DATA 642 Conference Paper
By Emily Eames & Aidan Hennessy
American University December 7, 2025

Title: US & India Financial Anomaly Time Series Comparative Analysis

Abstract:

This study analyzes financial market anomalies in the United States and India, using unsupervised and supervised machine learning methods. The study's dataset consists of daily stock market return data from the S&P 500 and Nifty 50 spanning from 2013 to 2019. The unsupervised learning methods of isolation forest and local outlier factor were first used with just returns alone and then again with an expanded feature set accounting for volatility and momentum. Contrary to initial assumptions, we found that the S&P 500 exhibits more frequent and prolonged anomalies than the Nifty 50, indicating that market development does not necessarily reduce susceptibility to market destruction. Incorporating volatility and momentum features significantly improved anomaly detection and revealed a bidirectional relationship between the United States and the stock markets. Our supervised long short-term memory (LSTM) model achieved a recall of 85.7% demonstrating a strong time series forecasting performance, even in an extremely imbalanced financial dataset.

Introduction:

The objective of this study was to compare the frequency, duration, and structure of financial anomalies within the stock market between a developed country, the United States, and a developing country, India. We began with two initial hypotheses. With our first hypothesis, we expected to see financial anomalies occur more often in the developing nation than in our developed nation. And with our second hypothesis, we expected to see those financial anonymity persist for longer periods in our developing nation as opposed to our developed nation. Both of these hypotheses were derived from the initial assumption that developing markets tend to have lower market efficiencies and higher volatility.

We chose to use the United States and India for this comparative analysis due to their demographic and economic similarities. More specifically, this choice was determined by market capitalization and the population size of each country. The United States is home to the world's largest developed economy and is the most populous developed nation. While India is home to one of the largest developing economies and is the world's most populous developing nation. The S&P 500 serves as the primary market benchmark within the United States and is typically regarded as the best indicator of large equity performances[1]. In addition, the S&P 500 covers approximately 80% of the total US market capitalization. India's Nifty 50 index represents the top 50 largest and actively traded companies on India's national stock exchange. It is also ranked the world's fourth-largest stock market. Given that the United States and India are two of the world's most populous nations and because they have comparable stock market indices, we felt confident that these two markets provided a meaningful basis for comparison.

Methodology/Data

Our data consists of daily closing prices for both the S&P 500 and Nifty 50 from January 1, 2013, to December 31, 2019. We split our data into training and testing sets using a 70/30 chronological split. With 70% of the training data consisting of earlier portions of the data set, while 30% of the testing data was reserved for the latter parts of our data. Before using our data in any of our models, we first cleaned out each of the NA values, which consisted of a very small minority of the observations in our dataset. These NA values were NA values primarily due to national holidays or any days when only one market was open. They were subsequently excluded from our dataset to ensure temporal alignment.

We began this study by using unsupervised learning methods - Isolation Forest (IS) and Local Outlier Factor (LOF) - to detect any anomalies within the testing set. This initial analysis only relied on daily stock market returns. Daily stock market returns are defined as the percentage change in a closing price relative to the previous trading day. The same procedure was repeated in our second round of experiments, where we used an expanded feature set that incorporated volatility and momentum measures. To further analyze financial anomaly structure and feature relevance, we used feature importance techniques such as K-Means clustering and Random Forest to detect further anomalies that may have been missed when we used just market returns. Finally, we used a Long Short-Term Memory (LSTM) neural network to implement supervised forecasting to attempt to predict future anomalies.

