



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85302>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Black Swan: Stock Crash Predictor

Mrs. Sreedevi S¹, Sanmitra B S², Shreyas G², Srujan M V⁴, Syed Muzain Ahmed⁵

¹Assistant Professor, ^{2,3,4,5}Seventh Semester Students, Department of Computer Science and Engineering, Jawaharlal Nehru New College of Engineering, Shivamogga, Visvesvaraya Technological University, Belagavi, Karnataka, India

Abstract: *Global financial markets have grown so interconnected and fast-moving that spotting a high-impact event before it hurts investors' portfolios has become a serious challenge. This project introduces "Black Swan: Stock Crash Predictor," an AI-driven system built to catch unpredictable global events early and gauge how they are likely to ripple through financial markets, using a combination of Generative AI and Agentic AI. Rather than waiting for a crisis to unfold, the system keeps a constant watch over news outlets, RSS feeds, financial reports, and live market feeds, flagging emerging risks such as geopolitical conflict, pandemics, economic downturns, and supply-chain breakdowns as they surface. Generative AI is used to turn this raw information into readable summaries, forward-looking narratives, and impact assessments, so investors can make sense of complicated situations quickly. Agentic AI adds a further layer of autonomy, allowing the system to reason independently, make decisions on its own, and keep adjusting to shifting market conditions. It also runs automated stress-testing simulations and risk assessments to gauge how markets and investments might behave under uncertainty. Where conventional monitoring tools tend to fall behind fast-changing events, this approach is built to keep pace, offering proactive alerts that give investors an early warning. Overall, the framework is aimed at sharpening key decision-making factors: response time, prediction accuracy, market awareness, and investment strategy planning.*

Keywords: *Black Swan Events, Stock Market Prediction, Generative AI, Agentic AI, Natural Language Processing, Risk Detection, Financial Market Analysis, Real-Time Alerts.*

I. INTRODUCTION

As global financial markets grow more complex and more retail investors enter the fray, the need for reliable risk-monitoring tools has never been greater. Most conventional investing strategies still lean heavily on historical data and manual analysis, an approach that struggles to keep up with volatile conditions and sudden global shocks. Consequently, many investors find themselves unable to spot, let alone respond to, unexpected events that send stock prices swinging.

Advances in artificial intelligence and real-time data processing are starting to close this gap. By continuously scanning large volumes of global news and other information streams, intelligent systems can pick up on potential risk events and estimate their likely effect on the markets, enabling decisions that are both faster and more accurate than what manual methods can achieve.

The Autonomous Black Swan Risk Monitor proposed here builds on these techniques to automatically flag high-impact global events, work out how they might affect specific companies and markets, and issue timely alerts. By converting a flood of raw, unstructured global information into insights that are actually usable, the system is designed to support proactive decision-making and strengthen risk management for investors navigating an increasingly unpredictable financial landscape.

A. Scope of the Project

Online trading platforms and instant access to market data have made investing more accessible than at any point in the past, drawing in a wave of retail participants eager to build wealth through the stock market. That accessibility comes at a cost, though — investors are now more exposed than ever to "Black Swan" events: rare, largely unforeseeable occurrences such as geopolitical conflict, pandemics, and supply-chain disruption that can send markets into sudden turmoil. Because these events defy the usual forecasting methods and tend to strike without much warning, retail investors — who typically lack access to sophisticated tools and often receive information later than institutional players — are left particularly exposed.

B. Motivation

The surge of retail participation in global financial markets has opened up new paths to wealth creation, but it has also left individual investors exposed to risks that are often difficult to foresee. Today's markets respond quickly to real-world developments — geopolitical friction, economic shocks, pandemics, and disruptions to global supply chains all leave their mark.

Yet most retail investors still depend on conventional strategies, backward-looking data analysis, or news that reaches them well after the fact. These methods simply aren't built to catch real-time risks emerging from a rapidly shifting global picture. Tracking world events by hand and estimating their financial fallout is slow and inefficient, and by the time an investor reacts, the moment has often already passed.

II. LITERATURE SURVEY

The growing complexity of global financial markets, paired with an explosion of real-time financial and textual data, has pushed researchers toward machine learning, natural language processing, and other intelligent techniques for stock prediction and anomaly detection. Much of the existing work centers on forecasting stock-price movement using historical data, financial news, sentiment signals, and technical indicators. This survey walks through prior work on anomaly detection, intelligent agents, financial news analysis, and machine-learning-based prediction, with the goal of surfacing the shortcomings of current approaches and building the case for a system that proactively detects high-impact events and gauges their effect on financial markets.

