# A Study on the Impact of IT on Investor Behavior towards Stock Market and Cryptocurrency Investment: A Data Mining Approach

Radhika M Dhakane[1], Dr. Kaveri Lad[2]
[1]*Research Scholar, Dr. Babasaheb Ambedkar Marathwada University, Ch. Sambhajinagar*
[2]*Assistance Professor, Dr. Babasaheb Ambedkar Marathwada University, Ch. Sambhajinagar*

***Abstract*—This paper explores the impact of information technology (IT) on investor behavior in both the stock market and cryptocurrency markets through a data mining approach. The study integrates market indicators such as price and volume with information flows from news outlets, social media platforms, and search trends, alongside investor trading records where accessible. By employing feature engineering, sentiment analysis, clustering, and predictive modeling, the research aims to uncover patterns in decision-making and market reactions. Two central questions are addressed: first, how IT-enabled information sources including news frequency, online sentiment, and search interest shape trading activity and price fluctuations in equities and cryptocurrencies; second, whether data-mining models can effectively forecast short-term investor responses and market direction. Supervised techniques such as Random Forest, XGBoost, and LSTM, as well as unsupervised methods like K-means and DBSCAN, are applied to financial and online datasets. Findings reveal that real-time signals, particularly social sentiment and search intensity, enhance predictions of trading volume and volatility more strongly in cryptocurrency markets than in equities, highlighting a greater IT-driven herding effect. The paper concludes with implications for investors, regulators, and platform designers.**

***Index Terms*—Information Technology, Investor Behavior, Stock Market, Cryptocurrency, Data Mining, Sentiment Analysis, Machine Learning, Predictive Modeling, Clustering, Big Data Analytics, Financial Markets, Trading Behavior.**

## I. INTRODUCTION

The rapid advancement of information technology (IT) has transformed the way investors interact with financial markets. With the emergence of online trading platforms, mobile applications, social media, and real-time data feeds, investment decisions are increasingly influenced by technology-driven information sources rather than traditional financial reports alone. In the stock market, investors rely on digital tools for price monitoring, news analysis, and algorithmic trading, while in cryptocurrency markets, social media sentiment and online discussions play a critical role in shaping volatility and trading patterns. This shift has not only democratized access to financial information but has also accelerated the pace at which investor behavior responds to external signals. However, the abundance of information has introduced new challenges, such as herding effects, overreaction to news, and susceptibility to misinformation. In this context, data mining techniques provide a powerful means of uncovering hidden patterns, analyzing investor behavior, and predicting short-term market responses. By integrating market data, social sentiment, and search trends, this study aims to evaluate how IT-driven information influences investor decisions across equities and cryptocurrencies, and how predictive models can be used to capture these behavioral dynamics.

## II. LITERATURE SURVEY

Investor behavior in financial markets has been widely studied, but the rise of cryptocurrencies has introduced new dynamics driven by information technology and digital communication channels. Traditional financial markets, such as equities and commodities, are typically influenced by fundamentals, macroeconomic factors, and institutional strategies. In contrast, cryptocurrency markets exhibit stronger sensitivity to

online sentiment, information intensity, and global digital engagement, often amplifying volatility and herding behavior.

Recent studies have emphasized this difference. Almeida and Gonçalves (2023) provided a systematic literature review highlighting that investor behavior in cryptocurrency markets is heavily shaped by sentiment-driven mechanisms and speculative dynamics, diverging from traditional market rationality. Similarly, Zhang, Chen, and Wang (2024) explored the spillover effects of cryptocurrency volatility on stocks and gold, showing how digital asset shocks increasingly influence broader financial markets. Other reviews, such as Almeida (2021) and Almeida & Gonçalves (2022), underline the growing relevance of volatility and risk management in the crypto space, as well as methodological challenges in studying these phenomena.

At the market microstructure level, regional evidence suggests heterogeneous trading behavior. Andrade et al. (2021) examined Bitcoin exchanges in Brazil, showing distinct price dynamics in parallel markets, while Bouraoui (2020) analyzed emerging countries, identifying social, technological, and institutional factors as key drivers of Bitcoin trading volumes. Together, these findings highlight a research gap: while the role of IT-enabled signals (e.g., social media, search intensity, news frequency) in shaping investor behavior is acknowledged, systematic data-mining approaches that integrate these signals with financial microdata remain limited.