Results and Analysis

Unsupervised Anomaly Detection: Isolation Forest & LOF - Returns Only

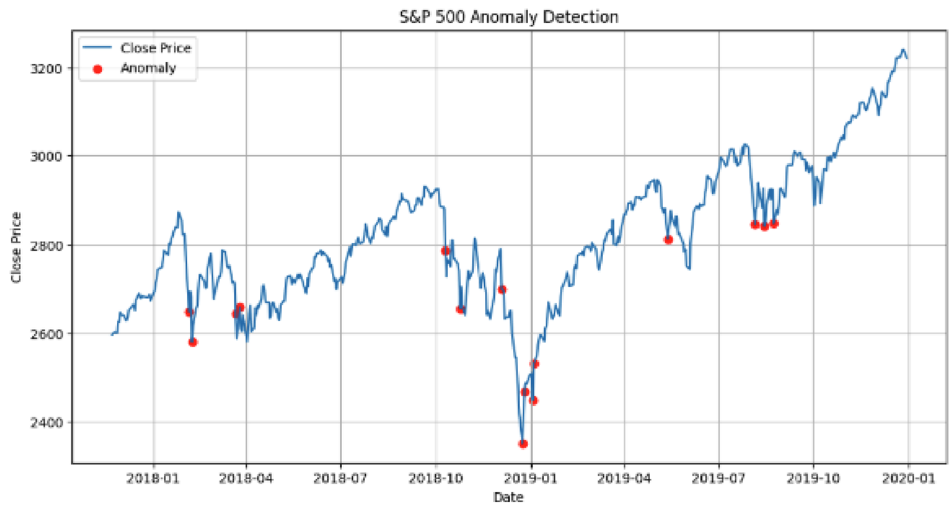
Using Isolation Forest on returns alone, our model identified approximately 32 anomalies - about 6%- in the S&P 500 and 11 anomalies - about 2% - in the Nifty 50. These findings directly contradicted our initial assumption that anomalies would occur more frequently in developing countries. This is clearly seen on Figures 1 and 2, where market anomalies are indicated as red

---

[1] *S&P 500®*. S&P Dow Jones Indices. (n.d.). https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview

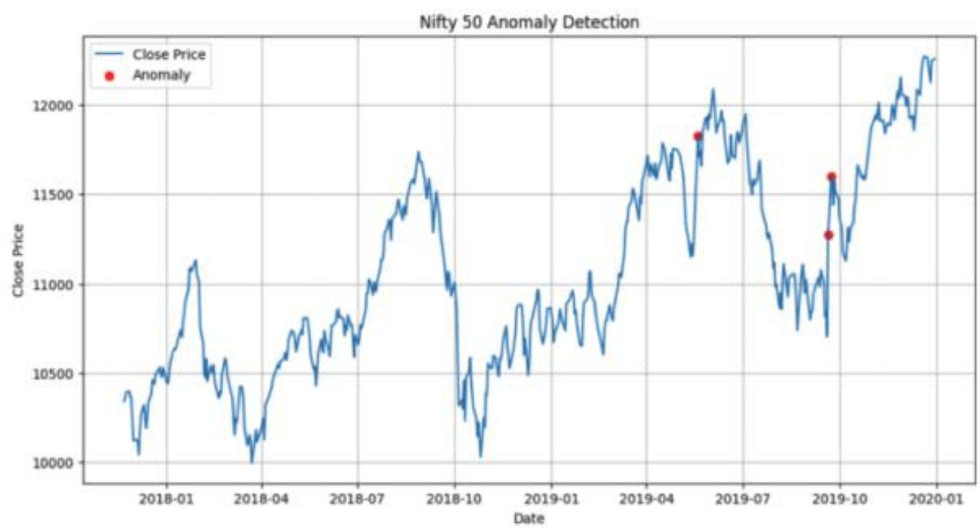
points on the graph. We clearly see a larger number of anomalies occurring with the S&P 500 dataset compared to India's Nifty 50.

