Stock market forecasting using financial graph network and deep learning techniques

Report submitted in fulfilment for the requirement of

B. Tech degree in Computer Engineering

Name of the Student	Roll Number
Adarsh Kumar	2020UCO1663
Rohit Sharma	2020UCO1697
Shubham Sharma	2020UCO1705

Under the supervision

Of

Dr. Preeti Kaur

Department of Computer Science and Engineering

Netaji Subhas University of Technology (NSUT)

New Delhi, India – 110078

Index

List of Figures

1 ABSTRACT

Stock prediction has been one of the major target problems of the artificial intelligence domain. Since its beginning, people have tried to predict the share price values of stocks based on various parameters. However, predicting the exact value of any share is a near impossible task as it is governed by infinitely many variables. Historical data, government policies, budgets, natural disturbances, to name a few of them. Hence, instead of predicting the exact value of a given share, we propose to predict the direction and intensity of change in the value of stock indices of India based on historical data. We propose to build a network of patterns seen in the 3 major stock indices of our country namely – SENSEX, NIFTY50 and NIFTY Consumption. After using the centrality measures of this network, we propose to use them as input variables to various classification methods such as KNN, SVM, certain deep learning strategies such as use of neural networks as well in order to classify the patterns of our test data. We also propose to use the results of these algorithms in real stock simulators in order to test their consistency using varying strategies.

2 INTRODUCTION

Stock market prediction has become a distinct field of research. Traditionally, market studies focused on manual data analysis and pattern identification in price fluctuations to forecast future market behaviour. However, the introduction of artificial intelligence has revolutionized this approach. While applying machine learning and deep learning techniques to stock prices may seem straightforward, the manner in which they are applied can greatly influence the results.

Our research focuses on graphs and networks, particularly in the context of the stock market. Unlike traditional approaches that aim to predict precise stock market movements, our study aims to forecast the potential behaviour of the market in the near future based on historical data. This task becomes a classification problem, where the goal is to determine whether the market will go up or down. The volatility parameter, which measures the dispersion of share price values, will be a central consideration in our study and is expected to significantly influence the results.

The primary objectives of this research are twofold:

- Experiment with various ensemble techniques and deep learning strategies for classifying patterns in stock prices data.
- Implement the developed strategy by investing on trading simulators to assess its accuracy and effectiveness in the real market.

3 LITERATURE SURVEY

Global stock market investment strategies based on financial network indicators using machine learning techniques

This paper combines the varying markets around the world to form a network and use all of them to perform a time series forecasting on stock data using some simple machine learning algorithms such as regression, random forests and SVM. The paper uses the parameter of volatility for forecasting the Z-score of each stock indices and then applies two strategies to find out which one performs better with each algorithm.

Forecasting stock crash risk with machine learning

This paper experiments with various features in order to find out which feature is responsible towards the financial distress of a stock. It also sheds light on the use of NLP techniques in order to extract data from news articles and find the features of stock market which has the highest variability in its SHAP score. It also uses distance-to-default parameter. It mainly focuses on news articles and business news in order to predict the directions of stock market and look for crashes.

Novel Method of Identifying Time Series Based on Network Graphs

This study experiments with various types of time series data and then converts them into graph. Each time series results in a separate kind of graph. The constant time series turns into a complete graph. The periodic time series like a sine graph turns into a regular graph and so on. The properties of the graph such as their centrality measures, clustering coefficient etc. gives information about the time series.

A hybrid supervised semi-supervised graph-based model to predict one-day ahead movement of global stock markets and commodity prices

This paper uses a semi-supervised approach by building a network of stock indices in the same time zone. The supervised portion of the model predicts the movement of stock market which then sends these results into the network. The research compares its results with the traditional classification methods such as KNN, SVM and Random Forests etc. with their model of HyS3 and Kruskal based graph construction.

Factors Affecting Stock Prices in the UAE Financial Markets

This research focuses on the development of the stock market in the United Arab Emirates (UAE) and aims to identify the key factors influencing stock prices in this emerging market. Covering the period from 1990 to 2005 and based on data from 17 companies, the study employs regression analysis with five independent variables, excluding oil price and dividend per share due to multicollinearity issues. Notably, the findings align with previous research, revealing a strong and positive impact of earnings per share (EPS) on UAE stock prices. Money supply and GDP exhibit expected positive coefficients, albeit statistically insignificant, while the consumer price index demonstrates a significant negative relationship with stock prices, particularly at the 1% confidence level, unlike the interest rate, which remains statistically insignificant.

