Mini Project Report

Entitled

Long Term Price Prediction of an IPO using ML

Submitted to the Department of Electronics Engineering in Partial Fulfilment for the Requirements for the Degree of

Bachelor of Technology
(Electronics and Communication)

: Presented & Submitted By :

Archit Sinojiya, Vandita Rawat, Khushi Solanki

Roll No. (U21EC012, U21EC046, U21EC071)

B. TECH. VI (EC), 6^{th} Semester

: Guided By :

Dr. Kishor Upla
Assistant Professor, SVNIT

(Year: 2023-24)

DEPARTMENT OF ELECTRONICS ENGINEERING
SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY
Surat-395007, Gujarat, INDIA.

Sardar Vallabhbhai National Institute Of Technology

Surat - 395 007, Gujarat, India

DEPARTMENT OF ELECTRONICS ENGINEERING

CERTIFICATE

This is to certify that the Mini-Project Report entitled "**Long Term Price Prediction of an IPO using ML**" is presented & submitted by Archit Sinojiya, Vandita Rawat, Khushi Solanki, bearing Roll No. U21EC012, U21EC046, U21EC071, of B.Tech. VI, 6^{th} Semester in the partial fulfillment of the requirement for the award of B.Tech. Degree in Electronics & Communication Engineering for academic year 2023-24.

They have successfully and satisfactorily completed their **Mini-Project** in all respects. We, certify that the work is comprehensive, complete and fit for evaluation.

Dr. Kishor Upla
Assistant Professor & Project Guide

Abstract

The report explores the intersection of India's burgeoning IPO market, the integration of machine learning in stock market analysis, and the hardware integration aspect using the Raspberry Pi board, highlighting the importance of accurate predictions for stakeholders. Dispelling the notion of stock prediction as random, the adoption of machine learning techniques stands out for discerning underlying patterns from historical market data, promising enhanced prediction accuracy and garnering widespread interest across research and industry sectors.

In addition to software-based analysis, the report underscores the significance of hardware integration, particularly with the Raspberry Pi board, in facilitating real-time data processing and predictive modeling. This integration empowers stakeholders with efficient and cost-effective solutions for accessing and analyzing market data, ultimately enabling informed investment strategies amidst the dynamic stock market landscape. Through rigorous validation and refinement, the predictive model derived from machine learning algorithms and Raspberry Pi hardware equips stakeholders with actionable insights, promising transformative advancements in stock market analysis and decision-making processes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IPO	Initial Public Offering
ML	Machine Learning
RPI	Raspberry Pi
GMP	Grey Market Price
LT Price	Long Term Price
HTTP	Hyper Text Transfer Protocol
GPIO	General Purpose Input Output
IoT	Internet of Things
SSH	Secure Socket Shell

Chapter 1
Introduction

India, as a developing nation, is witnessing a surge in companies opting to go public, exemplified by the launch of numerous Initial Public Offerings (IPOs).

1.1 Overview

Accurate predictions of stock performance hold significant potential for generating substantial profits, benefiting both sellers and brokers alike. Contrary to the notion that stock prediction is purely random, many argue that it follows chaotic patterns, implying that with careful analysis of historical market data, future trends can be anticipated. This understanding has led to the adoption of machine learning techniques in stock market analysis.

Machine learning offers a powerful approach to modeling and predicting stock market behavior. By analyzing vast amounts of historical data and identifying patterns, machine learning algorithms can generate predictions that closely approximate actual market values. This capability enhances the accuracy of stock predictions, enabling investors to make more informed decisions.

The introduction of machine learning into the realm of stock prediction has garnered considerable attention from researchers and practitioners alike. Its ability to provide efficient and precise measurements of market dynamics has led to widespread interest and adoption in the financial industry. As a result, machine learning continues to play an increasingly important role in shaping investment strategies and decision-making processes in the stock market.

An IPO denotes the process wherein a private company offers its shares to the general public in a new stock issuance. The price per share is determined based on the Grey Market Price (GMP), which represents the premium amount at which IPO shares are traded before being listed on the stock exchanges. GMP serves as an indicator of the company's perceived value prior to going public. If the GMP exceeds a certain threshold, potential buyers are advised to apply for shares; conversely, if it falls below a

specified threshold, buyers are advised to refrain from applying.

