Journal of Open Innovation: Technology, Market, and Complexity 10 (2024) 100180

Contents lists available at ScienceDirect

Journal of Open Innovation: Technology, Market, and Complexity

journal homepage: www.sciencedirect.com/journal/journal-of-open-innovation-technology-market-and-complexity

Forecasting stock prices of fintech companies of India using random forest with high-frequency data

Bharat Kumar Meher [a], Manohar Singh [b], Ramona Birau [c,*], Abhishek Anand [d]

[a] PG Department of Commerce & Management, Purnea University, Purnea, Bihar 854301, India
[b] Department of Commerce, Government Autonomous PG College, Chhindwara, Madhya Pradesh, India
[c] Faculty of Economic Science, University Constantin Brancusi of Tg-Jiu, Romania
[d] PG Department of Economics, Purnea University, Purnea, Bihar 854301, India

ARTICLE INFO

JEL classification:
C55
G17
G23

Keywords:
Forecasting
Random Forest
Fintech
High-frequency data
Python

ABSTRACT

The fintech segment is currently one of the most rapidly growing industries, attracting numerous investors who anticipate substantial returns in the future. Notably, not only individual retail investors but also mutual fund agencies are actively engaged in predicting stock prices within this sector to maximize their trading gains. The purpose of the study is to formulate stock forecasting models for top three Fintech Companies of India i.e., Policy Bazar, One 97 Communications Paytm Ltd., and Niyogin Ltd. Using Random Forest model with high-frequency data in Python. The literature review section also proves that this study is a novel piece of work as none of the existing research study focused on predicting stock prices of Fintech Companies of India using Random Forest model. The data is extracted from www.moneycontrol.com and www.kotaksecurities.com, for the period from 1st October, 2022–30 th September, 2023. The study deals about 293,280 data points i.e., 3 companies @ 97,760 each. It has been found that the forecasting model of random forest provides very successful results for prediction as the co-efficient of determination of all the selected companies is more than 95%.

1. Introduction

Technological advancements have a longstanding presence in the financial sector. Digital innovation, in particular, has ushered in significant enhancements in system connectivity, computational capabilities, cost efficiency, and the generation of actionable data. These improvements have led to the reduction of transaction costs and the emergence of novel business models and players in the financial landscape (Feyen et al., 2021). These new entrants are collectively referred to as "Fintech." In the digital age, numerous Fintech startups have proliferated, both in developed and developing nations, including India. Fintech, as the name suggests, is the fusion of finance and technology. It gained substantial momentum following the global financial crisis of 2008 and continues to evolve rapidly, with many unexplored opportunities remaining (Taujanskaitė and Kuizinaitė, 2022).

In this dynamic environment, many market participants leverage technology to streamline financial services, encompassing lending, insurance, investments, trading, budgeting, and more (Scardovi, 2017; Pazarbasioglu et al., 2020). This contributes to the seamless and efficient operation of financial services traditionally offered by banks and insurance companies (Alt et al., 2018; Breidbach et al., 2020). Fintech companies in India, such as Paytm, gained prominence during events like Demonetization and the COVID-19 pandemic when cashless transactions became the preferred choice for many (Jakhiya et al., 2020; Moid and Shankar, 2022; Khando et al., 2023). As the fintech sector expands, numerous players in India are narrowing their focus to niche areas. Consumer lending fintech firms constitute a significant portion, comprising 17% of the total fintech enterprises (Nenavath, 2022; Migozzi et al., 2023). The demand for credit in India continues to rise, prompting banks to collaborate with fintech companies to enhance service offerings. For example, Paytm simplifies payments by reducing the need for manual intervention in card and net banking transactions (Ganjoo et al., 2023). Angel investors are increasingly drawn to invest in fintech startups, recognizing the industry's substantial growth potential (Surana et al., 2020; Harris, 2021; Saura et al., 2021). Some fintech companies have even reached a point where they expand operations and become publicly listed on stock exchanges.

The fintech sector is currently one of the most rapidly growing industries, attracting numerous investors who anticipate substantial returns in the future (Mention, 2019; Zhang-Zhang et al., 2020; Arora

* Corresponding author.
E-mail addresses: ramona.f.birau@gmail.com (R. Birau), abhi2eco@gmail.com (A. Anand).

https://doi.org/10.1016/j.joitmc.2023.100180
Received 15 October 2023; Received in revised form 8 November 2023; Accepted 12 November 2023
Available online 18 November 2023

and Madan, 2023). Notably, not only individual retail investors but also mutual fund agencies are actively engaged in predicting stock prices within this sector to maximize their trading gains (Palmié et al., 2020; Bhatia et al., 2021). Various techniques, such as Exponential Moving Averages (EMA), AutoRegressive Integrated Moving Average (ARIMA), GARCH Models, and Holt-Winters Exponential Smoothing, have been employed to forecast stock prices using univariate data (Priyamvada & Wadhvani, 2017; Chatterjee et al., 2021). However, the Random Forest method emerges as a promising alternative. Random Forest is an ensemble learning approach that amalgamates multiple decision trees for making predictions (Asad, 2015; Thakur and Kumar, 2018). This ensemble strategy often yields more robust and accurate predictions compared to individual models, with the added advantage of accommodating non-linearities. While comparing the Random Forest model with the other benchmark model like ARIMA, random forest model can also consider non-linear patterns in stock prices movement, and can successfully handle seasonality and trends, outliers and anomalies and both short-term and long-term forecasting also.Moreover, it is versatile enough to handle both regression and classification tasks, making it suitable for various stock price forecasting scenarios (Valencia et al., 2019; Landis and Cha, 2020; Mohanta et al., 2020). In this study, we endeavor to predict stock prices in the burgeoning Indian FinTech sector using the Random Forest model, employing high-frequency data. The use of high-frequency data, specifically one-minute closing prices of Indian fintech companies, enhances prediction accuracy and precision. Innovation Dynamics research often explores how technological advancements impact various sectors, including finance. The use of such kind of high-frequency data and machine learning techniques like Random Forest to forecast stock prices of fintech companies suggests an application of technological innovation in the financial industry.Subsequently, the literature review section outlines key research endeavors related to stock price forecasting, high-frequency data, and Random Forest. This study not only substantiates the research gap but also underscores the novelty of our research within this field.

2. Review of literature

India is well-known as a thriving center for fintech, and as the Indian start-up ecosystem expands, more industries inspired by fintech use cases will start up and receive funding from different sources. Risky investments in the fintech segment of the equities market may pay off handsomely, according to a study that indicates a significant risk-return link in the Indian fintech industry (Mention, 2019; Brown and Wiles, 2020; Bhatnagar et al., 2022).

Akyildirim et al. (2023) test many machine learning strategies for their ability to foretell intraday excess returns. Prediction rates much above 50% are generated by machine learning analytics, and optimal profit ratios can go as high as 33%. The findings support the usefulness of analytics and machine learning techniques and prompt additional discourse on the market's moderate efficiency. Data from the Indian stock market shows that Levenberg-Marquardt (LM), Scaled Conjugate Gradient, and Bayesian regularization algorithms all achieve an accuracy of 99.9% when using tick data (Selvamuthu et al., 2019). When compared to the findings obtained using tick data, the accuracy over a 15-minute dataset lowers to 96.2% for LM, 97.0% for SCG, and 98.9% for Bayesian Regularization.