Figure 1: S&P 500 Anomaly Detection



*Anomalies are indicated by the red points on the graph

Figure 2: NIFTY 50 Anomaly Detection

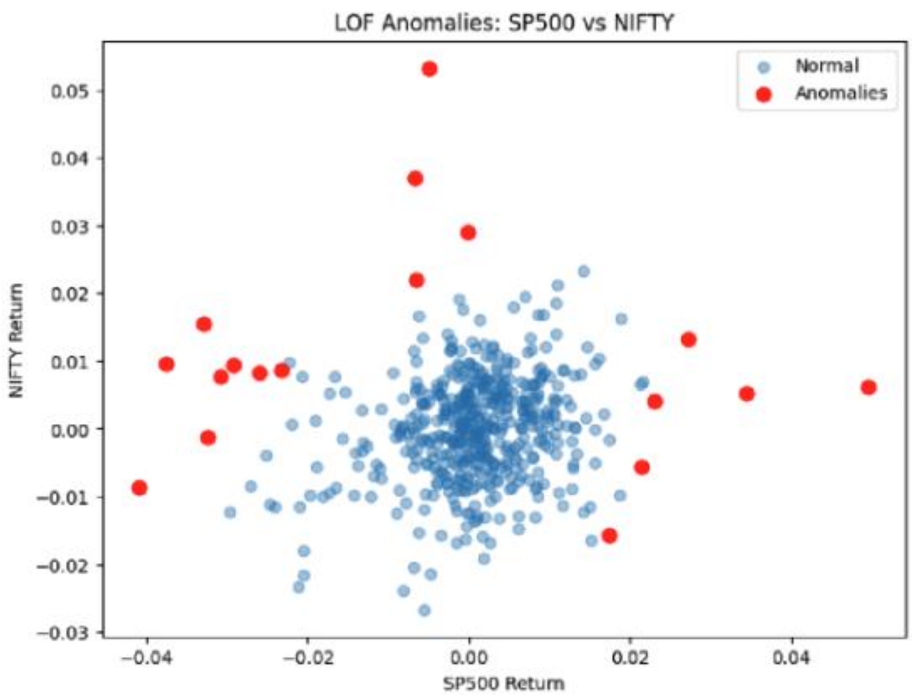


*Anomalies are indicated by the red points on the graph

Instead, what we saw was the US market exhibiting not only a large number of anomalies, but also more persistent periods of anomalies. Suggesting that market sophistication is not necessarily a main determinant for market extremes or unusual behaviors.

Our Local Outlier Factor analysis further challenged our assumptions once more. After aligning our datasets and excluding holidays unique to each country, our LOF model identified 18 anomalies across 500 testing observations. These anomalies can be seen in Figure 3, where normal occurrences are represented by blue points and anomalies are red points.

Figure 3: Local Outlier Factor Anomalies - S&P 500 vs. NIFTY 50



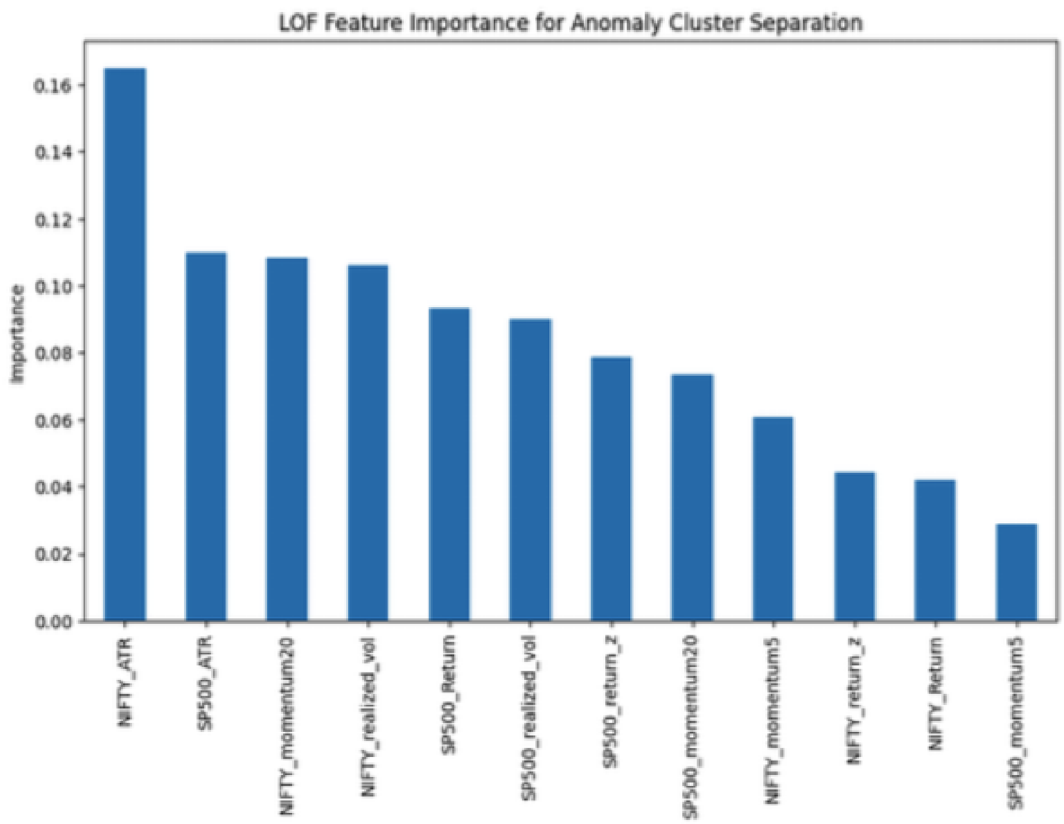
*Anomalies are indicated by the red points on the graph