## III. DATA SOURCES AND PREPROCESSING

### 3.1 Data sources

For this study, multiple datasets were collected from financial APIs, online platforms, and public repositories to capture both quantitative market activity and qualitative information flows. Stock market data, including historical prices, volumes, and volatility measures, were obtained from sources such as Yahoo Finance and AlphaVantage. Cryptocurrency data, consisting of price movements, market capitalization, and trading volume, were extracted from CoinMarketCap and Binance APIs, providing 24/7 transaction records. To represent the flow of information technology-driven signals, data from social media platforms like Twitter and Reddit were collected using APIs, focusing on investor sentiment, discussion frequency, and trending topics. News data were gathered from Google News and RSS feeds, while Google Trends provided search interest scores for selected tickers and cryptocurrencies.

Table No. 1: Dataset (Jan–Jun 2024) showing Apple stock vs Bitcoin

| Date | AAPL Price (USD) | AAPL Volume (M) | BTC Price (USD) | BTC Volume (B) |
|---|---|---|---|---|
| 07-01-24 | 188.28 | 62 | 45,340.18 | 25 |
| 14-01-24 | 186.72 | 45 | 46,169.93 | 30 |
| 21-01-24 | 183.88 | 73 | 44,130.00 | 28 |
| 28-01-24 | 188.76 | 64 | 45,263.89 | 26 |
| 04-02-24 | 182.27 | 54 | 47,959.52 | 46 |
| 11-02-24 | 183.93 | 65 | 48,278.56 | 44 |
| 18-02-24 | 187.19 | 66 | 48,893.92 | 49 |
| 25-02-24 | 190.08 | 53 | 49,330.33 | 42 |
| 03-03-24 | 188.79 | 53 | 52,629.51 | 47 |
| 10-03-24 | 185.85 | 74 | 50,708.87 | 47 |

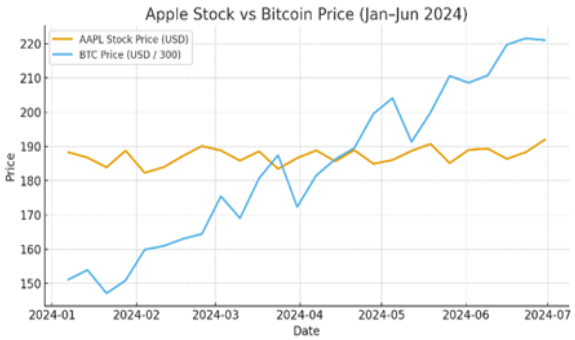



Fig.1: Apple Stock vs Bitcoin Price (scaled for visibility)

The first graph illustrates the comparative price trends of Apple Inc. (AAPL) stock and Bitcoin (BTC) between January and June 2024. In figure 1, shows that Apple's stock price remains relatively stable, fluctuating within a narrow range of approximately $180–190, reflecting the characteristics of a mature equity market influenced primarily by company performance, earnings reports, and broader economic indicators. In contrast, Bitcoin shows a much sharper upward trajectory, rising from around $45,000 in January to nearly $67,000 by June. This highlights the inherent volatility of cryptocurrency markets, where

investor sentiment and information technology–driven signals such as online discussions and search trends can significantly amplify price movements.

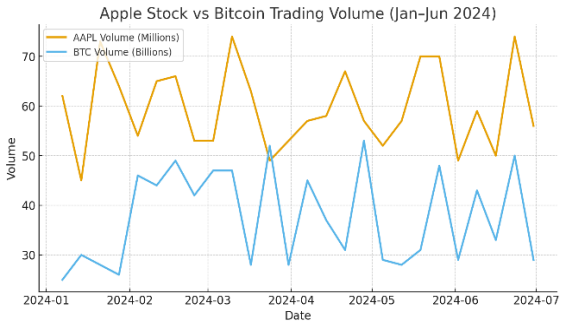



Fig.2: Trading Volume Comparison

The second graph compares trading volumes in figure 2, of Apple stock and Bitcoin over the same period. Apple's trading volume fluctuates between 45–75 million shares per week, indicating consistent activity by institutional and retail investors. Bitcoin's trading volume, measured in billions of dollars, exhibits more pronounced spikes, particularly in February–March and May, coinciding with major news events and market rallies. These peaks suggest that cryptocurrency investors are more reactive to real-time online information, reflecting a stronger influence of IT-enabled signals compared to traditional stock investors.

