



Machine Learning-Based Stock Market Behaviour Analysis Using K-Means Clustering

Somnath Hase¹ & Vikas Humbe²

Dept. of Computer Science, Smt. S. K. Gandhi College, Kada, India

School of Technology, SRTM University, Sub Center Latur, India

Corresponding Author – Somnath Hase

DOI - 10.5281/zenodo.19396653

Abstract:

Stock market behaviour involves complex interactions among various indicators, like market trends and diverse data. Traditional analytical tools often struggle to capture this complexity, especially with long-term relationships and multiple data sources. Machine learning offers a formal way to analyse these complexities by revealing hidden structures and key indicators that drive market dynamics. Systematically integrating different data sources can boost predictive reliability and support smart investments.

A machine learning model, using clustering and deep learning, examined stock market behaviour. K-Means clustering grouped similar market patterns, structuring the dataset into meaningful clusters. To consider long-term dependencies in financial information, a Long Short-Term Memory (LSTM) network was used. An attention feature was added to the LSTM framework to select signals that influence movements over time.

This hybrid LSTM–Attention framework enabled effective integration of various data sources and focused on relevant features affecting stock market movement. This hybrid approach identifies significant market patterns in a structured way, enhancing predictive accuracy and using multiple data streams to support data-driven investment and portfolio creation.

Keywords: *Machine Learning, Stock Market Analysis, K-Means Clustering, Long Short-Term Memory (LSTM), Attention Mechanism, Hybrid LSTM–Attention Model, Financial Data Prediction*

Introduction:

Stock trading used to be done on paper before computers changed everything. Back then, both buyers and sellers needed to be there to make a trade. Stock certificates were physical documents [1]. Algo trading is the automatic execution of high-frequency trades through mathematical models and established methods. It can quickly evaluate market conditions, find trading opportunities, and execute orders faster than a human can. This efficiency makes it very attractive [2]. Data and metrics are produced in massive volumes each day by financial markets that reflect the round-the-clock interactivity of economic variables, investor sentiment and market dynamics. So it has become an increasingly complex task to understand stock market behaviour. The nonlinear relationships and hidden structures are common in a financial dataset which is commonly missed by the traditional statistical and rule-based approaches [3]. With markets becoming increasingly complex, and with more and varied data sources available, there is need of new analytical techniques to make sense of patterns and enable robust decisions.

Machine learning has become an effective technique for financial data, since it can analyze larger datasets and extract hidden patterns that are not obvious through traditional methods. Machine learning models learn from historical data, allowing them to identify relationships between variables and trends which can help predict future market behaviour [4]. These offer great discovery potential that can be really

useful in stock market, where insights are relied upon to take decisions as the outperformance of timely and insightful decision take more importance.

K-Means clustering is one of the clusters analysis in machine learning that are essential to classify the structural pattern properties [5]. K-Means is an unsupervised learning algorithm that helps in clustering data points based on their similarity and enables analysts to segment financial data into meaningful clusters [6]. In stock market analysis, clustering can be used to uncover hidden market states, identify behavioural patterns, and reduce the dimensionality of large datasets for further predictive modelling. Clustering-based deep learning methods have been recognized as powerful model strategies in recent studies due to their ability to increase prediction performance. Long Short-Term Memory(LSTM) [7] networks is widely adopted to model sequential data by capturing the temporal dependencies, thus making them very apt for stock price movement modelling. The attention mechanisms enhance model performance by allowing the system to select most significant indicators from the dataset.

Combining these approaches creates a holistic framework for understanding patterns in stock market behaviour. Clustering is beneficial for revealing structural relationships within financial transactions; whereas, deep learning networks seize the temporal dependencies to recreate notable events that instigate a price change of an asset. This integrated approach could enhance predictive accuracy, facilitate better investment decisions, and contribute to the development of data-driven financial analytics solutions.

Methods:

K-Means clustering algorithm is used to study the stock market behaviour through unsupervised machine learning technique. It is aimed to determine the trends in the movement of stock prices over the period and cluster similar trading days in terms of price and volatility properties. The experiment involves the use of historical data of the Reliance industries ltd. which is listed in the National Stock Exchange (NSE).

