WORKING PAPER 270/2024

DETECTING AND FORECASTING FINANCIAL BUBBLES IN THE INDIAN STOCK MARKET USING MACHINE LEARNING MODELS

Mahalakshmi Manian
Parthajit Kayal

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

October 2024

Detecting and Forecasting Financial Bubbles in The Indian Stock Market Using Machine Learning Models

Mahalakshmi Manian
Research Scholar,

and

Parthajit Kayal
(corresponding author), Assistant Professor, Madras School of Economics, Chennai
parthajit@mse.ac.in

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

October 2024

WORKING PAPER 270/2024

October 2024

Price : Rs. 35

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

Phone: 2230 0304/2230 0307/2235 2157
Fax: 2235 4847/2235 2155
Email : info@mse.ac.in
Website: www.mse.ac.in

Detecting and Forecasting Financial Bubbles in The Indian Stock Market Using Machine Learning Models

Mahalakshmi Manian and Parthajit Kayal

Abstract

This research investigates the phenomenon of economic or financial bubbles within the Indian stock market context, characterized by pronounced asset price inflation exceeding the intrinsic worth of the underlying assets. Leveraging data from the NIFTY 500 index spanning the period 2003 to 2021, the study utilizes the Phillips, Shi, and Yu (PSY) method (Phillips et. al., 2015b), which employs a right-tailed unit root test, to discern the presence of financial bubbles. Subsequently, machine learning algorithms are employed to predict real-time occurrences of such bubbles. Analysis reveals the manifestation of financial bubbles within the Indian stock market notably in the years 2007 and 2017. Moreover, empirical evidence underscores the superior predictive efficacy of Artificial Neural Networks, Random Forest, and Gradient Boosting algorithms vis-à-vis conventional statistical methodologies in forecasting financial bubble occurrences within the Indian stock market. Policymakers should use advanced machine learning techniques for real-time financial bubble detection to improve regulation and mitigate market risks.

Keywords: *Financial Bubbles; Machine Learning; K-nearest Neighbour; Random Forest Classifier; Artificial Neural Network; Naïve Bayes*

JEL Codes: *G1, G2, G3, C1, C5*

Acknowledgement

The authors express their gratitude to the anonymous referees for their insightful comments, which were incorporated into the revised version of the paper, subsequently accepted for publication in the IIMB Management Review journal.

Mahalakshmi Manian
Parthajit Kayal

INTRODUCTION

The "Bubble Act" enacted by the British Parliament in 1720 marked the formal recognition and codification of the term "bubble." An economic or financial bubble manifests when asset prices, whether physical or financial, experience a rapid escalation that substantially surpasses the intrinsic value of the underlying asset (Banerjee and Kayal, 2022). A defining characteristic of bubbles is the divergence between asset price growth and fundamental value, leading to a decoupling phenomenon. This precipitous surge in prices typically precedes a significant downturn, culminating in a crisis (Singh *et. al.*, 2018). Economic bubbles exhibit diverse manifestations, encompassing stock market bubbles, asset bubbles, credit market bubbles, and commodity bubbles (Talin, 2022). The historical record reveals the recurrence of financial crises triggered by bubbles across various nations, with asset bubbles punctuating economic epochs at intervals often spanning several decades, invariably followed by periods of economic retraction and expansion.

Within the policy discourse surrounding bubbles, a central inquiry pertains to the appropriate course of action for policymakers when asset valuations experience rapid escalation devoid of commensurate alterations in anticipated dividend payouts. This scenario raises apprehensions among policymakers, as well as the broader public, as it potentially signifies the overvaluation of assets, thereby heightening the likelihood of a precipitous decline in value (Barlevy, 2018). Such circumstances prompt deliberations on policy interventions aimed at mitigating the risks associated with unsustainable asset price dynamics and averting potential systemic repercussions within the financial ecosystem.

The rupture of a bubble can precipitate the collapse of major financial institutions, precipitating sovereign insolvency and unleashing comprehensive financial and economic upheavals. Subsequent to such crises, governments find themselves compelled to allocate substantial

resources toward recovery endeavours and implementation of bailout measures. Moreover, enduring societal repercussions persist, with the restoration of public confidence in the market posing a formidable challenge. Individuals lacking extensive experience in navigating financial hazards are particularly vulnerable to adverse consequences (Galbraith *et. al.*, 2009). Such ramifications underscore the imperative for proactive regulatory and policy measures aimed at pre-empting and mitigating the adverse effects of bubble bursts on both financial stability and societal welfare.

Consequently, the assessment and anticipation of financial bubbles assume paramount importance for governmental entities and market regulatory bodies. Such endeavours facilitate the implementation of requisite measures aimed at mitigating the deleterious repercussions of financial bubbles on both the economy and society, particularly in the context of globalization and the seamless transmission of risks across markets (Tran *et. al.*, 2023). By proactively identifying and addressing burgeoning bubble dynamics, policymakers can enhance systemic resilience, foster market stability, and mitigate the potential contagion effects that could precipitate broader financial crises.

Hence, an asset exhibiting a trading price that diverges significantly from its intrinsic value aligns with the classification of a bubble as delineated by economists. Extant literature, including works by Banerjee and Kayal (2022), Chen *et. al.* (2023), Wöckl (2019), Gerdesmeier *et. al.* (2013), Jarrow and Kwok (2021), and Shi and Phillips (2022), has empirically delved into the identification of financial and economic bubbles across diverse financial markets. The PWY (Phillips, Wu, and Yu) technique, also known as the Sup Augmented Dickey–Fuller test (SADF), was proposed by (Phillips *et. al.*, 2011) as a methodological tool for ascertaining the presence of rational bubbles within financial markets. This approach hinges on the unit root null hypothesis, akin to the conventional right-tailed alternative hypothesis framework utilized in the Dickey–Fuller test.

The Generalized sup ADF (GSADF) test, introduced by Phillips *et. al.* (2015a), represents an advancement over the SADF methodology, commonly referred to as the Phillips, Shi, and Yu (PSY) process. This innovation addresses limitations inherent in the SADF approach. The GSADF test employs an iterative application of the right-tailed ADF test, leveraging a rolling-window framework to detect potentially explosive patterns within sample sequences. By systematically examining fluctuations in asset prices, the GSADF test enhances the capacity to identify and characterize emerging bubble dynamics within financial markets.

Due to its enhanced rolling window flexibility compared to the SADF method, the GSADF test serves as a valuable instrument for scrutinizing price explosion dynamics and confirming the presence of market bubbles. Empirical evidence from various theories supports the reliability of the GSADF test in detecting market bubbles (Tran *et. al.*, 2023). Consequently, in the context of this study aimed at identifying bubbles within the Indian stock market, we employ the PSY technique as a methodological framework.

The National Stock Exchange of India Limited (NSE) stands as a pivotal institution within the Indian financial landscape, serving as the primary platform for securities trading and exerting significant influence on the nation's financial system. Reflecting India's commitment to modernizing its financial infrastructure and its robust economic growth, the NSE symbolizes a cornerstone of the nation's financial progress. As India's economy continues to evolve, the NSE is poised to assume an increasingly prominent role for investors, regulators, and market participants alike. Facilitated by global trade networks and sophisticated electronic trading platforms, the NSE is intricately interconnected with major international stock exchanges and global financial markets. This linkage fosters the free flow of international capital, enhances market efficiency, and facilitates cross-border investment activities.