In [1], Zhang describes a network-wide anomaly detection and diagnosis scheme built around the perfSONAR framework for monitoring high-performance computing networks. The method combines Principal Component Analysis (PCA), Adaptive Plateau Detection (APD), and Q-statistics to pick out correlated and uncorrelated anomalies from multi-domain time-series measurements. Working from one-way delay measurements gathered across distributed perfSONAR archives, the system applies both temporal and spatial filtering to surface unusual patterns and trace possible root causes. Tests on synthetic and real-world datasets show strong detection accuracy, fewer false positives, and reliable anomaly localization, though the PCA processing carries a heavy computational cost and the approach leans heavily on how good and how frequent the measurement data is. While the study is built for network environments rather than financial risk detection, it makes a solid case for automated, continuous anomaly monitoring.

In [2], Siu et al. put forward a command-authentication framework based on a Multiagent System (MAS) aimed at detecting and mitigating False Data Injection Attacks (FDIAs) in economic dispatch systems. Individual generating units are each given an intelligent agent that independently works out optimal dispatch solutions and checks them against centralized control commands. Economic and stability impact factors, together with a truth-table validation mechanism, drive the detection and mitigation process. The technique was tested on 5-bus, 3-generating-unit, and 15-generating-unit systems through MATLAB simulations and Hardware-in-the-Loop experiments, and the results point to effective detection of stealth attacks while keeping the system stable and limiting financial losses. That said, authentication latency grows with system size, agents need secure channels to communicate, and scaling the approach up is far from trivial. The decentralized, agent-based design offers a useful reference point for building autonomous monitoring systems.

In [3], Kumar et al. build a hybrid stock-prediction framework around emotion analysis drawn from financial news and social media text, pulling data from sources such as Economic Times, Livemint, BSE India, NSE, Moneycontrol, Screener.in, and Twitter. After preprocessing, the text is converted into numerical form using trigram-based transformation and BERT-based feature extraction, and Support Vector Machine (SVM) regression is used to derive sentiment polarity scores. These scores are then paired with historical stock data — opening price, closing price, trading volume, and prior trends — and fed through Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, yielding a reported mean prediction error of roughly 13%. The approach does a good job of blending textual and historical signals, but its accuracy tends to drop during highly volatile stretches, and it needs a substantial amount of training data.

In [4], Ranjbaran et al. look at how the polarity of financial news affects stock-market prediction, drawing on more than 20,000 news articles collected from 240 websites. Relevant articles are first isolated through keyword filtering, then preprocessed and scored for sentiment using the TextBlob library; the resulting values are combined with stock-price changes recorded around each article's publication time. Several regression models — Support Vector Regression (SVR), Bayesian Ridge, LASSO, Decision Tree, and Random Forest — are then compared using RMSE and MAE. SVM-based models come out ahead of the alternatives, and sentiment-driven analysis proves more useful than random tagging, though performance still hinges on the quality of the underlying news and simple sentiment scoring can miss more nuanced emotional or market cues.

In [5], Sakphoowadon et al. introduce a Probabilistic Lexicon-Based Stock Market Prediction (PLSP) model that forecasts market direction from Thai financial news and closing prices. Their approach builds a probabilistic lexicon, ThaiFinLex, by pulling meaningful event terms out of financial news and linking them to the probability of specific price movements, following a pipeline of text preprocessing, event-term generation, price processing, news labeling, and lexicon construction.

Rate of Change (ROC) is used to classify movement as upward or downward, and testing against SVM, J48, and BayesNet reports accuracy as high as 96.64% when the event-term thresholds are well tuned. The model is, however, tightly bound to the Thai language, needs labeled news data, and involves a fairly involved event-term generation process.

In [6], Chen and Shih combine Chinese financial news, technical indicators, and Genetic Algorithms to predict stock trends. Their pipeline runs text preprocessing, TF-IDF vectorization, Chi-Square feature selection, and SVM classification, while technical indicators such as the Relative Strength Index (RSI) help sharpen the predictions further. A Genetic Algorithm-based Stock Trend Prediction (GASTP) routine then tunes parameters including document frequency, Chi-Square score percentage, and RSI thresholds. Tested against Taiwanese market news, the method improves both prediction accuracy and trading returns relative to more traditional strategies, though it demands large volumes of labeled news and adds computational overhead through the genetic optimization step.