The potential gains from accurate future forecasting have made it a goal of many societies and economies. With the help of AI, scientists will have access to more precise predictions than ever before. Over time, as technology and algorithms improve, they will become more precise. Overall, feature engineering proved to benefit the models (Alkhatib et al., 2022). When applied to models using Long Short-Term Memory, the new method yielded significant improvements. Predicting the Fintech index is useful for a number of different people since it can help investors create successful short-, medium-, and long-term investment plans and can direct financial regulators toward making accurate and effective regulatory rules. These findings show that the algorithm can be used as a more precise instrument to forecast the Fintech index (Liu et al., 2021). High frequency trading (HFT) algorithms are robust and effective, this is shown by the fact that the whole prediction system, which includes the deep learning block with RL framework corrections, can boost trend forecast accuracy to roughly 85% (Rundo, 2019). Additional insights on network connections may be gained by utilizing the high-resolution information contained in high-frequency intraday trading data sets. However, the asynchronicity, complicated dynamics, and non-stationarity of such data sets make them extremely difficult to model. Use of random forests, a cutting-edge machine learning approach that provides high prediction accuracy without the need for costly hyperparameter tweaking, to estimate financial networks and overcome these obstacles (Karpman et al., 2023).

Seven machine learning algorithms are compared in (Subasi et al., 2021) study across four stock index datasets (NASDAQ, NYSE, NIKKEI, and FTSE) with the goal of making investment risk mitigation easier. Furthermore, Random Forest were found to produce superior outcomes. Utilizing several types of random forests, including quantile random forests and extreme random forests (Demirer et al., 2022), demonstrate that risk aversion enhances the accuracy of realized volatility forecasts outside of the sampling frame. Realized skewness and kurtosis, as well as measures of jump intensity and leverage, have little to no effect on risk aversion's ability to anticipate future outcomes. Stock trend prediction accuracy can be effectively improved by using the random forest model and optimizing the various processes of stock research (Yin et al., 2023). Random forest do the best job of predicting both the overalltrend and the magnitude of future price changes (Akyildirim et al., 2022). Sadorsky (2021) finds that, compared to logit models, random forest methods are superior at predicting the direction of stock prices. For forecast horizons of 10 days or longer, random forests and tree-bagging models have a prediction accuracy of above 80%. Accuracy ratings of 85–90% are achieved by the tree bagging and random forest methods for a 20-day forecast horizon, whereas those of 55–60% are achieved by the logit models. By applying the artificial intelligence algorithm of ensemble random forest methods, moreover, (Lin et al., 2020) was able to get the probabilities of market reactions of start-up enterprises listed on the GISA equity crowdfunding platform and anticipate the degree of market reaction. Predicting actual data in market reactions using the GISA platform for startups has an accuracy of 65%. Similarly, according to (Luong and Dokuchaev, 2018), the random forest method is effective at predicting the direction of realized volatility. Empirical results on the S&P 200 indicate that the use of purified implied volatility and this machine learning technique resulted in an improvement of the pre-existing heterogeneous autoregressive model (HAR) framework.

2.1. Research gap

From the review of the existing literature, it can be observed that many researches have been done on forecasting but there are some studies related to forecasting stock prices using random forest but none of the study focused on forecasting stock prices of Fintech Companies of India using Random forest. Moreover, there is not enough studies that used high-frequency data while framing random forest model for forecasting. The study is fully focused on a new dimension by framing random forest model using the high-frequency data to forecast stock prices of the companies of most emerging sector of India i.e., FinTech. Hence, a study on this research gap is considered as a feasible one is definitely going to contribute to existing studies.Moreover, studying stock price forecasting in the context of fintech companies can provide insights into how market dynamics are influenced by innovation in the financial technology sector. Innovation Dynamics research may seek to understand the interactions between technological innovation and market behavior.

3. Objectives of the study

- To formulate stock forecasting models for top three Fintech Companies of India i.e., using Random Forest model with high-frequency data.
- To determine the efficacy of the formulated models in forecasting the future stock prices of fintech companies.

4. Research methodology

4.1. Research design

This research adopts a quantitative approach aimed at predicting stock prices of Fintech companies in India. The chosen methodology involves the application of a machine learning technique i.e. Random Forest, implemented using Python. This method is suitable for forecasting financial time series data as it accommodates non-linear relationships and handles high-dimensional data effectively.

4.2. Data collection

The primary data source for this study consists of high-frequency data i.e., historical one-minute daily open, high, low, close stock price of top three Fintech companies listed on the Indian stock exchanges namely Policy Bazar, One 97 Communications Paytm Ltd., and Niyogin Ltd. Such kind of high-frequency data are proved as rich in quality specifically in forecasting stock prices due its precision and use in algo-trading.High-frequency data in the context of the stock market refers to data that is recorded and updated at very short time intervals, often with sub-second or intra-second precision. This data can provide valuable insights for stock market research, trading, and analysis.The data is extracted from reputable financial databases, such as www.moneycontrol.com and www.kotaksecurities.com, for a period from 1st October, 2022–30 th September, 2023, ensuring the availability of a sufficient historical dataset to train and validate the model. The data set of each company has 97,760 data points. Hence, the study deals about 293,280 data points i.e., 3 companies @ 97,760 each.Innovation Dynamics research can involve studying new methods and tools for gathering and interpreting data to drive innovation and the use of high-frequency data and machine learning algorithms like Random Forest in this research indicates an innovation in data collection and analysis methods.

4.3. Data preprocessing

To prepare the data for model development, several preprocessing steps are employed. These include data cleaning, feature selection, handling missing values, and scaling. Stock price returns are calculated and used as the target variable, while the lags of opening price, high price and low price are used as independent variables.

4.4. Model development

Random Forest, a powerful ensemble machine learning algorithm, is chosen as the forecasting model. It is known for its ability to handle complex relationships and mitigate overfitting. The model is implemented using Python's scikit-learn library. It is trained on historical data, and hyperparameters are tuned through cross-validation to optimize performance.

4.5. Softwares and applications used

For the formulation of Random Forest Model to forecast the stock prices of Fintech Companies of India, PyCharm with Python Software with version 3.8 (6) has been used. Moreover, various library packages are installed namely, NumPy for numerical and mathematical operations, Pandas for reading and cleaning the financial data, Scikit-Learn (sklearn) that includes Random Forest Regressor, Matplotlib and Seaborn, openpyxl and xlrd for data visualization, High-frequency data source, and pycharm. The selected version of python can smoothly handle all the selected packages.

4.6. Evaluation techniques

To assess the model's forecasting accuracy, several evaluation metrics are employed. These metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2). Additionally, the out-of-sample forecasting accuracy is assessed to validate the model's generalization performance. The data is split into training and testing sets, with a rolling window approach to evaluate the model's performance over time. 70% of the data i.e., 68432 are training data points and remaining 30% i.e., 29,328 data points are used as testing sets.