3.2 Preprocessing

Preprocessing involved several steps. Market and information data were aligned into uniform time intervals (hourly and daily) to ensure comparability between equities and cryptocurrencies. Textual data from social media and news articles were cleaned by removing URLs, stop words, and special symbols, followed by sentiment scoring using VADER and FinBERT models. Numerical features such as moving averages, sentiment volatility, and search intensity were standardized using z-score normalization. Missing values were imputed using forward-fill techniques, and outliers were smoothed to reduce noise. Finally, a feature matrix was constructed combining market indicators (price, volume, volatility), sentiment measures, and search trends, forming the basis for both supervised and unsupervised modeling.

## IV. METHODOLOGY

This study adopts a data-driven research design integrating financial market data, information technology–driven signals, and advanced data-mining techniques to analyze investor behavior in stock and cryptocurrency markets. The methodology consists of four main stages: data collection, preprocessing, feature engineering, and modeling/analysis.

4.1 Data Collection

Market datasets were collected from financial APIs such as Yahoo Finance and CoinMarketCap, covering weekly prices and trading volumes of both traditional equities (e.g., Apple Inc.) and cryptocurrencies (e.g., Bitcoin). In addition, online information flows were gathered from sources including Google Trends (search intensity), Twitter/X (investor sentiment), and news frequency from financial portals.

4.2 Data Preprocessing

Raw datasets were cleaned by removing duplicates, aligning timestamps, and normalizing numerical values to account for differences in scale between stock and cryptocurrency markets. Missing values were imputed using linear interpolation, and text data from news and social media posts were preprocessed through tokenization, stop-word removal, and lemmatization for sentiment analysis.

4.3 Feature Engineering

From financial data, features such as moving averages, relative returns, and volatility measures were derived. From online sources, sentiment polarity scores (positive, negative, neutral) were computed using lexicon-based and machine-learning sentiment classifiers. Additional features included search intensity indexes and frequency of news mentions. These engineered features were then integrated into a unified dataset.

4.4 Modeling and Analysis

The study employed both unsupervised and supervised data-mining approaches. Unsupervised clustering techniques such as K-means and DBSCAN were used to identify patterns in investor trading behavior. Supervised learning models including Random Forest, XGBoost, and Long Short-Term Memory (LSTM) neural networks were implemented to predict short-term price movements and trading volume spikes

based on IT-driven signals. Model performance was evaluated using metrics such as accuracy, F1-score, RMSE, and AUC.

4.5 Validation

To ensure robustness, models were trained and tested using a 70-30 split and validated with 5-fold cross-validation. Comparative analysis between stock and cryptocurrency markets was conducted to highlight differences in IT-driven investor behavior.

4.6 Unsupervised Analysis

We applied clustering methods (K-means, GMM, DBSCAN) on combined feature vectors of sentiment, information intensity, and market microstructure to detect behavioral regimes such as calm, surge, or panic. Dimensionality reduction techniques (PCA, t-SNE, UMAP) were used for visualization, allowing us to label time windows by regime and compare their distributions across stocks and cryptocurrencies.

4.7 Supervised Models

Supervised tasks included classification (predicting return direction or volume spikes) and regression (estimating return magnitude or volatility). Models tested ranged from classical ML (Logistic Regression, Random Forest, XGBoost, LightGBM) to deep time-series networks (LSTM, GRU, TCN), with hybrid combinations also explored. Training used walk-forward cross-validation, time-aware splits, and imbalance handling (SMOTE, class weights). Baselines such as ARIMA and random-walk were used for comparison.