As one of the foremost stock exchanges in emerging economies, the NSE serves as a barometer of the Indian economy, offering insights into corporate dynamics, investor sentiment, and overall economic trajectory. Its performance, along with key indices such as the Nifty 50 and NIFTY 500, garners close attention from global investors, analysts, and policymakers seeking exposure to burgeoning market opportunities in developing economies. Moreover, the NSE serves as a conduit for directing foreign investment into various sectors of the Indian economy, thereby influencing foreign direct investment (FDI) and foreign institutional investment (FII) inflows.

The performance of the NSE is intrinsically linked to a multitude of factors, including prospects for economic development, regulatory changes, market stability, and investor confidence. Consequently, fluctuations in the NSE's performance are influenced by macroeconomic variables, geopolitical events, investor sentiment, and global market trends. Given these intricate linkages and the potential implications for market stability and investor protection, regulatory authorities, especially Securities and Exchange Board of India (SEBI) and Reserve Bank of India (RBI), must remain vigilant in promptly identifying and addressing emerging financial bubbles to safeguard the interests of investors and preserve overall market integrity.

In scholarly discourse, machine learning (ML) methodologies have emerged as viable alternatives to traditional statistical approaches for time series forecasting. While various definitions of ML abound in the literature, one of the most widely cited originates from Arthur L. Samuel, a trailblazer in artificial intelligence (AI). Samuel (1959) succinctly defines ML as "the field of study that gives computers the ability to learn without being explicitly programmed." Building upon this foundation, Masini *et. al.* (2023) aptly characterizes ML as a process that entails uncovering and elucidating latent patterns within vast and intricate datasets through the integration of robust statistical techniques with automated computational

algorithms. Consequently, the statistical underpinning of ML derives from the principles of statistical learning theory.

Within this framework, this study endeavours to contribute to the existing body of literature by undertaking an examination aimed at discerning the existence of financial bubbles within the Indian stock market. Leveraging the NIFTY 500 index, which encompasses approximately 93 percent of free float market capitalization, this research seeks to illuminate the prevalence of financial bubbles within the Indian stock market spanning the period from 2003 to 2021. Subsequently, the paper endeavours to forecast future occurrences of financial bubbles by integrating macroeconomic variables into the analysis. To achieve this objective, we employ the PSY process as a methodological framework to detect market phases characterized by bubble-like attributes and to facilitate predictions utilizing ML algorithms.

Furthermore, we applied the Synthetic Minority Over-Sampling Technique (SMOTE) to rectify data imbalances within the dataset, thereby enhancing the robustness of our forecasting outcomes. Our overarching objective is to discern the optimal performing model among the various approaches considered. This endeavour holds the potential to furnish investors and governmental bodies with early warning signals, thereby empowering them to make well-informed financial decisions.
The subsequent sections of this research paper are organized as follows: Section 2 delineates our data sources and methodology, encompassing the PSY process and ML methodologies utilized. Section 3 encapsulates our study findings, accompanied by an analysis of their ramifications. Finally, Section 4 offers a succinct conclusion and summary, encapsulating the primary findings and conclusions derived from the study.

DATA AND METHODOLOGY

Phase 1 - Bubble Detection

Data Collection and Pre-Processing

The data collection process proceeded in two stages. Initially, for the identification of financial bubbles within the NIFTY 500, dividend yield data spanning from January 2003 to December 2021 was acquired. This dataset was sourced from Trendlyne[1], a reputable and widely trusted stock market analytics platform. Subsequently, the daily dividend yield of the NIFTY 500 was aggregated and averaged to obtain monthly data over the same period. The calculated dividend yield was then utilized to compute the price-dividend ratio, a crucial variable employed in the modeling process. This study period encompasses a diverse array of economic events and global fluctuations, notably during the periods of 2006-2008 and 2017-2018, as documented in various studies (Banerjee and Kayal, 2022; Singh *et. al.*, 2018). The dataset comprises 252 months, of which 8 months have been identified to exhibit characteristics indicative of a financial bubble.

Research Design - PSY Method for Bubble Detection

The PSY method was deployed to scrutinize the presence of a unit root under an alternative right-tailed hypothesis. This methodology, aimed at detecting multiple periods of bubbles within time series data, was initially introduced by (Phillips *et. al.*, 2015b). This original right-tailed unit root test has proven instrumental in delineating the onset and culmination dates of asset price bubbles, serving as an early warning mechanism across diverse financial markets for detecting bubble-like phenomena (Hu, 2023).

The PSY method represents an advancement over the PWY approach, designed to identify periodically collapsing bubbles. While the PWY method is effective in detecting singular instances of bubble

[1] Trendlyne website can be accessed at: https://trendlyne.com/markets-today-loggedin/

episodes, its utility is limited to such occurrences. In contrast, the PSY method enhances the modelling framework by enabling the detection of multiple episodes of bubbles within the dataset. This augmentation empowers the analytical framework to capture the nuanced dynamics of bubble phenomena across various temporal intervals, thereby enhancing the accuracy and comprehensiveness of the analysis.

Essentially, the rejection of the null hypothesis serves as both statistical and empirical evidence indicating the occurrence of a financial bubble within this testing methodology. Critical values for these tests are established through Monte Carlo simulations, the outcomes of which play a pivotal role in delineating and identifying the commencement and conclusion dates of financial bubbles. This approach leverages rigorous statistical techniques to ascertain significant deviations from expected market behaviour, thereby facilitating the precise identification of bubble phenomena within the dataset.

The null hypothesis is a martingale with an asymptotic drift specified as in Eq. 1.

$$H_0 : y_t = dT^{-\eta} + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma^2) \tag{1}$$

d is a constant, the localizing coefficient η is greater than 1/2 and T is the sample size (Hu, 2023).

$$H_1 : y_t = \delta_T y_{t-1} + \varepsilon_t \tag{2}$$

where, $\delta_T = 1 + cT^{-\theta}$, $c > 0$ and $\theta \in (0,1)$

The objective is to ascertain statistical metrics pertaining to the right tail of the Augmented Dickey-Fuller (ADF) test in relation to a designated time series. Under the null hypothesis, the time series data is assumed to conform to a random walk process characterized by a minute drift coefficient, as computed through the following regression analysis in Eq. 3.

$$\Delta y_t = \mu + \sigma y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-1} + e_t \quad (3)$$

where y_{t-1} represents stock prices at time t; m is the intercept; p is the maximum lag; ϕ_i are the regression coefficients corresponding to different lags; and e_t is the error term (Tran *et. al.*, 2023).

The BSADF serves as the test statistic utilized, defined as the supremum value of the ADF statistic sequence. It represents the maximum ADF test statistic within the right tail (see Eq. 4).

$$BSADF_{r2}(r_0) = sup_{r1\in[0,r2-r0]}\left(ADF_{r_1}^{r_2}\right) \quad (4)$$

In order to pinpoint the onset and cessation dates of the bubbles, the BSADF statistic and its corresponding critical values are employed, as defined below. (see Eq. 5-6).

$$r_e = inf_{r_2\epsilon[r_0,1]}\left(r_2 : BSADF_{r_2} > cv_{r_2}^{\beta_T}\right) \quad (5)$$

$$r_f = inf_{r_2\epsilon[r_e,1]}\left(r_2 : BSADF_{r_2} > cv_{r_2}^{\beta_T}\right) \quad (6)$$

For preliminary estimation purposes, it is imperative that the minimum window size r_0 be adequately large, yet not excessively so as to potentially overlook an early bubble episode. As per the recommendation in PSY (Phillips *et. al.*, 2015b), the minimum window size r_0 is determined as $0.01 + 1.8\sqrt{T}$.