By following the above approach, the study aims to develop an accurate and robust forecasting model that can assist investors, analysts, and policymakers in making informed decisions in the dynamic and rapidly evolving Fintech sector of the Indian stock market.

5. Need of the study and managerial implications

A research study focused on forecasting stock prices of Fintech companies in India using the Random Forest algorithm holds significant potential for societal benefit. Accurate stock price predictions for these companies are essential not only for investors but also for the broader society. Firstly, such research aids investors in making informed decisions. Stock market investments are integral to many individuals' financial planning, and the Fintech sector's dynamism offers both opportunities and risks. Reliable forecasts can guide investors, helping them allocate their resources effectively and mitigate potential losses. This, in turn, promotes financial literacy and stability among the populace. Secondly, a successful prediction model for Fintech stocks can stimulate investment in this sector. Fintech companies drive innovation, financial inclusion, and economic growth. If investors have confidence in their ability to forecast stock prices accurately, they are more likely to invest in these companies, fostering innovation and job creation. The success or failure of fintech companies can have a significant economic impact. Innovation Dynamics research often delves into the economic consequences of innovation. Analyzing stock price forecasting in fintech can contribute to understanding how innovation in this sector affects the broader economy.Furthermore, a well-developed forecasting model can serve as a valuable tool for regulators and policymakers. It enables them to monitor market stability and respond proactively to emerging challenges, ultimately safeguarding the interests of both consumers and investors. This research on forecasting stock prices of Fintech companies using Random Forest is a critical endeavor that benefits society by empowering investors, fostering innovation, and aiding regulatory oversight in a sector poised for transformative growth.

6. Limitations of the study

- Limited Generalizability: The findings of the paper may be limited in their applicability to fintech companies in India. Stock price prediction models developed for one market or time period may not generalize well to different markets or time frames. Moreover, it may not explore the robustness of the model to changing market conditions, and whether it continues to perform well in different market regimes.
- Data Quality and Availability: The accuracy and reliability of stock price data, especially high-frequency data, can be a limitation. Data may contain errors, gaps, or inconsistencies that can affect the results and generalizability of the model. Moreover, such High-frequency data often comes with a lag, which can affect the practicality of

real-time trading strategies based on the model's forecasts. The paper may not address the implications of this lag.

- Model Parameter Tuning: Random Forest models have hyper-parameters that need to be tuned for optimal performance. The paper may not explore the sensitivity of the model's performance to different hyperparameter settings.
- Model Interpretability: Random Forest models are often considered as "black-box" models, making it challenging to interpret the reasons behind specific predictions. The paper may not delve into model interpretability techniques or insights into the importance of specific features.
- Market Dynamics and External Factors: The paper is not adequately considered external factors such as economic conditions, regulatory changes, or geopolitical events that can significantly impact stock prices, particularly in the volatile fintech sector.

7. Analysis and discussion

The study employs the Random Forest (RF) algorithm, originally introduced by Breiman in 2001 as an enhanced version of the decision tree, to predict stock prices within the fintech sector in India. At its core, Random Forests consist of multiple decision trees. During training, Random Forests construct numerous individual decision trees, and the predictions from these trees are combined to make the final prediction. This aggregation is achieved by considering the mode of the classes for classification tasks or the mean prediction for regression tasks. Two key parameters in the RF model can significantly impact its performance: the number of trees (n_{tree}) and the number of candidate variables randomly selected at each split (n_{try}). A recommended value for n_{try} is $\frac{p}{3}$, where p represents the number of input variables (Dudek, 2015).

To forecast a time-series Yt, the study employs the autoregressive random forest (AR-RF) model, denoted as AR-RF(p), where p signifies the number of autoregressive lags. In contrast to various other machine learning models, Random Forests offer superior precision and excel in handling large datasets with numerous variables, often extending into the thousands. Furthermore, Random Forests possess the capability to automatically balance datasets, especially when one class is less frequent than others in the data.

7.1. Steps of applying random forest in predicting stock prices of fintech companies

- In this study all the data sets of the fintech companies which are non-stationary in nature are converted into stationary for the application of random forest.
- Again, the data set of each company i.e., 97,760 data points, is splitted into two parts, the first part is the 70% contains 68,432 data points and remaining 30% contains 29,328 data points.
- The first part containing 68,432 data points is considered as training set and the remaining part containing 29,328 data points is considered as testing set. The training set is used for framing the random forest model and the testing set is where the formulated model is applied for predicting.
- The strength and suitability of the random forest model can be judged by values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2)which can be calculated by using the following formulae:-

Mean Squared Error	**Root Mean Squared Error**
$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$	$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$
Mean Absolute Error	**Coefficient of Determination**
$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}\rvert$	$R^2 = 1 - \frac{\Sigma(y_i - \hat{y})^2}{\Sigma(y_i - \bar{y})^2}$

Where $\widehat{y} = predicted \quad value \quad of \quad y$ and $\bar{y} = mean \quad value \quad of \quad y$

- After the application of the formulated model in the testing test, the outcomes of the model in the form of predicted stock prices and the existing observed (actual) stock prices are represented in graph to mark out the differences between the those two.
- As the testing set contains huge data i.e., 29,328 which may lack in clear depiction of differences between the actual and predicted stock prices if presented through graphs. Hence, a deviation graph has been used to show the deviations between the actual and predicted stock prices.

7.2. Parameters for applying random forest

This research used “scikit-learn” Python packagefor formulating the Random Forest model. There are various “hyperparameters” which control the behavior of the algorithm and form the random forest model. Research used default “hyperparameters” given in “scikit-learn” Python package.

- Tree Count (n_estimators): Set to scikit-learn's default value of 100, which strikes a balance between model performance and computational economy.
- Trees' Maximum Depth (max_depth): Left at ‘None’. This will let the trees to grow until all of their leaves are pure or contain less samples than min samples split. This helps in fully representing the Complexity of the data, with further regularization coming from other hyperparameters to reduce overfitting.
- Minimum Samples for Split (min_samples_split): Set to 2, which is the bare minimum needed to create a new node. As a result, the dataset might be segmented finely, enhancing the model's capacity to learn from the training set.
- Minimum Samples (min_samples_leaf) at Leaf Nodes: Set to 1 allows for the most detailed class definitions per leaf, especially helpful for datasets with a complex decision boundary.
- Maximum Features (max_features): By default, "auto" is set in “scikit-learn” Python package and the same is used, which chooses the square root of the feature count. this parameter is essential in order to diversify the individual trees and advance model generalization.
- Bootstrap Samples: By default, this feature is enabled in “scikit-learn” Python package, this enables every tree to undergo training using a bootstrapped sample of the data. By adding randomness to the model, this lessens overfitting and improves the ensemble's capacity to generalize.

7.3. Brief about the selected fintech companies

7.3.1. Niyogin fintech Ltd

Niyogin Fintech Ltd. operates as a non-banking finance company, specializing in providing loans, financing, investments, and related services to micro, small, and medium enterprises in India. The company places a strong emphasis on superior execution, utilizing advanced technology, innovative risk management, and establishing robust on-ground connections. Niyogin's primary objective is to offer small businesses an efficient and cost-effective support system through cutting-edge technology and a dedicated network of partners. They aspire to become the leading organization in India that caters to the needs of small businesses, empowering their customers with a comprehensive ecosystem of products, partnerships, technology, and exceptional customer experiences.

