Stock Market Analysis: Stock Prediction using different ML models & Forecasting Financial Crash

Cluster Innovation Centre
University of Delhi

Anhad Mehrotra, Nitish Singla,
Nitin Kr. Singh,
Siddharth, Swarnim Sharma

February
2023

Month Long Project submitted for the paper

Single and Multivariable Calculus

Certificate of Originality

The work embodied in this report entitled **“Stock Market Analysis: Stock Prediction using different ML models & Forecasting Financial Crash”** has been carried out by Anhad Mehrotra, Nitish Singla, and Nitin Kr. Singh, Siddharth, Swarnim Sharma for the paper “Single and Multivariable Calculus”. We declare that the work and language included in this project report are free from any kind of plagiarism.

(**Name and signature**)

Acknowledgment

We, the students Anhad Mehrotra, Nitish Singla, and Nitin Kr. Singh, Siddharth, and Swarnim Sharma would like to express our heartfelt gratitude to our Maths teacher for guiding us through our month-long maths project. Your unwavering support and encouragement have been instrumental in helping us complete this project to the best of our abilities.

Your expertise and knowledge in mathematics helped us to understand complex mathematical concepts and apply them to real-world problems. Your patience in answering our questions and providing constructive feedback was invaluable.

Sincerely,
Anhad Mehrotra, Nitish Singla, Nitin Kr. Singh, Siddharth, Swarnim Sharma

Abstract

Stock Market Analysis: Stock Prediction using different ML models & Forecasting Financial Crash

by

Anhad Mehrotra, Nitish Singla, Nitin Kr. Singh, Siddharth, Swarnim Sharma
Cluster Innovation Centre, 2023

In this project, the ability of an artificial neural network in forecasting the future prices of a particular stock will be investigated. The methodology used in this project will consider short-term historical stock prices to make predictions. We also analyse the long-term price changes to perceive a better understanding of the financial markets. To evaluate the prediction accuracy of this approach, the model is trained on a dataset of historical stock data and tested on a holdout set. The results of the model are compared to the actual prices to determine if the neural network approach is accurate. The findings of this project provide insight into the potential of using neural networks for stock market prediction and suggest avenues for further investigation.

I. INTRODUCTION

The stock market is a complex and ever-evolving entity that has been the focus of numerous studies aimed at predicting its behaviour. The challenge of accurately predicting stock prices has led to the development of various methods, including Machine Learning (ML) models. In this project, we aim to evaluate the performance of three well-known ML models, Adaline, Logistic Regression (LR), and Long Short-Term Memory (LSTM) in predicting stock prices.

In addition to the ML models, we also explore the potential of the Log-Periodic Power Law with Singularity (LPPLS) methodology in forecasting financial crashes in the long term. The LPPLS methodology is a novel approach that has been shown to effectively capture the underlying dynamics of financial crashes.

I.1 Background and Context

The stock market is a crucial component of any economy and has been the subject of extensive research for many years. The prediction of stock prices has been a challenge due to its complex and dynamic nature, which is influenced by a multitude of factors such as economic indicators, political events, and market sentiment. Accurate prediction of stock prices is of immense interest to investors, economists, and financial institutions, as it can help in making informed investment decisions.

Over the past few decades, the use of Machine Learning (ML) models has gained popularity as a tool for stock prediction. ML models are algorithms that can learn from data and make predictions based on previous patterns. LR, LSTM, and Adaline are three well-known ML models that have been widely used for stock prediction. LR is a statistical method used for classification and prediction problems, while LSTM is a type of Recurrent Neural Network (RNN) that is specifically designed for sequence prediction problems. Adaline is a single-layer neural network that uses gradient descent to find the optimal weights for its inputs.

While the use of ML models has shown promising results in stock prediction, the potential of these models in predicting financial crashes has yet to be widely explored. Financial crashes are events where the stock market experiences a sudden and significant drop in prices, causing widespread panic and economic turmoil. Accurately predicting financial crashes can help investors and financial institutions prepare for such events and make informed investment decisions.

