Predicting Stock Price Bubbles in China Using Machine Learning

Yunxi Wang[1*], Tongjai Yampaka[2*]
Chakrabongse Bhuvanarth International Institute for Interdisciplinary Studies (CBIS),
Rajamangala University of Technology Tawan-OK, Bangkok, Thailand[1, 2]
School of Finance, Guangzhou Huashang College, Guangzhou, China[1]

***Abstract*—Financial bubbles have long been a focus of researchers, particularly due to the severe negative impacts following the bursting of financial bubbles. Therefore, the ability to effectively predict financial bubbles is of paramount importance. The aim of this study is to measure and predict the stock market price bubble in China from January 2015 to December 2023. To achieve this, we utilized the GSADF test, currently the most effective, to identify and measure the situation of the stock market price bubble in China. Subsequently, we selected inflation rate, consumer confidence index, stock yield, and price-earnings ratio as explanatory/predictive variables. Finally, four machine learning methods were employed to forecast the stock market price bubble in China. The results indicate that a price bubble occurred in the Chinese stock market during the first half of 2015, before the outbreak of the COVID-19 pandemic in China in January 2020. Furthermore, the comparison reveals that among the machine learning methods, logistic regression is the most suitable and effective for China, while other methods such as deep learning and decision trees also hold certain value.**

***Keywords*—Stock price bubbles; machine learning; Chinese stock market**

I. INTRODUCTION

Asset price bubbles refer to asset prices that exceed their fundamental values, and their occurrence has consistently had significant impacts on the economies of nations and the lives of their citizens [1]. Whether considering global instances, such as the Japanese real estate and stock market bubbles during 1986 to 1991, the late 1990s dot-com bubble in the United States, or from the perspective of China, such as the 2009s Chinese stock market bubble that occurred following the U.S. subprime mortgage crisis, it is evident that financial bubbles exert considerable influence on economies, particularly with regard to adverse effects. When a financial bubble bursts, they can precipitate the collapse of financial institutions and push nations to the brink of bankruptcy. Moreover, they not only impact the development of a single country but sometimes also trigger global financial crises or induce worldwide economic downturns [2]. Generally, following the occurrence of these crises, governments are compelled to allocate substantial resources and implement a variety of measures to attempt to stabilize and salvage the national economy.

Furthermore, for investors and the public, the negative consequences of financial bubbles make it difficult for confidence to be restored in the market. Most of the public lacks experience and risk management abilities, and they are most heavily affected by the bursting of financial bubbles. When they go bankrupt, it causes societal upheaval [3]. Therefore, studying and forecasting financial bubbles are of paramount importance for governments and regulatory authorities. Such endeavors enable governments to implement appropriate economic policies at the right juncture to mitigate the adverse effects of financial bubbles. Moreover, in the current context of economic globalization, where nations and various types of markets are interconnected, the detrimental impacts of financial bubbles can have broader repercussions. China that is the world's second-largest economy possesses unique characteristics and complexities in its stock market. The emergence of a stock market bubble in China not only affects its domestic economy but also has ramifications for the global economy. Consequently, accurate prediction of stock market bubbles in China holds positive implications for both the Chinese and global economies. Such predictions can offer valuable guidance for investors, provide early warnings for financial institutions, and prompt regulatory authorities to take necessary actions to deal with the existence of bubbles.

The Chinese stock market was established in 1990 with the founding of the Shanghai Stock Exchange. From the establishment of the Chinese stock market in 1990 to 1996, there were four price bubbles in the early stage of the Chinese stock market, and each price fluctuation was extremely violent. In 1999, China promulgated the Securities Law, which created a favorable environment for the further development of the Chinese stock market, attracting more investors to participate in the stock market. However, it also led to the re-emergence of stock market price bubbles. Subsequently, in 2001, China's accession to the World Trade Organization (WTO) resulted in a surge of foreign capital inflows, providing significant impetus for the rapid growth of the Chinese economy. This also led to the rapid development of the Chinese stock market, with an expansion in market size and increased trading activity, attracting more investors. Concurrently, the Chinese government implemented a series of reform and opening-up policies, including financial market reform and state-owned enterprise reform, promoting further development and healthy growth of the Chinese stock market. During this period, the Chinese stock market reached a historical high of 2245 points, representing a cumulative increase of 66.7%. Subsequently, it experienced a slow bear market, with the stock index falling to a low of 998 points. From 2007 to 2008, amidst favorable global economic development and China's hosting of the 2008

Olympic Games, the stock index reached a new high of 6124 points in 2007, soaring more than fivefold. However, with the outbreak of the global financial tsunami triggered by the U.S. subprime mortgage crisis, the stock market plummeted to 1664 points in October 2008. This time, the stock market bubble burst rapidly.In the early 2010s, driven by economic growth and increased participation of domestic and foreign investors, the Chinese stock market experienced rapid expansion. However, this period also witnessed market turbulence, especially the stock market crash in 2015, prompting government intervention to stabilize the market. Subsequently, the Chinese stock market underwent further reforms aimed at improving market efficiency and sustainability. Measures such as the introduction of the Science and Technology Innovation Board (STAR Market) and the implementation of IPO registration system aimed to promote innovation and enhance the quality of listed companies. As of the end of 2023, the market capitalization of the Chinese stock market was approximately 85.54 trillion Yuan, while China's GDP in 2023 reached 126.06 trillion Yuan, accounting for approximately 67.86% of China's GDP [4]. There are a total of 5,346 listed companies in the Chinese domestic stock market, with the industries of manufacturing, information transmission, software and information technology services, and wholesale and retail trade ranking among the top three in terms of the number of listed companies [4]. Since its establishment in 1990, the Chinese stock market has experienced rapid development over the past 30 years. However, along with this rapid growth, the Chinese stock market has also encountered a series of issues, particularly manifested in the frequent occurrence of stock market bubbles. The Chinese market is typically sensitive to various rumors, leading to price manipulation of many stocks by rumor mongers. The main reasons for these issues lie in the lack of transparency in the market information and fluctuations in investor sentiment. Therefore, the government and regulatory authorities should remain vigilant at all times to detect financial bubbles promptly and formulate corresponding policies to protect stock market investors, especially retail investors, and stabilize the economic market.

