a specific social media platform, and it is not certain whether the use of different social network would lead to results that are comparable to our findings. Finally, a thorough assessment of our results might include additional tests, such as the Diebold–Mariano [82] or Pesaran–Timmermann [83] tests.

Further research to improve the performance of neural networks by introducing a fuzzy component [84] and incorporating the predictions to construct optimal portfolios [85] with stocks and cryptocurrencies is currently high on our agenda.


Funding: This research received no external funding.

Data Availability Statement: Data are publicly retrievable on the Yahoo Finance and Twitter websites.

Acknowledgments: The authors would like to thank two anonymous reviewers for their comments and suggestions, which considerably improved the quality and the clarity of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References
6. Li, S.; Huang, Y. Do cryptocurrencies increase the systemic risk of the global financial market? China World Econ. 2020, 28, 122–143. [CrossRef]


28. Wen, Z.; Bouri, E.; Xu, Y.; Zhao, Y. Intraday return predictability in the cryptocurrency markets: Momentum, reversal, or both. *N. Am. J. Econ. Financ.* 2022, 62, 101733. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.