In [7], Kaya and Karsligil build a stock-prediction system around financial news and data mining, drawing on a year's worth of Microsoft-related news and price data. The pipeline covers data collection, delta-price calculation, preprocessing, feature selection, classification, and evaluation. Rather than using single words as features, the study introduces "word couples" — noun-verb pairs appearing in the same sentence — and applies Chi-Square statistics to narrow down the most useful features before training SVM classifiers. Using 10-fold cross-validation, the word-couple approach reaches about 61% accuracy with an RBF-kernel SVM. It's an improvement on single-word features, but the accuracy ceiling remains fairly low, and the method takes no account of external political or global events that could be moving the market.

In [8], Sharma et al. put together an integrated machine-learning framework that pairs historical stock prices with sentiment drawn from market news. Historical data comes from platforms like Yahoo Finance and Alpha Vantage, while news is sourced from Bloomberg and Reuters, and the raw text goes through cleaning, normalization, tokenization, lemmatization, and stop-word removal. Technical indicators — Moving Average (MA), Exponential Moving Average (EMA), and RSI — are computed, and VADER along with BERT-based sentiment analysis convert the news into numerical signals. LSTM handles the sequential stock-price side, while BERT processes the contextual news information. Measured against MAE, RMSE, and R-squared, the framework outperforms traditional linear models, ARIMA, and standalone LSTM, especially in volatile conditions, though the hybrid setup is computationally demanding and needs large amounts of both historical and textual data.

In [9], Ishica et al. study how sentiment analysis and machine learning can be used to forecast stock-price changes from financial news, processing daily headlines through text cleaning, tokenization, CountVectorizer, and TF-IDF-based feature extraction. The resulting sentiment scores feed into a Random Forest Classifier that predicts whether prices will rise or fall, with performance measured via accuracy, precision, recall, F1-score, and a confusion matrix. The paper reports strong results overall, though the accuracy figure shifts somewhat between the abstract (85.97%) and the detailed results and conclusion (84.39%). The authors also flag the need to bring in historical prices, macroeconomic indicators, real-time data, and more advanced NLP methods such as BERT to sharpen contextual understanding and forecasting.

In [10], Wijaya et al. examine the short-term effect of financial-news sentiment on Indonesian stock prices using transformer-based deep learning, drawing on roughly 2,000 articles from IDXChannel alongside high-frequency market data. Their feature set blends textual sentiment with contextual market signals — momentum, index movement, trading-volume anomalies, and volatility — and uses FinBERT and IndoELECTRA as the core transformer models, with a LightGBM-based stacking ensemble combining their outputs to predict whether prices will move up or down within an hour of a news release. The results land close to random chance (MCC near zero, Brier scores around 0.25), underscoring how hard it is to extract reliable short-term signals from public news once market noise, neutral financial language, external factors, and the market's own speed at absorbing information are taken into account.

In [11], Alstad and Davulcu build a directional stock-prediction framework that combines online financial news with breaking-news signals inferred from Twitter activity, drawing on more than 50,000 NASDAQ articles gathered over five years for 30 Dow Jones stocks alongside more than 750,000 tweets collected over six months. Hourly spikes in tweet volume are flagged using the three-sigma rule, and news tied to those spikes is analyzed with text-mining techniques, extracting 1-gram and 2-gram features, applying Chi-Square selection, and testing sparse logistic regression and SVM classifiers through 10-fold cross-validation. The authors find that news tied to breaking Twitter activity meaningfully improves hourly directional prediction compared with using all available news indiscriminately, though document-level sentiment on its own adds little once other textual keyword features are already in the model.

A. Summary of Literature Survey

Taken together, the literature traces a clear evolution — from traditional machine learning and text-based methods toward sentiment analysis, deep learning, and hybrid intelligent models. Early work established that techniques such as Support Vector Machines (SVM), Naive Bayes, Principal Component Analysis (PCA), and various feature-extraction methods could support both anomaly detection and financial prediction. Later studies folded in financial news, social media signals, and technical indicators to sharpen stock-movement forecasts, and hybrid architectures pairing SVM, LSTM, GRU, and BERT have since improved the ability to capture temporal patterns and contextual meaning in financial data. Even so, most of this work still centers on predicting price movement rather than on detecting the real-world events that trigger sudden market disruption in the first place.

