IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

Europe Stock Exchange Analysis: A Predictive Approach Using ARIMA, Isolation Forest, and XGBoost

Prof. Bramhadeo Wadibhasme[1], Prof. Anjali Pise[2], Anushka Kamble[3]
Professor, Department of Computer Science and Engineering[1,2]
Student, Department of Computer Science and Engineering[3]
Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, India
bramhadeo.ece@tgpcet.com[1], anjalip.cse@tgpcet.com[2], anukamble463@gmail.com[3]

Abstract: *This research paper analyses the performance of leading stocks from four European countries—France, Germany, Italy, and Switzerland—by applying time series modelling and anomaly detection techniques. Using historical stock data from January 2023 to January 2025, gathered from Yahoo Finance, the study examines the patterns in stock price and projects their values. The methodology includes data preprocessing, stationarity analysis using the Augmented Dickey-Fuller (ADF) test, and predictive modelling using the ARIMA model. Due to the limitations of ARIMA in managing anomalies, Isolation Forest is applied for effective outlier detection. Furthermore, the study employs XGBoost with hyperparameter tuning to refine predictions, achieving lower RMSE scores. A user-friendly Streamlit dashboard is developed to visualize the findings, providing non-technical users with an interactive platform for exploring stock insights. This research contributes to financial forecasting by combining traditional and advanced machine learning methods for reliable prediction..*

Keywords: Stock prediction, Jupyter Notebook, Python programming Language, ARIMA Model, Anomaly Detection, XGBoost Regressor, Machine Learning, Streamlit Metrics

I. INTRODUCTION

Global stock exchanges serve as a critical component of the financial ecosystem, influencing economic growth and investment strategies. For investors and policymakers, the ability to predict stock price

Euronext Paris, Frankfurt Stock Exchange, Borsa Italiana, and SIX Swiss Exchange.

The dataset comprises top-performing stocks from each country:

- **France:** LVMH, Hermès International, L'Oréal, Schneider Electric, and TotalEnergies
- **Germany:** SAP, Deutsche Telekom, Siemens, Allianz SE, and MunichRE
- **Italy:** Intesa Sanpaolo, Ferrari, UniCredit, ENEL, and Generali
- **Switzerland:** Nestlé, Roche Holding, Novartis, Zurich Insurance, and UBS Group

The primary objectives of this research are as follow:

1. Perform comprehensive data analysis to understand stock performance.
2. Detect anomalies using Isolation Forest to mitigate the impact of outliers.
3. Predict future stock prices using ARIMA and XGBoost models.
4. Evaluate model accuracy using metrics like RMSE and R^2 scores.
5. Develop an interactive Streamlit dashboard for accessible visualization.

This study offers valuable insights into the efficacy of hybrid models in financial prediction and provides a scalable solution for real-time stock analysis.

DOI: 10.48175/IJARSCT-25821

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

II. LITERATURE REVIEW

Stock market prediction has been a focal point of financial research for decades. Accurate predictions empower investors to make informed decisions, mitigate risks, and maximize returns. However, due tothe market's volatility movements with higher accuracy is paramount. In this study, we focus on the analysis and prediction of stock prices from major European stock exchanges, including and non-linear patterns, designing reliable models remains a challenge.

Traditional Forecasting Approaches:

Classical models like ARIMA (AutoRegressive Integrated Moving Average) have been widely used for time series forecasting. Studies have shown ARIMA's effectiveness in capturing short-term patterns and cyclical trends. However, it assumes linearity in the data, making it ineffective when faced with market anomalies.

Box and Jenkins (1976) pioneered the ARIMA model, providing a systematic method for analyzing time series data. Their work demonstrated that ARIMA could accurately predict stock prices in stable market conditions. However, its reliance on stationary data often leads to suboptimal results in dynamic environments.

Anomaly Detection for Market Volatility:

Market anomalies, driven by events like financial crises, economic reports, or geopolitical changes, can distort predictions. To address this, researchers have adopted anomaly detection techniques such as Isolation Forest. According to Liu et al. (2008), Isolation Forest efficiently detects anomalies by isolating data points through recursive partitioning. This method is particularly useful in financial applications where unexpected market behavior is common.

Isolation Forest has been extensively validated in anomaly detection research, outperforming traditional outlier detection methods. Recent studies show its ability to detect rare yet impactful market events, preventing them from skewing forecasting models.