The comprehensive process elucidated above for bubble detection has been encapsulated and integrated into the 'psymonitor' package[2] in R. This package has been utilized in the present study to detect financial bubbles within the NIFTY 500 index, employing the PSY procedure as proposed by (Phillips *et. al.*, 1984).

[2] Details of this package is available at: https://itamarcaspi.github.io/psymonitor/

Phase 2 - Bubble Prediction using ML Algorithms

Data Collection and Pre-Processing

The analysis conducted during Phase 1 served as the primary dataset for processing in the subsequent phase. In this phase, the presence of a financial bubble, as identified by the PSY approach in phase 1, was designated as the target variable for each month. Specifically, months characterized by the presence of a financial bubble were labelled as 1, while those without a bubble were labelled as 0. This binary response variable was then utilized for training and testing purposes alongside the NIFTY 500 price, NIFTY 500 price growth, and 42 additional macroeconomic variables, as outlined in the Table A1 in Appendix. These macroeconomic indicators, encompassing metrics such as GDP growth, Consumer Price Index, and Lending Interest Rate, were sourced from the World Bank data spanning the same time frame of 2002 to 2021.

The purpose of this paper is to explore the extent and possibility of predicting financial bubbles using macroeconomic variables, making them machine-learnable for easier future prediction with available data. The macroeconomic variables analyzed capture the characteristics of the economy and its various aspects. To emphasize the importance of some of these variables, GDP represents the value of products and services produced within a country's borders during a specific period, serving as a broad indicator of the economy's state. GDP growth, on the other hand, is a key indicator of economic health, influencing investor sentiment and stock market performance (Başoğlu Kabran and Ünlü, 2020). While GDP provides a total value, GDP growth offers a more accurate reflection of the economy's trajectory and future trends. The balance of payments accounts for all transactions between a country's citizens, corporations, and government and those of other countries. It reflects the economy's position relative to international economies, providing insights into trade deficits or surpluses and overall economic performance.

Interest rates are another crucial indicator of economic strength. Decision-makers closely monitor interest rate movements to understand

market price trends, which significantly influence financial decisions by investors, borrowers, and traders (Başoğlu Kabran and Ünlü, 2020). Inflation, defined as the price increase of goods and services over time, directly impacts stock prices by affecting the present value of future cash flows, making it a critical variable for study. Foreign Direct Investment (FDI) also plays a significant role in shaping investors' financial decisions. FDI often signals a growing market with potential opportunities. A country receiving substantial FDI generally indicates strong economic performance and growth potential. High levels of FDI suggest political and economic stability, making the country a safer investment destination by reducing risks such as currency fluctuations and policy changes.

Notably, the macroeconomic data obtained from the World Bank were originally annual in nature and were subsequently transformed into monthly data using the Cubic Spline Interpolation approach, a method recognized for preserving the characteristic features of the data (Ajao *et. al.*, 2012; Tran *et. al.*, 2023).

Another crucial aspect of data processing involved addressing class imbalance within the collected and processed dataset. Financial datasets commonly exhibit class imbalance, where the occurrence of certain events is sporadic compared to non-occurrence. This imbalance can adversely affect the efficacy of classifiers, particularly in accurately predicting events from the minority class. While predictive accuracy is a commonly used metric for evaluating ML systems, it may not be appropriate in the presence of data imbalance or when the consequences of different errors vary significantly (Chawla *et. al.*, 2002). To mitigate this challenge, the SMOTE was employed to balance the dataset. By generating synthetic samples for the minority class while preserving the features of the data distribution, SMOTE effectively rectifies the imbalance.

Accurate identification of financial bubbles holds paramount importance in financial markets, and SMOTE contributes to this endeavour by reducing bias, enhancing model generalization, and improving performance metrics such as precision and recall. Subsequently, the dataset was divided into training and test subsets in a 75:25 ratios to ensure an adequate amount of data for both model development and evaluation. The 75:25 ratio for dividing financial time series data into training and test subsets is commonly used because it strikes a balance between having enough data to effectively train the model and a sufficient amount to accurately evaluate its performance (Akhtar *et. al.*, 2022; Kamalov *et. al.*, 2021). This ratio ensures the model can learn from a substantial portion of the data while still providing a reliable assessment of how well it generalizes to unseen data, helping to mitigate overfitting and ensuring robust results in financial forecasting. Following this, the data was standardized using the 'standardscaler' package in Python. ML algorithms were then constructed and evaluated using Python and its associated libraries, including seaborn, scikit-learn, and imbalance-learn, among others.

Research Design - ML Algorithms to Predict Financial Bubbles

The focal variable under examination pertains to the manifestation of a financial bubble within the NIFTY 500 market. Designated as the target variable, the presence or absence of a financial bubble is respectively represented by the binary labels 1 and 0. This delineates a binary classification conundrum, wherein the primary objective is to prognosticate the emergence of said bubble, leveraging the 44 pre-defined features outlined earlier in this study.

For training and prediction purposes, seven binary classification ML Algorithms have been employed utilizing the Python scikit-learn package. These algorithms encompass Logistic Regression, Random Forest, K Nearest Neighbour, Support Vector Machine, Gradient Boosting, Artificial Neural Network, and Naive Bayes, all of which are well-suited for conducting binary classification analyses. Below, a succinct overview

delineates these models along with their respective methodologies for deployment.

Logistic Regression

Logistic regression serves as a statistical technique designed for modelling the probability of a discrete outcome contingent upon input variables. Typically, logistic regression models are tailored to binary outcomes, as highlighted in prior literature ("Research Methods for Cyber Security," 2018). Widely embraced for predictive modelling endeavours, especially in scenarios where the response variable is binary, logistic regression aligns seamlessly with the objective of the present study, which revolves around forecasting financial bubbles and crises. In this context, the model furnishes a probability estimate regarding the occurrence of a financial bubble, predicated upon the 44 input features integrated into the analysis. Notably, logistic regression models ascertain the likelihood of the default class through the utilization of the following equation.

$$P_n(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}} \quad (7)$$

Hence, logistic regression yields a linear classifier, establishing a decision boundary denoted by the solution to $\beta_0 + x.\beta = 0$, effectively partitioning the two projected classes. Beyond delineating the location of this class boundary, logistic regression, through the aforementioned equation, also elucidates the manner in which class probabilities fluctuate relative to their proximity to the boundary. Remarkably, as the norm of ||β|| escalates, probabilities tend to converge more rapidly towards the extremes of 0 and 1. Furthermore, following training on labelled data, the model is adept at predicting outcomes for novel, unseen data instances.

Random Forest (RF)

Derived from the decision tree paradigm, the random forest technique was pioneered by Breiman (2001). In RS, decision trees are constructed based on subsets of randomly selected attributes. This method ensures diversity among the constituent classifiers through the stochastic sampling of both data instances and features.