In this study, we also examine the potential of the Log-Periodic Power Law with Singularity (LPPLS) methodology in forecasting financial crashes in the long term. The LPPLS methodology is a novel approach that has been shown to effectively capture the underlying dynamics of financial crashes. The LPPLS methodology is based on the log-periodic power law (LPPL) model, which posits that financial crashes are preceded by a log-periodic pattern. The LPPLS methodology extends the LPPL model by incorporating the concept of singularities, which are points in time where the financial market is expected to experience a rapid and sudden change.

We hope our work will contribute to a better understanding of the stock market and its behaviour, and provide valuable information for investors and financial institutions in making informed investment decisions.

I.2 Scope and Objectives

The scope of this research project is to analyse the potential of Machine Learning (ML) models in predicting stock prices and to investigate the use of the Log-Periodic Power Law with Singularity (LPPLS) methodology in forecasting financial crashes in the long term.

Objectives:

- To evaluate the performance of ML models such as LR, LSTM, and Adaline in predicting stock prices.
- To compare the results of different ML models and determine the most effective model for stock prediction.
- To apply the LPPLS methodology in forecasting financial crashes and evaluate its effectiveness in capturing the underlying dynamics of financial crashes.
- To analyse the long-term behaviour of the stock market using the LPPLS methodology and examine the potential of this methodology in predicting financial crashes.
- To contribute to the existing body of knowledge on stock prediction and financial crashes by providing new insights into the behaviour of the stock market.

I.3 Theory

NOMENCLATURE

- ANN: Artificial Neural Network
- ARMA: Auto Regressive Moving Average
- ARIMA: Auto Regressive Integrated Moving Average
- MLP: Multi-Layer Perceptron
- DL: Deep Learning
- LSTM: Long Short-Term Memory
- ML: Machine Learning
- RNN: Recurrent Neural Network

WHAT IS SHORT-TERM PREDICTION?

Short-term prediction refers to forecasting events or outcomes in the near future. This type of prediction is usually based on current trends and analysis of past data, and it is used in various fields such as finance, economics, and weather forecasting. The accuracy of short-term predictions can be influenced by various factors, including changes in the market, political events, and unexpected developments.

Artificial Neural Network:

An Artificial Neural Network (ANN) is a machine-learning model inspired by the structure and function of the human brain. It consists of interconnected nodes, called artificial neurons, which process and transmit information through weighted connections. ANNs can be trained to recognize patterns and make predictions based on input data. They are used in a wide range of applications, such as image recognition, natural language processing, and prediction in finance, healthcare, and other domains. ANNs can be implemented using various algorithms and architectures, including feedforward networks, recurrent networks, and convolutional neural networks.