In the absence of tailored datasets pertaining to companies that have not yet gone public, the emphasis of this project centers on the forecast of Long Term Prices (LT Prices) specifically for publicly traded companies. This approach acknowledges the inherent challenge of accessing comprehensive data for private entities and redirects attention to leveraging available information regarding publicly listed firms. By concentrating on this subset of companies, the project aims to employ existing market data to develop predictive models for estimating LT Prices, thereby circumventing the limitations posed by the unavailability of data for private entities. Linear Regression Analysis is employed for this purpose.

Linear regression analysis serves as a pivotal tool for predicting the Long Term (LT) Price of stocks by examining various independent variables. This method aims to establish a mathematical relationship between the dependent variable, which is the LT Price, and independent variables such as IPO Size, Subscription, GMP, IPO Price, Estimated Price, and Listing Price. Through the formulation of a linear equation, the model attempts to quantify the impact of each independent variable on the LT Price, represented by coefficients.

Once the model is validated, it can be utilized for prediction purposes. New LT Prices can be forecasted based on different combinations of independent variable values. This predictive capability enables stakeholders to gain insights into potential future stock performance, thereby facilitating informed decision-making in investment strategies.

Moreover, the integration of hardware components, such as the Raspberry Pi board, further enhances the capabilities of machine learning models in stock prediction. The Raspberry Pi, a versatile and affordable single-board computer, provides a platform for deploying machine learning algorithms in real-world applications. By integrating machine learning models with the Raspberry Pi board, investors can access computational power and flexibility, enabling them to perform predictive analytics tasks efficiently and effectively. This integration not only facilitates faster decision-making but also enhances accessibility, allowing investors to leverage advanced analytics tools even in resource-constrained environments.

1.2 Organization of Report

The report is divided into 4 parts. Each chapter contains sections and sub- sections covering different aspects of the topic. Firstly, the overview of the topic is given.Highlighting the reasons for selecting this project. In the second chapter, the software aspects of the project are discussed. Integration of backend and frontend, along with the brief of the linear regression model. Thereafter, hardware interfacing is covered, where Raspberry Pi Board and GPIO is introduced. At last, the future scope is mentioned as well as the references used to understand this topic.

Chapter 2
Software Implementation

In this chapter we go into the details of the Software implementation of the Linear regression model, where the pre-trained model is loaded in the backend, using flask.

2.1 Backend

In this project, Flask serves as the backend framework responsible for handling communication between the website's frontend and the Raspberry Pi module. Flask, a lightweight and versatile Python web framework, provides the necessary infrastructure for creating web applications with ease. It enables us to define routes that handle HTTP requests, such as receiving input data from the website's form submission[1]. Additionally, Flask facilitates data processing and interaction with the pre-trained linear regression model loaded on the Raspberry Pi.

Within Flask, routes are defined to listen for specific HTTP methods, such as POST requests containing input data from the website. Upon receiving this data, Flask extracts and prepares it for processing by the linear regression model[3]. Flask's simplicity and flexibility make it an ideal choice for this project, as it streamlines the backend logic without unnecessary complexity. By utilizing Flask, we establish a seamless connection between the website's frontend and the Raspberry Pi module, enabling efficient prediction of the LT Price and subsequent visualization of the final verdict through LED control.

2.2 Frontend

In this project, HTML, CSS, and JavaScript collectively play crucial roles in creating an interactive and visually appealing user experience on the website frontend.

HTML (HyperText Markup Language) serves as the backbone of the webpage, providing the structural foundation for displaying content and user interface elements. The HTML code defines the layout of the webpage and includes elements such as forms, input fields, labels, and divs[5]. These elements are essential for creating a user-friendly

interface where users can input the IPO data required for LT Price prediction. By structuring the content with HTML, we ensure that the webpage is well-organized and easy to navigate, enhancing the overall user experience.

CSS (Cascading Style Sheets) is utilized to style the HTML elements and control their appearance. Through CSS, we can apply colors, fonts, spacing, and other visual attributes to the webpage elements, and functions well across various devices and screen sizes. In this project, CSS is used to style the form inputs, labels, buttons, and other elements, providing a polished and professional appearance to the webpage.

JavaScript adds interactivity and dynamic behavior to the webpage, enhancing its functionality beyond static content. In this project, JavaScript is used to handle form submission events and send data to the Flask backend via AJAX (Asynchronous JavaScript and XML) requests[1]. It listens for form submission events, retrieves the input data entered by the user, and sends it to the Flask backend for LT Price prediction. Furthermore, JavaScript processes the response from the backend and dynamically updates the webpage with the predicted LT Price and final verdict, providing real-time feedback to the user without requiring a page refresh. This interactive behavior improves user engagement and facilitates a smooth and seamless user experience on the website frontend(refer Figure 2.1).