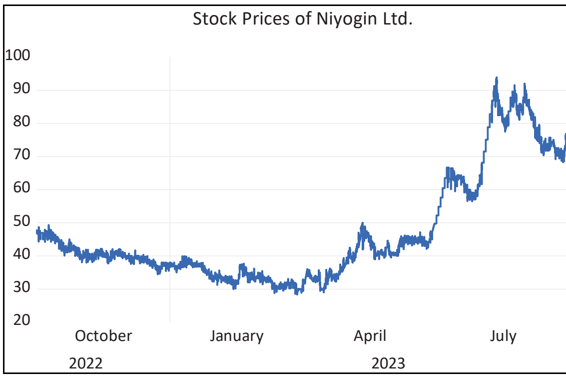
One 97 Communications Ltd (Paytm).

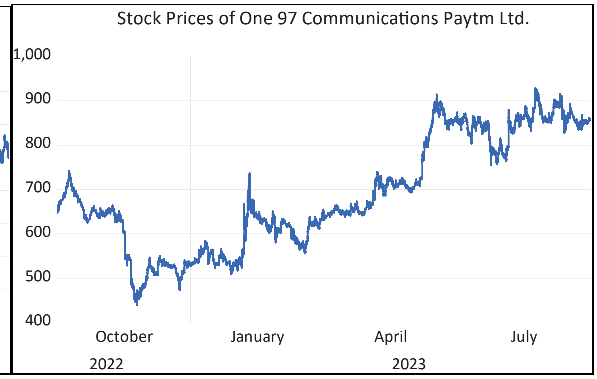
According to RedSeer, One 97 Communications Ltd., better known as Paytm, is India's most advanced e-commerce platform. As of June 30, 2021, Paytm served 337 million clients and over 21.8 million registered merchants with its vast suite of services, which included payment solutions, e-commerce services, cloud services, and financial services. Paytm, a digital payment platform that was initially released in 2009 with a focus on mobile devices, ushered in a new era of cashless transactions in India. Paytm Wallet first appeared as a means of making cellphone recharges and bill payments. Based on RedSeer's analysis of consumer volume, merchant count, transaction volume, and income, Paytm had become India's largest payments platform by March 31,

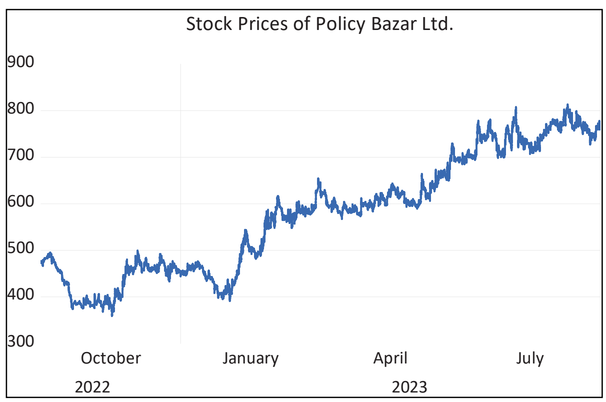
2021. Paytm's brand value of US$6.3 billion, according to the Kantar BrandZ India 2020 Report, confirms its place as India's most valuable payments brand and makes it the go-to option for transactions across multiple channels.

policies sold. Also, when looking at the total number of policies sold online in India, both insurance firms and distributors, Policybazaar accounted for 65.3%. Following this section is a visual representation of the real stock values of a few India-based fintech companies.

Graphical Representation of actual Stock Prices of Selected Fintech Companies.







7.3.2. PB fintech Ltd. (policybazar)

Policybazaar's parent business, PB Fintech Ltd., launched its flagship platform in 2008 to satisfy customers' demands for improved insurance education, selection, and disclosure. With the launch of Paisabazaar in 2014, the company set out to streamline the process of securing personal loans and credit cards for individuals in India by putting an emphasis on simplicity, speed, and openness. According to research from Frost & Sullivan, in terms of disbursals in Fiscal 2021, Paisabazaar held a dominant 53.7% market share in India's digital consumer credit marketplace. In addition, Policybazaar overtook all other online insurance distributors in Fiscal 2020 to become the largest digital insurance marketplace, with a market share of 93.4% based on the number of

.

Source: Authors' Construction of Graphs using EVIEWS.

The graphs of the stock prices of all the selected fintechs seem non-stationary, hence, it is necessary to convert the non-stationary data into stationary data for the application of random forest for which log returns have been computed and plotted using line graphs which are shown in the succeeding section.

Graphical Representation of Log Returns of Selected Fintech Companies.

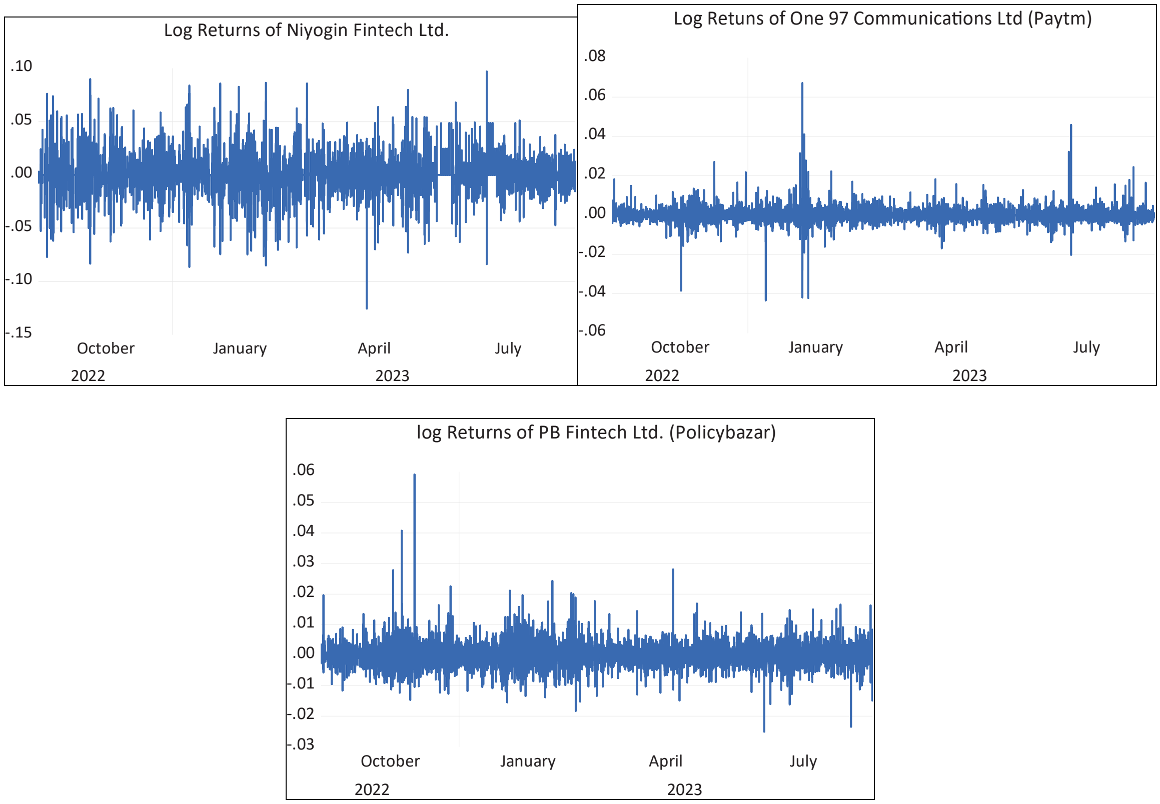
.