Within the RF framework, each decision tree exclusively accesses a random subset of the training data points and poses inquiries based on a randomly selected subset of features. This approach fosters diversity within the forest, enhancing the reliability of collective predictions and warranting the moniker "random forest." During prediction, the RF amalgamates the estimates from individual decision trees, leveraging their collective average. The prevailing class is determined by a majority vote, a mechanism that not only engenders more precise projections but also mitigates the risk of overfitting, thereby fortifying the model against spurious generalizations.

K Nearest Neighbour (KNN)

In scenarios where the distribution of data lacks substantial prior knowledge, the KNN classification technique emerges as a fundamental and pragmatic option, as advocated by Peterson (2009). Functioning within the realm of supervised classification, KNN leverages the class labels inherent in the training data. Operating on the principle of similarity metrics, KNN endeavours to categorize new data instances while retaining the characteristics of existing instances in the dataset. Essentially, KNN assigns classification based on the collective classification of its nearest neighbours, rendering it a method primarily reliant on the proximity of data points within the feature space.

The selection of the parameter K in KNN classification is a critical aspect that significantly impacts the accuracy of the model. A lower value of K leads to a more intricate model, while a higher value tends to produce a simpler model. Therefore, careful experimentation is

imperative to identify the optimal K value. Various techniques such as cross-validation, grid search, and the elbow method are commonly employed to ascertain the most suitable K value. These methodologies aim to minimize prediction error and enhance model accuracy by identifying the optimal balance between model complexity and performance. Once the ideal value of K has been determined, the KNN algorithm is trained and subsequently deployed to generate predictions.

Support Vector Machine (SVM)

SVM stands as a potent ML algorithm utilized for addressing multifarious tasks encompassing classification, regression, and outlier detection within supervised learning paradigms. SVM achieves this through optimal data transformations, delineating boundaries between data points predicated upon predefined classes, labels, or outputs. Primarily designed for binary classification tasks, SVMs are adept at discerning complex decision boundaries, rendering them invaluable across a spectrum of applications in the realm of machine learning.

SVM leverages the concept of hyperplanes to partition observations within high-dimensional feature spaces. These hyperplanes are positioned to maximize the margin, ensuring the optimal separation between the classes of interest. The classification decision is determined by employing the Eq. 8:

$$y_i = \begin{cases} +1, if\ b + \alpha^T x \geq +1 \\ -1, if\ b + \alpha^T x \leq -1 \end{cases} \quad (8)$$

where b is the bias.

Mathematically, SVM constitutes a suite of ML techniques employing kernel methodologies to transform data characteristics through kernel functions. These functions are predicated on the notion of mapping complex datasets into higher dimensions, thereby facilitating the separation of data points (Kanade, 2022). Notably, one of the primary

advantages of SVMs lies in their resilience against imbalanced distributions and their ability to mitigate overfitting concerns, particularly in scenarios characterized by limited sample sizes.

Gradient Boosting

Boosting stands as a highly efficacious ensemble technique within the realm of ML. Diverging from conventional models that learn from data in isolation, boosting amalgamates predictions from multiple weak learners to yield a singular, more precise strong learner. Among boosting methodologies, gradient boosting occupies a prominent position, finding extensive application across regression and classification tasks alike. This technique engenders a prediction model comprising an amalgamation of weak prediction models, frequently manifested as decision trees. Notably, gradient boosting emerges as a formidable strategy, adept at harnessing the collective strength of numerous weak learners to foster the emergence of robust, high-performing models.

At the core of this technique lies the principle of constructing new base-learners that exhibit maximum correlation with the negative gradient of the loss function associated with the entire ensemble (Natekin and Knoll, 2013). Employing gradient descent, the method iteratively trains each new model to minimize the loss function inherited from the preceding model, typically comprising metrics like mean-squared error or cross-entropy. Throughout each iteration, the technique calculates the gradient of the loss function relative to the predictions of the current ensemble and subsequently trains a new weak model to minimize this gradient. The predictions generated by the new model are then integrated into the ensemble, perpetuating this iterative process until a predefined stopping criterion is satisfied (Tran *et. al.*, 2023).

Artificial Neural Network (ANN)

An ANN, commonly referred to as a neural network, is another ML methodology inspired by the intricate structure and interconnections among neurons in the human brain. This framework endeavours to tackle

complex problems by mimicking the organizational structure of the brain. Comprising multiple layers of artificial neurons interconnected with one another, an artificial neural network encompasses input, output, and hidden layers within each stratum. Notably, in contrast to conventional regression methods, artificial neural networks excel in capturing and simulating intricate nonlinear relationships inherent in complex datasets. ANNs, drawing inspiration from the intricate network of biological neurons, manifest as expansive and intricately interconnected systems comprised of basic processors that function in extensive parallel. These models strive to embody certain organizational principles purportedly present in human cognition. Within one variant of neural network architecture, nodes are conceptualized as "artificial neurons" (Kaur and Gupta, 2020).

Via mathematical modelling, these artificial neurons emulate the functionality exhibited by their biological counterparts. Each artificial neuron receives an input signal, denoted as $x_1, x_2, \dots, x_j$, typically comprising binary values (0 or 1). Subsequently, leveraging the respective weights associated with these signals— $w_1, w_2, \dots, w_j$ —the neuron computes the weighted sum of the received signals. If the cumulative weight of the received signals surpasses a predefined threshold (see Eq. 9), the neuron transmits the signal to the subsequent artificial neuron in the network (Tran *et. al.*, 2023).

$$y_i = output \begin{cases} 0 \; if \; \sum_j w_j x_j \; \leq threshold \\ 1 \; if \; \sum_j w_j x_j \; > \; threshold \end{cases} \tag{9}$$

The neural network undergoes training utilizing a dataset comprising input-output pairs, wherein the weights and thresholds of artificial neurons are iteratively adjusted until the desired level of accuracy is achieved. Through this iterative learning process, the neural

network acquires the ability to accurately predict outcomes for new input data. Kwong (2001) underscores the considerable utility and efficacy of neural networks, particularly ANNs, in the domain of financial forecasting. Empirical evidence cited in his work highlights the significant contributions of neural networks across crucial areas such as Bankruptcy Failure Prediction, Bond rating, and Commodities market analysis (Wilson and Sharda, 1994; Fadlalla and Lin, 2001; Wong and Selvi, 1998).

Naïve Bayes

The Naïve Bayes classifier represents a different supervised ML approach extensively employed for text categorization and various classification tasks. It belongs to the generative learning algorithm family, aiming to replicate the input distribution associated with a specific class or category. Utilizing Bayes' theorem, the Naïve Bayes classifier facilitates probabilistic classification. By observing a set of features or parameter values (input data), the classifier calculates the likelihood that the input data pertains to a particular class (de Souza *et. al.*, 2021).

The naive Bayes classifier operates under the assumption that the features employed to predict the target are independent of each other. This classifier overlooks the interdependence among features in real-world data that collectively influence the target outcome. Despite its name, the independence assumption, although seldom entirely accurate in real-world scenarios, often holds empirically true in practice, hence the term "Naive" (Gamal, 2021).