More recent research points to the value of real-time information, sentiment analysis, event-based features, and intelligent decision-making for improving both prediction and risk assessment. It's clear from these studies that financial news carries real signal about market behavior, and that multi-agent architectures show promise for autonomous monitoring and response. Still, notable gaps remain — a reliance on historical data, only limited continuous monitoring of global news and RSS feeds, sensitivity to volatility, and essentially no proactive detection of Black Swan events. Autonomous reasoning and early-warning capability are likewise underdeveloped in existing systems. These gaps are exactly what motivate the proposed Black Swan Risk Monitor, which brings together real-time data collection, Natural Language Processing (NLP), Agentic AI, event detection, impact analysis, stress-testing simulations, and real-time alerts into a single framework for proactive financial risk assessment.

B. Table of Summary

Paper No.	Title	Authors	Techniques Used	Focus Area	Remarks
[1]	Network-Wide Anomaly Event Detection and Diagnosis With perfSONAR	Zhang (2020)	PCA, APD, Q-statistics, temporal and spatial filtering	Network anomaly detection and root-cause diagnosis	Accurate real-time detection; computationally intensive
[2]	Command Authentication Using Multiagent System for Attacks on the Economic Dispatch Problem	Siu et al. (2022)	Multiagent System, anomaly detection, impact factors	FDIA detection and mitigation	Agents improve reliability; complex to scale
[3]	Emotion Analysis of News and Social Media Text for Stock Price Prediction using SVM-LSTM-GRU Composite Model	Kumar et al. (2022)	SVM, LSTM, GRU, BERT, sentiment analysis	Stock prediction using news, social media, and historical data	Combines sentiment and sequence data; weaker in volatile markets
[4]	Analyzing Effect of News Polarity on Stock Market Prediction: A Machine Learning Approach	Ranjbaran et al. (2021)	TextBlob, SVR, Bayesian Ridge, LASSO, Decision Tree, Random Forest	Financial news sentiment and stock prediction	SVM-based models perform well; dependent on news quality

[5]	Probabilistic Lexicon-Based Approach for Stock Market Prediction: A Case Study of The Stock Exchange of Thailand	Sakphoowadon et al. (2018)	ThaiFinLex, text preprocessing, ROC, probabilistic event terms	Event-term-based stock direction prediction	High accuracy reported; language-dependent, needs labeled data
[6]	A Stock Trend Prediction Approach based on Chinese News and Technical Indicator Using Genetic Algorithms	Chen and Shih (2019)	TF-IDF, Chi-Square, SVM, RSI, Genetic Algorithm	Stock trend prediction using news and technical indicators	Combines textual and technical features; optimization adds complexity
[7]	Stock Price Prediction Using Financial News Articles	Kaya and Karsligil (2019)	Text mining, word couples, Chi-Square, SVM	Stock movement prediction from financial news	Word-couple features help; global events not considered
[8]	Optimizing Stock Market Prediction through Integration of Market News and Stock Prices with Machine Learning Models	Sharma et al. (2024)	LSTM, BERT, VADER, MA, EMA, RSI	Integrated stock-price and news-based prediction	Better in volatile markets; computationally expensive
[9]	Utilizing Sentiment Analysis and Machine Learning to Forecast Stock Price Changes from Financial News	Ishica et al. (2024)	NLP, CountVectorizer, TF-IDF, sentiment analysis, Random Forest, BERT-based extensions	Prediction of stock-price direction using news sentiment	Strong performance; needs historical prices, macro data, richer context
[10]	Analyzing the Short Term Impact of News Sentiment on Indonesian Stock Prices using Deep Learning Models	Wijaya et al. (2025)	FinBERT, IndoELECTRA, LightGBM stacking, sentiment analysis, contextual market features	One-hour stock direction prediction using Indonesian news	Performance near random; short-term signal weak, noisy
[11]	Directional Prediction of Stock Prices using Breaking News on Twitter	Alostad and Davulcu (2015)	Twitter-volume spikes, 1-gram/2-gram features, Chi-Square, logistic regression, SVM, sentiment analysis	Hourly stock-direction prediction using breaking news and Twitter	Breaking-news periods help; sentiment alone adds little

III. OBJECTIVES

- To design and develop an intelligent system for detecting Black Swan events in financial markets.
- To continuously monitor global news, RSS feeds, and real-time data sources.
- To identify affected companies and analyze their potential market impact.
- To integrate Generative AI for summarization and predictive impact analysis.
- To automate risk detection and reduce manual monitoring efforts.
- To provide real-time alerts and a user-friendly dashboard for investors.
- To improve the accuracy, speed, and reliability of risk assessment.
- To perform stress-testing simulations for better decision-making.