These anomalies largely captured major return shifts rather than more subtle irregularities. This highlights the limitations of using just returns alone for anomaly detection. Further, these results contradicted the hypothesis that anomalous events would last longer in developing nations as opposed to developed nations.

K-Means Clustering and Feature Importance

K-means clustering and random forest feature importance analysis were then conducted on the combined anomaly dataset. We ultimately found that the optimal number of clusters was four. This can be seen in Figure 4, in the first four bars. We also observed that volatility-related measures seem to dominate cluster separation. We also found that across both markets, the Average True Range (ATR) emerged as the most influential feature when it came to explaining anomalous behavior.

Figure 4: LOF Feature Importance for Anomaly Cluster Separation



Though both markets shared ATR as the most influential feature for indicating anonymous behavior, some market-specific predictors differed between both countries. For the S&P 500, the daily raw returns remained the strongest predictor among anomalies (Figure 5). While for the Nifty 50, short-term momentum (Momentum 5) played the dominant role (Figure 6). What these differences suggest is that while globally volatility cycles tend to influence both markets, when it comes to how anomalies manifest within each market, local dynamics play the dominant role.

Figure 5: S&P Feature Importance with IF Anomalies Highlighted

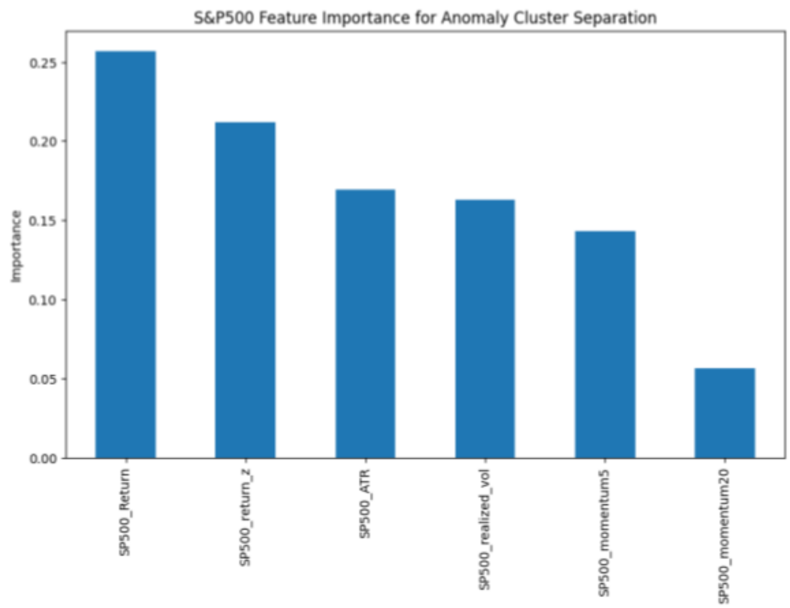
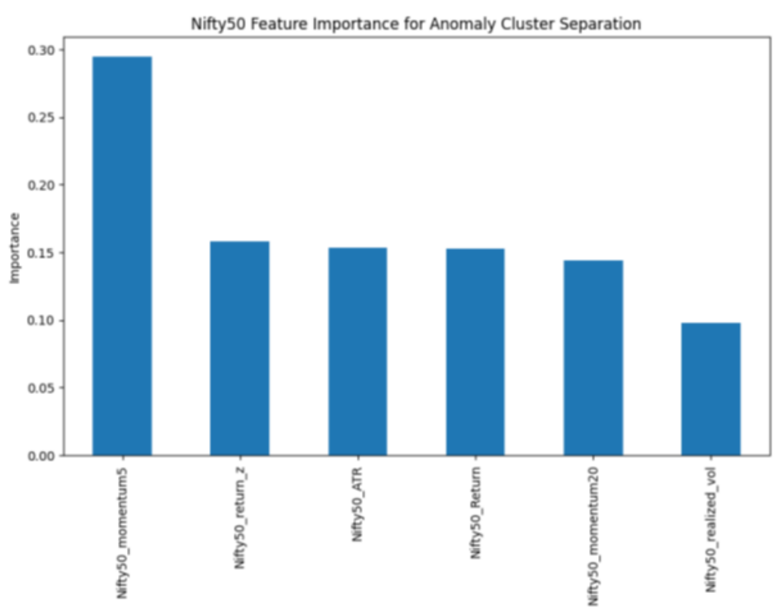


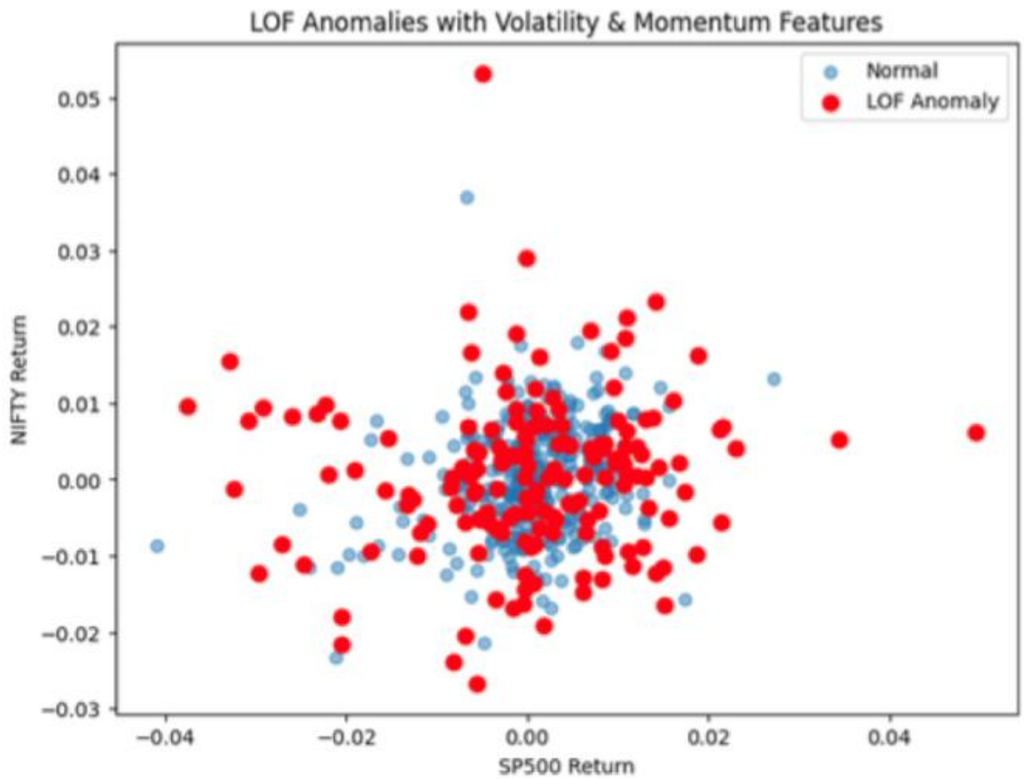
Figure 6: Nifty50 Feature Importance for Anomaly Cluster Separation