Table 1

Values of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination ($\mathbf{R^2}$) using Python.

Statistics	Niyogin Fintech Ltd.	One 97 Communications Ltd (Paytm)	PB Fintech Ltd. (Policybazar)
Mean Squared Error (MSE)	0.00598	0.1502	0.1486
Root Mean Squared Error (RMSE)	0.07736	0.3875	0.3855
Mean Absolute Error (MAE)	0.01148	0.2337	0.2336
Coefficient of Determination (R^2)	0.99998	0.99999	0.99999

Source: Authors' Computation using Python



Source: Authors' Construction of Graphs using EVIEWS.

The line graphs of log returns seem stationary in nature. Moreover, the stationary of data should be examined through statistical test. Hence, the stationarity of log returns series of the above fintech companies have been examined with the help of a unit root test named Augmented Dickey Fuller Test with the inclusion of test equation as Intercept, Trend and Intercept and None and found stationary and now prepared to use it for random forest model.

Table 1 depicts the calculated Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) of all the selected fintechs. MSE measures the average squared difference between the predicted values and the actual values. A lower MSE indicates that the model's predictions are closer to the actual values. The MSE of Niyogin Fintech Ltd., One 97 Communications Ltd (Paytm) and PB Fintech Ltd. (Policybazar) is 0.00598, 0.1502 and 0.1486 respectively which suggests that, on average, the squared difference between the predicted and actual stock prices is quite small. This is a positive sign for the model's accuracy.

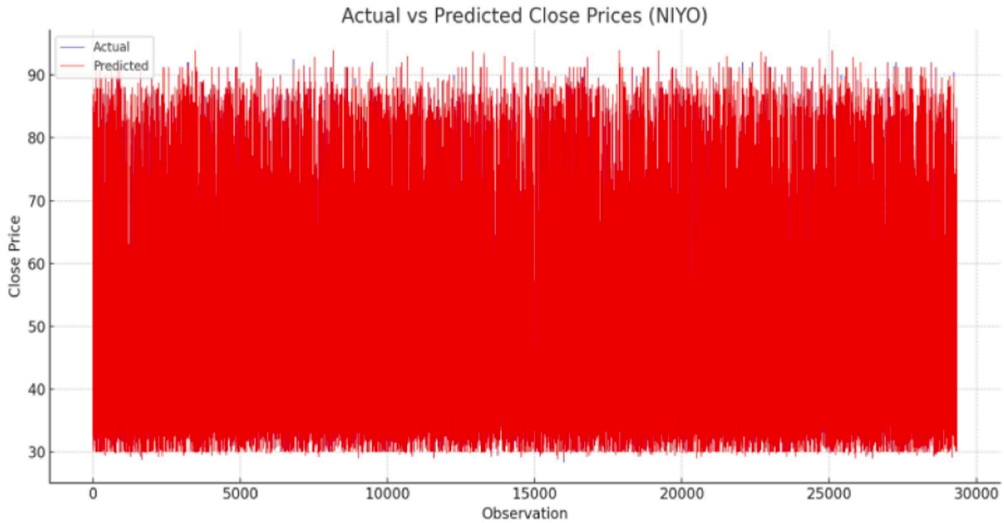
Table 1RMSE is the square root of the MSE. It provides a measure of the average magnitude of errors in the same units as the predicted variable. An RMSE of 0.07736, 0.3875 and 0.3855 of the three fintechs respectively indicates that, on average, the model's predictions are approximately 0.07736, 0.3875 and 0.3855 units away from the actual stock prices of the three fintechs respectively. In a suitable model a lower RMSE is always desirable.

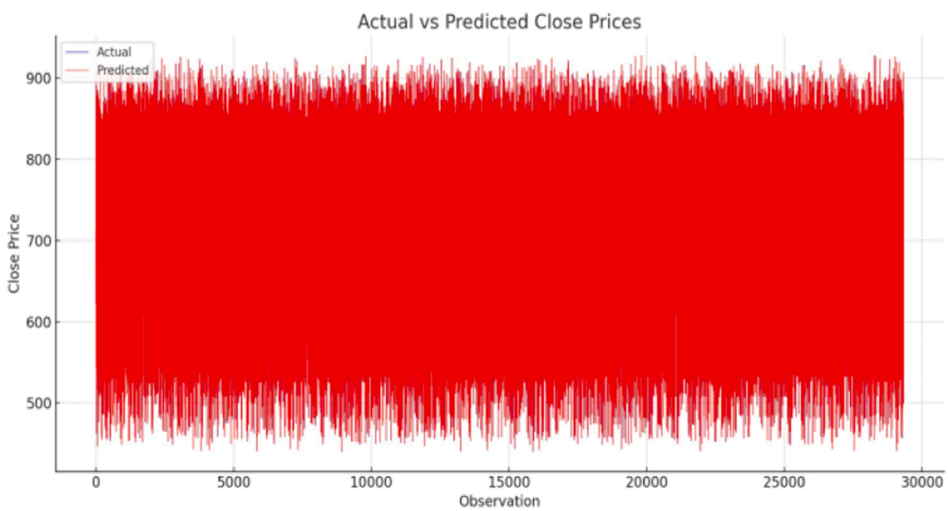
Similarly, the MAE measures the average absolute difference between the predicted and actual values. Like MSE and RMSE, a lower MAE indicates better model accuracy and this is another indicator of good model performance.