A method for automatic stock trading combining technical analysis and nearest neighbour classification

This paper uses a nearest neighbour classifier and checks whether considering only historical data can be feasible in analysis of stock market or not. It uses technical indicators such as stop loss, stop gain, RSI filter as parameters for its own trading strategy. It compares the results of the traditional buy-and-hold strategy with its own. The variable to analyse here was profit which turned out to be better than the buy-and-hold strategy's profit.

4 PROBLEM STATEMENT

To classify stock price variation patterns using complex network and machine learning and deep learning techniques.

5 METHODOLOGY AND WORK DONE

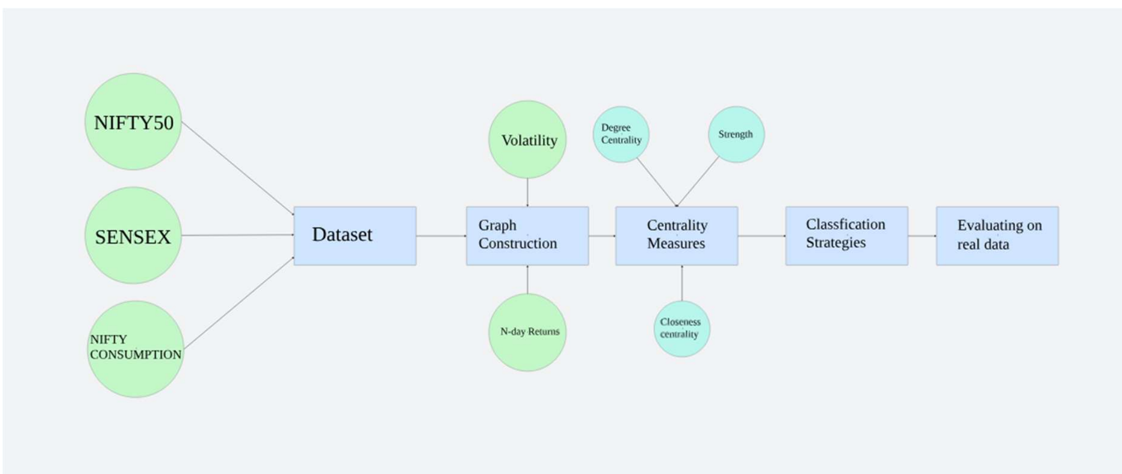

Figure 1: Workflow of research

5.1 DATASET

The dataset used in this research consists of the closing prices of 3 stock indices of India namely SENSEX, NIFTY50 and NIFTY Consumption every day from 01-01-2014 to 31-12-2023 (10 years). SENSEX is a free-float market capitalization consisting of 30 most traded and relatively liquid stocks which contribute towards the balance of the country's equity market. NIFTY50 on the other hand is a benchmark index of 50 companies. NIFTY

Consumption reflects the performance of companies in the domestic consumption sector. The data for SENSEX and NIFTY50 is taken from MarketWatch and NIFTY Consumption is taken from Yahoo Finance.