Unsupervised Anomaly Detection: Isolation Forest & LOF with Feature Expansion

After recognizing the limitations of a return-only based analysis, we expanded the feature space to include market return Z-Scores, short-term momentum (Momentum 5), medium-term momentum (Momentum 20), average true range (ATR), and realized volatility. Ultimately, we saw that incorporating these features led to a dramatic increase and overall anomaly detection across both markets. Momentum 5 is momentum over the course of a 5-day period. Momentum 20 is momentum over the course of a 20-day period.

Figure 7: LOF with Momentum and Volatility Feature

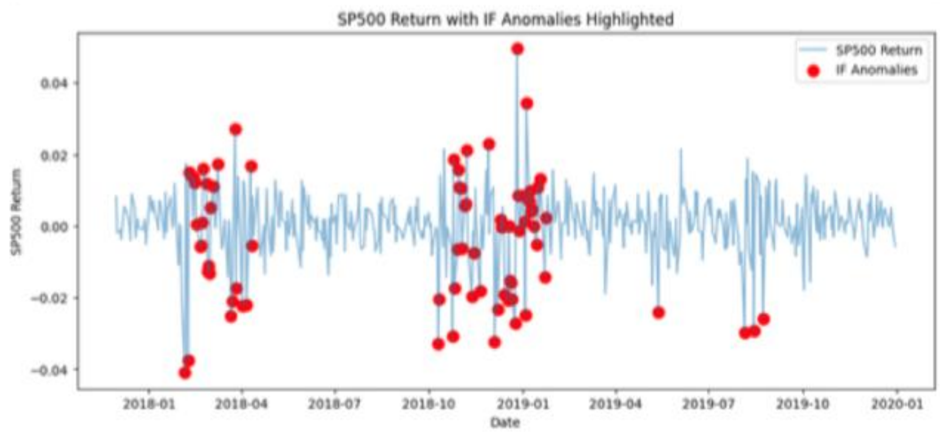


*Anomalies are indicated by the red points on the graph

With our expanded feature set, LOF identified 164 anomalies in total (see Figure 5), which illustrates that returns alone account for a significant portion of anomalous behavior within the market. Isolation forest detected 75 anomalies within the S&P 500 (see Figure 6) and 18 anomalies in the Nifty 50 (see Figure 7), with anomalous clusters concentrated around the years

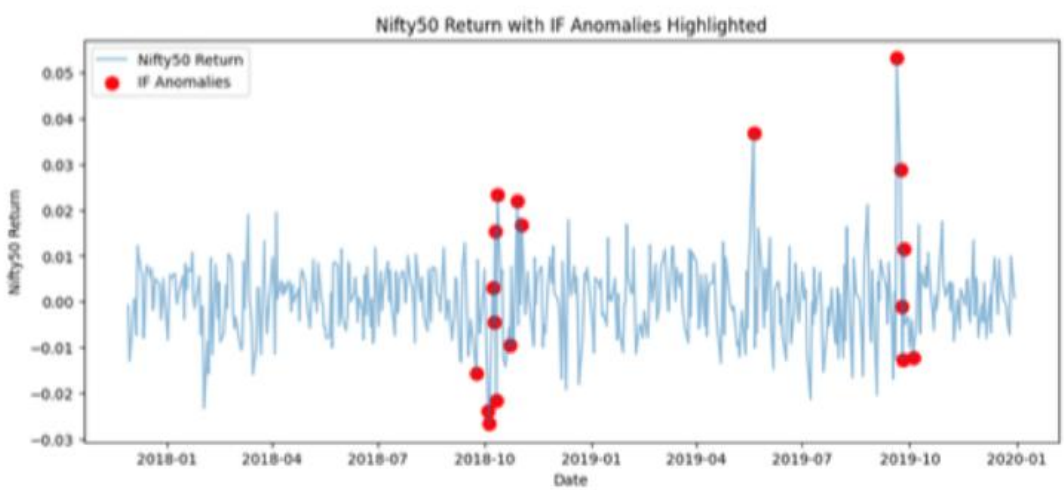
2018 and late 2019. These results strongly indicate that volatility and momentum-driven dynamics are crucial to understanding financial shocks within the market.