IV. CONCLUSIONS

Black Swan: Stock Crash Predictor offers a proactive, intelligent way to catch unexpected events before they take a serious toll on financial markets. By keeping a constant eye on global news, RSS feeds, and live data sources, the system is designed to pick up on potential Black Swan events early and estimate how they might play out across the markets. Bringing Artificial Intelligence, Generative AI, and Agentic AI together allows raw, unstructured information to be turned into insight that investors can actually act on, cutting down the need for manual monitoring and sharpening overall risk awareness.

Beyond detection, the system helps identify which companies are likely to be affected, produces predictive impact analysis, sends out real-time alerts, and runs stress-testing simulations to support better decision-making in uncertain conditions. By pairing automated event detection with intelligent financial analysis, the framework moves past the limitations of conventional tools that lean mainly on historical data and basic indicators. The project is deliberately scoped around risk awareness and decision support rather than direct trading or portfolio management, positioning it as a scalable foundation for proactive financial risk monitoring and early-warning systems.

V. FUTURE SCOPE

There's considerable room to grow this work further. One avenue is strengthening the system's learning ability so it can continuously absorb newly emerging events, market reactions, and updated financial data, helping it adapt as conditions change and sharpening both event detection and impact analysis over time. The system could also be broadened to pull from a wider set of global news sources, financial platforms, and real-time market feeds for more thorough risk coverage.

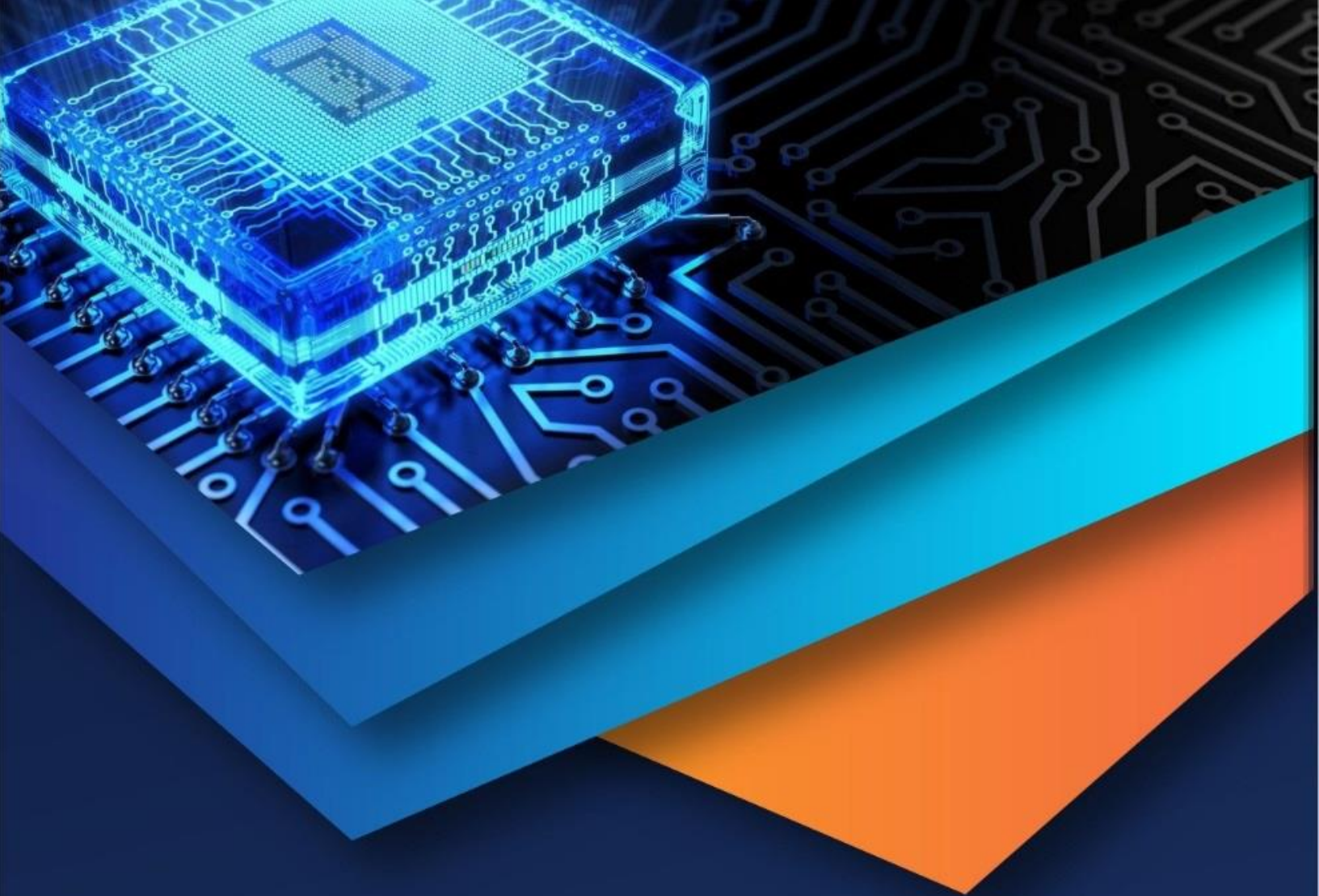
Another promising direction is deepening the Agentic AI side for more autonomous analysis and decision support — better identification of affected companies and sectors, more detailed predictive impact assessments, and more advanced stress-testing across different market scenarios. The real-time alert system and dashboard could also become more personalized and actionable for individual investors. Looking further ahead, the framework could be scaled up and integrated with broader financial applications while staying true to its core purpose: risk awareness and decision support, rather than trading itself. Developments along these lines could turn this into a genuinely robust, reliable early-warning system for unexpected disruptions in financial markets.