The linear regression model used in this project aims to predict the Long-Term (LT) Price of Initial Public Offerings (IPOs) based on several input features. Here's a detailed description of each step involved in implementing this model[3].

2.2.1 Introduction to Linear Regression

In the context of the provided code, linear regression serves as a predictive modeling technique used to estimate the Long-Term (LT) Price of Initial Public Offerings (IPOs) based on several input features. Linear regression assumes a linear relationship between the independent variables (features) and the dependent variable (LT Price), implying that changes in the features are linearly related to changes in the LT Price.

The dependent variable, LT Price, represents the target variable we aim to predict. It is the variable of interest in our analysis, as it represents the expected long-term performance or valuation of an IPO.

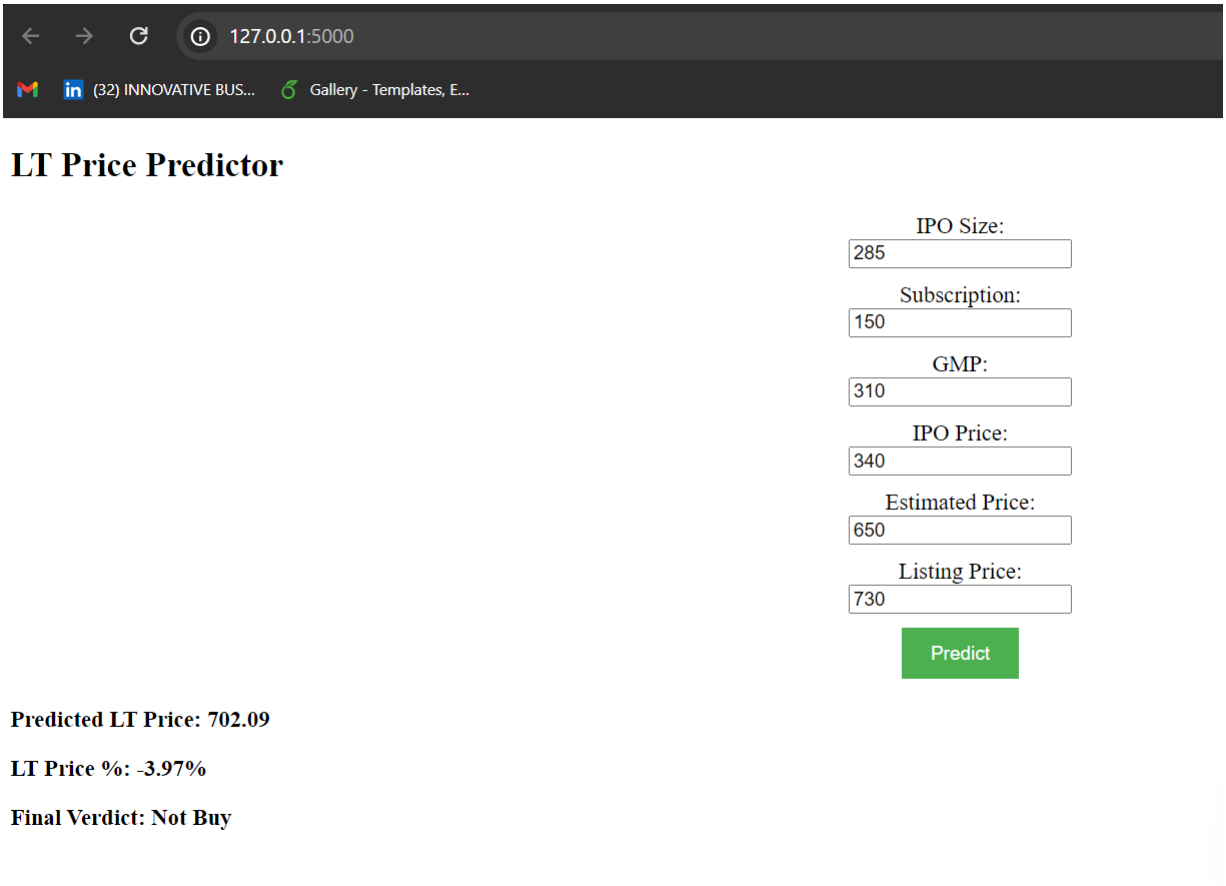

Figure 2.1: Software Implementation

On the other hand, the independent variables (features) are the input factors that influence the LT Price. These features include parameters such as IPO Size, Subscription, GMP (Grey Market Premium), IPO Price, Estimated Price, and Listing Price. Each of these features provides information that may influence the eventual LT Price of an IPO.