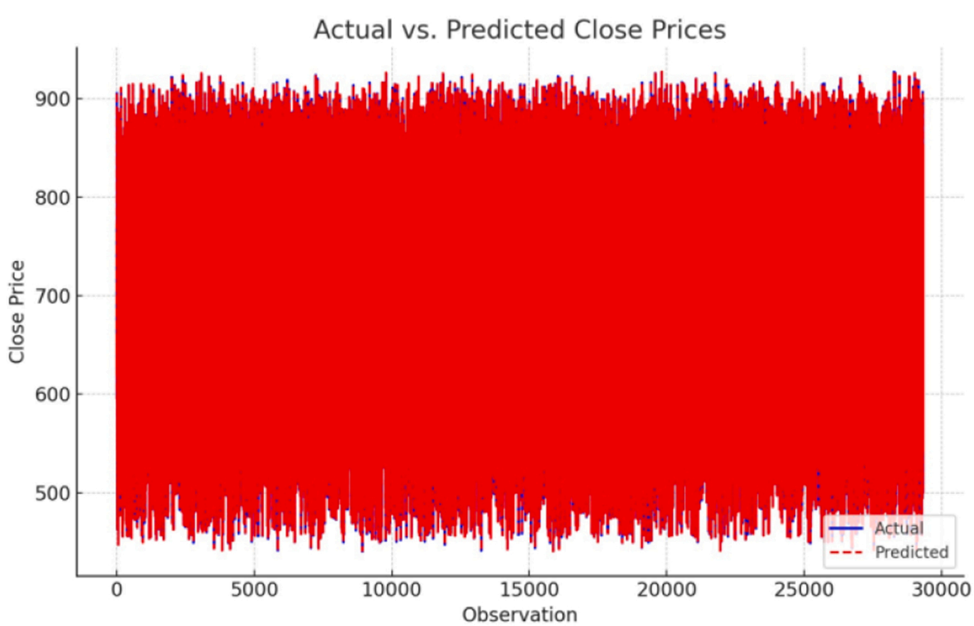
Moreover, the R^2 of all selected fintech are very close to 1, suggesting that the models explains almost all of the variance in the stock prices. This is an exceptionally high R-squared value and indicates an excellent fit of the model to the data. In summary, based on the provided statistical results, it appears that the Random Forest model used for forecasting stock prices of all Fintech companies are performing exceptionally well. The low MSE, RMSE, and MAE values indicate that the model's predictions are very close to the actual values. Additionally, the high R-squared value (close to 1) suggests that the model is an excellent fit for the data.

Again, by using the formulated models the predicted stock prices are calculated and shown in the line graphs along with the actual stock prices for comparison.

Graphical Representation of Actual and Predicted Stock Prices of Selected Fintech Companies on testing data using the formulated Random Forest Model.


Actual vs Predicted Close Prices (NIYO)
Actual
Predicted
Close Price
Observation


Actual vs Predicted Close Prices
Actual
Predicted
Close Price
Observation


Actual vs. Predicted Close Prices
Actual
Predicted
Close Price
Observation

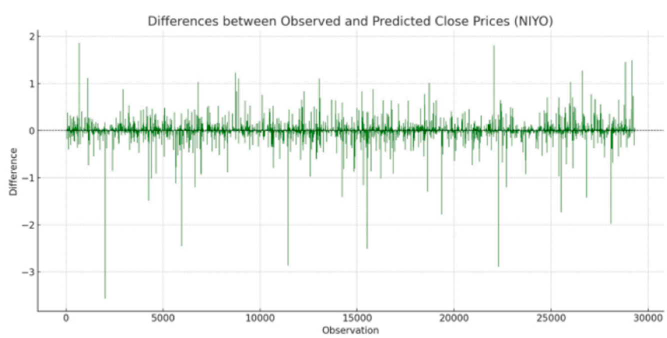
.

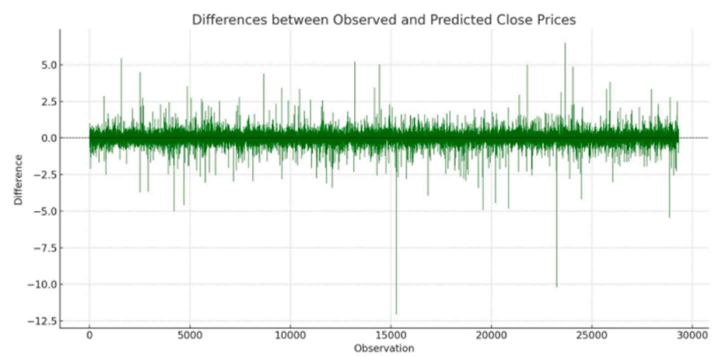
Source: Authors' Construction of Graphs using Python.

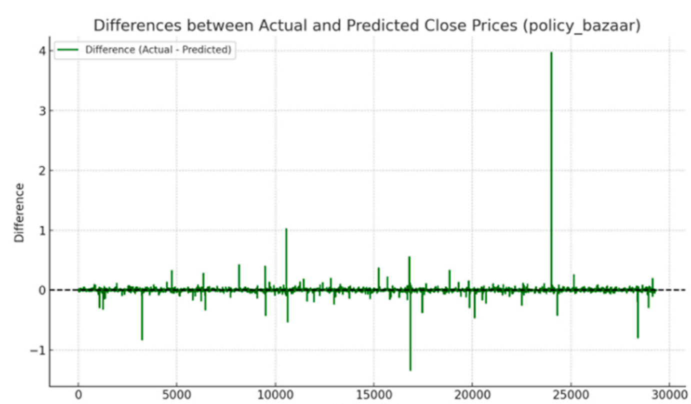
Since the above graphs consist nearly 29,328 observations used for prediction, the trend line of actual and predicted stock prices of the companies cannot be precisely seen.

Hence, for simplicity and clear understanding, deviations plot (the difference between actual and predicted stock prices) are generated that could depict a single trend line. If the model's predictions are perfect, the difference will be zero, and no spike can be seen on that particular date and time. Otherwise, the line will oscillate around zero, showing the upward spike where predictions are high and downward spike where predictions are low. The succeeding section shows the deviations plot of all the three selected companies.

Graphical Representation of Deviations plot of Selected Fintech Companies.







.

Source: Authors' Construction of Graphs using Python.

From the deviation graphs it can be observed that there are very few and small spikes in the Policy Bazar graph representing much accuracy in the prediction and a very few i.e., around 5–6 spikes that oscillated between –1.25 to + 4 against the average price 576.70 in last year. Again, some spikes can be seen in the Niyogin Ltd. graph with a maximum oscillation of –3 to + 2 where the average price of Niyogin Ltd. in last one year is about 48, but still it can be considered as accurate forecasting as most of the spikes in particular timestamp are near to zero. Lastly, in Paytm Graph, too many spikes can be seen but the most of the spikes are oscillating between –2.5 to + 2.5 and very few spikes between –12.5 to + 5 where the average price in last one year is about 688 and again it can be considered as optimally accurate forecasting.

