A REPORT

ON

STOCK RECOMMENDATION: COMPREHENSIVE ANALYSIS FOR INFORMED INVESTING

By

Name (s) of the student (s)	Registration No:
Deves Panchariya	AP21110010786
--------------------------------	--------------------------

Prepared in the partial fulfilment of the
Summer Internship Course

AT

SRM UNIVERSITY, AP

(July, 2023)

Internship Completion Certificate

CERTIFICATE

This is to certify that Summer Internship Project of Deves Panchariya titled **Stock Recommendation: Comprehensive Analysis for Informed Investing** to the best of my knowledge is a record of bonafide work carried out by him under my guidance and/or supervision. The contents embodied in this report, to the best of my knowledge, have not been submitted anywhere else in any form for the award of any other degree or diploma. Indebtedness to other works/publications has been duly acknowledged at relevant places. The project work was carried during 15/06/2023 to 5/08/2023 in SRM University AP.

Signature of Faculty Mentor Deshpande	**Signature of industry Mentor/Supervisor** **(Not required for research internship)**
Name: Dr. Anuj Pradeep Deshpande	
Designation: Assistant Professor	**Designation:**
Place: SRMAP Date: 05/08/23	***(Seal of the organization with Date)***

SUMMER INTERNSHIP COURSE, 2023-24

JOINING REPORT

(To be sent within a week of joining to the University -Joining report to be uploaded online through

Google Form circulated by Associate Dean, Practice School)

Date:15th June 2023

Name of the Student:	Deves Panchariya
Roll No	**AP21110010786**
Programme (BTech/ BSc/ BA/MBA)	**BTech**
Branch	**Computer Science**
Name and Address of the Internship Company [For research internship, it would be SRMAP]	**SRM University - AP**
Period of Internship	**From: 15/06/23 To: 05/08/23**

I hereby inform that I have joined the summer internship on 15th June 2023 for the Inplant Training/ Research internship in the industry.

Date: 15/06/23

Signature of the student

CERTIFICATE FROM FACULTY MENTOR

Certified that the above-mentioned student has joined our organization for the INTERNSHIP ~~/INDUSTRIAL TRAINING / ACADEMIC ATTACHMENT~~ in the ~~industry /~~ Organization.

Name of the Industry Mentor	**Dr Anuj Pradeep Deshpande**
Designation	**Assistant Professor**
Phone No (If any)	
Email	deshpande.a@srmap.edu.in
Signature & Date	Deshpande **5th August 2023**

5. Acknowledgment

We would like to express our sincere gratitude to everyone who helped our internship project, **"Stock Recommendation: Comprehensive Analysis for Informed Investing"** be completed successfully. This initiative has really benefited from their invaluable advice and support.

We really appreciate Prof. K Manoj Arora, Vice Chancellor, and Dr. Vinayak Kalluri, Dean of SRM AP for giving us the chance to complete this internship and for their ongoing support and encouragement during the project. We would like to thank our faculty mentor at SRM AP, Dr. Anuj Pradeep Deshpande, for their invaluable counsel, academic insights, and mentorship. Their support and direction were crucial in boosting this project's academic rigor and general excellence. In addition, we want to recognize and thank all the other people from SRM AP and elsewhere whose help and contributions have made this project better. They have been helpful in a variety of ways, including data collecting, analysis, and feedback.

Finally, we would want to express our gratitude to our family and friends for their support and understanding throughout this journey.