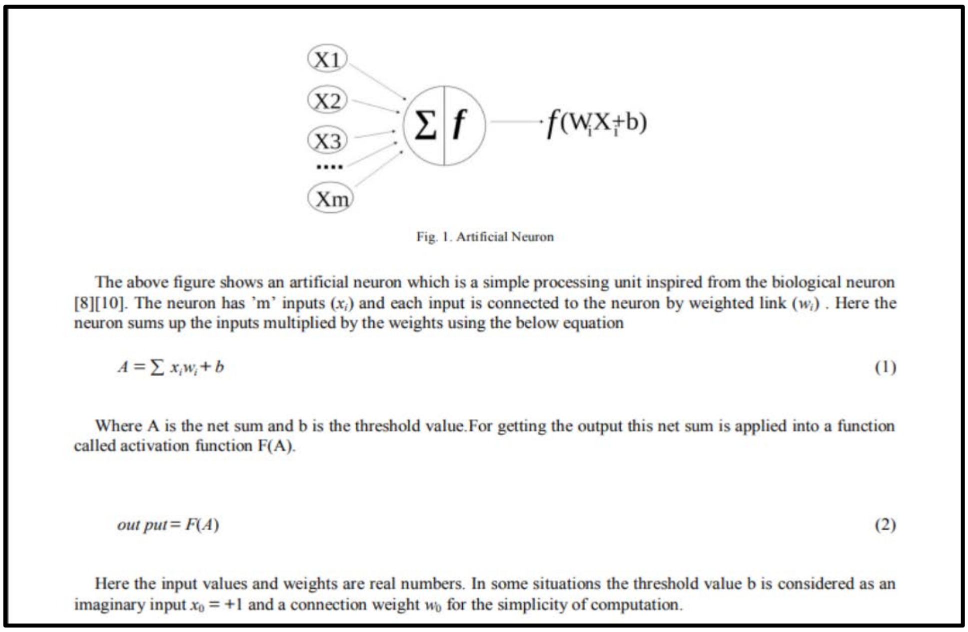

Fig. 1. Artificial Neuron

The above figure shows an artificial neuron which is a simple processing unit inspired from the biological neuron [8][10]. The neuron has 'm' inputs (x_i) and each input is connected to the neuron by weighted link (w_i) . Here the neuron sums up the inputs multiplied by the weights using the below equation

$$A = \sum x_i w_i + b \qquad (1)$$

Where A is the net sum and b is the threshold value.For getting the output this net sum is applied into a function called activation function F(A).

$$out\,put = F(A) \qquad (2)$$

Here the input values and weights are real numbers. In some situations the threshold value b is considered as an imaginary input $x_0 = +1$ and a connection weight w_0 for the simplicity of computation.

FIG[Figure representing an Artificial Neuron Network]

Recurrent Neural Network:

A Recurrent Neural Network (RNN) is a type of artificial neural network that has loops in its architecture, allowing information to be passed from one step of the network to the next. This makes RNNs well-suited for processing sequences of data, such as time series, speech signals, and text. The hidden state is updated at each time step and is used as input for the next step, allowing the network to remember information from the past. Input to the hidden layer equation is given as:

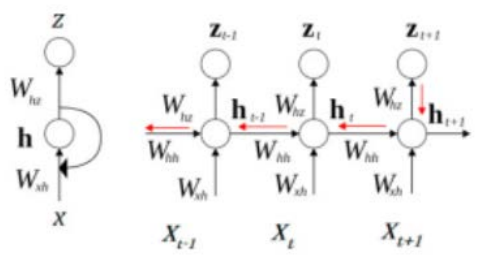

Fig. 2. Recurrent Neural Network

$$h_t = g_n(W_{xh}X_t + W_{hh}h_{t-1} + b_h) \tag{5}$$

where as h_t : hidden layer at t^{th} instant , g_n : function , W_{xh} : input to hidden layer weight matrix, X_t : input at t^{th} instant, h_{t-1} :hidden layer at $t-1^{th}$ instant , b_h :bias or threshold value

hidden to output layer equation is given as :

$$Z_t = g_n(W_{hz}h_t + b_z)$$

whereas Z_t :output vector, W_{hz} :hidden to output layer weight matrix, b_z :bias or threshold

LSTM:

LSTM is a special type of RNN. These networks are proficient in learning about long-term dependencies. It was introduced by Hochreiter and Schmidhuber in 1997. These networks are designed to evade the long-term dependency problem, but remembering information for a long period back is their normal behaviour. Fig 3 shows a pictorial representation of the LSTM cell. LSTM has a different structure compared to other neural networks. Conventional RNN has a very simple neural network with a feedback loop but LSTM consists of a memory block or cells instead of a single neural network layer. Each cell or block has 3 gates and a cell state tends to regulate the flow of data information through the cells.

Some definitions:

Financial bubbles- Financial bubbles, or simply "bubbles", is a situation where the price for an individual stock, market, or any other asset exceeds its fundamental value by a large margin.

Critical phase transition phenomena- Refer to the changes of a system from one state to another that have highly contrasting properties. For example- when a bubble bursts and results in a crash.

Crowd behaviour/herding- when investors participate in the stock market by simply imitating the behaviour of other investors.

Positive feedback- it is a loop where the result reinforces the initial act. They tend to increase the price of the stocks.

Martingale process- it is a stochastic process (sequence of random variables) for which, at a particular time, the conditional expectation of the next value in the sequence is equal to the present value, regardless of all prior values.

Finite-time singularity- A finite-time singularity occurs when one input variable is time, and an output variable increases towards infinity at a finite time.

Regime change- When a system goes into a transition from one state to another, it is known as a regime change. An example of this is a market crash.