Research on stock price bubbles typically addresses several key questions and objectives, which can be categorized into three main areas. First, it evaluates the factors that contribute to the formation of stock price bubbles. Second, it identifies the early stages of stock price bubbles using the GSADF method. Finally, it develops and validates effective machine learning models and techniques for early detection of stock price bubbles, aimed at improving the accuracy of bubble identification. By addressing these research questions and objectives, this study seeks to provide a comprehensive understanding of stock price bubbles, with a particular focus on the Chinese market, and to offer practical recommendations for enhancing market stability and investor decision-making.

II. Theoretical Literature Review

A. Theoretical Literature Review about Measuring Stock Market Bubble

In research conducted throughout history, there has been a wealth of studies devoted to measuring financial market bubbles. These studies encompass various types of markets, such as stock markets, real estate markets, cryptocurrency markets, and others. Given that this paper focuses on the domain of stock markets and specifically examines price bubbles within this context, it provides a concise overview of measuring bubbles in the stock market domain, with particular emphasis placed on studies employing statistical models applied to time-series data.

Dickey (1979) proposed the Augmented Dickey-Fuller (ADF) test in 1979 to examine whether time series data exhibit unit roots, indicating non-stationarity [5]. In the realm of finance, the ADF test is also utilized to investigate the presence of asset price bubbles. This method, grounded in unit root testing, entails regression analysis of time series data to assess the presence of unit roots within the sequence. The existence of a unit root suggests non-stationarity and the potential existence of a price bubble; conversely, the absence of a unit root indicates stationarity and a lower likelihood of a price bubble. The significance of the test results is typically determined by setting thresholds, thereby ascertaining the presence or absence of a price bubble. Wang (2020) employed the ADF test to evaluate the existence of bubbles in the Chinese stock market [6].

Cheung (1995) introduced the Supremum Augmented Dickey-Fuller (SADF) test as an enhancement to the Augmented Dickey-Fuller (ADF) test [7]. Similar to the ADF test, the SADF test is employed to examine whether time series data possess unit roots, thereby determining the presence of non-stationarity. However, the SADF test introduces the concept of "supremum," allowing for testing across multiple lag lengths and identifying the optimal lag length. By doing so, the SADF test can more accurately ascertain the non-stationarity of time series data and provide more precise unit root test results. Consequently, the SADF test is considered a more reliable method than the ADF test in some cases, particularly when dealing with long or unstable time series data. Homm and Breitung (2012) utilized this test to detect stock market bubbles and, through a process of simulation and comparison of evaluation criteria, determined the SADF test to be the most optimal among the methods employed [8]. While effective in identifying single bubble events, the SADF test may encounter challenges in practical applications where multiple bubbles occur in sufficiently large samples. Although successful in identifying notable historical bubbles, the SADF test failed to detect the bubble associated with the 2007 – 2008 debt crisis.

Phillips et al. (2011) proposed the Generalized Supremum Augmented Dickey-Fuller (GSADF) test as an advancement and refinement of the Supremum Augmented Dickey-Fuller (SADF) test [9]. Similar to the SADF test, the GSADF test is utilized to examine whether time series data exhibit unit roots, thereby determining non-stationarity. However, the GSADF test introduces the Maximized Average Power (MAP) statistic, which allows for the testing of unit root presence and location at each stage, rendering it more flexible in determining the existence and location of unit roots. By considering the possibilities across multiple lag lengths, the MAP statistic enhances the flexibility of the test, leading to a more accurate determination of non-stationarity in time series data. This

enables the GSADF test to be applicable to a wider range of hypotheses and more flexible in determining the existence and location of unit roots. Through the utilization of the GSADF test, researchers can more accurately identify non-stationarity in time series data. Phillips et al. (2015b) employed both the SADF and GSADF tests to empirically apply them to Standard & Poor's 500 stock market data spanning from January 1871 to December 2010[10]. The new GSADF method successfully identified historical events of prosperity and collapse during this period, such as the Panic of 1873 (October 1879 to April 1880) and the Dot-com bubble (July 1997 to August 2001).

Based on the comprehensive review of methods for measuring the stock market domain, we have found that the Generalized Supremum Augmented Dickey-Fuller (GSADF) measurement is currently the most effective among the detection methods. Therefore, in our study of measuring price bubbles in the Chinese stock market, we will utilize the GSADF method.

B. Theoretical Literature Review about Machine Learning in Financial Field

In recent years, machine learning methods have garnered increasing attention from scholars, whether in forecasting financial crises [11], predicting financial bubbles [12][13], or anticipating stock price trends [14] [15]. They all have provided researchers with a novel set of tools and solutions for investigation.

Ouyang and Lai (2021) utilized machine learning algorithms to assess systemic risk warnings in China [11],. Their study revealed that the Attention-Long Short-Term Memory (Attention-LSTM) neural network model within the machine learning algorithms demonstrated higher accuracy compared to other models. This suggests that in the context of China, the Attention-LSTM neural network model holds significant value for systemic risk assessment and early warning.

Başoğlu Kabran and Ünlü (2021) employed machine learning techniques to forecast financial bubbles [12]. They utilized the Support Vector Machine (SVM) algorithm within the domain of machine learning for predicting financial bubbles and compared this approach against alternative methods, concluding that the Support Vector Machine exhibited superior effectiveness in forecasting financial bubbles. The study focused on predicting bubbles within the Standard & Poor's 500 Index.

Tran et al. (2023) employed machine learning methods to predict financial bubbles in the Vietnamese stock market from 2001 to 2021 [13]. They utilized six different algorithms within machine learning to forecast these financial bubbles and compared these algorithm results. Their findings concluded that the Random Forest and Artificial Neural Network algorithms outperformed traditional statistical methods in predicting financial bubbles in the Vietnamese stock market.

Gu et al. (2020) applied machine learning methods to empirical asset pricing [14]. They found that decision trees and neural networks exhibited the best predictive performance among machine learning algorithms. The outstanding predictive capability of these two algorithms primarily stems from their ability to capture complex nonlinear interactions among predictive variables, a task often challenging for other algorithms. Using these two machine learning algorithms yielded performance twice as high as traditional statistical methods. Furthermore, this study identified return reversal and momentum, stock liquidity, stock volatility, and valuation ratios as the most influential factors in asset pricing among the predictive variables.

Zhou et al. (2023) utilized a Deep Neural Network (DNN) model within the domain of machine learning to forecast stock premiums [15] .The research spanned from December 1950 to December 2016, employing monthly data. Stock premiums were computed as the difference between the logarithmic returns of the Standard & Poor's 500 Index (including dividends) and those of risk-free assets. The investigation compared the DNN model from machine learning against the Ordinary Least Squares (OLS) model and Historical Average (HA) model from traditional statistical analysis, ultimately revealing the superior predictive efficacy of the DNN model. Researchers enhanced the predictive capability of the DNN model by incorporating 14 predictive variables. They attributed the DNN model's superior predictive performance primarily to its ability to automatically extract high-dimensional features from data and identify various predictive patterns within the dataset.