Figure 8: S&P Returns with IF Anomalies Highlighted



*Anomalies are indicated by the red points on the graph

Figure 9: NIFTY 50 Returns with IF Anomalies Highlighted



*Anomalies are indicated by the red points on the graph

Further, the expanded feature analysis revealed a bidirectional relationship between the United States and the Indian stock market. We found that in 2018, anomalies in India were followed shortly by anomalies in the United States. While in late 2019, we see that direction reversed, where anomalies in the United States precede those in India. This can be seen when comparing 2018-2019 in Figures 5 and 6. This bidirectional relationship suggests cross-market interdependence rather than isolated domestic shocks, which is consistent with prior literature on international financial spillovers[2].

What is unique about our findings, however, is that in prior literature, the presence of a bidirectional relationship between stock market markets is only theorized between two developed nations[3]. Typically, we see unidirectional relationships between developed countries and developing countries, an idea our research clearly contradicts.

Supervise Forecasting with Long Short-Term Memory

To do supervised forecasting, a Long Short-Term Memory (LSTM) model was trained using a 30-day rolling window, incorporating both engineered features and prior labeled anomalies. Due to extreme class imbalance, we see that true anomalies consisted of fewer than 3% of our total observations. To address this, we used weights to prioritize recall.

The LSTM model achieved a recall of 0.857, meaning it was successfully capturing 85.7% of true anomalies within the testing set. This far exceeded our goal performance rate of 80% for financial anomaly detection. Our precision was predictably low at 0.19, reflecting our model's tendency to over-predict anomalies. The resulting F1 score of 0.321 is also consistent with our high recall model and does not indicate weak performance since we prioritized recall over precision, due to the fact that it is more costly to miss a financial anomaly than to over-predict.

We found that the optimal decision threshold was approximately 0.038, which is substantially lower than the conventional 0.5 cutoff. What this lower threshold reflects is the rarity of anomalies and enables our model to detect subtle precursors to anomalous events. Though increasing sensitivity, improved recall, it did come at the cost of precision. Again, we find this to be an acceptable trade-off since the goal and objective are early detection of potential market disruptions.

Conclusion

Our study demonstrates that volatility and momentum are far more informative determinants when identifying financial anomalies than raw returns alone. Feature engineering increased the detection of anomalies from 18 to 165, confirming that the return-only approaches significantly

---

[2] Mishra, Alok, et al. "Volatility Spillover between Stock and Foreign Exchange Markets: Indian Evidence." *International Journal of Business*, vol. 12, 2007, www.researchgate.net/publication/228637821_Volatility_Spillover_between_Stock_and_Foreign_Exchange_Markets_Indian_Evidence. Accessed 2 Dec. 2025.

[3] Demir, I. (2025, January 6). *The Role of Stock Markets in Economic Growth: Empirical evidence from panel data analysis*. Research Working Paper: The role of stock markets in economic growth. https://www.world-exchanges.org/our-work/articles/role-stock-markets-economic-growth

underestimated abnormal market behavior. Contrary to our initial hypotheses, the S&P 500 exhibited far more frequent and prolonged anomalies than the Nifty 50, indicating that developed markets remain highly susceptible to systemic shocks.