Materials:

Tools and datasets used in the research are the following:

- Stock price history of Reliance industries Ltd (NSE)
- Attributes used:
 - Open price
 - High price
 - Low price
 - Close price
 - Trading volume

Software Tools:

- Python software programming
- Jupyter Notebook / Google Colab.
- Libraries:
 - Pandas
 - NumPy
 - Matplotlib

- Scikit-learn
- yfinance (to download NSE data)

Data Preprocessing:

Dataset is pre-processed to enhance the quality of the data before the clustering algorithm is used.

Steps include:

- Removing missing values
- Calculating daily returns
- Calculating volatility
- Normalization of the features with Standard scaler

Feature engineering is performed using:

Daily Return:

$$Return_d = \frac{Close - Open}{Open}$$

Volatility:

$$Volatility_d = High - Low$$

These characteristics are used to capture the movement of the market and fluctuation of prices.

K-Means Clustering Algorithm:

K-Means algorithm is an algorithm that divides dataset into K clusters that are similar.

Algorithm steps:

1. Select number of clusters K
2. Randomly sample cluster centres.
3. Each data point is allocated to the closest centroid.
4. Recalculate centroid locations.
5. Repeat until convergence

In this study, K = 3 clusters which are:

- Bullish market behaviour
- Stable market behaviour
- Bearish market behaviour

Data Analysis and Visualization:

The visualization of results is performed after clustering on the scatter plots and the results are observed in terms of relation between returns and volatility. The clusters are the representations of various market conditions.

This will assist the investors in knowing how the stocks perform and what opportunities are available in the market.