Based on the literature discussed above, it is evident that machine learning algorithms exhibit superior performance in classification and time series regression problems. However, it is important to note that the predicted results may vary significantly among different models depending on the dataset utilized, and there is no universally applicable method to ensure consistently superior performance.

Drawing upon the synthesized literature, it becomes apparent that the utilization of machine learning for predicting financial bubbles in the stock market is a relatively novel approach, with limited research attention received thus far. So far, only Başoğlu Kabran and Ünlü (2021) employed machine learning methods to predict bubbles in the S&P 500 index, as well as Tran et al. (2023) in forecasting bubbles in the Vietnamese stock market from 2001 to 2021, as mentioned earlier in the text. Research in the financial domain primarily focuses on predicting financial crises and stock price trends [12] [13]. There is a dearth of corresponding studies in China regarding the prediction of stock market price bubbles, particularly concerning the utilization of machine learning algorithms. To the best of our knowledge, there have been no studies utilizing machine learning methods to forecast stock market price bubbles in China. Therefore, the purpose of this study is to measure and predict the price bubble in the Chinese stock market, and compare the performance of the machine learning algorithms used to select the most suitable model for the price bubble in the Chinese stock market.

III. RESEARCH DESIGN

The primary objective of our study is to measure the stock market price bubbles in China from January 2015 to December 2023, with January 2020 serving as the demarcation point [16], dividing the time period into pre-China COVID-19 and post-China COVID-19 phases, and using carefully selected four

explanatory variables — namely, the inflation rate in macroeconomic factors, the consumer confidence index in sentiment factors, and stock yield and price-to-earnings ratio in market factors to predict the occurrence of stock market price bubbles in China. In this research, we employ the Generalized Supremum Augmented Dickey-Fuller (GSADF) test to identify and measure the presence of stock market price bubbles in China and select four machine learning algorithms for prediction. Ultimately, by comparing the performance results of the four machine learning algorithms, we find the best model for predicting stock market price bubbles in China. This study theoretically contributes empirical evidence to the application of machine learning in forecasting financial bubbles and practically offers early warnings to investors and decision-makers, enabling them to make appropriate financial decisions.

IV. Data and Methodology

A. Data

We utilized the stock market index data of China (Shanghai Composite Index) from January 2015 to December 2023 and employed the Generalized Supremum Augmented Dickey–Fuller (GSADF) method to identify price bubbles in the Chinese stock market during this period. The Chinese stock market index or the Shanghai Composite Index refers to the capitalization-weighted index of all companies listed on the Shanghai Stock Exchange. The monthly dataset of the Chinese stock market comprises 108 data points, while the weekly dataset comprises 470 data points. Among these, it was observed that price bubbles occurred in the Chinese market for 6 months and 25 weeks respectively. The measurement method for price bubbles in the Chinese stock market is the Generalized Supremum Augmented Dickey–Fuller method, which is elaborately described in Section II.A.

In employing machine learning methods, for the convenience of training and testing, we opted for four machine learning algorithm models. We divided the data into two datasets: one for training and the other for testing. Specifically, the training dataset comprises weekly data from January 2015 to December 2023, while the testing dataset comprises monthly data from the same period. The training dataset comprises stock market bubble conditions derived from weekly data publicly disclosed on the official website of the Shanghai Stock Exchange. In contrast, the testing dataset consists of stock market bubble conditions derived from monthly weighted average data disclosed on the same website. Due to the disparate sources of weekly and monthly data, the datasets for training and testing during the same time periods are not identical. However, both datasets all cover the period from January 2015 to December 2023. For instance, the training dataset for January 2015 consists of four weekly data points from that month, whereas the testing dataset consists of the monthly data for January 2015.

The daily and intraday data are unsuitable for this research due to the insufficient labeling of bubbles in the Chinese stock market. Using daily data results in significant classification issues with the labels. To mitigate this problem, the study shifts to analyzing weekly and monthly observations [17]. The selection of evaluation metrics must align with the nature of the classification problem. For such tasks, pertinent metrics include AUC, F-measure, accuracy, precision, and sensitivity [18].

This essentially ensures that the condition of stock price bubbles in the training dataset is four times that in the testing dataset. The purpose of this arrangement is to ensure that both the training and testing datasets contain sufficient data for model development and application in machine learning.

In employing machine learning methods, we incorporated four explanatory variables into the algorithmic model. These four explanatory variables consist of the inflation rate from macroeconomic factors, the consumer confidence index from sentiment factors, and the stock yield and price-earnings ratio from market factors. Within the time frame selected from January 2015 to December 2023, they were also segregated into a testing dataset comprising solely monthly data and a training testing dataset comprising solely weekly data.

The measurement of price bubbles in the Chinese stock market (Shanghai Composite Index) was obtained through the Generalized Supremum Augmented Dickey–Fuller (GSADF) program in the EViews software. For data analysis, we utilized corresponding algorithms in machine learning tools — specifically, logistic regression, deep learning, decision tree, and support vector machine—via the RapidMiner software.

B. Methodology

This study is divided into two parts. The first part involves the detection of price bubbles in the Chinese stock market, while the second part involves the use of four machine learning algorithms to predict the occurrence of price bubbles in the Chinese stock market.

In the first part, we utilized monthly and weekly stock market price data from China spanning from 2015 to 2023 to identify financial bubble occurrences. During this timeframe, with January 2020 marking the dividing line, we segmented the data into pre-COVID-19 pandemic and post-COVID-19 pandemic periods. In January 2020, the Chinese government officially declared the emergence of the COVID-19 pandemic in China and implemented nationwide controls [16]. The purpose of this section of the study using detection methods is to identify the occurrence of price bubbles in the Chinese stock market on a monthly and weekly basis during this period. The monthly and weekly data of the Chinese stock indices (Shanghai Composite Index) were obtained through web scraping from the official website of the Shanghai Stock Exchange.