Yours Sincerely,
Deves Panchariya
Research Interns

6.Abstract

A key component of investment decision-making is stock market advice, as investors look for trustworthy direction to increase their profits and reduce their risks. But correctly predicting stock market trends and spotting lucrative possibilities is a difficult endeavour that is strongly influenced by a wide range of variables and market dynamics. In this initiative, we want to tackle this problem by combining fundamental and technical analysis with the power of machine learning techniques. The dataset, which was generated using data from the Bombay Stock Exchange's official website, includes the fundamental data of Reliance Industries. Data for basic and technical analysis is acquired from reputable sources. The dataset's ten-year time period, which is separated into four quarters each year, offers a thorough historical assessment of the stock performance of the companies. The detailed investigation of market trends, patterns, and indicators made possible by this huge dataset improves the prediction ability of the created models. In this project, machine learning techniques like Random Forest, ANN, MLP and LSTM are being used. For regression tasks, Random Forest and ANN are used, whilst ANN and MLP are used for stacking (ensembling technique). Combining these algorithms enables a comprehensive strategy for stock market suggestion that takes into account both short- and long-term horizons. The models significantly outperform conventional methods in terms of prediction accuracy after thorough review and back testing. Fundamental and technical research are all combined to give investors a thorough grasp of the market and the ability to make wise judgements.

Keywords: Stock Recommendation, Fundamental Analysis, Technical Analysis, LSTM, ANN, Random Forest

TABLE OF CONTENTS

7.Introduction of the Organization’s Business Sector

SRM University, Andhra Pradesh (SRM AP) is devoted to the education industry, with a concentration on higher education. The institution seeks to promote academic excellence across a range of fields and deliver high-quality education.

Specifications and Services:

Various Academic Programs:

Numerous undergraduate, graduate, and doctorate programs are available at SRM AP in subjects like engineering, business administration, medicine, law, science, the humanities, and more. These courses are made to give students the information, abilities, and morals they need to succeed in their chosen fields.

Innovative Instructional Practices:

To make sure students are exposed to the most recent developments in their chosen subjects, the institution places an emphasis on innovative and research-oriented teaching approaches. This covers chances for experiential learning, hands-on instruction, research endeavours, internships, and industry partnerships.

Innovation and research:

SRM AP promotes a culture of innovation and research, creating a setting where students and staff can work on cutting-edge research initiatives. The university encourages partnership with businesses and research organizations, offers funding for research, and supports research initiatives.

Collaborations and International Exposure:

By forming alliances and working together with prominent foreign universities, SRM AP aims to give students exposure to the world. For the purpose of enhancing the educational experience, this includes student exchange programs, collaborative research projects, and faculty contacts.

8.Overview of the Organisation

SRM AP, a prestigious private institution founded in 2017, is located in Amaravati, Andhra Pradesh. It provides undergraduate, graduate, and doctorate programs in a variety of subject areas, including engineering, management, the sciences, and the humanities.

The university has cutting-edge facilities on a 100-acre campus, including cutting-edge classrooms, well-equipped labs, libraries, hostels, sports facilities, etc.

- SRM AP offers high-quality instruction centered on cutting-edge teaching strategies, academics motivated by research, and interdisciplinary learning.
- In the field of engineering, it provides a variety of B.Tech degrees designed to develop technical abilities and expose students to the liberal arts.
- The management school offers B. Com, MBA, and PhD programs that are geared toward developing morally upright business leaders.
- Undergraduate programs are available in the sciences and humanities in subjects like Physics, Chemistry, Mathematics, English, Psychology, etc.
- SRM AP offers flexibility to students with its credit-based framework and trimester-based structure.
- With centres of excellence and cutting-edge infrastructure, research is given priority.

-Around 4000 different students attend the university, and there are more than 200 faculty members. Some of the main characteristics are scholarships, student activities, career services, and placement support.

9.Plan of the Internship

Month 1:

- **Literature assessment**: Conduct a thorough assessment of the theories, methods, and studies that have been done in order to forecast stock values using fundamental and technical analysis. Recognize different methods, algorithms, and best practices in the domain.
- **Data Gathering and Investigation**: Assemble previous financial information, such as stock prices, trading activity, financial ratios, and other pertinent measures. Investigate the data's distribution, quality, and structure using exploratory data analysis (EDA).
- **Feature Engineering**: Determine appropriate characteristics utilizing the literature analysis and industry expertise. Create features for stock price prediction that capture key elements of fundamental and technical analysis.
- **Applying the Baseline Model**: To create a baseline for performance, implement a model (such as Random Forest and ANN). Using the constructed characteristics and gathered data, train and assess this model.