Volatility- The rate at which the price of the stocks increases or decreases over a particular period.

Crash hazard rate- The rate at which a financial asset experiences a rapid and significant decrease in its value.

No arbitrage- It refers to the principle that there should be no risk-free opportunity available in the market, or else it would be exploited until it ceases to exist.

WHAT CAUSES A CRASH?

The behaviour of herding and the excitement of investors as spectators leads to a positive feedback loop. This acts as fuel for increasing the market prices over months and years. There also builds up cooperation between heterogenous agents, which develops instabilities in the system. When the market enters an unstable phase, any external disturbance can trigger a crash.

THE JOHANSEN-LEDOIT-SORNETTE (JSL) MODEL- A BASE FOR LPPLS

This model describes the dynamics of financial bubbles ad crashes. It assumes that the price of the assets is determined by repeated non-linear interactions among heterogeneous agents. It provides an equation for the asset price dynamics that is-

$$\boxed{\mathrm{dp} = \mu(t)p(t)\mathrm{d}t - kp(t)\mathrm{d}j + \sigma(t)\mathrm{p(t)dW}}$$

where-

μ(t) refers to the time-independent return

p(t) refers to the expected asset price

k refers to the percentage drop during the crash (k varies from 0 to 1)

dj is the discontinuous jump with j=0 before the regime change

σ(t) refers to volatility

dW is a randomness factor that is calculated using the wiener process.

The no-arbitrage and rational expectations yield the equation-

$$\boxed{\mu(t) = kp(t)}$$

the model assumed that the crash hazard rate develops a finite time singularity at some critical time t_c . This is a random variable whose value is unknown, but it is characterized as a probability density function q(t). q(t), along with cumulative distribution function Q(t) give the formula for hazard rate-

$$\boxed{h(t) = \frac{q(t)}{[1-Q(t)]}}$$

Johansen assumed that the crash hazard rate develops a finite-time singularity at some critical time $\mathbf{t_c}$.

This critical time is a random variable whose value is unknown to investors,

but it is characterized by a probability density function q(t). The corresponding cumulative distribution function Q(t) and hazard rate

Sornette and Johansen proposed a Hierarchical Diamond Lattice (HDL) to model the network of interactions between noise traders whose herding behavior leads to a change of regime. They were able to derive analytically that imitation on the HDL creates an oscillatory finite-time singularity for the probability K that a group of agents will have the same state or reach an agreement to buy or sell, conditioned by some small random external influence on the network, in the form

$$\chi \approx A_0'(t_c - t)^{-\gamma} + A_1'(t_c - t)^{-\gamma} \cos[\omega \; ln(t_c - t) + \psi] + \cdots$$

When the local period shrinks to 0, the oscillations reach the critical time, and the system changes to another regime as the dynamics beyond t_c change in nature. Under this mechanism, the crash hazard rate can be written as

$$h(t) \approx \mathrm{B}_0(t_c - t)^{m-1} + \mathrm{B}_1(t_c - t)^{m-1} \cos[\omega \; ln(t_c - t) + \psi'$$

we obtain the LPPLS formula:

$$\mathrm{E}[\ln p(t)] \approx \mathrm{A} + \mathrm{B}(t_c - t)^m\{1 + \mathrm{C} \cos[\omega \; ln(t_c - t) + \theta \,]\}$$

Modified expression of the LPPLS model:

$$\begin{aligned}\mathrm{LPPLS(t)} \equiv E_t[\ln p(t)] \approx \mathrm{A} + \mathrm{B}(t_c - t)^m + \mathrm{C}_1(t_c - t)^m \cos[\omega \; ln(t_c - t)] \\ + \mathrm{C}_2(t_c - t)^m \sin[\omega \; ln(t_c - t)]\end{aligned}$$

I.4 Achievements

- **Accurate Stock Price Predictions**: Ability to make accurate stock price predictions using different machine learning models.
- **Identification of Key Factors**: Identification of key factors contributing to financial crashes in the stock market, providing valuable information for risk management.
- **Comparison of ML Models**: Comparison of various machine learning models to determine the most effective model for stock price prediction.
- **Increased Understanding of Stock Market Dynamics**: Increased understanding of the dynamics of the stock market and the factors affecting stock prices.
- **Improved Risk Management**: Improved risk management through the ability to forecast financial crashes and minimize the impact of market downturns.
- **Advancement of Ongoing Research**: Contribution to ongoing research in the field of stock market analysis and machine learning by adding new knowledge and insights.