In the second part, we utilized four machine learning algorithms to forecast price bubbles in the Chinese stock market and employed four explanatory variables to predict the occurrence of price bubbles in the Chinese stock market. The dependent variable is the occurrence of monthly/weekly price bubbles in the Chinese stock market, with the outcomes being the results obtained from the first part of the study. When price bubbles occurred in the Chinese stock market, we assigned a value of 1 to the corresponding month/week, and when price bubbles did not occur, we assigned a value of 0 to the

corresponding month/week. The explanatory variables we employed include the inflation rate from macroeconomic factors, the consumer confidence index from sentiment factors, and the stock yield and price-earnings ratio from market factors. The monthly data for these explanatory variables were sourced from the official website of the National Bureau of Statistics of China and the Wind financial database website. Overall, we selected inflation rate, consumer confidence index, stock yield, and price-earnings ratio, these four significant economic indicators, to forecast price bubbles in the Chinese stock market using their data. Since most of these data are monthly, we utilized the EViews tool to convert monthly data into weekly data.

1) The Generalized Supremum Augmented Dickey–Fuller (GSADF) method for measuring price bubbles in the Chinese stock market: In the first part, the method utilized for measuring price bubbles in the Chinese stock market involved employing the currently most effective time series measurement technique, specifically tailored for detecting asset price bubbles—the Generalized Supremum Augmented Dickey–Fuller (GSADF) test [19]. This method was initially proposed by Philialps et al. (2015b) in 2015 and evolved from the augmented Dickey–Fuller (ADF) test and supremum augmented Dickey–Fuller (SADF) test. It utilizes recursive regression techniques to investigate the presence of unit roots when faced with an alternative right-tail explosion hypothesis, enabling the identification of multiple bubble periods within a time series dataset. Rejection of the null hypothesis during the test indicates the existence of asset price bubbles. In the Generalized Supremum Augmented Dickey–Fuller (GSADF) test, critical values for the test statistics are typically obtained through 2000 Monte Carlo simulations [20], aiding in determining the onset and conclusion of asset price bubbles.

The aim of Generalized Supremum Augmented Dickey–Fuller (GSADF) test was to analyze statistical properties on the upper end of the Augmented Dickey–Fuller (ADF) test concerning a time series. By comparing the maximum values generated from the test statistics with predetermined threshold values obtained from the distribution, analysts can make conclusions about the volatility of the observed values.

Phillips et al. (2015b) proposed a more generalized version of the Supremum Augmented Dickey–Fuller (SADF) test, known as the Generalized Supremum Augmented Dickey–Fuller (GSADF) test [10]. Unlike the original SADF test, which involves fixing the starting point of the sample and progressively recursing through minimum sub-samples to the entire sample, the GSADF test allows for both the starting and ending points of the sample to be flexible. It involves recursively regressing the equation for SADF by simultaneously shifting the starting and ending points of the sample forward. Subsequently, the upper bound of the Augmented Dickey–Fuller (ADF) test is obtained based on this, followed by taking the upper bound of a series of SADF statistics.

The fundamental steps of the Generalized Supremum Augmented Dickey–Fuller (GSADF) test are as follows: first, determine the minimum sample window size k_0. Then, allow the starting point of the sub-sample k_1 and the ending point of the sub-sample k_2 to vary between $[0, k_2 - k_0]$ and $[k_0, T]$, respectively. For each sub-sample in this series, conduct an Augmented Dickey–Fuller (ADF) test to obtain a series of ADF statistics. The formula for constructing the GSADF statistic is shown below in Eq. (1).

$$GSADF(k_0) = \sup_{k_1 \in [0, k_2 - k_0]} \sup_{k_2 \in [k_0, T]} \left\{ ADF_{k1}^{k2} \right\} \quad (1)$$

The Generalized Supremum Augmented Dickey–Fuller (GSADF) test is based on regressing the same equation over a series of sub-samples of the time series data. Its null hypothesis and alternative hypothesis are identical. Therefore, the obtained statistic is compared with the critical value on the right side based on a certain significance level. If the statistic exceeds the critical value, the null hypothesis is rejected, and the alternative hypothesis is accepted: a bubble exists.

For estimating the timing of bubble onset and collapse, given the complex evolution of asset prices, Phillips et al. (2015b) represent the three stages of asset price dynamics with the following Eq. (2) [10]:

$$p_t = p_{t-1} I\{t < \tau_e\} + \rho_n p_{t-1} I\{\tau_e \ll t \ll \tau_f\} + \left(\sum_{k=\tau_f}^{t} \varepsilon_t + P_{\tau_f}^*\right) I\{t > \tau_f\} + \varepsilon_k I\{t \geq\} \varepsilon_k \sim iid(0, \sigma^2) \quad (2)$$

Among them Pn>1, $P_{\tau_f}^* = p_{\tau_e} + p^*$, p_{τ_e} represent asset prices before the formation of a bubble, $P^* = \sum_{t=1}^{\tau_f - \tau_e} \varepsilon_t$ indicates the deviation of prices from pre-bubble levels after the bubble forms, τ_f indicates the moment of bubble burst. when $t < \tau_e$, the asset price sequence p_t follows a unit root process, indicating the absence of bubbles in prices. when $\tau_e << t << \tau_f$, Pn > 1, the asset price series exhibits an explosive process. When $t > \tau_f$, the asset price series reverts to a unit root process. The BSADF statistic is calculated based on recursive selection of samples for upper-bound unit root testing. The Eq. (3) for BSADF is provided below.

$$BSADF_{k_2}(k_0) = \sup\left\{BADF_{k_1}^{k_2}\right\} \; k_1 \in [0, k_2 - k_0], k_2 \in [k_0, T] \quad (3)$$

When the statistic first exceeds its corresponding right-tailed unit root test critical value, it indicates the onset of a bubble. Subsequently, when the statistic first falls below its corresponding right-tailed unit root test critical value, it indicates the collapse of the bubble. However, it is important to note that as the recursive testing selects an increasing sample size, the sample critical values also exhibit an increasing trend. Therefore, it necessitates significant computational effort to calculate finite sample critical values for each sub-sample based on Monte Carlo simulation.