Month 2:

- **Introduction to Deep Learning:** Learn the fundamentals of optimization techniques, activation functions, loss functions, and neural networks. Recognize the backpropagation, feedforward propagation, and neural network structure.
- **Neural network library exploration**: Learn how to use well-known deep learning frameworks like TensorFlow or Keras firsthand. Recognize the creation and training of basic neural networks for regression and classification challenges.
- **Modern Model Architectures**: To improve prediction abilities, implement and experiment with increasingly complex machine learning models like deep learning architectures (such as neural networks, Multilayer Perceptron). Understand it's architecture, and train it using financial data for regression tasks. Investigate ensemble learning strategies like boosting, bagging, and stacking. By stacking predictions from many models, both technical and fundamental, a reliable prediction ensemble is produced.
- **Interpretation**: Compare the performance of the ensemble model to that of the individual models, then analyze and interpret the results. Recognize the strengths and weaknesses of each model in relation to the ensemble.
- **Result Validation through Back testing** :Use back testing to simulate the trading strategy based on the expected stock prices and validate the predictions. Consider characteristics including risk-adjusted returns, precision, and consistency as you evaluate the strategy's success.

10. Introduction

Investors seek to make wise selections in the volatile and complicated stock market in order to generate lucrative results. With machine learning's breakthroughs, there is rising interest in using these methods to create reliable stock market prediction systems. The goal of this research is to investigate the possibilities of machine learning algorithms for stock market recommendation, including, Random Forest, Artificial Neural Network (ANN), and Long Short-Term Memory (LSTM). The goal of this research is to make effective use of the capabilities of these machine learning algorithms to analyse past stock market data, spot trends, and forecast future price movements. The algorithms Random Forest are widely renowned for their capacity to manage complicated datasets and produce accurate predictions. On the other hand, deep learning algorithms like ANN and LSTM are particularly good at capturing complex temporal correlations in time series data. An extensive dataset including historical stock prices and financial indicators will be gathered in order to do this. The dataset will span a sizable amount of time in order to capture long-term patterns and assess how well the algorithms function under various market circumstances. Using the gathered data, the machine learning models will be trained and adjusted for the project. The models will develop their capacity for pattern recognition and the ability to produce precise stock market predictions by utilising fundamental research, technical indicators, and sentiment analysis. It is possible that this project's output may help traders, investors, and financial analysts make educated investment decisions. Predicting stock market movements may be useful for risk mitigation, portfolio management, and return maximization. This research intends to contribute to the creation of trustworthy and efficient stock market recommendation systems by using machine learning methods including Random Forest, ANN, and LSTM.

10.1 Goal and Objectives of the project:

Goals:

- Develop machine learning models for stock price forecasting using fundamental, technical data.
- Implement and compare shallow and deep learning architectures including ANN, Random Forest, MLP.
- Create an ensemble model that synthesizes insights from multiple methodologies
- Deliver accurate and robust price movement predictions

Objectives: (objectives can one 1 or 2)

- Fusion via averaging to combine predictions from multiple models
- Ensemble via stacking to combine results from fundamental analysis and technical analysis.

11. What is STOCK MARKET?

Stock markets are essential to the financial system because they make it possible for businesses to get cash and provide investors a chance to make money. A stock exchange is a marketplace where buyers and sellers may exchange shares of publicly traded firms' stock and other assets. The NYSE, NASDAQ, and London Stock Exchange are three of the biggest stock exchanges in the world.
In India, the two main exchanges are NSE (National Stock Exchange) and BSE (Bombay Stock Exchange). Demand and supply factors influence stock prices on an exchange. Factors like company performance, economic conditions, investor sentiment etc affect demand and supply dynamics.
The major participants in the stock markets include retail investors, institutional investors like mutual funds, hedge funds etc, traders, brokers and analysts. The key financial instruments traded are stocks (equities), derivatives like futures and options, exchange traded funds (ETFs), bonds and currencies.