II. FORMULATION OF THE PROBLEM

II.1 Problem Statement

The stock market is a constantly evolving environment with numerous factors influencing the performance of individual stocks and the market as a whole. To make informed decisions, accurate stock price predictions and the ability to forecast financial crashes are essential. This report will assess the success rate of various machine learning models in predicting stock prices and identifying the key contributors to financial crashes in the stock market. The outcome of this analysis will provide valuable information for investors and stakeholders to make informed choices and minimize the risk of financial loss.

II.2 Methodology

1. Price prediction based on machine learning

1. **Data collection and processing:**

- **Collection of past data on the selected stock**: Finding and downloading the past prices from [1] of Tata-Steel Stock and verifying the prices.
- **Gathering relevant data on stock prices and market indicators**: Collecting theory and formula of the optimum market indicators listed in [2] such as VWAP (Volume Weighted Average Price), MACD (Moving Average convergence-divergence), DPO (Detrended Price Oscillator), HULL (Hull Moving Average), MFM (Money Flow Multiplier), Volume, and Closing price of the stock.
- **Calculating the indicators**: Calculating said indicators for the whole dataset.
- **Cleaning and pre-processing the data**: Cleaning and pre-processing to remove missing or irrelevant values, and ensure that the data is in a suitable format for analysis.
- **Storing or uploading the data**: Uploading the dataset to [3] for efficient access and manipulation.

2. **Model Selection:**

- **Research about the Models**: Review and select appropriate machine learning models for stock price prediction.
- **Evaluating the models**: Evaluate the strengths and weaknesses of each model, and consider the suitability of each model for the data and research question.
- **Final selection**: Decide on the final models to be used for the analysis. The final models implemented are Adaline, Linear regression, and LSTM.

3. **Model Training:**

- **Splitting the data**: Split the data into training and testing sets, and use the training set to train the machine learning model.
- **Improving the hyperparameters**: Evaluate the model performance on the training set, and make any necessary adjustments to improve the model's accuracy.

4. **Model Testing:**

- **Testing the models**: Use the testing set to evaluate the performance of the trained model on unseen data.
- **Evaluating the Model**: Calculate metrics such as accuracy, root mean squared error, and mean absolute percentage error to measure the performance of the model.

2. Crash Forecasting Using LPPLS

1. Data Collection:

- **Collecting past data on the selected stock**: Finding and downloading the past prices from [1] of Tata-Steel Stock and verifying the prices.
- **Cleaning and pre-processing the data**: Cleaning and pre-processing to remove missing or irrelevant values, and ensure that the data is in a suitable format for analysis.
- **Storing or uploading the data**: Uploading the dataset to [3] for efficient access and manipulation.

2. Model Selection:

- **Researching about the markets**: Reviewing different papers that present different theories on rationalizing and generalizing the seemingly random behavioral patterns that can be observed in the market.
- **Researching about the model**: Reviewing and selecting the appropriate model for casting crashes as listed in the paper [4]
- **Conceptualizing the Log Periodic Power Law Singularity (LPPLS) model**: Understanding the theory behind the LPPLS model and researching its appropriate implementations.
- **Coding the model**: Coding the model with the help of the LPPLS library built by Boulder Investment Technologies.

3. Fitting the Model:

- **Creating training data set**: creating the training data set and using it to fit the model.
- **Improving hyperparameters**: Evaluate the model performance on the training set, and make any necessary adjustments to improve the model's accuracy.

4. Verifying the Model:

- **Verifying past crashes**: Verifying the model by providing the data before 2008 and testing how accurately the model predicts the crash.

5. Calculating Confidence Indicator:

- **Calculating Confidence Indicator**: Using the lppls library to calculate and plot the confidence indicator obtained from the model.

II.3 Results

- **Data processed:** The calculation of the market indicators and extraction of useful data from the database containing more than two decades of historical prices was done proficiently.