2) Machine learning approaches to forecasting price bubbles in the Chinese stock market

a) Logistic regression: Logistic regression is a statistical method used to model binary classification problems, typically employed to predict the probability of an event occurrence. In this study, we will utilize logistic regression to forecast the presence of price bubbles in the Chinese stock market. Four explanatory variables will be inputted into the model, which ultimately generates the probability of price bubble occurrences in the Chinese stock market. This probability is derived using the Eq. (4) presented below.

$$P(y=1|x)=\frac{1}{1+e^{-(\beta_0-\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n)}} \quad (4)$$

Logistic regression is generally considered a fundamental method in machine learning. This algorithm is relatively easy to understand and implement, making it accessible to a broad spectrum of users. Consequently, due to its excellent interpretability, logistic regression is frequently employed in practical applications within financial institutions.

b) Deep learning: Deep learning is typically regarded as an advanced algorithm within the realm of machine learning. It is a machine learning method based on artificial neural networks, which utilize multi-layered neural network architectures for feature learning and representation learning, thereby achieving the learning and prediction of complex data patterns. The core idea of deep learning involves gradually extracting abstract features from input data through multiple layers of non-linear transformations, enabling the solution of higher-level tasks such as image recognition, speech recognition, and natural language processing.

The advantages of deep learning algorithms lie in their powerful adaptability, capable of learning complex non-linear relationships and applicable to various types of data. They automatically learn feature representations from input data without the need for manual feature engineering. Deep learning also exhibits strong generalization capabilities, enabling learned patterns from the training set to be generalized to unseen data, thereby enhancing the reliability and stability of models in practical applications.

c) Decision tree: The decision tree algorithm is also considered a fundamental machine learning algorithm. It learns and extracts a series of decision rules based on a given dataset through a tree-like structure. This algorithm utilizes metrics such as Gini coefficient or entropy to determine the optimal allocation of each split, ensuring the maximization of purity for each split. In decision trees, decision rules are presented in a tree structure, starting from the root node and traversing through a series of internal nodes to reach the leaf nodes, where each leaf node represents a category or output result.

The advantages of the decision tree algorithm lie in its simplicity, ease of implementation, and interpretability. It also offers flexibility in data handling, hence finding wide application in many fields. However, the performance of the decision tree algorithm may be limited when dealing with complex data and high-dimensional feature spaces.

d) Support vector machine: The Support Vector Machine (SVM) algorithm is an advanced method in machine learning. It belongs to the category of supervised learning algorithms, primarily used for classification and regression analysis. The core idea of SVM is to find a hyperplane that maximizes the margin between classes, thus optimizing classification performance. Alternatively, its fundamental principle is to identify an optimal hyperplane in the feature space that maximally separates samples of different classes while maintaining the maximum margin between classes. The classification in Support Vector Machines is conducted using Eq. (5).

$$f(x)=sign(w\cdot x+b) \quad (5)$$

Where x represents the feature vector of a given new input sample, w is the normal vector to the hyperplane, b is the bias term, and sign( ) denotes the sign function. When w∗x+b is greater than 0, the result is 1, and when it is less than 0, the result is -1. Ultimately, this function result informs us about the class membership of sample x.

The advantages of Support Vector Machines (SVMs) include effective handling of small sample sizes, high-dimensional, and non-linear datasets. For high-dimensional and non-linear data, SVMs can utilize kernel functions to map low-dimensional non-linear separable problems into high-dimensional spaces for linear classification.

Popular machine learning algorithms such as logistic regression, deep learning, decision trees, and support vector machines have shown considerable promise in detecting and predicting stock price bubbles due to their ability to analyze extensive datasets, identify patterns, and adapt to new information. For instance, logistic regression can estimate the probability of a bubble by analyzing historical data. Deep learning methods are particularly effective for anomaly detection, as they learn the typical patterns in data and identify deviations that could signal bubble formation. Decision trees and random forests excel in handling non-linear relationships and interactions between features, making them proficient at recognizing conditions indicative of bubbles. Support vector machines can classify similar data points and detect outliers, which may also suggest bubble formations [21]. Together, these algorithms offer valuable insights into market dynamics and potential bubble developments.

V. Empirical Results and Discussion

A. The Results of Measuring Price Bubbles in the Chinese Stock Market using the Generalized Supremum Augmented Dickey-Fuller (GSADF) Method

In this study, we employed the Generalized Supremum Augmented Dickey-Fuller (GSADF) method to measure the presence of price bubbles in the Chinese stock market from January 2015 to December 2023. The monthly average data and the weekly average data publicly released by the Shanghai Stock Exchange served as the source of the Chinese stock market index (Shanghai Composite Index) for this research. When executing the GSADF procedure using the Eviews software, the study adhered to the program's setting specifying a minimum window of 14 observations. The measurement process commenced from January 2015.

TABLE I. STATISTICAL DATA ON THE OCCURRENCE OF PRICE BUBBLES IN THE CHINESE STOCK MARKET

Serial number	*Bubble occurrence time*	*SSECI price*	*Peak*
1	2015/01/05-2015/01/09	3258.63	0.826567
2	2015/01/12-2015/01/16	3258.21	0.8910738
3	2015/01/19-2015/01/23	3189.73	0.9233272
4	2015/01/26-2015/01/30	3347.26	0.9555806
5	2015/02/02-2015/02/06	3148.14	0.987834
6	2015/02/09-2015/02/13	3063.51	1.2599515
7	2015/02/16-2015/02/17	3206.14	1.532069
8	2015/02/23-2015/02/27	3256.48	1.8041865
9	2015/03/02-2015/03/06	3332.72	2.076304
10	2015/03/09-2015/03/13	3224.31	2.50627625
11	2015/03/16-2015/03/20	3391.16	2.9362485
12	2015/03/23-2015/03/27	3640.10	3.36622075
13	2015/03/30-2015/04/03	3710.61	3.796193
14	2015/04/06--2015/04/10	3899.42	3.6476376
15	2015/04/13--2015/04/17	4072.72	3.4990822
16	2015/04/20--2015/04/24	4301.35	3.3505268
17	2015/04/27-2015/04/30	4441.93	3.2019714
18	2015/05/04-2015/05/08	4441.34	3.053416
19	2015/05/11-2015/05/15	4231.27	2.545104
20	2015/05/18-2015/05/22	4277.90	2.036792
21	2015/05/25-2015/05/29	4660.08	1.52848
22	2015/06/01-2015/06/05	4633.10	1.020168
23	2015/06/08-2015/06/12	5045.69	0.68747
24	2015/06/15-2015/06/19	5174.42	0.354772
25	2015/06/22-2015/06/26	4471.61	0.022074

Table I presents the results of identifying price bubbles in the Chinese stock market. This table provides the time occurrence of price bubbles in the Chinese stock market, the overall market prices of the Chinese stock market (Shanghai Composite Index prices) for each period, and the peak values calculated for each bubble period. In Fig. 1, we visually illustrate the time periods during which price bubbles occurred in the Chinese stock market from January 2015 to December 2023. The blue line in the Fig. 1 represents the GSADF statistic sequence, while the orange line denotes the asymptotic critical values obtained from 2000 Monte Carlo simulations using the EViews software tool. By comparing the GSADF statistic sequence (blue line) with the 95% critical value sequence (orange line), the timing of overall market price bubbles in the Chinese stock market (represented by the Shanghai Composite Index prices) is identified. During this period, there were six months with price bubbles on a monthly basis and 25 weeks experiencing financial bubbles on a weekly basis. Notably, we observe a prolonged financial bubble in the first half of 2015. Following the identification of price bubbles in the Chinese stock market from January 2015 to December 2023, we designate months/weeks with identified occurrences of stock market price bubbles as 1, while months/weeks without price bubbles are marked as 0. This preparation aims to facilitate the subsequent creation of datasets for the four machine learning prediction stages.