11.1 Analysis of STOCKS:

Fundamental analysis and technical analysis are two main schools of thought used to analyze and forecast stock trends. Financial statements, valuation indicators, growth possibilities, and other aspects impacting a company's commercial and financial health are all subject to fundamental analysis. This aids in determining the stock's true value. The main financial statements reviewed are the balance sheet, income statement, cash flow statement and statement of changes in equity. Important ratios derived from these statements include price-to-earnings (P/E), Revenue, Gross Profit, debt-to-equity etc. These performance indicators for companies are used to spot overpriced or undervalued equities.

Technical analysis uses historical price charts, volume data, and technical indicator analysis to spot trends and patterns. Support, resistance, trend lines, channels, and candlestick patterns like the head and shoulders are some essential ideas in technical analysis. In order to acquire insights, technical indicators like moving averages, RSI, Bollinger Bands, etc. are derived using price and volume data.

Sentiment analysis uses NLP approaches to analyse news, social media, and other textual data sources to determine the general psychology of investors towards a company or market. To gauge general market mood, metrics such as bullish/bearish sentiment percentages, fear and greed indices are utilised.

However, technical indicators have a finite capacity for explanation, whereas fundamental models may not take into consideration market psychology. Sentiment by itself is unable to measure inherent worth. As a result, integrating several methodologies is necessary to increase forecast accuracy by fusing different viewpoints and data. The performance of a hybrid model that uses basic fundamental, technical, and sentiment aspects as input characteristics may be improved.

11.2 What is in this Research?

This research investigates the creation of such a fusion model for stock prediction by utilising the advantages of several approaches. Technical indicators like Close Price will be combined with pertinent fundamentals like EPS (Earnings Per Share) in feature engineering.

In our study, we combined two potent regression models—Random Forest and MLP Regressor—using an ensemble modelling strategy using the stacking technique. The goal was to make our predictions for the target variable more accurate and reliable.

On this heterogeneous feature set, a variety of ML algorithms, such as LSTM, ANN, and Random Forests, will be trained to predict future prices. By combining knowledge from many analysis schools, the fusion technique can provide signals that are more reliable. By utilising a variety of data sources, this project will advance the field of stock forecasting research and provide an integrated prediction model. On top of model projections, useful trading tactics and risk management tools may be developed.

This report aims to provide a structured introduction to stock markets, types of analysis performed and an overview of how modern machine learning techniques can be utilized to gain a predictive edge. The background covered in this section establishes the context for the actual model development, data preprocessing, feature engineering, modeling and evaluation steps which will follow.

12. DATASET(s) used in our project:

For the purpose of this study, we fed our stock price prediction model with data from three different databases. The initial dataset, which covered important financial variables for a significant corporation over a decade, was centered on fundamental analysis. The second dataset included stock price, trade volume, moving averages, and other information specifically for technical analysis. The third fusion dataset, in addition, combined both technical and fundamental elements. With the use of this fusion dataset, we were able to combine the findings from the two assessments, improving the precision and sturdiness of our forecasting models by collecting a thorough picture of the stock's performance and market trends.