The linear regression model seeks to estimate the coefficients of a linear equation that best fits the relationship between the independent variables and the LT Price as shown in Equation 2.1.

Table 2.1: Dataset

IPO Size	Subscription	GMP	IPO Price	Estimated Price	Listing Price	LT Price	LT %age
834	51.14	28	282	310	303	294	4.43
490.33	23.86	48	792	840	792	1109	40.04
463	77	16	60	76	71	68	13.58
1701.44	7.61	24	324	348	330	475	46.64
1900	41.69	160	648	808	829	781	20.54

$$\begin{aligned} LTPrice = &(-0.09 * IPO_Size) + (1.68 * Subscription) \\ &+ (GMP * -0.57) + (IPOPrice * 1.37) \\ &+ (EstimatedPrice * 0.79) + (ListingPrice * -0.57) \\ &+ 92.67 \end{aligned} \tag{2.1}$$

2.2.2 Libraries Used

- NumPy: For numerical computations and data manipulation.
- Pandas: For data manipulation and preprocessing, specifically for importing the dataset into a DataFrame.
- Scikit-learn: For implementing the linear regression model, preprocessing, model selection, and evaluation.

2.2.3 Importing the Dataset

The dataset containing the IPO data is imported into a Pandas DataFrame. This dataset includes features like IPO Size, Subscription, GMP, IPO Price, Estimated Price, and Listing Price, along with the target variable, LT Price.

2.2.4 Preprocessing

- Handling Missing Values: If there are missing values in the dataset, they are handled using techniques like imputation or deletion.

- Handling Missing Values: If there are missing values in the dataset, they are handled using techniques like imputation or deletion.
- Splitting Data: The dataset is split into training and testing sets to train the model on a subset of the data and evaluate its performance on unseen data.

2.2.5 Model Selection

The appropriate features are selected to include in the model. This step may involve techniques like feature selection or regularization methods (such as LASSO or Ridge regression) to prevent overfitting and improve the model's generalization ability.

2.2.6 Fitting Data into Model

The training data is used to estimate the coefficients of the linear equation that best fits the relationship between the independent variables (features) and the target variable (LT Price). This is done using the 'fit()' method provided by the linear regression model object from scikit-learn.

2.2.7 Model Prediction

Once the model is trained, it can be used to make predictions on new data. This involves passing the new data through the model's 'predict()' method, which calculates the predicted values of the target variable based on the learned coefficients.
By following these steps, the linear regression model can effectively predict the LT Price of IPOs based on the given input features, allowing investors to make informed decisions about whether to buy or not.

Chapter 3
Interfacing Hardware

This chapter highlights the key implementation of the Linear regression model on RaspberryPi 4B. The first section involves the introduction to the microprocessor based mini computer, RaspberryPi 4B, dwelling further in dumping the model on the board for processing[5].

3.1 Introduction to Raspberry PI

Raspberry Pi 4 is a versatile and affordable single-board computer that offers significant potential for implementing machine learning models like linear regression. With its enhanced processing power, ample memory, and GPIO pins, Raspberry Pi 4 provides a suitable platform for various data science and machine learning projects, including value prediction using linear regression[6]. Linear regression is a fundamental technique in machine learning used to model the relationship between two or more variables by fitting a linear equation to observed data. Implementing a linear regression model on Raspberry Pi 4 can be beneficial for applications such as predictive analysis, forecasting, and sensor data analysis in IoT projects.

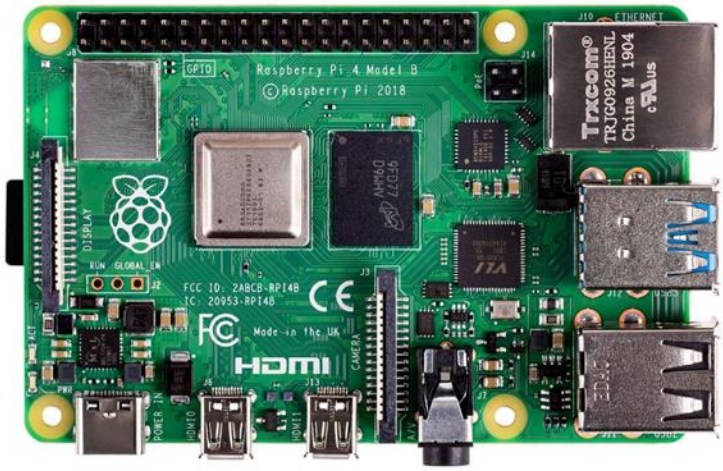

Figure 3.1: Raspberry Pi 4 board

3.2 Connecting RPi

Uploading the python code and csv dataset file on the board requires the following steps to be followed:

3.2.1 Establishing SSH connection

Ensure that both your PC and Raspberry Pi are connected to the same network. Use ssh command to establish the connection.