8. Conclusion

From observation of above analysis and discussion it is clear that the results obtained from this study provide valuable insights into the accuracy and performance of the predictive model. The statistical metrics presented in Table 1, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R2), serve as crucial indicators of the model's effectiveness. It is evident that the Random Forest model has demonstrated exceptional predictive capabilities for all the selected fintech companies. First, the low MSE values for Niyogin Fintech Ltd., One 97 Communications Ltd (Paytm), and PB Fintech Ltd. (Policybazar) indicate that the model's predictions are consistently close to the actual stock prices. This reflects the accuracy of the model in capturing the underlying patterns in the stock price movements. The RMSE values further confirm the model's accuracy, with values of 0.07736, 0.3875, and 0.3855 for the three fintechs, respectively. These results indicate that, on average, the model's predictions deviate by a relatively small margin from the actual stock prices. Lower RMSE values are generally preferred in forecasting models, and in this case, they signify the model's reliability in making precise predictions. Additionally, the MAE values for all selected fintechs are low, reaffirming the model's ability to make accurate predictions with minimal absolute errors. This is a crucial aspect of model performance, as lower MAE values imply better accuracy in predicting stock prices. Perhaps the most compelling evidence of the model's proficiency is the high Rˆ2 values, which are very close to 1 for all fintech companies. An R2 value near 1 indicates that the model explains almost all of the variance in the stock prices, demonstrating an exceptional fit of the model to the data. This is a remarkable achievement, as it signifies that the Random Forest model effectively captures the underlying factors influencing the stock prices of these fintech companies. Furthermore, the deviation graphs provided in the analysis offer visual confirmation of the model's accuracy. The small and infrequent spikes in the Policy Bazar graph, the limited oscillations in the Niyogin Ltd. graph, and the relatively manageable spikes in the Paytm graph all contribute to the overall assessment of accurate forecasting. Moreover, the findings of this research paper strongly support the effectiveness of the Random Forest model in forecasting stock prices of fintech companies in India using high-frequency data. The combination of low MSE, RMSE, and MAE values, along with the high R2 values and visually confirmed accuracy, underscores the model's exceptional performance. These results have significant implications for investors, financial analysts, and decision-makers, as they can rely on this model to make informed decisions in the dynamic fintech sector. This study contributes to the growing body of knowledge in financial forecasting and reinforces the value of machine learning techniques in stock price prediction.But there are also certain limitations of the findings of this study, like the results and findings are based on the lagged open, high and low stock prices only and does not consider any macro-economic factors that might affect the stock prices of fintech companies in India.Moreover, in last few years the financial inclusion has been significantly increased. Many people of India have started using fintech services especially the digital banking transactions after demonetization and during COVID-19, which ultimately shoot up the demand for services of fintech companies that leads to increment of retail investment on such companies, but such effect are also not considered as regressor.Hence, the future researchers may conduct some study which could overcome such limitations too. Furthermore, this study could also act as a guide for future researchers to employ varied nature of random forest using high-frequency data on unexplored sectors and areas of global stock markets to forecast the stock prices or values of market indices.

Author contributions

All authors contributed equally to this research work. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Ethical statement/approval

No applicable because the study does not include research involving animal or human subjects.

Funding

This research received no external funding.

CRediT authorship contribution statement

Conceptualization: Bharat Kumar Meher. Data curation: Bharat Kumar Meher, Manohar Singh, Abhishek Anand, Ramona Birau. Formal analysis: Bharat Kumar Meher, Manohar Singh, Ramona Birau, Investigation: Bharat Kumar Meher, Manohar Singh, Ramona Birau. Review of Literature and Research Gap: Abhishek Anand, Bharat Kumar Meher. Methodology: Bharat Kumar Meher, Manohar Singh, Abhishek Anand. Project administration: Bharat Kumar Meher, Ramona Birau. Resources: Ramona Birau. Software: Manohar Singh, Bharat Kumar Meher. Validation: Bharat Kumar Meher, Abhishek Anand. Visualization: Manohar Singh, Ramona Birau. Writing – original draft: Bharat Kumar Meher, Abhishek Anand. Writing – reviewing & editing: Ramona Birau, Bharat Kumar Meher, Abhishek Anand, Manohar Singh.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Akyildirim, E., Bariviera, A.F., Nguyen, D.K., Sensoy, A., 2022. Forecasting high-frequency stock returns: a comparison of alternative methods. Ann. Oper. Res. 313 (2), 639–690. https://doi.org/10.1007/s10479-021-04464-8.
Akyildirim, E., Nguyen, D.K., Sensoy, A., Šikić, M., 2023. Forecasting high-frequency excess stock returns via data analytics and machine learning. Eur. Financ. Manag. 29 (1), 22–75. https://doi.org/10.1111/eufm.12345.
Alkhatib, K., Khazaleh, H., Alkhazaleh, H.A., Alsoud, A.R., Abualigah, L., 2022. A new stock price forecasting method using active deep learning approach. J. Open Innov.: Technol., Mark., Complex., 8(2), Artic. 2. https://doi.org/10.3390/joitmc8020096.
Alt, R., Beck, R., Smits, M.T., 2018. FinTech and the transformation of the financial industry. Electron. Mark. 28 (3), 235–243. https://doi.org/10.1007/s12525-018-0310-9.
Arora, S., Madan, P., 2023. Conceptual framework depicting the drivers for the fintech growth: an outlook for India. In: Grima, S., Sood, K., Özen, E. (Eds.), Contemporary Studies of Risks in Emerging Technology, Part A. Emerald Publishing Limited, pp. 197–220. https://doi.org/10.1108/978-1-80455-562-020231014.
Asad, M., 2015. Optimized Stock market prediction using ensemble learning. 9th Int. Conf. Appl. Inf. Commun. Technol. (AICT) 2015, 263–268. https://doi.org/10.1109/ICAICT.2015.7338559.
Bhatia, A., Chandani, A., Atiq, R., Mehta, M., Divekar, R., 2021. Artificial intelligence in financial services: a qualitative research to discover robo-advisory services. Qual. Res. Financ. Mark. 13 (5), 632–654. https://doi.org/10.1108/QRFM-10-2020-0199.
Bhatnagar, M., Özen, E., Taneja, S., Grima, S., Rupeika-Apoga, R., 2022. The dynamic connectedness between risk and return in the fintech market of India: evidence using the GARCH-M approach. Article 11 Risks 10 (11). https://doi.org/10.3390/risks10110209.
Breidbach, C.F., Keating, B.W., Lim, C., 2020. Fintech: research directions to explore the digital transformation of financial service systems. J. Serv. Theory Pract. 30 (1), 79–102. https://doi.org/10.1108/JSTP-08-2018-0185.
Brown, K.C., Wiles, K.W., 2020. The growing blessing of unicorns: the changing nature of the market for privately funded companies. J. Appl. Corp. Financ. 32 (3), 52–72. https://doi.org/10.1111/jacf.12418.
Chatterjee, A., Bhowmick, H., Sen, J., 2021. Stock price prediction using time series, econometric, machine learning, and deep learning models. IEEE Mysore Sub Sect. Int. Conf. (MysuruCon) 2021, 289–296. https://doi.org/10.1109/MysuruCon52639.2021.9641610.
Demirer, R., Gkillas, K., Gupta, R., Pierdzioch, C., 2022. Risk aversion and the predictability of crude oil market volatility: a forecasting experiment with random forests. J. Oper. Res. Soc. 73 (8), 1755–1767. https://doi.org/10.1080/01605682.2021.1936668.
Dudek, G., 2015. Short-term load forecasting using random forests. In: Filev, D., Jab\lkowski, J., Kacprzyk, J., Krawczak, M., Popchev, I., Rutkowski, L., Sgurev, V., Sotirova, E., Szynkarczyk, P., Zadrozny, S. (Eds.), Intelligent Systems'2014. Springer International Publishing, pp. 821–828.
Feyen, E., Frost, J., Gambacorta, L., Natarajan, H., & Saal, M. (2021). Fintech and the digital transformation of financial services: Implications for market structure and public policy. https://www.bis.org/publ/bppdf/bispap117.htm.
Ganjoo, D., Mukherjee, S., Mukhopadhyay, S., 2023. Razorpay: providing payment convenience to disruptors. Indian Inst. Manag. Ahmedabad 1–19. https://doi.org/10.1108/CASE.IIMA.2023.000009.
Harris, J.L., 2021. Bridging the gap between 'Fin' and 'Tech': the role of accelerator networks in emerging FinTech entrepreneurial ecosystems. Geoforum 122, 174–182. https://doi.org/10.1016/j.geoforum.2021.04.010.
Jakhiya, M., Mittal Bishnoi, M., Purohit, H., 2020. Emergence and growth of mobile money in modern india: a study on the effect of mobile money. Adv. Sci. Eng. Technol. Int. Conf. (ASET) 2020, 1–10. https://doi.org/10.1109/ASET48392.2020.9118375.
Karpman, K., Basu, S., Easley, D., Kim, S., 2023. Learning financial networks with high-frequency trade data. Data Sci. Sci. 2 (1), 2166624 https://doi.org/10.1080/26941899.2023.2166624.
Khando, K., Islam, M.S., Gao, S., 2023. The emerging technologies of digital payments and associated challenges: a systematic literature review. Future Internet 15 (1). https://doi.org/10.3390/fi15010021.
Landis, W., Cha, S., 2020. Towards high performance stock market prediction methods. IEEE Cloud Summit 2020, 156–160. https://doi.org/10.1109/IEEECloudSummit48914.2020.00030.
Lin, C.-S., Lin, C.-Y., Reynolds, S., 2020. Applying the random forest model to forecast the market reaction of start-up firms: case study of GISA equity crowdfunding platform in Taiwan (Scopus). WSEAS Trans. Bus. Econ. 17, 241–259. https://doi.org/10.37394/23207.2020.17.26.
Liu, C., Fan, Y., Zhu, X., 2021. Fintech index prediction based on RF-GA-DNN algorithm. Wirel. Commun. Mob. Comput. 2021, e3950981 https://doi.org/10.1155/2021/3950981.
Luong, C., Dokuchaev, N., 2018. Forecasting of realised volatility with the random forests algorithm. J. Risk Financ. Manag. 11 (4), 4 https://doi.org/10.3390/jrfm11040061. Article 4.