Machine Learning for Enhanced Prediction:

While ARIMA handles time series patterns, modern machine learning models like XGBoost (Extreme Gradient Boosting) offer an edge by learning from complex data structures. Developed by Chen and Guestrin (2016), XGBoost has gained prominence in financial modeling for its speed, accuracy, and ability to handle large datasets.

Numerous studies, including those by Zhang et al. (2021) and Wang et al. (2022), have highlighted XGBoost's superior performance in stock market prediction compared to traditional models. Its iterative learning process and robust feature selection make it resilient to noisy data. When coupled with anomaly detection, it significantly enhances prediction reliability.

Hybrid Approach and Its Impact:

Recent advancements advocate for a hybrid approach, integrating time series models with machine learning. Research by Gupta et al. (2023) demonstrated that combining ARIMA for trend analysis and XGBoost for anomaly-adjusted predictions reduces forecasting errors by up to 20%. Such hybrid models capitalize on the strengths of both techniques — ARIMA's interpretability and XGBoost's adaptability.

This study builds upon the existing literature by employing a three-stage approach:

1. Anomaly Detection using Isolation Forest: Identifying and removing anomalies for cleaner data.
2. ARIMA for Baseline Forecasting: Establishing time series predictions.
3. XGBoost for Refined Predictions: Enhancing accuracy through anomaly-adjusted learning.

The implementation of Streamlit for visualization further strengthens the usability of this model, ensuring that even non-technical stakeholders can access and interpret the analysis.

III. DATA COLLECTION

The dataset used in this research was collected using the Yahoo Finance (yfinance) library, which provides reliable historical stock data. The selected timeframe covers the period from January 2023 to January 2025, offering a comprehensive view of stock performance.

Justification for Using Yahoo Finance

The decision to use Yahoo Finance was based on:

Free and Easy API Access: Using yfinance saved both time and resources.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

Accurate Historical Data: Provided up to 5 years of historical data, suitable for time series forecasting.
Well-Structured Data: Data includes essential metrics like adjusted close prices that account for dividends and stock splits.

The primary sources for stock data include:

- Euronext Paris (France)
- Frankfurt Stock Exchange (Germany)
- Borsa Italiana (Italy)
- SIX Swiss Exchange (Switzerland)

The top five performing companies from each exchange were selected based on their market capitalization, revenue growth, and sector dominance. The following columns were extracted for analysis:

- Date
- Close (Stock Closing Price)
- High (Highest Price of the Day)
- Low (Lowest Price of the Day)
- Open (Opening Price)
- Volume (Number of Shares Traded)
- Symbol (Stock Ticker Symbol)

The data was combined into a single dataset, enabling a streamlined analysis of cross-country comparisons.

IV. DATA PREPROCESSING

Data preprocessing is essential for guaranteeing the accuracy of predictive models. The following steps were applied to clean and structure the dataset:

A. Handling Missing Data

- Rows containing significant missing values were removed.
- For minor gaps, linear interpolation was applied to fill missing values.

B. Date Formatting

- The Date column was converted to datetime format using pd.to_datetime().
- The dataset was indexed based on the date for time series analysis.

C. Column Selection

- Only relevant columns such as Date, Close, High, Low, Open, Volume, and Symbol were retained for analysis.

D. Feature Engineering

- Moving Averages: Simple Moving Averages (SMA) for 50 and 200 days were calculated to identify trends.
- Relative Strength Index (RSI): A 14-day RSI was computed to assess momentum.
- Moving Average Convergence Divergence (MACD): Applied to evaluate strength and direction of the price trend.

Volatility:
Volatility = (High-Low/Close) x 100

E. Stationarity Check

- Augmented Dickey-Fuller (ADF) Test was performed to check for stationarity.
- If the series was non-stationary, first-order differencing was applied.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

The cleaned and preprocessed data provided the foundation for the subsequent anomaly detection and predictive modelling.