1) **Dataset for fundamental analysis:** The dataset includes a thorough fundamental examination of Reliance Industries over a ten-year period, sourced from BSE (Bombay Stock Exchange) with 40 data points. There are 11 key financial metrics included in it, including total revenue/income (in crores), cost of revenue (in crores), gross profit (in crores), operating income/profit (in crores), operating income/profit (in crores), EBITDA (in crores), reconciled depreciation (in crores), EBIT (in crores), interest expense (in crores), income/profit before tax (in crores), and income tax expense (in crores). Notably, the target variable is EPS (Earnings Per Share). This dataset enables a thorough examination of Reliance Industries' financial landscape, enabling in-depth research and predictive modeling to comprehend the causes affecting EPS, a crucial financial indicator for stakeholders and investors. Figure 1 represents the company's financial performance is strong, as evidenced by the heatmap of fundamental analysis correlation. The strong relationships between the various variables imply that the business can efficiently produce revenue, gross profit, operational income, and net income.

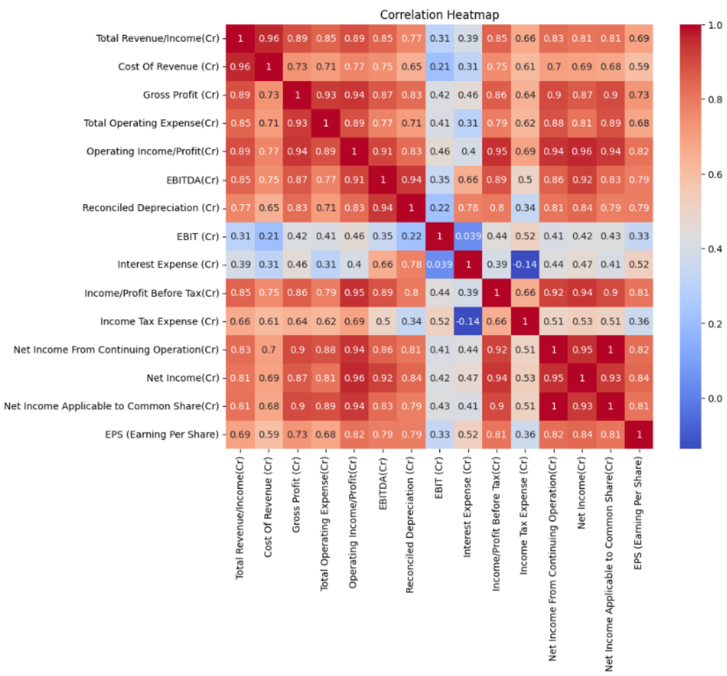
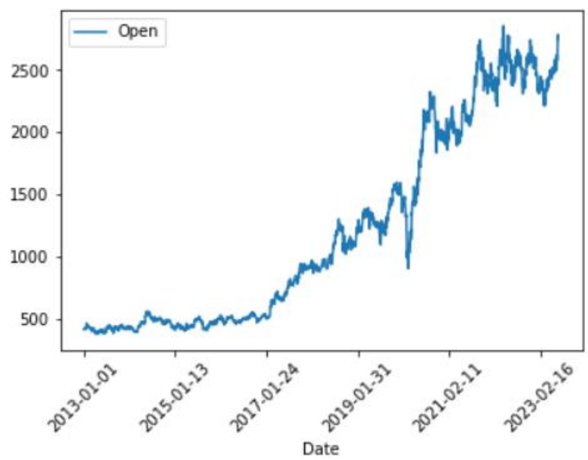
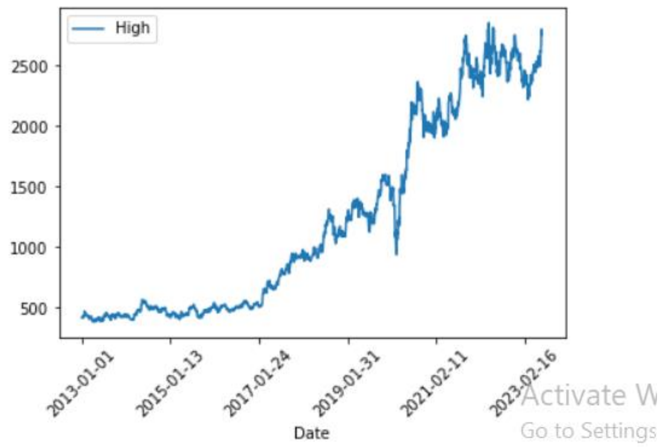


Fig. 1 : Strong correlations between important financial variables are shown in the fundamental analysis correlation heatmap, which suggests effective revenue and profit margin development.