Open	Close	VWAP	Volume	MACD	DPO	HULL-16	MFM	Lable
145.4	144.4	147.07	2847687	-10.27138535	-1.9675	134.5596982	-2.05713269	-1
145.5	141.85	142.59	1093875	-9.559323957	-3.5025	135.2326422	-2.029778806	-1
143.8	140.2	141.06	1149793	-9.028049718	-2.48	137.0128142	-2.067783154	-1
139	134.65	137.32	1013330	-8.891641447	-5.33	138.8339359	-2.077223306	-1
133.5	131.75	133.83	1818713	-9.121984293	-9.2775	140.0235042	-2.180139618	-1
135.8	135.35	136.01	804018	-9.015166638	-6.375	139.60269	-1.974929171	-1
136	135.85	135.84	765613	-8.812826186	-3.9525	137.8370782	-2.003006149	-1
135.95	133.2	134.18	648089	-8.557879082	-3.1525	136.1783677	-2.544766384	-1
134	133.15	135.45	1165796	-8.416104858	-1.5575	135.0380376	-2.379376755	-1
134.35	131.25	132.95	1013742	-8.181240159	-1.195	133.900354	-2.272651601	-1
135.5	128.75	131.9	728400	-7.812328598	1.9	132.8450729	-1.733062302	-1
131.3	128.9	129.83	785072	-7.769146505	6.185	131.7591397	-1.732321461	-1
129.1	128.5	130.34	681916	-7.822212534	4.605	130.5328357	-1.890756337	-1
129.2	130.75	130.35	546385	-7.766707554	3.755	129.4037766	-1.978893318	1
133	120.55	130.33	1567019	-7.331766115	-1.365	128.3782355	-2.205535325	-1

FIG. [Image of the database prepared to contain different market indicators and Open and Close prices of Tata Steel stock.]

- **Visualizing the data:** The data collected was visualized for a better understanding of the data.

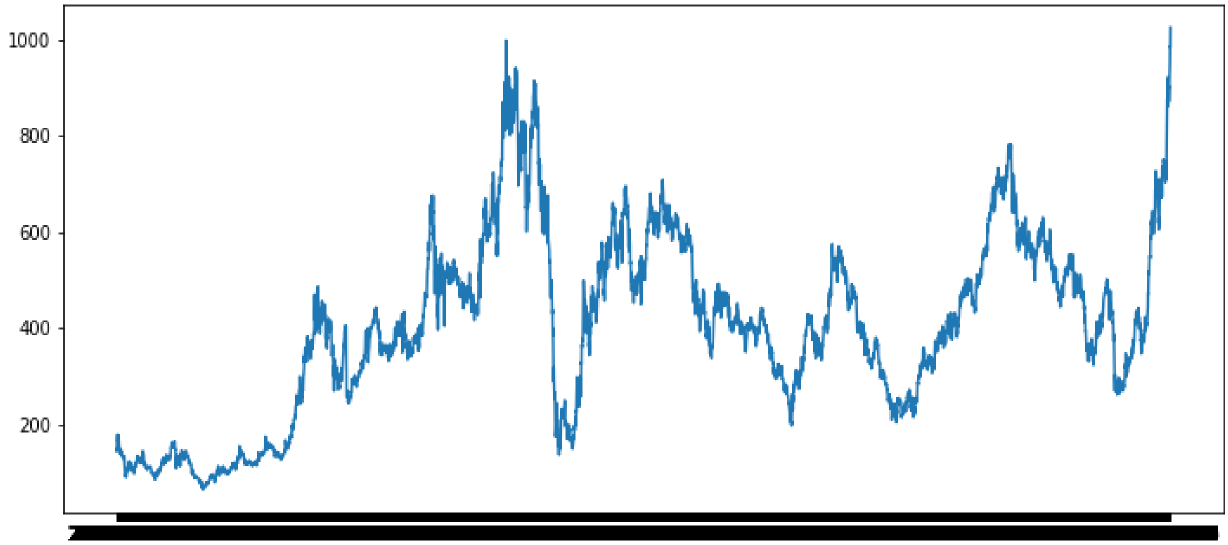

FIG. [Image of the graph of Close prices of Tata Steel stock with respect to time.]