Evaluating the Algorithms

This paper closely follows the evaluation framework established by Tran *et. al.* (2023), wherein we assess the performance of each model based on metrics such as Accuracy, Precision, F1 Score, and Area Under the Curve (AUC), complemented by the examination of confusion matrices.
A confusion matrix is a tabular representation utilized to assess the efficacy of a classification algorithm. Studies such as those conducted by Hagemann and Wohlmann (2019) and Öğüt *et. al.* (2012) leverage

confusion matrices to elucidate and synopsize the performance of machine learning models in financial contexts. The summary and visualization of a classification algorithm's performance are encapsulated within a confusion matrix, as depicted in Table 1.

Table 1: A Confusion Matrix Structure

		Predicted Value	
		Positive	Negative
Actual value	Positive	True Positive (TP)	True Negative (TN)
	Negative	False Positive (FP)	False Negative (FN)

The additional metrics for comparison mentioned above are defined and calculated as follows:

Accuracy: Accuracy measures the frequency with which a machine learning model accurately predicts outcomes. It is computed as the ratio of correctly predicted observations to the total number of observations (see Eq. 10). A higher accuracy score indicates a more proficient model.

$$Accuracy = \frac{True\ Positives\ +\ True\ Negatives}{Total\ Observations} \tag{10}$$

Precision: Precision quantifies the proportion of accurately predicted true values relative to the total number of observed positive predictions. It is calculated as in Eq. 11:

$$Accuracy = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives} \tag{11}$$

F1 Score: The F1 score, which is the harmonic mean of precision and recall, ensures that the score is higher only when both precision and recall are higher. It is computed as in Eq. 12:

$$Accuracy = 2 * \frac{Precision * Recall}{TPrecision + Recall} \tag{12}$$

Receiver Operating Characteristic (ROC) Curve: This curve illustrates the performance of each classification model by plotting the true positive rate against the false positive rate.

Area Under the Curve (AUC): The AUC characterizes the ROC curve by representing the area beneath the curve. It reflects the area enclosed by the lower-right segment of the ROC curve.

RESULTS AND DISCUSSION

Financial Bubble Detection using PSY Method

The paper employed the PSY method to identify financial bubbles in the Indian Stock Market, specifically focusing on the National Stock Exchange's NIFTY 500 index. The study covers the period from January 2003 to December 2021, incorporating major economic events. The analysis involved calculating the NIFTY 500 dividend yield and the price-dividend ratio, as detailed in Eq. 13 and following Caspi (2023).

$$\text{Price} - \text{Dividend Ratio} = \frac{1}{Dividend\ Yield} \tag{13}$$

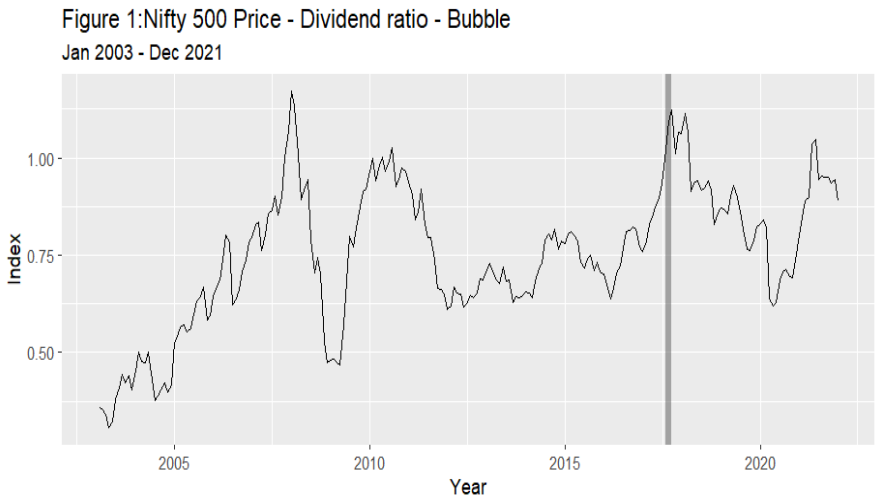
The NIFTY 500 data underwent aggregation, with daily data averaged over monthly periods. The PSY procedure adhered to a minimal window determined by the rule $t_{min} = 0.01 + 1.8\sqrt{T}$, ensuring a minimum of 38 observations for analysis.

Table 2 presents the outcome of the aforementioned monitoring procedure, delineating the commencement and conclusion dates of each identified financial bubble. The analysis has detected the occurrence of financial bubbles spanning 8 months concerning the NIFTY 500 index. This finding is visually depicted in Figure 1, where dark lines highlight the presence of bubbles. Notably, the bubble observed in 2007 persisted for a very brief duration and may not be prominently visible in the graph. Furthermore, distinct phases featuring multiple bubbles emerged, notably one at the end of 2007 and the beginning of 2008, and another

during the period of 2017–2018. Remarkably, these findings align closely with observations made by Tran *et. al.* (2023), who conducted similar PSY monitoring for the Vietnamese stock exchange over a comparable timeframe.

Several economic experts and studies, including Ray (2009), Berger *et. al.* (2017), Kanojia and Malhotra (2021), and Khan and Suresh (2022), corroborate the aforementioned findings. These sources support the observation of overvalued stock prices in the market during the identified periods, as confirmed by financial and economic specialists.

Figure 1: NIFTY 500 Price-Dividend Ratio - Bubble



Notes: The solid line is the price-to-dividend ratio and the shaded areas are the periods where the PSY statistic exceeds its 95% bootstrapped critical value.

Source: Computed by author

Table 2: Nifty 500 Price- Dividend Bubble Dates

Start	End
2007-11-01	2007-12-31
2017-07-01	2017-09-29

Source: Computed by author

Following the identification of financial bubbles, the months characterized by their occurrence were labelled as 1, while those without were labelled as 0. This dataset served as the foundation and constituted the response variable for subsequent research and processing employing supervised machine learning techniques.

Table 3 presents the descriptive statistics of the NIFTY 500 index throughout the study period, categorized according to the presence or absence of financial bubbles. The data indicates that the average index price during bubble months surpasses that of non-bubble months, a trend similarly reflected in the quartile ranges and their respective values. While the minimum value during bubble months notably exceeds that of non-bubble months, the same cannot be observed for the maximum value, consistent with findings reported by Tran *et. al.* (2023).

Table 3: Descriptive Statistics for NIFTY500

	Overall	Bubble	Non-Bubble
Count	252.00	8.00	244.00
Mean	6443.97	7258.98	6417.25
Std. Dev.	4252.99	1849.19	4308.4
Min	698.90	4859.7	698.90
25th	3449.55	5251.72	3418.64
50th	4798.65	8361.92	4725.98
75th	9048.95	8696.58	9161.5
Max	18061.8	8815.25	18061.8

Source: Computed by author

Forecasting Financial Bubbles using Machine Learning

Seven classifier models, namely Logistic Regression, Random Forest, KNN, SVM, Gradient Boosting, ANN, and Naive Bayes, were implemented with the processed dataset, incorporating macroeconomic variables as features (see Table A1 in Appendix). The dataset was partitioned such that 75 percent served as training data, while the remaining 25 percent

constituted the test data. The performance of each model is summarized below in Table 4 and a comparison is made in Table 5.