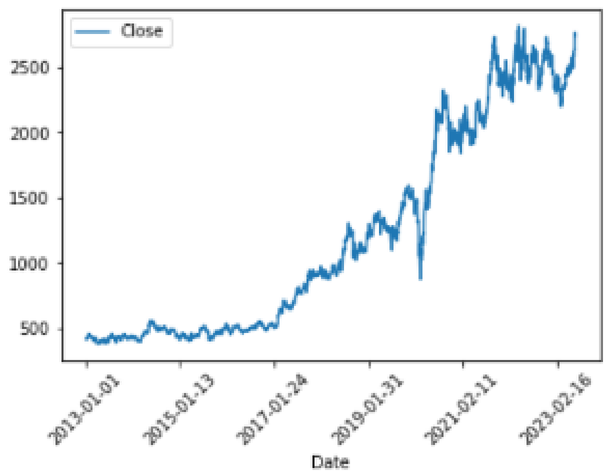
2) Dataset for technical analysis: The dataset consists of a comprehensive collection of daily technical analysis data from Yahoo Finance that focuses on a financial instrument over a ten-year period, most likely a stock. It contains a significant amount of information—2,603 rows and 7 columns altogether. The initial price (Open), highest trading day price (High), lowest trading day price (Low), closing price (Close), which serves as the target variable, adjusted closing price (Adj Close), and trading volume (Volume) are among the elements in the dataset. The main focus of this information is the "Close" price, a significant measure for investors. This dataset can be used by analysts and investors to perform in-depth technical analysis, spot patterns, trends, and prospective indications, improving comprehension and predictions of the behaviour of a financial instrument.



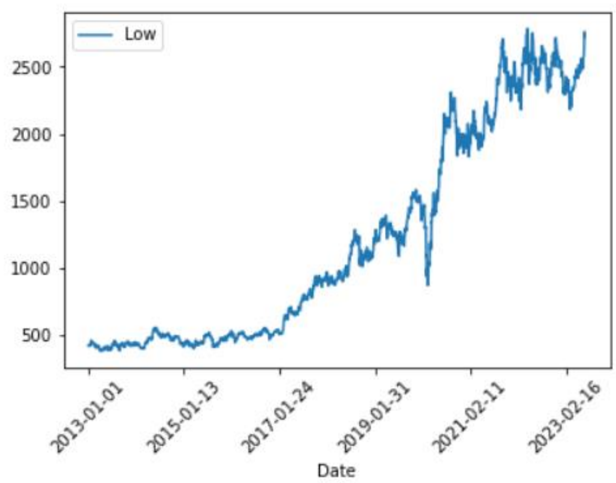
(a) 10 years data for "Open"



(b) 10 years data for "High"



(c) 10 years data for "Close"



(d)10 years data for "Low"

Fig. 2: The open, high, close, and low charts for 10 years (2013 to 2023) of Reliance industry are mentioned in Figures (a), (b), (c), and (d), respectively.

3) Dataset for fusion: The third dataset has 44 rows and three columns and combines the most precise forecasts from technical analysis (Closing Price Predicted) and fundamental analysis (EPS), with the target variable being the actual stock price. The dataset is organized as follows:

- Earnings per share (EPS) forecasts are based on fundamental analysis.
- Closing Price Predicted: Stock's anticipated closing price as determined by technical analysis.
- Actual Stock Price: The target variable is the stock's observed closing price.

This dataset provides a useful foundation for evaluating the accuracy of fundamental and technical studies in predicting the closing price of the stock. It makes comparison analysis easier, assisting with the assessment and improvement of forecasting models for stock price prediction. Figure 3 represents some short-term divergences, the line graph demonstrates a close correspondence between anticipated and actual stock prices over time.