REFERENCES

- [1] A. Sharma, B. Devender, S. Kumar, V. S. Nishok, C. Shyamsunder, and K. S. Sidhu, "Optimizing Stock Market Prediction through Integration of Market News and Stock Prices with Machine Learning Models," 2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2024.
- [2] H. Alostad and H. Davulcu, "Directional Prediction of Stock Prices using Breaking News on Twitter," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2015.
- [3] M. Kaya and M. E. Karsligil, "Stock Price Prediction Using Financial News Articles," IEEE International Conference on Information and Financial Engineering (ICIFE), 2010.
- [4] J. Bollen, H. Mao, and X. Zeng, "Twitter Mood Predicts the Stock Market," Journal of Computational Science, vol. 2, no. 1, pp. 1-8, 2011.
- [5] A. K. Nassiroussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text Mining of News Headlines for Forex Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics and Sentiment," Expert Systems with Applications, vol. 42, no. 1, pp. 306-324, 2015.
- [6] Y. Mao, W. Wei, B. Wang, and B. Liu, "Correlating S&P 500 Stocks with Twitter Data," Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, 2012.
- [7] Y. Mao, W. Wei, and B. Wang, "Twitter Volume Spikes: Analysis and Application in Stock Trading," Proceedings of the 7th Workshop on Social Network Mining and Analysis, 2013.
- [8] H. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of TimeSeries," IEEE Access, vol. 7, pp. 28299-28308, 2019.



- [9] Ishica, Sneha, S. K. Jha, Suraj, Ujala, and R. Chaudhary, "Utilizing Sentiment Analysis and Machine Learning to Forecast Stock Price Changes from Financial News," 2024 First International Conference on Data, Computation and Communication (ICDCC), IEEE, 2024, doi: 10.1109/ICDCC62744.2024.10961794.
- [10] A. A. Wijaya, G. M. Sutarman, B. A. Jabar, and A. Wijaya, "Analyzing the Short Term Impact of News Sentiment on Indonesian Stock Prices using Deep Learning Models," 2025 4th International Conference on Digital Transformation and Applications (ICDXA), IEEE, 2025, doi: 10.1109/ICDXA69105.2025.11329903.
- [11] H. Alstad and H. Davulcu, "Directional Prediction of Stock Prices using Breaking News on Twitter," 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2015, doi: 10.1109/WI-IAT.2015.82.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)