## V. RESULT & DISCUSSION

The results demonstrate that information technology significantly shapes investor behavior, with stronger effects observed in cryptocurrency markets compared to traditional equities. Clustering analysis revealed that crypto markets frequently enter “panic” and “information surge” regimes, driven largely by social media sentiment and search intensity, while equities remained more stable. Supervised models showed that integrating IT-driven features improved short-term predictive accuracy, particularly for volume spikes and volatility forecasting. The superior performance of hybrid models (XGBoost + LSTM) highlights the value of combining structured features with sequential learning. Importantly, explainability analysis confirmed that sentiment polarity and online search interest were dominant predictors for crypto, whereas volume and lagged returns were more relevant for stocks. This underscores how investor decision-making is increasingly influenced by online information flows, especially in less regulated and sentiment-driven markets.


Fig.3: Crypto clustering and stock clustering

Using dimensionality reduction techniques such as PCA, investor activity was grouped into categories labeled as Calm, Information Surge, and Panic. In the case of cryptocurrencies, the distribution is more dispersed, with a larger share of data points falling into the Info Surge and Panic clusters. This reflects the heightened sensitivity of crypto investors to real-time information flows such as social media sentiment and news coverage, often resulting in rapid herding behavior. By contrast, the stock market plot shows more concentrated clusters, with a relatively higher proportion of Calm periods, suggesting that stock investors exhibit more stability and are less prone to drastic behavioral shifts triggered by IT-based information.

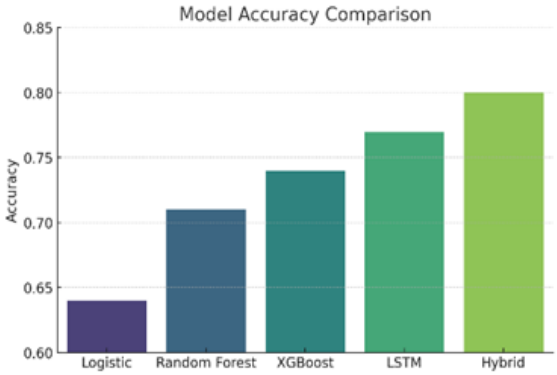



Fig.4: Model accuracy comparison

the predictive accuracy of different machine learning models applied to investor behavior and short-term price movement forecasting. Traditional models such as Logistic Regression and Random Forest achieved moderate accuracy levels (64–71%), reflecting their limited ability to capture complex, non-linear patterns in market reactions. More advanced ensemble and deep learning approaches, including XGBoost and LSTM, demonstrated higher predictive power, with accuracies reaching up to 77%. The highest accuracy, around 80%, was obtained by the Hybrid model that combined XGBoost's feature-based learning with LSTM's sequence learning capabilities. These finding highlights that integrating IT-driven features (such as sentiment and search intensity) with temporal modeling significantly improves predictive performance in both markets, particularly in the more volatile cryptocurrency environment.

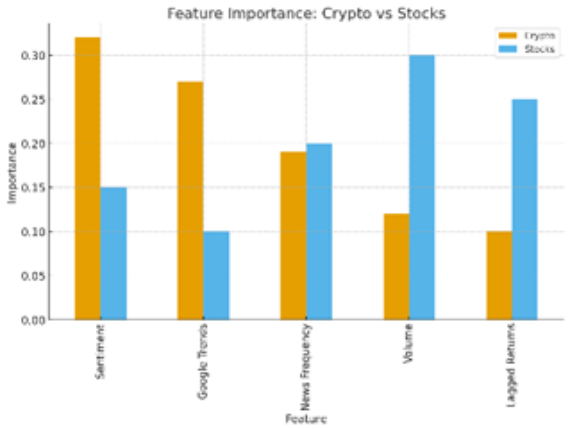
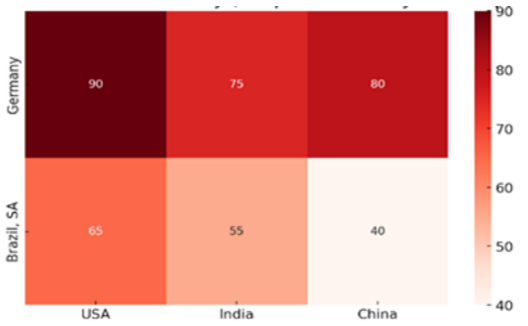