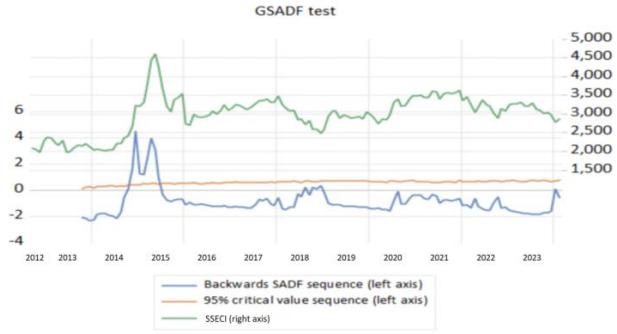

Fig. 1. Chinese Stock market price bubbles from January 2015 to December 2023.

B. The Result of Predicting the Price Bubble in the Chinese Stock Market Using Four Machine Learning Algorithms

In this study, we will employ four machine learning algorithms to predict the occurrence of price bubbles in the Chinese stock market. These four machine learning algorithms are logistic regression, deep learning, decision tree, and support vector machine.

For machine learning models, we optimize the models using the hyperparameter of Area Under the Curve (AUC). In machine learning, particularly in evaluating binary classification models, AUC typically refers to the area under the Receiver Operating Characteristic (ROC) curve. AUC quantifies the entire two-dimensional area underneath the ROC curve, providing a single measure to assess the classifier's performance across various thresholds, with values ranging between 0 and 1. A higher AUC value indicates better model performance, while an AUC value closer to 0.5 suggests performance closer to random guessing. Throughout this process, various hyperparameter values are experimented with to enhance the model. Post-training, AUC is computed using the test dataset. The hyperparameter combination resulting in the highest AUC is designated as optimal. This approach ensures the selection of hyperparameters based on the model's classification performance, with AUC serving as the key metric.

The following Table II presents the performance of the four machine learning algorithm models utilized in our study.

TABLE II. THE PERFORMANCE RESULTS OF THE FOUR MACHINE LEARNING ALGORITHM MODELS

Algorithm	*AUC*	*F Measure*	*Accuracy*	*Precision*	*Sensitivity*
Logistic Regression	1	86.7%	98.5%	90.0%	90.0%
Deep Learning	1	79.3%	96.3%	70.0%	100.0%
Decision Tree	0.992	80.0%	97.8%	90.0%	80.0%
Support Vector Machine	0.558	Not Available	94.8%	Not Available	0.0%

From Table II, we observe that the performance of the logistic regression model surpasses that of other algorithms in terms of AUC, F-measure, accuracy, and precision. The logistic regression model achieves an AUC of 1, an F-measure of 86.7%, accuracy of 98.5%, and precision of 90.0%, all of which are the highest among all algorithms. This indicates its capability to accurately classify the presence of bubbles in the Chinese stock market. However, in terms of sensitivity, the logistic regression model exhibits a lower value compared to the deep learning model, at 90.0%, suggesting a relatively weaker ability of the logistic regression model to correctly identify positive instances. This outcome suggests that while the logistic regression model demonstrates excellent performance in predicting instances of stock market bubbles in China, it may lack flexibility in handling certain types of data and feature representations, leading to relatively lower performance in identifying positive instances.

Furthermore, the deep learning model exhibits perfect performance in terms of AUC and sensitivity, with values of 1 (100.0%), indicating that the model can perfectly distinguish between positive and negative instances at all possible thresholds, without any misclassifications. It can perfectly identify all positive instances without missing any. However, the results for F-measure (79.3%), accuracy (96.3%), and precision (70.0%) suggest that the deep learning model, in predicting the occurrence of bubbles in the Chinese stock market, strikes a compromise between precision and recall, resulting in a certain number of misclassifications overall, with a higher rate of false positives when predicting positive instances. In contrast, the decision tree algorithm performs moderately across all aspects and can serve as a baseline for evaluating the performance of these four machine learning models. Meanwhile, the support vector machine model either performs the worst in all aspects or yields results that are not available, indicating its unsuitability for predicting the occurrence of bubbles in the Chinese stock market.

Solely based on the AUC scores of model performance, we observe that within the machine learning algorithms utilized, both the logistic regression model and the deep learning model achieved perfect scores of 1. This score signifies their ability to maintain a low false positive rate while achieving a high true positive rate. Essentially, this value indicates their proficiency in distinguishing periods of stock market bubbles from those without. However, upon considering the other four performance metrics, overall, the logistic regression model outperforms. Solely based on AUC scores, the other two models — the decision tree model and the support vector machine model — exhibit relatively lower scores. While the decision tree model's score (0.992) demonstrates some competitiveness, the score of the support vector machine model (0.558) indicates relatively poor performance in classification tasks, akin to random guessing. Although the latter two models — the decision tree model and the support vector machine model — may offer some insights, their performance in analyzing the occurrence of stock market bubbles in China lags behind that of logistic regression and deep learning.

Based on the above results, we compared the outcomes of four machine learning methods. These four machine learning algorithms are commonly employed approaches for addressing classification problems in finance. Logistic regression and decision tree are considered fundamental machine learning methods, while deep learning and support vector machine are classified as advanced machine learning methods. The findings indicate that the fundamental machine learning methods (logistic regression and decision tree) outperform the advanced machine learning methods (deep learning and support vector machine) in terms of F-measure, accuracy, and precision. Overall, this suggests that in the specific domain of predicting stock market price bubbles in China, simple fundamental machine learning methods may be more suitable, and there may be no need to blindly pursue complex advanced algorithms, as doing so may yield counterproductive outcomes.