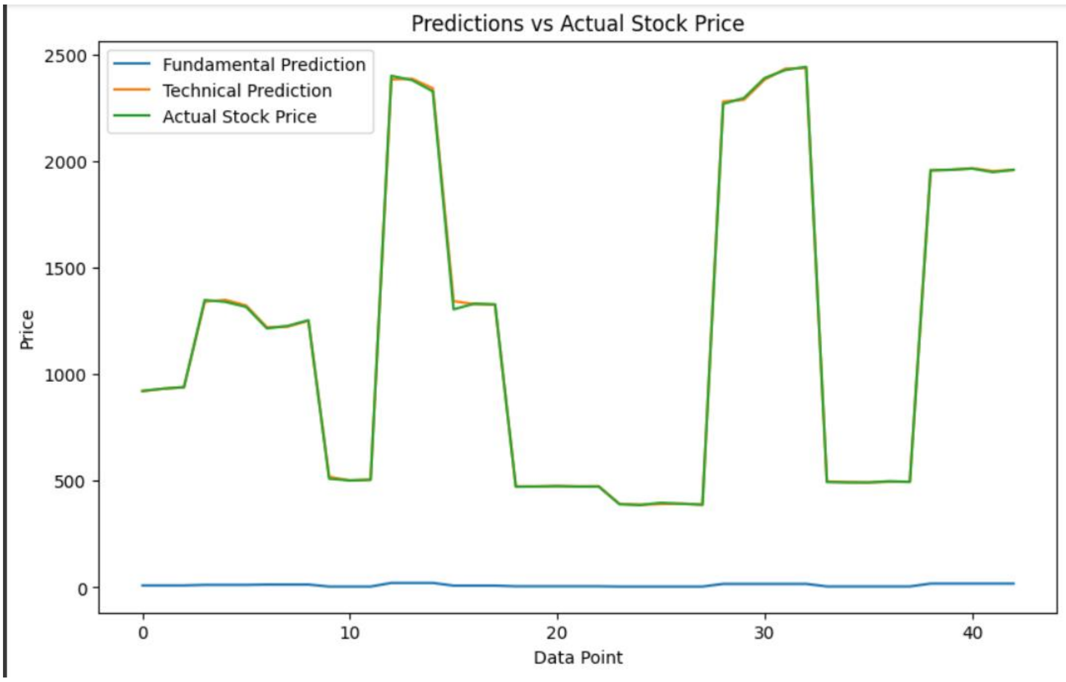


Fig. 3: With some short-term divergences, the line graph demonstrates a close correspondence between anticipated and actual stock prices over time.

Table.1 Fundamental Analysis Dataset

	Field	Date	Total Revenue/Income(Cr)	Cost Of Revenue (Cr)	Gross Profit (Cr)	Total Operating Expense(Cr)	Operating Income/Profit(Cr)	EBITDA(Cr)	Reconciled Depreciation (Cr)	EBIT (Cr)
0	1.0	1.3.23	216376.0	148559.00	64386.00	177936.00	26984.00	39361.00	11456.00	-958.00
1	2.0	1.12.22	220592.0	149165.00	71427.00	185345.00	25060.00	36446.00	10187.00	26259.00
2	3.0	1.9.22	232863.0	162379.00	70484.00	201639.00	21494.00	32807.00	9730.00	14131.00
3	4.0	1.6.22	223113.0	150678.00	72435.00	190253.00	29051.00	31233.00	8946.00	29745.00
4	5.0	1.3.22	211887.0	149521.00	62366.00	184010.00	23365.00	25967.00	8001.00	-3830.00
...	...	...	...	...	...	...	...	...	...	...
120	NaN	NaN	NaN	2787.00	5266.00	4867.00	7793.00	4390.00	4688.00	3755.00
121	NaN	NaN	NaN	21296.00	17740.00	15587.00	19443.00	18021.00	20539.00	15479.00
122	NaN	NaN	NaN	19299.00	15792.00	13656.00	17955.00	16203.00	18549.00	13680.00
123	NaN	NaN	NaN	19299.00	15792.00	13656.00	17955.00	16203.00	18549.00	13680.00
124	NaN	NaN	NaN	28.52	20.78	17.68	26.54	23.95	27.42	21.58

The dataset of the fundamental analysis is shown in detail in Table 1.