Our analysis also revealed a clear bidirectional relationship between the United States and Indian stock markets. Where we see anomalies in one market often precede anomalies in the other. Feature importance analysis also reinforced this interdependency, as ATR dominated anomaly detection in both markets. This suggests that shared global volatility cycles drive market disruptions across nations.

Our analysis on which features are most successful at detecting stock market disruptions were different for S&P 500 than Nifty 50 when we analyzed them separately. Returns do the best job detecting S&P 500 stock market disruptions. While Momentum 5 does the best job detecting Nifty 50 market disruptions. This showcases the inadequacy of using just one feature to detect market disruptions in both stock markets. And that local dynamics unique to one nation matter.

Lastly, our supervised LSTM forecasting significantly improved time series predictive performance. Our recall of 85.7% in an extreme imbalance financial dataset also underscores the importance of combining historical anomaly patterns with feature engineering. Overall, integrating unsupervised detection, feature engineering, cluster analysis, and deep learning techniques provides us with a robust framework for not only forecasting, but understanding financial anomalies in global markets.

References

Agyemang, E. F. (2024). Anomaly detection using unsupervised machine learning algorithms: A simulation study. *Scientific African*, *26*. https://doi.org/10.1016/j.sciaf.2024.e02386

Barazandeh, I., Haratizadeh, S., & Sermpinis, G. (2025). A temporal graph-based contrastive approach for Financial Time Series forecasting. *Engineering Applications of Artificial Intelligence*, *153*, 110834. https://doi.org/10.1016/j.engappai.2025.110834

Hu, Haolan. "A Global Comparative Study of Financial Market Anomalies." *Advances in Economics Management and Political Sciences*, vol. 153, no. 1, 7 Jan. 2025, pp. 16–21, scispace.com/papers/a-global-comparative-study-of-financial-market-anomalies-7e1fby0wsd3j, https://doi.org/10.54254/2754-1169/2024.19468. Accessed 26 May 2025

Demir, I. (2025, January 6). *The Role of Stock Markets in Economic Growth: Empirical evidence from panel data analysis*. Research Working Paper: The role of stock markets in economic growth. https://www.world-exchanges.org/our-work/articles/role-stock-markets-economic-growth

Jain, Lakshya. "Momentum vs. Value Investing: A Comparative Analysis of the S&P 500 and Nifty 50." International Journal of Innovative Science and Research Technology, 22 May 2025, pp. 818–821, www.researchgate.net/publication/392001189_Momentum_vs_Value_Investing_A_Comparative_Analysis_of_the_SP_500_and_Nifty_50, https://doi.org/10.38124/ijisrt/25may210.

Mary, Britney Johnson. *The Comparative Study of NIFTY and S&P 500 Index Using GARCH Models*. 9 Dec. 2024, www.researchgate.net/publication/386571730_The_Comparative_Study_of_NIFTY_and_SP_500_Index_Using_GARCH_Models.

Mishra, Alok, et al. "Volatility Spillover between Stock and Foreign Exchange Markets: Indian Evidence." *International Journal of Business*, vol. 12, 2007, www.researchgate.net/publication/228637821_Volatility_Spillover_between_Stock_and_Foreign_Exchange_Markets_Indian_Evidence. Accessed 2 Dec. 2025.

Muguto, Lorraine, and Paul-Francois Muzindutsi. "A Comparative Analysis of the Nature of Stock Return Volatility in BRICS and G7 Markets." *Journal of Risk and Financial Management*, vol. 15, no. 2, 18 Feb. 2022, p. 85, https://doi.org/10.3390/jrfm15020085.

Samuel Kwaku Agyei, et al. "Spillovers and Contagion between BRIC and G7 Markets: New Evidence from Time-Frequency Analysis." *PLOS ONE*, vol. 17, no. 7, 27 July 2022, pp. e0271088–e0271088, https://doi.org/10.1371/journal.pone.0271088. Accessed 21 Apr. 2023.