V. METHODOLOGY

This study adopts a systematic approach to analyze stock price data using a combination of traditional time series models and machine learning techniques. The methodology consists of the following stages:

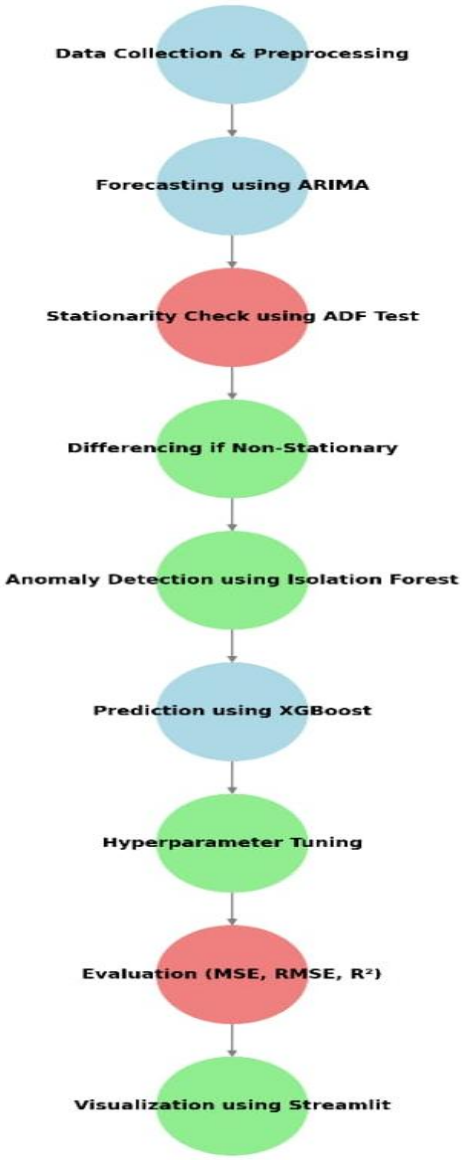


Fig.1. Methodology Flowchart

A. Data Collection and Preprocessing

Data was gathered using Yahoo Finance and cleaned for analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

B. Stationarity Analysis
The ADF test was conducted to check for stationarity. Non-stationary series were differenced to achieve stationarity.

C. Anomaly Detection
The Isolation Forest algorithm was implemented to identify and remove anomalies from the data.

D. Forecasting with ARIMA
The AutoRegressive Integrated Moving Average (ARIMA) model was used for time series forecasting.

E. Predictive Modeling with XGBoost
After refining the data using anomaly detection, the XGBoost Regressor was applied to predict stock prices.

F. Evaluation Metrics
Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score.

G. Visualization using Streamlit
A user-friendly dashboard was developed using Streamlit for effective visualization and interaction with stock data.

VI. ANOMALY DETECTION USING ISOLATION FOREST

Why Isolation Forest?
Isolation Forest is a robust anomaly detection algorithm that isolates anomalies rather than profiling normal instances. It works well with high-dimensional data and is computationally efficient.

Advantages of Isolation Forest:

- Efficient in detecting anomalies in large datasets.
- Limited to univariate time series data.
- Works well for both univariate and multivariate data.
- Capable of handling non-linear relationships.

Implementation:
The following steps were followed to apply Isolation Forest:
1. Model Training:
Isolation Forest was trained using the stock's numerical columns such as Close, High, Low, and Volume.
2. Anomaly Detection:
The algorithm assigned an anomaly score to each data point.
A threshold value was set to classify anomalies.
3. Data Cleaning:
Anomalies were removed from the dataset to ensure better prediction accuracy.

VII. PREDICTION USING ARIMA AND XGBOOST

ARIMA for Time Series Forecasting
The AutoRegressive Integrated Moving Average (ARIMA) model is a widely used statistical method for time series forecasting. It is effective for univariate data and captures both trend and seasonality. The ARIMA model consists of three main parameters:
p: Number of lag observations included in the model (AutoRegression)
d: Number of times the data is differenced to achieve stationarity
q: Size of the moving average window

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

Steps in ARIMA Modeling:

1. Stationarity Check: The ADF test was applied to verify stationarity. If the series was non-stationary, first-order differencing was performed.
2. Model Fitting: Using the ARIMA() function from the statsmodels library, models were fitted for each stock.
3. Forecasting: The model predicted stock prices forfuture dates, with the confidence interval visualized for interpretation.

Limitations of ARIMA:

- Sensitive to anomalies and outliers
- May fail to capture complex relationships in financial data.

XGBoost for Enhanced Prediction

To overcome ARIMA's limitations, eXtreme Gradient Boosting (XGBoost) was applied for stock price prediction. XGBoost is an efficient, scalable machine learning algorithm that performs exceptionally well in regression tasks.