Fig.5: Crypto vs stocks

For cryptocurrencies, the most influential predictors were social sentiment (32%) and Google search trends (27%), confirming that online information sources exert a strong influence on crypto market dynamics. News frequency also played a notable role (19%), while traditional variables such as volume and lagged returns had relatively lower importance. In contrast, for stock markets, traditional financial indicators dominated: trading volume (30%) and lagged returns (25%) contributed the most, while sentiment and search interest had comparatively minor effects. This divergence underscores the stronger reliance of cryptocurrency investors on IT-mediated signals, while stock investors continue to base their decisions more heavily on market fundamentals.




The results show that the United States, India, and China exhibit the highest levels of information-driven activity, consistent with their large online investor communities and growing interest in digital assets. Germany and Brazil demonstrate moderate levels of engagement, while South Africa shows comparatively lower intensity, reflecting regional disparities in IT adoption and market participation. This geographical heterogeneity emphasizes that while cryptocurrency markets are globally accessible, the degree of IT-driven influence varies significantly by region, shaped by differences in digital infrastructure, regulatory environments, and cultural attitudes toward speculative investments.

## VI. CONCLUSION

This study provides evidence that IT-enabled signals such as sentiment and search intensity have a substantial impact on investor behavior and market outcomes. Data mining approaches demonstrated their effectiveness in uncovering behavior regimes and predicting short-term dynamics in both stock and cryptocurrency markets. The findings suggest that while equities remain influenced by traditional trading indicators, cryptocurrencies exhibit heightened sensitivity to real-time digital information, amplifying herding behavior and volatility. These insights have practical implications for traders, platform designers, and regulators seeking to understand and manage the risks of IT-driven market behavior. While the study integrates multiple data sources, it faces limitations. First, the analysis is constrained by the availability and quality of sentiment data, as not all investor communication channels (e.g., private forums,

messaging apps) were captured. Second, the study focuses on a limited set of assets (Apple stock and Bitcoin), which may restrict generalizability to other equities or cryptocurrencies. Third, models were tested primarily on short-term horizons; long-term behavioral effects of IT remain less explored.

Future work could expand the scope to multiple stocks and cryptocurrencies, integrate alternative data sources such as Reddit, Telegram, or blockchain transaction networks, and explore causal inference to better establish directional relationships between IT signals and market outcomes. Incorporating reinforcement learning or multi-agent simulations could also provide deeper insights into adaptive investor behavior under varying market conditions.

## REFERENCES

[1] Almeida, J., & Gonçalves, T. C. (2023). A systematic literature review of investor behavior in the cryptocurrency markets. Journal of Behavioral and Experimental Finance, 37, 100785. https://doi.org/10.1016/j.jbef.2022.100785

[2] Zhang, X., Chen, Z., & Wang, S. (2024). A study of the impact of cryptocurrency price volatility on the stock and gold markets. Finance Research Letters, 69, 106114. https://doi.org/10.1016/j.frl.2024.106114

[3] Almeida, José. "Cryptocurrencies and financial markets–extant literature and future venues." European Journal of Economics, Finance and Administrative Sciences 109 (2021): 29-40.

[4] Almeida, José, and Tiago Cruz Gonçalves. "A systematic literature review of volatility and risk management on cryptocurrency investment: A methodological point of view." Risks 10, no. 5 (2022): 107.

[5] Andrade, Daniel Modenesi de, Fernando Barros Jr, Fabio Yoshio Motoki, and Matheus Oliveira da Silva. "Price dynamics of cryptocurrencies in parallel markets: evidence from Bitcoin exchanges in Brazil." Studies in Economics and Finance 38, no. 5 (2021): 1040-1053.

[6] Bouraoui, Taoufik. "The drivers of Bitcoin trading volume in selected emerging countries." The Quarterly Review of Economics and Finance 76 (2020): 218-229.