3.2.2 Finding Raspberry Pi's IP address

Find IP address by running ifconfig on your Raspberry Pi terminal or using a tool like nmap on your PC. Look for the IP address under inet for the interface you're using (usually eth0 for Ethernet or wlan0 for Wi-Fi).

3.2.3 Transferring the file using SCP

Use the scp command to transfer the files from the pc to the board by providing the path on the pc as well as the IP address of the rpi board.

3.2.4 Verifying Transfer and Accessing

ls command can be used to ensure the transfer of files. The files can be accessed by navigating the directories through cd command.

3.3 GPIO

The General-Purpose Input/Output (GPIO) pins on the Raspberry Pi 4 (RPI4) are a set of physical pins that allow the Raspberry Pi to interact with the outside world. These pins can be configured as either inputs or outputs, enabling the Raspberry Pi to read data from sensors, control motors, LEDs, and other electronic components, and interface with a wide range of external devices. The Raspberry Pi 4 has a total of 40 GPIO pins arranged in two rows of 20 pins each. The GPIO pins operate at 3.3 volts. In addition to general-purpose digital I/O, many GPIO pins also support serial communication protocols such as SPI, I2C, and UART, allowing the Raspberry Pi to communicate with a wide range of sensors, displays, and other devices.

Chapter 4
Future Scope and Conclusion

The current report highlights the individual components as hardware and software implementation of the Long term price prediction using the given data set. Moving further there can be 2 major improvisations-

- **Integrating the backend to send user data** - It is possible for the user to enter data on the website, the website will send the input data to the Rpi microprocessor for computation of the LT price. This would require an active wifi-module and communication setup efficiently between all the nodes in the network. The Rpi will give the final verdict and light up the LED further as well as display the numeric data on the frontend of the website.
- **API Fetching** - This will use High frequency trading in the current implementation, where just by typing the company's name, all the data is automatically fetched from a standard API, the data is then sent to the model for predicting the future stock prices. We can further extend this to use the current market trends, by subscribing to the license.

In conclusion, the project involving linear regression for predicting the Long-Term (LT) Price of Initial Public Offerings (IPOs) based on various features is significant in today's world for several reasons[7].

Firstly, accurate prediction of the LT Price of IPOs is crucial for investors, by leveraging machine learning techniques like linear regression, investors can make more informed decisions regarding whether to invest in an IPO or not, thus minimizing investment risks and maximizing returns.

Furthermore, the project highlights the importance of data-driven decision-making in the financial sector. With vast amounts of data available, companies can leverage machine learning models to extract valuable insights and patterns, enabling them to make smarter investment choices and stay competitive in today's dynamic market environment.

Overall, the project underscores the significance of applying machine learning techniques like linear regression in finance to improve decision-making processes, enhance investment strategies, and ultimately drive better outcomes in the current world of finance.

References

[1] S.Shukla and B.Shah's "Stock Price Prediction Using Multiple Linear Regression Models"International Journal of Engineering Science Invention (IJESI). Published in October 2018.

[2] N.Rouf and S.Singh's "Stock Price Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions" MDPI. Published in November 2021.

[3] V. Gururaj, S. V R and A.karthick's "Stock Price Prediction Using Linear Regression and Support Vector Machines".International Journal of Applied Engineering Research. Published in 2019.

[4] M.Vijh, D.Chandola's "Stock Closing Price Prediction using Machine Learning Techniques". International Conference on Computational Intelligence and Data Science (ICCIDS).

[5] K.Daniel's "A Study on Stock Price Prediction Using Linear Regression and Random Forest Regression" (2022).

[6] I.Kumar, C.Utreja's "A Comparative Study of Supervised Machine Learning Algorithms for Stock Market Trend Prediction". IEEE Conference. (2018).

[7] V.K. Sai Reddy's "Stock market prediction using machine learning". International Research Journal of Engineering and Technology (IRJET). Published in 2018.