The performance of the seven machine learning models in predicting financial bubbles was assessed using various metrics. The RF achieved perfect accuracy with a score of 1.000, correctly classifying all 110 instances without any errors[3]. Similarly, Gradient Boosting also attained perfect accuracy, predicting all 110 cases correctly. ANN model had an accuracy of 0.945 and a test loss of 0.1732, indicating strong prediction performance with 104 correct predictions out of 110, and 5 false negatives. Logistic Regression also achieved a high accuracy of 0.945, accurately predicting 104 out of 110 cases but with 6 false negatives and 6 false positives. SVM model had an accuracy of 0.920, correctly predicting 102 cases, with 8 false positives and no false negatives. KNN model recorded an accuracy of 0.900, achieving 99 correct predictions, with 8 false positives and 3 false negatives. In contrast, the Naive Bayes model had the lowest accuracy of 0.727, correctly predicting 80 cases and recording 14 false positives and 16 false negatives[4]. These results underline the varying effectiveness of each model, with Random Forest and Gradient Boosting demonstrating the highest performance.

The application of the PSY method to the NIFTY 500 index reveals significant periods of financial bubbles, specifically from late 2007 to early 2008 and during 2017–2018. These findings underscore the

[3] The significance of the features used in predicting financial bubbles is depicted in Table A.1 in Appendix. In RF model, the NIFTY 500 index demonstrates significance at approximately 13 percent in predicting the response variable, with all other indicators bearing comparable or greater weightage. Consequently, all selected indicators were retained as features for further analysis and processing.

[4] The accuracy score of this model is comparatively lower than that of the other processed models. This discrepancy could be attributed to the potential multicollinearity among the features utilized for analysis. One of the primary assumptions of the Naive Bayes model is the independence between the features considered for study, given the target class. Addressing this multicollinearity and removing features that may potentially impact each other could substantially enhance the efficacy and accuracy of this model.

utility of the PSY method in identifying overvalued stock phases, aligning with prior research and adding empirical weight to the theory of financial bubbles. Furthermore, the machine learning models, particularly Random Forest and Gradient Boosting, demonstrated high accuracy in predicting these bubbles, emphasizing the effectiveness of advanced algorithms in capturing complex financial dynamics. This underscores their ability to accurately and precisely predict and classify periods based on the occurrence of financial bubbles. Both of these algorithms belong to the ensemble learning category, leveraging multiple models for training and prediction, thereby achieving superior results. These findings suggest that policymakers could leverage these insights to enhance market surveillance and implement more responsive regulatory measures during suspected bubble periods.

Table 4: Confusion Matrices for Machine Learning Models

Algorithm	TP	FP	FN	TN	Comments
Random Forest Classifier	53	0	0	57	Perfect prediction of all 110 instances
Gradient Boosting	53	0	0	57	Perfect prediction of all 110 instances
Artificial Neural Network	45	1	5	56	104 correct predictions out of 110, test loss of 0.1732
Logistic Regression	47	6	6	57	104 correct predictions out of 110
Support Vector Machine	45	8	0	57	102 correct predictions out of 110
K Nearest Neighbour	45	8	3	54	99 correct predictions out of 110
Naive Bayes	37	14	16	43	80 correct predictions out of 110

Source: Computed by author

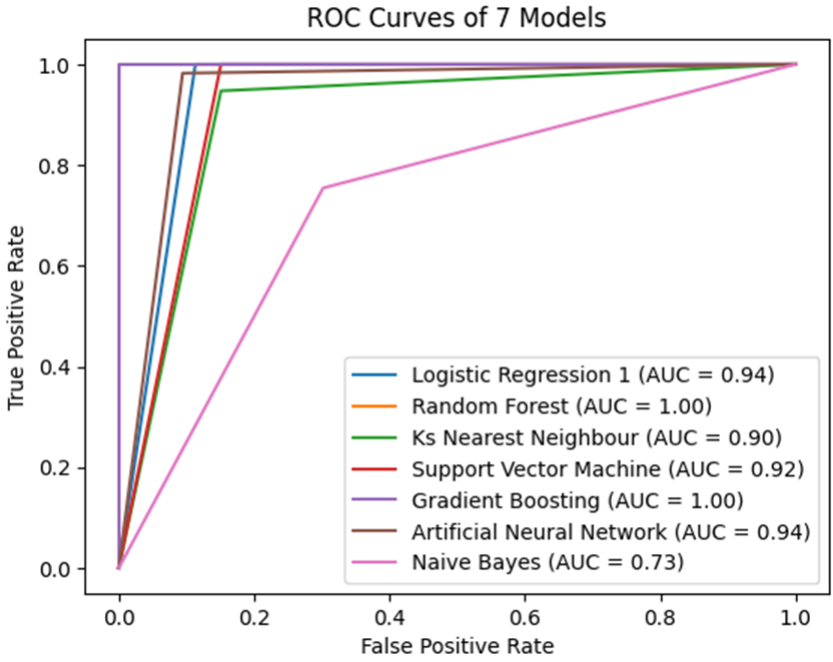
Following these highly accurate models, the ANN exhibits a high precision of 0.92. Conversely, the Naive Bayes model underperforms in comparison to every other model across all parameters, likely due to multicollinearity, as previously mentioned.

Table 5: Performance Comparison of all Classifier Machine Learning Algorithms

Algorithm	AUC	F1 Score	Accuracy	Precision
Random Forest Classifier	1.00	1.00	1.00	1.00
Gradient Boosting	1.00	1.00	1.00	1.00
Artificial Neural Network	0.94	0.95	0.95	0.92
Logistic Regression	0.94	0.95	0.95	0.90
Support Vector Machine	0.92	0.93	0.93	0.88
K Nearest Neighbour	0.90	0.91	0.90	0.87
Naive Bayes	0.73	0.74	0.72	0.73

Source: Computed by author

Figure 2: ROC Curves of all 7 Models



Source: Computed by author

Figure 2 summarizes the results in the form of a graph, depicting the ROC curves of each model. As previously mentioned, Gradient Boosting and RF exhibit the highest AUC scores, indicative of their capability to achieve a high true positive rate while maintaining a low false positive rate. This ability underscores their effectiveness in differentiating between stock market bubbles and non-bubble periods. As outlined by Tran *et. al.* (2023) in finance, models like logistic regression, decision trees, and support vector machines are frequently employed to address classification problems. However, the findings of this study indicate that sophisticated ML techniques—specifically, Random Forest, Gradient Boosting, and Neural Networks—outperform traditional techniques. This observation aligns with existing literature and underscores the potential of machine learning algorithms to accurately capture the complex relationships associated with financial bubbles.

CONCLUSION

This paper applies the PSY method of bubble detection to identify financial bubbles in the Indian Stock market, utilizing the NIFTY 500 index data spanning from 2003 to 2021. The analysis reveals the presence of financial bubbles in 2007 and 2017, subsequently subjecting the dataset to testing with seven machine learning algorithms using a 75 percent training and 25 percent test data split.

Among the employed machine learning models, advanced techniques such as Gradient Boosting, Random Forest Classifier, and Artificial Neural Networks demonstrated superior performance, exhibiting very high accuracy and precision.