Mention, A.-L., 2019. The future of fintech. Res. Technol. Manag. 62 (4), 59–63. https://doi.org/10.1080/08956308.2019.1613123.

Migozzi, J., Urban, M., Wójcik, D., 2023. You should do what India does": FinTech ecosystems in India reshaping the geography of finance. Geoforum, 103720. https://doi.org/10.1016/j.geoforum.2023.103720.

Mohanta, B., Nanda, P., Patnaik, S., 2020. Management of V.U.C.A. (Volatility, Uncertainty, Complexity and Ambiguity) using machine learning techniques in industry 4.0 paradigm. In: Patnaik, S. (Ed.), Big Data & Cyber Physical Systems (pp, New Paradigm of Industry 4.0: Internet of Things. Springer International Publishing, pp. 1–24. https://doi.org/10.1007/978-3-030-25778-1_1.

Moid, S., & Shankar, N. (2022). Creating Value Proposition for Rural Banking Customers in Emerging Markets: Adoption of Mobile Banking Technology Induced by Disruptive Events in India. In A. Thrassou, D. Vrontis, L. Efthymiou, Y. Weber, S. M. R. Shams, & E. Tsoukatos (Eds.), Business Advancement through Technology Volume I: Markets and Marketing in Transition (pp. 47–72). Springer International Publishing. https://doi.org/10.1007/978–3-031–07769-2_3.

Nenavath, S., 2022. Impact of fintech and green finance on environmental quality protection in India: By applying the semi-parametric difference-in-differences (SDID). In: Renewable Energy, 193, pp. 913–919. https://doi.org/10.1016/j.renene.2022.05.020.

Palmié, M., Wincent, J., Parida, V., Caglar, U., 2020. The evolution of the financial technology ecosystem: An introduction and agenda for future research on disruptive innovations in ecosystems. Technol. Forecast. Soc. Change 151, 119779. https://doi.org/10.1016/j.techfore.2019.119779.

Pazarbasioglu, C., Mora, A.G., Uttamchandani, M., Natarajan, H., Feyen, E., & Saal, M. (2020). DIGITAL FINANCIAL SERVICES.

Priyamvada, Wadhvani, R., 2017. Review on various models for time series forecasting. 2017 Int. Conf. Invent. Comput. Inform. (ICICI) 405–410. https://doi.org/10.1109/ICICI.2017.8365383.

Rundo, F., 2019. Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems. Appl. Sci. 9 (20), 20. https://doi.org/10.3390/app9204460. Article 20.

Sadorsky, P., 2021. A random forests approach to predicting clean energy stock prices. J. Risk Financ. Manag. 14 (2), 2 https://doi.org/10.3390/jrfm14020048. Article 2.

Saura, J.R., Reyes-Menéndez, A., deMatos, N., Correia, M.B., 2021. Identifying startups business opportunities from UGC on twitter chatting: an exploratory analysis. J. Theor. Appl. Electron. Commer. Res. 16 (6), 1929–1944. https://doi.org/10.3390/jtaer16060108.

Scardovi, C., 2017. Digital Transformation in Financial Services. Springer,.

Selvamuthu, D., Kumar, V., Mishra, A., 2019. Indian stock market prediction using artificial neural networks on tick data. Financ. Innov. 5 (1), 16 https://doi.org/10.1186/s40854-019-0131-7.

Subasi, A., Amir, F., Bagedo, K., Shams, A., Sarirete, A., 2021. Stock market prediction using machine learning. Procedia Comput. Sci. 194, 173–179. https://doi.org/10.1016/j.procs.2021.10.071.

Surana, K., Singh, A., Sagar, A.D., 2020. Strengthening science, technology, and innovation-based incubators to help achieve sustainable development goals: lessons from India. Technol. Forecast. Soc. Change 157, 120057. https://doi.org/10.1016/j.techfore.2020.120057.

Taujanskaitė, K., Kuizinaitė, J., 2022. Development of FinTech business in Lithuania: driving factors and future scenarios. Bus. Manag. Econ. Eng. 20 (1), 1 https://doi.org/10.3846/zenodo.2022.16738.

Thakur, M., Kumar, D., 2018. A hybrid financial trading support system using multi-category classifiers and random forest. Appl. Soft Comput. 67, 337–349. https://doi.org/10.1016/j.asoc.2018.03.006.

Valencia, F., Gómez-Espinosa, A., Valdés-Aguirre, B., 2019. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy 21 (6), 6. https://doi.org/10.3390/e21060589.

Yin, L., Li, B., Li, P., Zhang, R., 2023. Research on stock trend prediction method based on optimized random forest. CAAI Trans. Intell. Technol. 8 (1), 274–284. https://doi.org/10.1049/cit2.12067.

Zhang-Zhang, Y., Rohlfer, S., Rajasekera, J., 2020. An eco-systematic view of cross-sector fintech: the case of Alibaba and Tencent. Sustainability 12 (21). https://doi.org/10.3390/su12218907. Article 21.