Our research findings differ from Başoğlu Kabran and Ünlü (2021), who utilized machine learning methods to predict the S&P 500 index and concluded that SVM was the best approach [12]. There are two reasons for these discrepancies. First, differences exist in the explanatory variables selected for the input models. Second, variations in the sizes of the datasets utilized by both studies contribute to these disparities. However, it is noteworthy that our study is the first to employ a comparative approach involving multiple machine learning methods to forecast market bubbles in China, the second-largest economy globally. Moving forward, we plan to conduct broader empirical research within the Chinese market context.

Our research results differ from those of Tran et al. (2023), who applied machine learning methods to predict the Vietnamese stock market from 2001 to 2021, concluding that random forest and artificial neural network algorithms outperformed traditional statistical methods in forecasting financial bubbles in the Vietnamese stock market [13]. There are three main reasons for these discrepancies. Firstly, differences exist in the explanatory variables selected as inputs to the models. Secondly, disparities in the time periods of the datasets used in both studies contribute to the variations observed. Lastly, discrepancies arise from the distinct machine learning methods employed in each study. In contrast, our study represents the first comprehensive application of multiple machine learning methods to predict stock market bubbles in China, the world's second-largest economy. Looking ahead, we plan to conduct broader empirical research in diverse market contexts and with a wider array of machine learning methodologies.

C. Robustness Test

In order to ensure the accuracy and reliability of the machine learning models obtained, we conducted robustness tests on them. For this purpose, we divided the data into two equal parts, the first part covering the period from January 2015 to December 2018, and the second part covering the period from January 2020 to December 2023. The main reason for this division is that in January 2020, the Chinese government officially announced the emergence of the COVID-19 pandemic in China and implemented nationwide controls [16]. We utilized the two best-performing machine learning models, namely the logistic regression model and the Deep Learning model, to predict the occurrence of stock market price bubbles during these two data set periods.

Subsequently, we trained and tested the models within their respective data sets. Afterwards, we evaluated the performance of the obtained models using relevant metrics, including accuracy, AUC, and sensitivity. Finally, we analyzed the results of the robustness tests conducted for each time period to compare the performance of the models in different time periods.

From Table III, we can observe that the accuracy of both models remains stable across the two time periods, with the logistic regression model averaging 94.05% and the deep learning model averaging 97.75%. Regarding the AUC, both logistic regression and deep learning models maintain stability across the two time periods, with average values of 0.978 and 1, respectively. However, we note a sensitivity decline in the logistic regression model towards the dataset, particularly during the period from January 2020 to December 2023. In contrast, the deep learning model demonstrates more consistent performance in sensitivity. Overall, both the logistic regression and deep learning models exhibit robustness.

TABLE III. THE ROBUSTNESS TEST RESULT

	Logistic regression			***Deep learning***		
	January 2015 - December 2018	***January 2020 - December 2023***	***Average***	***January 2015 - December 2018***	***January 2020 - December 2023***	***Average***
Accuracy	93.3%	94.8%	94.05%	100.0%	95.5%	97.75%
AUC	1	0.956	0.978	1	1	1
Sensitivity	100.0%	90%	95%	100.0%	100.0%	100.0%

The stability testing method employed in this study ensures the reliability of the predictive models for detecting stock market price bubbles in China, allowing for a clear understanding of variations in model performance over time. This facilitates making practical decisions in real-world financial applications.

Logistic regression is ideally suited for binary outcomes, making it an excellent option for identifying the presence or absence of a bubble. Given the small size of the dataset, deep learning models tend to underperform relative to logistic regression. However, it's important to note that larger datasets come with their own set of challenges.

D. Summary of discussion

Stock price bubbles prediction applying advanced machine learning techniques is potentially extends existing financial theories. It offers empirical evidence that can either support or challenge traditional models of bubble formation and economic cycles. The adaptability and continuous learning capability of machine learning models underscore the dynamic nature of financial bubbles and economic cycles.

The explanatory variables we employed include the inflation rate from macroeconomic factors (IR), the consumer confidence index from sentiment factors (CCI), the stock yield (SY), and price-earnings ratio from market factors (RET).

Table IV displays the relative importance of different attributes (variables) in predicting an outcome, likely in a logistic regression model. The inflation rate from macroeconomic factors (IR) has the highest weight, indicating it is the most important predictor in the model. Its relative importance value of 0.539 suggests it contributes significantly more to the prediction compared to the other attributes. The consumer confidence index from sentiment factors (CCI) attribute is the second most important predictor. Its weight of 0.154 indicates that while it is less influential than IR, it still plays a substantial role in the model. The stock yield (SY) attribute has a weight of 0.116, making it the third most important predictor. Its contribution is notable but less significant compared to IR and CCI. The price-earnings ratio from market factors (RET) attribute has the smallest weight of 0.018, indicating it has the least influence on the prediction. Its relative importance is minimal compared to the other attributes. The model relies heavily on the IR attribute for its predictions, which means understanding and accurately measuring this variable is critical. While CCI and SY are important, their contributions are secondary. Adjustments or improvements in measuring these variables could still enhance model performance.

TABLE IV. THE RELATIVE VARIABLE IMPORTANCE VALUES IN THE CHINESE STOCK MARKET (SHANGHAI COMPOSITE INDEX)

Variable	***Weights (Importance Value)***
inflation rate(IR)	0.539
consumer confidence index(CCI)	0.154
stock yield (SY)	0.116
price-earnings ratio (RET)	0.018

VI. CONCLUSIONS

In this study, we employed the widely acknowledged Generalized Supremum Augmented Dickey-Fuller (GSADF) method to identify the presence of price bubbles in the stock market and utilized data spanning from January 2015 to December 2023 to forecast the occurrence of price bubbles in the Chinese stock market. The findings reveal that a price bubble occurred in the Chinese stock market during the first half of 2015, before COVID-19, while no financial bubbles were observed at other times. Among the predictive models, the logistic regression model demonstrated the best performance with an F-measure score of 86.7%, followed by the deep learning model and the decision tree model, which exhibited slightly inferior yet respectable performance, with F-measure scores of 79.3% and 80.0%, respectively. From a practical standpoint, these results furnish valuable machine learning models for real-time detection and prediction of stock market price bubbles, thereby enabling governmental decision-makers, regulatory authorities, and market oversight agencies to formulate and implement corresponding economic policies aimed at mitigating the adverse effects stemming from financial bubbles. From a theoretical perspective, the utilization of diverse machine learning algorithms in predicting

financial bubbles in this study holds significant reference and generalization implications for the application of machine learning techniques in financial market research. Moreover, the macroeconomic factor (inflation rate), investor sentiment factor (consumer confidence index), and market factors (stock yield and price-earnings ratio) into the machine learning prediction models enables us to delve further into the complex mechanisms underlying the emergence of financial market bubbles and advance the predictive understanding of such phenomena.