The study's findings highlight the effectiveness of advanced machine learning models, particularly Random Forest and Gradient Boosting, in detecting financial bubbles. Policymakers should integrate these sophisticated tools into regulatory frameworks to improve real-time monitoring and early detection of market anomalies. This integration can

help in crafting timely responses to mitigate risks associated with financial bubbles, adjusting policies to enhance market stability, and guiding investors to make more informed decisions. Embracing such advanced models can thus significantly bolster financial oversight and stability.

These results furnish regulatory agencies and central banks with increased capacity to manage capital flows and devise appropriate monetary policies, thereby curbing speculative activities in financial asset markets and upholding systemic stability. Additionally, investors equipped with the ability to identify price bubbles can safeguard their investment portfolios by making more informed decisions regarding asset selling or purchasing. Moreover, bubbles present arbitrageurs with opportunities to profit in the short term by selling assets in overvalued markets.

Despite the promising results, this study has several limitations. Firstly, the analysis is based on historical data, and the identified bubbles are constrained by the period of study, potentially overlooking recent or emerging bubbles. Secondly, the machine learning models, while effective, are limited by their reliance on past data patterns, which may not fully capture future market dynamics or structural changes in the financial system.

Future research could expand the dataset to include more recent data, incorporating other parameters, other stock indices or financial markets to enhance the robustness of bubble detection. Additionally, exploring alternative machine learning algorithms and hybrid models could offer further improvements in accuracy and predictive power. Sentiment analysis may also provide a more comprehensive understanding of bubble formation and evolution. Finally, real-time application and validation of these techniques in various market conditions would further strengthen their practical utility in financial regulation and decision-making.

REFERENCES

Ajao, I. O., A. G. Ibraheem, and F. J. Ayoola (2012), "Cubic Spline Interpolation: A Robust Method of Disaggregating Annual Data to Quarterly Series", *Journal of Physical Sciences and Environmental Safety*, *2*(1), 1-8.

Akhtar, M. M., A. S. Zamani, S. Khan, A. S. A. Shatat, S. Dilshad, and F. Samdani (2022), "Stock Market Prediction Based on Statistical Data Using Machine Learning Algorithms", *Journal of King Saud University-Science*, 34(4), 101940.

Banerjee, A., and P. Kayal (2022), "Bubble Run-Ups and Sell-Offs: A Study Of Indian Stock Market", *Review of Behavioral Finance*, 14(5), 875-885.

Barlevy, G. (2018), "Bridging Between Policymakers' and Economists' Views on Bubbles", *Economic Perspectives*, *42*(4), 1-21.

Başoğlu Kabran, F., and K. D. Ünlü (2021), "A Two-Step Machine Learning Approach to Predict Sandp 500 Bubbles", *Journal of Applied Statistics*, *48*(13-15), 2776-2794.

Berger, A., J. Leininger, and D. Messner (2017), "The G20 In 2017: Born in A Financial Crisis—Lost in A Global Crisis?", *Global Summitry*, *3*(2), 110–123.

Breiman, L. (2001), "Random Forests", *Machine Learning*, *45*(1), 5–32.

Caspi, I. (2023, December 29), *Itamarcaspi/Psymonitor*. *Github*. Https://Github.Com/Itamarcaspi/Psymonitor.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002), "SMOTE: Synthetic Minority Over-Sampling Technique", *Journal of Artificial Intelligence Research*, *16*(16), 321–357.

Chen, Y., P. C. Phillips, and S. Shi (2023), "Common Bubble Detection in Large Dimensional Financial Systems", *Journal of Financial Econometrics*, *21*(4), 989-1063.

De Souza, G. F. M., A. H. D. A. Melani, M. A. D. C. Michalski, and R. F. Da Silva (Eds.) (2021), "Reliability Analysis and Asset Management Of Engineering Systems"*, Elsevier*.

Fadlalla, A., and C. H. Lin (2001), "An Analysis of The Applications of Neural Networks In Finance", *Interfaces*, 31(4), 112-122.

Galbraith, J. K., S. Hsu, and W. Zhang (2009), "Beijing Bubble, Beijing Bust: Inequality, Trade, and Capital Inflow into China", *Journal of Current Chinese Affairs*, *38*(2), 3–26.

Gamal, B. (2021, April 11), "Naïve Bayes Algorithm", *Analytics Vidhya*. Available At: Https://Medium.Com/Analytics-Vidhya/Na Percentc3 Percentafve-Bayes-Algorithm-5bf31e9032a2.

Gerdesmeier, D., H. E. Reimers, and B. Roffia (2013), "Testing For the Existence Of A Bubble in the Stock Market", (No. 01/2013), *Wismarer Diskussionspapiere*.

Hagemann, D., and M. Wohlmann (2019), "An Early Warning System to Identify House Price Bubbles", *Journal of European Real Estate Research*, *12*(3), 291–310.

Hu, Y. (2023), "A Review of Phillips-Type Right-Tailed Unit Root Bubble Detection Tests", *Journal of Economic Surveys*, *37*(1), 141-158.

Öğüt, H., M. M. Doğanay, N. B. Ceylan, and R. Aktaş (2012), "Prediction Of Bank Financial Strength Ratings: The Case of Turkey", *Economic Modelling*, *29*(3), 632-640.

Jarrow, R. A., and S. S. Kwok (2021), "Inferring Financial Bubbles from Option Data", *Journal of Applied Econometrics*, *36*(7), 1013–1046.

Kamalov, F., I. Gurrib, and K. Rajab (2021), "Financial Forecasting with Machine Learning: Price Vs Return", Kamalov, F., Gurrib, I. and Rajab, K.(2021), "Financial Forecasting with Machine Learning: Price Vs Return", *Journal of Computer Science*, 17(3), 251-264.

Kanade, V. (2022, September 29), "What Is A Support Vector Machine? Working, Types, and Examples", *Spiceworks*. Available At: Https://Www.Spiceworks.Com/Tech/Big-Data/Articles/What-Is-Support-Vector-Machine/

Kanojia, S., and D. Malhotra (2021), "A Case Study of Stock Market Bubbles in the Indian Stock Market", *Indian Journal of Finance*, *15*(2), 22-48.

Kaur, J., and N. Gupta (2020), "Artificial Neural Network: A Review", Available At: Https://Ijtrs.Com/Uploaded_Paper/ ARTIFICIAL Percent20neural Percent20network Percent20a Percent20 review.Pdf.

Khan, T., and G. Suresh (2022), "Do All Shocks Produce Embedded Herding and Bubble? An Empirical Observation of The Indian Stock Market", *Investment Management and Financial Innovations*, *19*(3), 346-359.

Ray, K. K. (2009), "Investment Behavior and the Indian Stock Market Crash 2008: An Empirical Study of Student Investors", *IUP Journal of Behavioral Finance*, *6* (3/4), 41.

Kwong, K. (2001), "Financial Forecasting Using Neural Network or Machine Learning Techniques", *University of Queensland*, *13*, 221-228.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2023), "Machine Learning Advances for Time Series Forecasting", *Journal of Economic Surveys*, *37*(1), 76-111.

Natekin, A., and A. Knoll (2013), "Gradient Boosting Machines, A Tutorial", *Frontiers In Neurorobotics*, *7,* 21.

Orhan, A., and N. Sağlam (2023), "Financial Forecast in Business and An Application Proposal: The Case of Random Forest Technique" *Journal of Accounting and Finance/Muhasebe Ve Finansman Dergisi*, (99), 171.