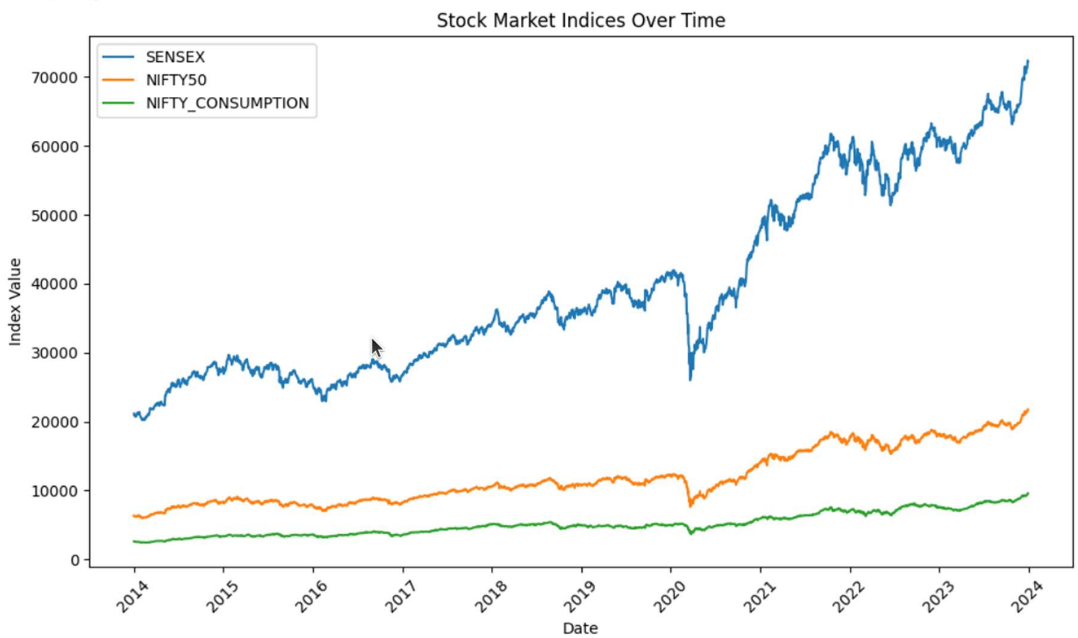

Figure 2 : Stock market indices graph over time for SENSEX , NIFTY50 and NIFTY consumption

5.2 Creation of Graph

The graph is constructed using the study of Cao, Lin et. al. who uses *N*-day volatility *V*, and *N*-day return *R* in order to divide the movement of stock index in 4 separate variations.

$$R = ln\left(\frac{Close(t)}{Close(t-N)}\right)$$

Here, *t* refers to the current day in consideration and *N* refers to the number of continuous trading days (generally a week if there is no national holiday) and *Close(t)* refers to the closing price of the stock index on t^{th} day. In order to find out *V*, we need to find one-day return which is r which is given by

$$r = ln\left(\frac{Close(t)}{Close(t-1)}\right)$$

After calculating *r*, we can calculate *V* for *N* days is given by

$$V = S.D.\left(r_1, r_{2,\dots} r_N\right) * \sqrt{N}$$

Where $S.D.(r_1, r_{2,\dots} r_N)$ refers to the standard deviation of $r_1, r_{2,\dots} r_N$.

We can now calculate the average Volatility of the entire stock index in question by simply averaging over the entire time series.

$$V' = \frac{1}{N}\sum V$$

Now, we can classify the changes in any stock index on the basis of these parameters in the following way

$$P = \left\{ \begin{array}{ll} P1, & when\ R \geq 0\ and\ V \geq V'\ (sharp\ rise) \\ P2, & when\ R \geq 0\ and\ V < V'\ (stable\ rise) \\ P3, & when\ R\ < 0\ and\ V \geq V'\ (sharp\ fall) \\ P4, & when\ R < 0\ and\ V < V'(stable\ fall) \end{array} \right\}$$

The classification is done for all the 3 indices and the combination of the patterns formed represents a node of a graph. Since the total number of combinations can be $4^3 = 64$, we used a 4-base number system as nodes u and v for the graph.

The graph is constructed for 60 days although experimenting with other window sizes is still a future prospect of this research.

days	SENSEX	NIFTY50	NIFTY Consumption	Combined Pattern
1	P3	P2	P1	P3P2P1 Window 1
2	P4	P1	P4	P4P1P4 Window 2
3	P2	P1	P3	P2P1P3
4	P3	P2	P1	P3P2P1
5	P1	P2	P3	P1P2P3
6	P2	P1	P3	P2P1P3
7	P3	P2	P1	P3P2P1
…				
60	P1	P2	P4	P1P2P4
61	P2	P3	P3	P2P3P3

Figure 3: Sample patterns for 7 days of data of SENSEX, NIFTY50 and NIFTY Consumption

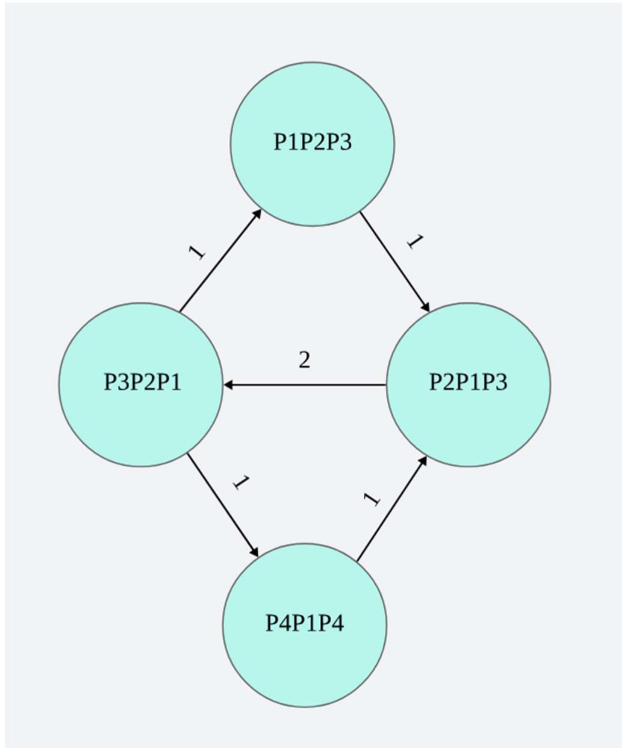

Figure 4: Example Weighted pattern graph for the data mentioned in Table 1 for the first 7 days