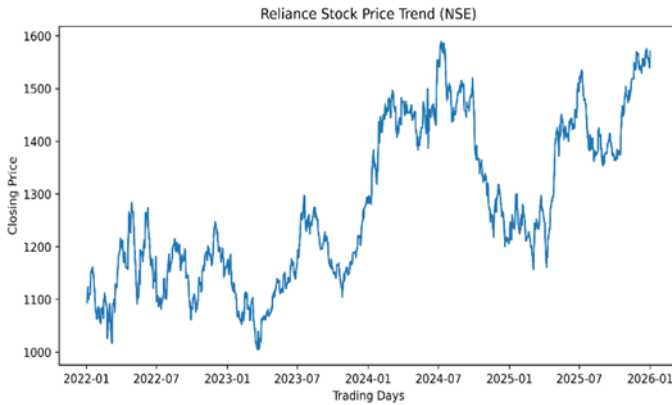


Fig. 1 Reliance Stock Price Trend

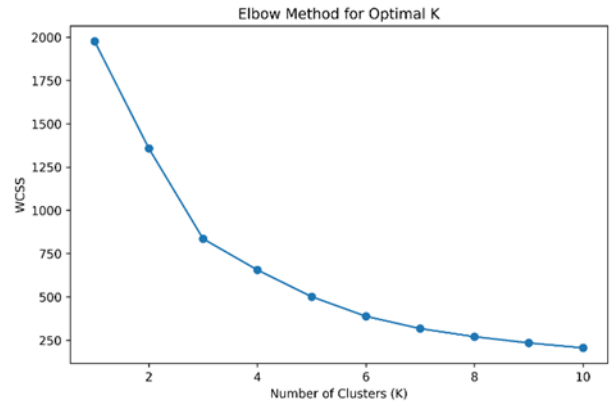


Fig. 2 Elbow Method for Optimal K

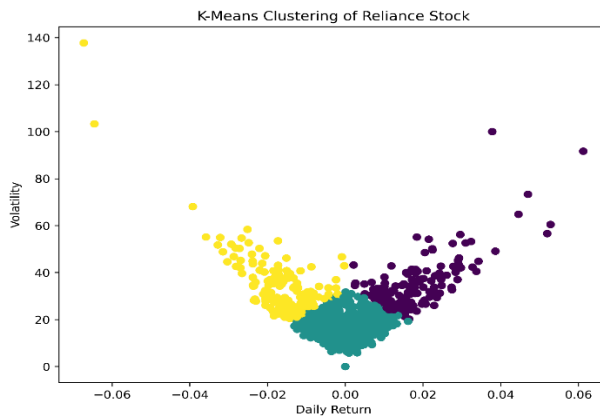


Fig. 3 K- Means Clustering of Reliance Stock

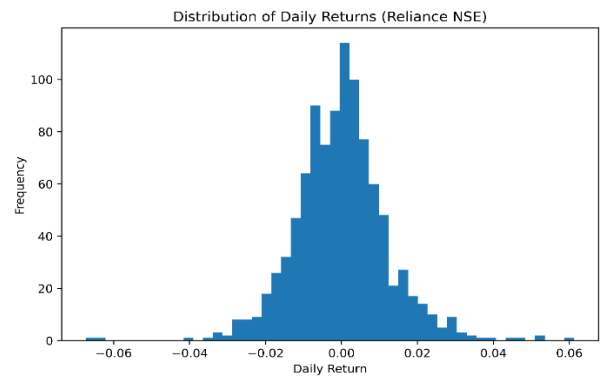


Fig. 4 Distribution of Daily returns

Conclusion:

This study explored the use of the K-Means clustering algorithm to analyze stock market behaviour using historical data of Reliance Industries listed on the National Stock Exchange (NSE). Data preprocessing and normalization were performed to improve the effectiveness of the clustering process. Several graphical analyses were used to better understand the data. The stock price trend graph showed the overall movement of the stock during the selected period. The return distribution histogram helped in observing the variation and spread of daily returns, indicating potential risk and stability levels. The elbow method was applied to determine the optimal number of clusters for the K-Means algorithm. Based on the analysis, three clusters were selected to categorize different trading patterns.

The clustering results grouped trading days with similar return and volatility characteristics. These clusters represent different market behaviours such as stable periods, high volatility conditions, and possible bullish or bearish movements. Although K-Means clustering does not directly forecast future prices, it helps identify hidden structures in financial data.

Overall, the study shows that machine learning techniques can assist in understanding stock market behaviour and support better financial analysis and decision-making.

References:

1. S. Hase and V. Humbe, "Python-Powered ETF Trading: Unleashing Algorithmic Trading Strategies," in *2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON)*, 2024, pp. 1–4. doi: 10.1109/DELCON64804.2024.10866662.

2. S. Hase and V. Humbe, “Kite Connect, Uplink and Smart API: A Comprehensive Study of Python API Libraries,” in *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies - Volume 5: ICRDICCT'25*, SciTePress, 2025, pp. 814–819. doi: 10.5220/0013944200004919.
3. A. Fawaid Ridwan and S. Supian, “IDX30 Stocks Clustering with K-Means Algorithm based on Expected Return and Value at Risk,” *International Journal of Quantitative Research and Modeling*, vol. 2, no. 4, pp. 201–208, 2021.
4. B. Aslam, “Identifying optimistic stocks with K-means clustering algorithm,” *International Review of Economics and Finance*, vol. 104, Dec. 2025, doi: 10.1016/j.iref.2025.104579.
5. A. G. Suteja, “Machine Learning-Based Customer Segmentation and Behavioral Analysis Using K-Means Clustering,” *JoCoSiR Journal of Computer Science Research Analysis Using K-Means Clustering Journal of Computer Science Research (JoCoSiR) with CC BY NC SA license*, doi: 10.65126/jocosir.v3i2.
6. T. Renugadevi, R. Ezhilarasie, M. Sujatha, and A. Umamakeswari, “Stock market prediction using hierarchical agglomerative and K-means Clustering Algorithm,” *Indian J. Sci. Technol.*, vol. 9, no. 48, Dec. 2016, doi: 10.17485/ijst/2016/v9i48/108029.
7. S. Hase and V. Humbe, “A Hybrid LSTM-Attention Framework for Multivariate Stock Forecasting with Macroeconomic Indicators,” *MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal*, vol. 36, no. 1, pp. 1831–1834, Jan. 2026, [Online]. Available: <https://mswmanagementj.com/>