This study has made significant contributions both theoretically and practically, particularly in utilizing machine learning, a novel tool, to forecast price bubbles in the stock market of China that is the world's second-largest economy, providing empirical evidence.The research findings highlight the suitability of the foundational algorithm in machine learning, the logistic regression model, for predicting price bubbles in the Chinese stock market. Nevertheless, other machine learning algorithms such as deep learning and decision tree algorithms also exhibit potential in the domain of financial bubble prediction.This study highlights to policymakers and regulators the significance of promptly enacting policies to reduce both the probability and ramifications of financial bubbles.For central banks and regulatory bodies, utilizing advanced machine learning tools to measure financial bubbles facilitates the formulation of appropriate monetary policies to regulate capital behavior in the economy, thereby reducing speculative activities in financial asset trading and stabilizing the entire financial system.For investors, based on the insights gleaned from this study, they can more effectively allocate their investment portfolios by leveraging machine learning algorithms ability to predict financial price bubbles, deciding opportune moments for long and short positions. Moreover, during market bubble occurrences, i.e., when market prices are excessively high, investors can seize suitable opportunities to sell assets and generate corresponding profits.

In subsequent research, other scholars can utilize the machine learning methods employed in this study to forecast bubble situations in varying locales markets, such as Hong Kong, Singapore, the United States, and others. These machine learning methods can likewise be harnessed to anticipate bubbles across diverse market categories, including but not limited to the real estate market, cryptocurrency market, etc. Furthermore, analysis can be conducted on the interplay of financial bubbles between dissimilar markets, such as the stock market and real estate market, both of which hold significant economic sway. Understanding these dynamics can enable regulatory authorities to implement effective financial policies, thereby preventing the formation of financial bubbles or controlling their occurrence, ultimately fostering a healthy and robust market investment environment. This study utilized a limited number of machine learning models, which may have resulted in certain limitations in the obtained results. In the future, employing a wider variety of machine learning models can further advance research in the prediction of stock market bubbles. In addition, speculative bubbles across different markets exhibit both common characteristics and distinct features and impacts. Investigating the relationships between speculative bubbles in various markets and understanding the mechanisms of bubble transmission to develop more effective regulatory strategies presents a significant challenge for future research. Furthermore, addressing the issue of insufficient datasets is a critical concern that needs to be prioritized.

ACKNOWLEDGMENT

This article is part of the Doctor of Philosophy in Digital Transformation and Business Innovation program at the Chakrabongse Bhuvanarth International Institute for Interdisciplinary Studies (CBIS) at Rajamangala University of Technology Tawan-ok in Thailand. The researchers would like to thank all the cited experts and reviewers involved in this study.Yunxi Wang, the author of this article,would like to express special gratitude to her PhD advisor-- Tongjai Yampaka, for the assistance and support provided during her studies.

REFERENCES

[1] Xiong W., Yu J. L. 2011. The Chinese warrants bubble. American Economic Review 101(6), 2723–2753.

[2] Miao J. J., Wang P. F. 2018. Asset bubbles and credit constraints. American Economic Review, 108(9), 2590–2628.

[3] Galbraith, James K., Sara Hsu, and Wenjie Zhang. 2009. Beijing bubble, Beijing bust: Inequality, trade, and capital inflow into China. Journal of Current Chinese Affairs 38: 3–26.

[4] National Bureau of Statistics. 2024. China economic annual report 2023. China: National Bureau of Statistics.

[5] Dickey D. A., Fuller W. A. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74(366a), 427-431.

[6] Wang Y., Wu C. 2020. Testing for bubbles in Chinese stock market: A study based on ADF test. International Journal of Finance & Economics 25(4), 530-544.

[7] Cheung Y. W., Lai K. S. 1995. Lag order and critical values of the augmented Dickey-Fuller test. Journal of Business & Economic Statistics 13(3), 277-280.

[8] Homm, Ulrich, and Jörg Breitung. 2012. Testing for speculative bubbles in stock markets: A comparison of alternative methods. Journal of Financial Econometrics 10: 198–231.

[9] Phillips, Peter C. B., Yangru Wu, and Jun Yu. 2011. Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? International Economic Review 52: 201–26.

[10] Phillips, Peter C. B., Shuping Shi, and Jun Yu. 2015b. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International Economic Review 56: 1043–78.

[11] Ouyang, Zi-sheng, and Yongzeng Lai. 2021. Systemic financial risk early warning of financial market in China using Attention-LSTM model. The North American Journal of Economics and Finance 56: 101383.

[12] Başoğlu Kabran, Fatma, and Kamil Demirberk Ünlü. 2021. A two-step machine learning approach to predict S&P 500 bubbles. Journal of Applied Statistics 48, 2776–2794.

[13] Tran K.L., Le H.A., Lieu C.P., and Nguyen D.T. 2023. Machine Learning to Forecast Financial Bubbles in Stock Markets: Evidence from Vietnam. International Journal of Financial Studies 11(4), 133.

[14] Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33: 2223–73.

[15] Zhou, Xianzheng, Hui Zhou, and Huaigang Long. 2023. Forecasting the equity premium: Do deep neural network models work? Modern Finance 1: 1–11.

[16] Sweeny K., Rankin K., Cheng X., Hou L., Long F., Meng Y., ... and Zhang W. 2020. Flow in the time of COVID-19: Findings from China. PloS One 15(11), e0242043.

[17] Li X., Wang Z. 2021. Challenges and solutions in financial market labeling and classification. Journal of Financial Data Science 3(4), 47-59.

[18] John G. H., Langley P. 1995. Performance measures for classification problems. Machine Learning 33(1), 103-139.

[19] Shimizu R., Weber E. 2020. Identifying Asset Price Bubbles Using the Generalized Supremum Augmented Dickey-Fuller Test. Journal of Financial Econometrics 18(4), 679-707.

[20] Hansen B. E. 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of Econometrics 93(2), 345-368.

[21] Krauss C., Do X., Huck N. 2017. Deep neural networks for financial prediction: A comparison of deep learning approaches. European Journal of Operational Research 256(1), 185-200.