5.3 CENTRALITY MEASURES AS INPUT VARIABLES

The significance of this graph is that the denser this graph is, more is the dispersion and hence more care needs to be taken by investors while investing. Hence, we consider certain centrality measures in order to feed them as characteristics of our graph for classification on unseen data.

The measure we are considering for this research are degree centrality, strength (as described by Cao, Lin et. al.), closeness centrality and betweenness centrality. So far, we have applied the KNN algorithm for degree centrality and strength of the network.

5.3.1 Network Average Degree Centrality

In undirected networks, the average degree centrality of the network reflects the level of connection between one node and other nodes in the network, that is, whether one node is connected with the other nodes or not [17]. The formula is as follows:

where is the number of nodes in the network and is the value of the adjacency matrix of an undirected network. if node and node are connected, otherwise . The adjacency matrix of an undirected network is a symmetric matrix. However, does not mean that in a directed network. In the directed network, we must consider the out-degree and in-degree. We connect the nodes in time order so that the in-degree and the out-degree are the same except for the first node and the last node. Therefore, we only select the in-degree for analysis, and calculate the average in-degree centrality as follows:

$$\rho = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N} a_{ij}$$

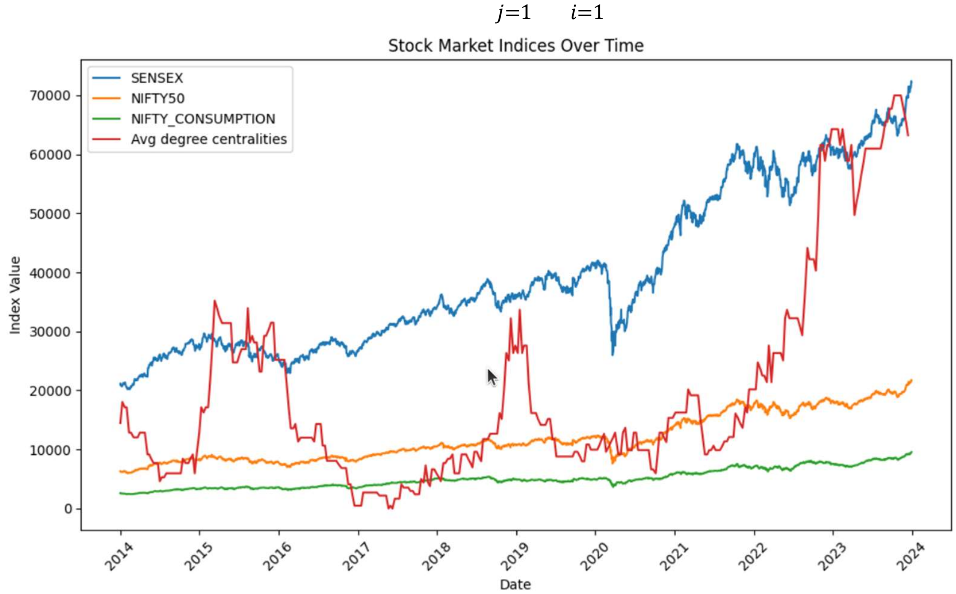

Figure 5: Normalized average degree centralities over the time

5.3.2 Average Network Strength

In a network, the strength of the connection from node to node is the weight of the directed edge from node to node . Similar to the in-degree and out-degree of the directed network, the strength of the directed weighted network can also be divided into in-strength and out-strength. In this study, we describe average out-strength as average network strength:

The greater the average strength of the network, the fewer the number of network nodes, the simpler the composition of the price volatility patterns, the smaller the complexity of the network, and the higher the frequency of the same node. The simpler price patterns reflect the fact that the consistency of price changes of different stocks is stronger and lasts for longer.