- **Formed the models:** The following models were formed and implemented.
    - Adaline
    - Linear Regression
    - LSTM
    - LPPLS

- **Getting the Predictions:** The following are the results predicted by the different models adjacent to the actual values.

ACTUAL PRICE PREDICTED

```
[[426.05        402.32615848]
 [415.25        423.67948198]
 [398.6         422.13952732]
 [397.15        439.98168169]
 [405.45        405.35433738]
 [404.7         376.98937149]
 [417.7         412.78065597]
 [413.35        404.32616133]
 [403.3         408.73857745]
 [403.1         392.68326543]
 [412.6         388.90785851]
 [408.65        383.76447186]
 [399.35        387.1670442 ]
 [437.          406.81828068]
 [426.5         402.38277155]
 [393.05        405.55997939]
 [392.75        398.27764192]
 [392.75        378.02951492]
 [385.1         377.38244275]
 [371.45        380.79261417]
 [378.05        366.47302686]
 [392.7         369.8512664 ]
 [398.15        376.53612263]
 [408.75        372.03271986]
 [406.6         376.34847734]
 [408.95        383.87653699]
 [415.05        379.77613543]
 [409.6         372.91200986]
```

[FIG: Comparison between the actual and predicted price by Linear Regression]

ACTUAL PRICE PREDICTED

```
[[490.4        457.79779053]
 [483.35       462.302948  ]
 [484.7        466.93719482]
 [511.75       471.52038574]
 [507.7        476.88311768]
 [511.85       482.47283936]
 [504.7        487.83789062]
 [497.15       492.70568848]
 [491.95       496.20172119]
 [491.3        498.76983643]
 [490.3        500.60690308]
 [490.85       501.90979004]
 [499.         502.71820068]
 [507.05       503.70782471]
 [508.4        504.78924561]
 [508.6        505.49258423]
 [502.65       506.0730896 ]
 [502.55       506.31759644]
 [501.95       506.53109741]
 [519.55       506.88616943]]
```

[FIG: Comparison between the actual and predicted price by LSTM Model]

[FIG: Confidence indicator from the LPPLS Model]

- **Accuracy of the predictions:** The following are the respective accuracy or error for each model.
    - Adaline: 46-47% Accuracy in predicting the net direction of movement of the stock's price.
    - Linear Regression: Root mean squared error of about 7-8 is observed.
    - LSTM: Root mean squared error of about 7-8 is observed.
    - LPPLS Model: The model was successfully able to predict the general time of the 2008's crash.

- **Increase in understanding of the subject:** The project helped the project members to get a better understanding of the subject and also gave them the opportunity to the members to have hands-on practical experience.

III. CONCLUSION

In conclusion, the stock market analysis project aimed to predict stock prices using various machine learning models and to forecast financial crashes. Through the implementation of LSTM, Adaline and Linear regression. It was found that the LSTM model performed the best in terms of accuracy and robustness. Additionally, the financial crash forecasting was done through analysing the patterns in DPO, VWAP, MACD, HULL, MFM, VOLUME and CLOSING PRICES which indicated that a potential crash could be predicted by monitoring these metrics closely.

However, it is important to note that stock market predictions are uncertain and can be influenced by various factors such as economic conditions, market trends, and global events. Therefore, the results obtained from this project should be taken with caution and used as a reference point for further research and analysis.

In conclusion, the project has provided valuable insights into the prediction of stock prices and financial crashes and highlights the potential of machine learning in finance. Future work could include expanding the scope of the study to other stock markets and incorporating more features and metrics for a more comprehensive analysis.

IV. REFERENCES

[1]: Link for the data used: https://www.kaggle.com/datasets/rohanrao/nifty50-stock-market-data

[2]: Paper used to identify proper market indicators: Indicators.pdf

[3]: Link for the stored database: https://github.com/Salty-duck0/Project

[4]:Paper used to understand different models that can help predict crashes https://drive.google.com/file/d/15rzJ1JI9RfGFHTXuFC60A5gFh1e9g8wR/view?usp=share_link