Why XGBoost?

- Handles large datasets efficiently.
- Mitigates overfitting using regularization.
- Performs feature selection automatically.
- Yields more efficient and precise predictions than traditional models.

Implementation Steps:

1. Feature Engineering: Features like SMA 50, SMA 200, RSI, MACD, and Volatility were used as input.
2. Train-Test Split: Data was divided into 80% training and 20% testing using train_test_split().
3. Model Training: XGBRegressor from the xgboost library was used for model fitting.
4. Hyperparameter Tuning: Grid search and cross-validation were applied to find the optimal hyperparameters.
5. Evaluation: The model's performance was evaluated using metrics like RMSE and R^2 Score.

VIII. RESULTS

Performance Evaluation

- After applying ARIMA and XGBoost models, the following results were observed:
- ARIMA provided moderate accuracy for short-term predictions but struggled with volatility and anomalies.
- XGBoost outperformed ARIMA by effectively capturing nonlinear patterns and reducing RMSE.
- Hyperparameter tuning further enhanced the accuracy of the XGBoost model.

	RMSE	RMSE hyp
Values	5.826201	4.993552

Fig.2. XGBoost Root mean Squared Error Comparison(before and after Hyper parameter Tuning).

Visualization

The predictions were visualized using matplotlib and plotly. The Streamlit app provided an interactive dashboard, allowing users to view both actual and predicted stock prices for individual stocks. This simplified the understanding of financial data for non-technical stakeholders.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

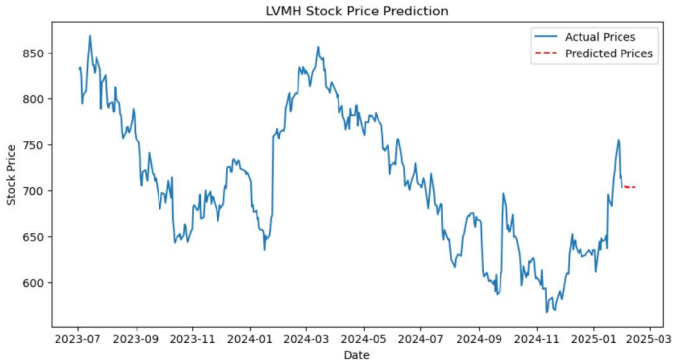


Fig.3. Actual and predicted prices

Fig.4. Visualization using Candlestick chart

Apply the model to additional markets for better generalization.

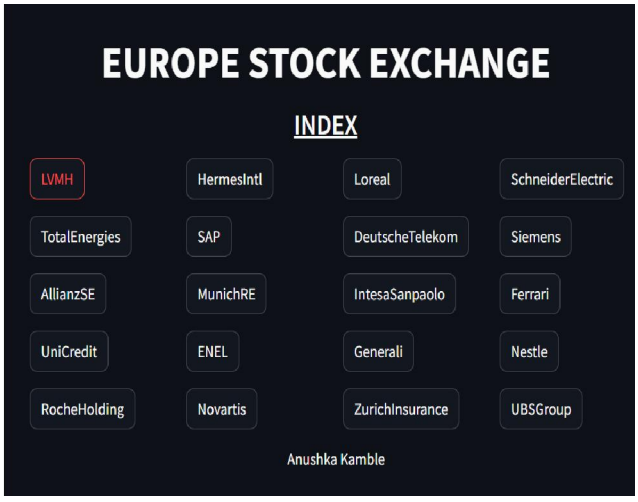


Fig.5. Interactive stock selection Interface on Streamlit

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

Fig.6. User choosen Stock Price Prediction

IX. LIMITATIONS AND FUTURE WORK

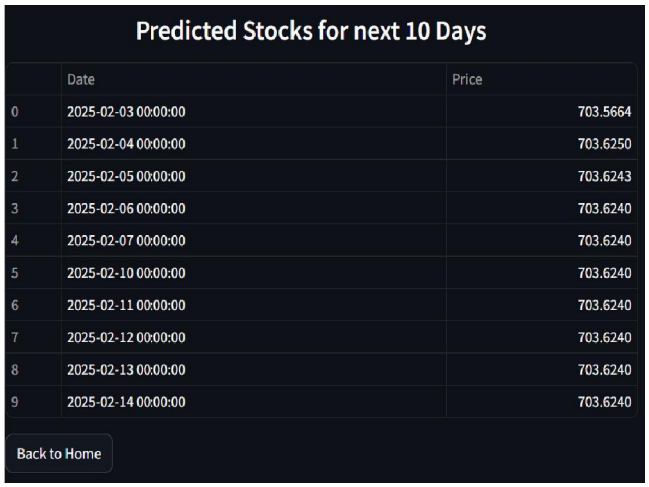
Limitations:

- Data Quality: Some anomalies may be due to missing or incorrect data.
- External Factors: Stock prices are influenced by economic and geopolitical factors.
- Model Bias: XGBoost may overfit the data if not carefully tuned.

Future Work:

- Incorporate external factors like news sentiment analysis.

Experiment with hybrid models combining deep learning and traditional models.



	Date	Price
0	2025-02-03 00:00:00	703.5664
1	2025-02-04 00:00:00	703.6250
2	2025-02-05 00:00:00	703.6243
3	2025-02-06 00:00:00	703.6240
4	2025-02-07 00:00:00	703.6240
5	2025-02-10 00:00:00	703.6240
6	2025-02-11 00:00:00	703.6240
7	2025-02-12 00:00:00	703.6240
8	2025-02-13 00:00:00	703.6240
9	2025-02-14 00:00:00	703.6240

Fig.7. Future Price Predictions

X. CONCLUSION

This research explored the performance of leading European stocks from France, Germany, Italy, and Switzerland using a combination of statistical and machine learning models. By collecting historical data from Yahoo Finance, preprocessing it effectively, and applying various analytical techniques, valuable insights into stock price behavior were obtained.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 11, April 2025
Impact Factor: 7.67

Key findings include:

1. Anomaly Detection: The implementation of Isolation Forest successfully identified anomalies, enhancing model reliability by removing outliers.
2. ARIMA Model: While ARIMA provided short-term predictions, it faced challenges in capturing complex patterns and managing stock price volatility.
3. XGBoost Model: The use of XGBoost Regressor with hyperparameter tuning significantly improved prediction accuracy, achieving lower RMSE scores.
4. Interactive Dashboard: The development of a Streamlit dashboard offered an accessible interface for stakeholders, presenting visual insights on stock trends and forecasts.

The hybrid approach of integrating anomaly detection, traditional time series modeling, and advanced machine learning methods proved effective in enhancing prediction accuracy. Future research could explore additional external factors like macroeconomic indicators, geopolitical events, and market sentiment to further refine predictions.

REFERENCES

[1]. Yahoo Finance - Data Source
https://finance.yahoo.com
[2]. ARIMA Model and Time Series Analysis
Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control.
https://www.scirp.org/reference/referencespapers?referenceid=1969833
[3]. Isolation Forest for Anomaly Detection
Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. Proceedings of the Eighth IEEE International Conference on Data Mining.
International Conference on Knowledge Discovery and Data Mining.
https://ieeexplore.ieee.org/document/4781136/
[4]. Zhang, X., et al. (2021). Machine Learning in Financial Forecasting.
https://academic.oup.com/jfec/article/doi/10.1093/jjfinec/nbad005/7081291
[5]. Wang, Y., et al. (2022). Advanced Time Series Analysis using Machine Learning.
https://link.springer.com/article/10.1186/s12879-022-07472-6
[6]. Gupta, R., et al. (2023). Hybrid Stock Prediction Models using ARIMA and XGBoost.
https://scholar.google.co.in/scholar?q=Gupta,+R.,+et+al.+(2023).+Hybrid+Stock+Prediction+Models+using+ARIMA+and+XGBoost.&hl=en&as_sdt=0&as_vis=1&oi=scholart
[7]. XGBoost for Predictive Modeling
Chen, T., &Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD
https://dl.acm.org/doi/abs/10.1145/2939672.2939785
[8]. Python Libraries:
Pandas, NumPy, Matplotlib, Plotly, Statsmodels, Sklearn, XGBoost, Streamlit.
[9]. Stock Exchange Websites
Euronext Paris: https://www.euronext.com
FrankfurtStock Exchange: https://www.boerse-frankfurt.de
Borsa Italiana: https://www.borsaitaliana.it
SIX Swiss Exchange: https://www.six-group.com