$$S = \frac{1}{N}\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}$$

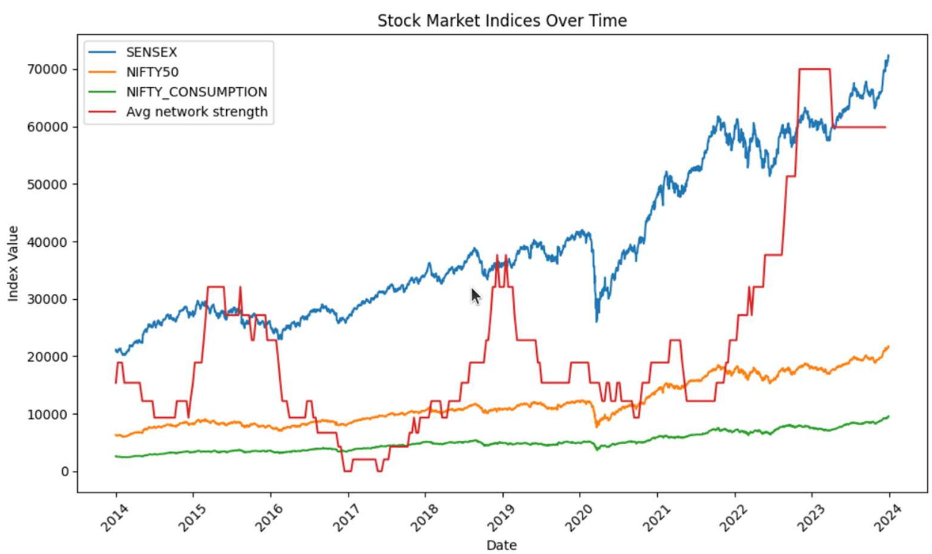

Figure 6: Normalized average network strength over the time

6 Future Work

The following are the major contributions of this research which are yet to come in the future:

1. Applying various classification methods including SVM, random forest, deep learning strategies including neural networks in order to compare the strategies and find out the best one. We will also consider ensemble techniques by combining multiple classification strategies together.
2. Currently, we worked with only two centrality measures. In the future, we will work with more centrality measures such as betweenness centrality and closeness centrality in order to increase the input variables for classification algorithms.
3. We will also practically apply these machine learning techniques by using them through various investment strategies in trading simulators to better evaluate our results.
4. Currently, a window size of 60 days is only considered. However, graphs with other window sizes will also be considered on which these algorithms will be applied again. This will generate more results.

7 References

[1] Lucas Lacasa, Bartolo Luque, Fernando Ballesteros, Jordi Luque and Juan Carlos Nun. From time series to complex networks: The visibility graph

[2] Yung-Keun Kwon, Sung-Soon Choi and Byung-Ro Moon. Stock Prediction based on Financial Correlation.

[3] Ying Li, Hongduo Cao and Yong Tan. Novel Method of Identifying Time Series Based on Network Graphs. Department of Management Science, Business School, Sun Yat-Sen University, Guangzhou 510275, China

[4] Minggang Wang, Ying Chen, Lixin Tian, Shumin Jiang, Zihao Tian, Ruijin Du. Fluctuation behaviour analysis of international crude oil and gasoline price based on complex network perspective. Accepted at Elsevier, May 2016.

[5] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels, Ruben Gryp. Evaluating multiple classifiers for stock price direction prediction. Accepted at Elsevier, May 2015.

[6] Lamartine Almeida Teixeira, Adriano Lorena Inácio de Oliveira. A method for automatic stock trading combining technical analysis and nearest neighbour classification. Accepted at Elsevier.

[7] Hongduo Cao, Tiantian Lin, Ying Li, and Hanyu Zhang. Stock Price Pattern Prediction Based on Complex Network and Machine Learning. Accepted at Wiley, May 2019

[8] Arash Negahdari Kiaa, Saman Haratizadeha, Saeed Bagheri Shourakib. A hybrid supervised semi-supervised graph-based model to predict one-day ahead movement of global stock markets and commodity prices. Accepted at Expert Systems with Applications, March 2018

[9] Jan Grudniewicz, Robert Slepaczuk. Application of machine learning in algorithmic investment strategies on global stock markets. Accepted at Elsevier, July 2023.

[10] Tae Kyun Leea, Joon Hyung Chob, Deuk Sin Kwonb, So Young Sohnb. Global stock market investment strategies based on financial network indicators using machine learning techniques. Accepted at Expert Systems with Applications, September 2018.

[11] Hussein A. Hassan Al-Tamimi, Ali Abdulla Alwan & A. A. Abdel Rahman. Factors Affecting Stock Prices in the UAE Financial Markets. Accepted at Transnational Management, March, 2011.