Peterson, L. (2009), "K-Nearest Neighbor", *Scholarpedia*, *4*(2), 1883.

Phillips, P. C., S. Shi, and J. Yu (2015a), "Testing for Multiple Bubbles: Limit Theory of Real-Time Detectors", *International Economic Review*, *56*(4), 1079-1134.

Phillips, P. C., S. Shi, and J. Yu (2015b), "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the Sandp 500", *International Economic Review*, *56*(4), 1043-1078.

Phillips, P. C., Y. Wu, and J. Yu (2011), "Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?", *International Economic Review*, *52*(1), 201-226.

Phillips, P. C., S. Shi, I. Caspi, and M. I. Caspi (1984), "Package 'Psymonitor'", *Biometrika*, *71*, 599-607.

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

Sadorsky, P. (2021), "A Random Forests Approach to Predicting Clean Energy Stock Prices", *Journal Of Risk and Financial Management*, *14*(2), 48.

Samuel, A. L. (1959), "Some Studies in Machine Learning Using the Game of Checkers", *IBM Journal of Research and Development*, *3*(3), 210–229.

Shi, S., and P. C. Phillips (2022), "Econometric Analysis of Asset Price Bubbles", Available At: Https://Elischolar.Library.Yale.Edu/ Cowles-Discussion-Paper-Series/2688/.

Singh, S., N. Vats, P. Jain, and S. S. Yadav (2018), "Rational Bubbles in the Indian Stock Market: Empirical Evidence from the NSE 500 Index", *Finance India*, *32*(2), 457-478.

Talin, B. (2022, September 10), "Economic Bubble – Definition, Types and 5 Stages of Financial Bubbles", *Morethandigital*. Available At: Https://Morethandigital.Info/En/Economic-Bubble-Definition-Types-And-5-Stages-Of-Financial-Bubbles/

Wöckl, I. (2019), "Bubble Detection in Financial Markets-A Survey of Theoretical Bubble Models and Empirical Bubble Detection Tests", Available At SSRN 3460430.

Wong, B. K., and Y. Selvi (1998), "Neural Network Applications in Finance: A Review and Analysis of Literature (1990–1996)", *Information And Management*, 34(3), 129-139.

Wilson, R. L., and R. Sharda (1994), "Bankruptcy Prediction Using Neural Networks", *Decision Support Systems*, 11(5), 545-557.

APPENDIX

Table A1: Significant Features

Feature	Significance
NIFTY500	0.131323
Debt Service on External Debt	0.035999
Net Trade in Goods and Services	0.035317
CAB of GDP	0.031609
Lending Interest Rate	0.030294
Multilateral Debt Service	0.030133
Net Primary Income	0.030053
Nifty500_Growth	0.028972
GDP	0.028060
IBRD Loans and IDA Credits	0.027822
Official Exchange Rate	0.027278
Short-Term Debt	0.027027
Primary Income Payments	0.026768
GCF Growth	0.026302
GNE Of GDP	0.022341
FCE Growth	0.021863
GNI	0.021630
Unemployment	0.021419
CPI	0.021329
Exports of GDP	0.021133
GGFCE percent Growth)	0.020181
Final Consumption Expenditure	0.019452
Current Account Balance	0.019091
FCE 2015	0.018945
Imports percent Growth	0.01829
GCF	0.018258
Inflation	0.017867
Personal Remittances of GDP	0.017334
Handnpishs FCE	0.017161
GNI Growth	0.016888
FDI	0.016861
Use Of IMF Credit	0.016417
Personal Remittances	0.014555

Feature	Significance
External Debt Stocks	0.014518
GDP Growth	0.012561
Imports of GDP	0.012537
Total Debt Service	0.012468
Goods Exports	0.012420
Imports	0.012397
Trade GDP	0.011660
Gross National Expenditure	0.010126
Exports Constant 2015	0.009463
Exports Growth	0.007604
GDS of GDP	0.006273

Source: Computed by author

MSE Monographs

* Monograph 34/2015
Farm Production Diversity, Household Dietary Diversity and Women's BMI: A Study of Rural Indian Farm Households
Brinda Viswanathan

* Monograph 35/2016
Valuation of Coastal and Marine Ecosystem Services in India: Macro Assessment
K. S. Kavi Kumar, Lavanya Ravikanth Anneboina, Ramachandra Bhatta, P. Naren, Megha Nath, Abhijit Sharan, Pranab Mukhopadhyay, Santadas Ghosh, Vanessa da Costa and Sulochana Pednekar

* Monograph 36/2017
Underlying Drivers of India's Potential Growth
C.Rangarajan and D.K. Srivastava

* Monograph 37/2018
India: The Need for Good Macro Policies *(4th Dr. Raja J. Chelliah Memorial Lecture)*
Ashok K. Lahiri

* Monograph 38/2018
Finances of Tamil Nadu Government
K R Shanmugam

* Monograph 39/2018
Growth Dynamics of Tamil Nadu Economy
K R Shanmugam

* Monograph 40/2018
Goods and Services Tax: Revenue Implications and RNR for Tamil Nadu
D.K. Srivastava, K.R. Shanmugam

* Monograph 41/2018
Medium Term Macro Econometric Model of the Indian Economy
D.K. Srivastava, K.R. Shanmugam

* Monograph 42/2018
A Macro-Econometric Model of the Indian Economy Based on Quarterly Data
D.K. Srivastava

* Monograph 43/2019
The Evolving GST
Indira Rajaraman

MSE Working Papers
Recent Issues

* Working Paper 261/2024
  Assessment of Urban Road Transport Sustainability in Indian Metropolitan Cities
  B. Ajay Krishna & K.S. Kavi Kumar
* Working Paper 262/2024
  Empowerment of Scheduled Tribes and other Traditional Forest Dwellers for Sustainable Development of India
  Ulaganathan Sankar
* Working Paper 263/2024
  How Green (performance) are the Indian Green Stocks – Myth Vs Reality
  Saumitra Bhaduri & Ekta Selarka
* Working Paper 264/2024
  Elementary Education Outcome Efficiency of Indian States: A Ray Frontier Approach
  Jyotsna Rosario & K.R Shanmugam
* Working Paper 265/2024
  Are the Responses of Oil Products Prices Asymmetrical to Global Crude Oil Price Shocks? Evidence from India
  Abdhut Deheri & Stefy Carmel
* Working Paper 266/2024
  Drivers and Barriers to the Adoption of Renewable Energy: Investigating with the Ecological Lens
  Salva K K & Zareena Begum Irfan
* Working Paper 267/2024
  A Multi-Criteria Decision-Making Model to Determine the Share of Variable Renewable Energy Sources
  Salva K K & Zareena Begum Irfan
* Working Paper 268/2024
  Determinants of Renewable Energy in Asia: Socio-Economic and Environmental Perspective
  Salva K K & Zareena Begum Irfan
* Working Paper 269/2024
  Adaptive Analysis of 3E Factors (Economy, Energy, and Environment) for Renewable Energy Generation in the South and South-East Asian Region
  Salva K K & Zareena Begum Irfan

---

* Working papers are downloadable from MSE website http://www.mse.ac.in

$ Restricted circulation