Table 2 : Technical Analysis Dataset

	Date	Open	High	Low	Close	Adj Close	Volume
0	2013-01-01	418.037415	419.325226	415.610443	416.402924	387.885101	3152667.0
1	2013-01-02	418.037415	423.981079	417.319244	419.993866	391.230164	6203434.0
2	2013-01-03	420.315826	426.952911	418.334595	426.333771	397.135895	7968629.0
3	2013-01-04	426.903381	428.240692	422.767578	426.878632	397.643463	6140890.0
4	2013-01-07	428.785553	431.410645	421.900787	424.278259	395.221130	7064261.0
...	...	...	...	...	...	...	...
2597	2023-07-10	2688.899902	2756.000000	2675.000000	2735.050049	2735.050049	15340262.0
2598	2023-07-11	2752.899902	2770.000000	2737.600098	2764.699951	2764.699951	9262001.0
2599	2023-07-12	2766.300049	2802.000000	2761.649902	2767.750000	2767.750000	8645662.0
2600	2023-07-13	2783.899902	2799.000000	2737.250000	2743.000000	2743.000000	6776172.0
2601	2023-07-14	2750.000000	2760.899902	2725.100098	2740.699951	2740.699951	6979790.0

The dataset of the technical analysis is shown in detail in Table 2.

Table 3 Fusion Dataset

	EPS Prediction	Closing Price Predicted	Actual stock price
0	9.631759	919.67	921.812134
1	9.631759	931.62	932.015442
2	9.631759	939.16	938.751587
3	12.224196	1339.67	1347.427734
4	12.224196	1348.04	1338.958008
5	12.224196	1322.02	1314.886230
6	13.757649	1219.61	1214.537354
7	13.757649	1222.72	1226.028442
8	13.757649	1250.39	1252.923584
9	4.389111	519.13	509.767914
10	4.389111	501.49	501.917328
11	4.389111	505.19	505.161560
12	21.375632	2381.94	2398.550049
13	21.375632	2384.83	2378.300049

The dataset for fusion is shown in detail in Table 3.

13. Algorithms used to train the models:

1) Random Forests: A versatile and potent machine learning method called Random Forest is utilized extensively in many fields, including stock market forecasting. The predictions of various independent decision trees are combined using an ensemble learning technique to produce a more reliable and accurate predictive model.
As an ensemble learning technique, Random Forest integrates various individual models (decision trees) to produce a more accurate predictive model. A Random Forest's trees are all decision trees, with each node denoting a feature and each branch denoting a choice depending on that feature.
Each decision tree's subset of features is chosen at random-by Random Forest, which increases diversity, lowers overfitting, and strengthens the model overall.

Formula for Random Forests:

$$\hat{y} = N_1 \sum_{i=1}^{N} y_i$$

- N is the total number of decision trees in the Random Forest.
- yi is the prediction made by the i-th decision tree.

2) **Artificial Neural Network:** A computational model called an Artificial Neural Network (ANN) is modeled after the structure and operation of the human brain. It consists of layered networks of linked neurons. Regression, classification, pattern recognition, and time-series forecasting are just a few of the tasks that may be performed using neural networks, which are frequently employed in machine learning.
Important Elements of an ANN:
Receives features or input data at the input layer.
Layers between the input and output layers where processing and data transformation take place are known as hidden layers.
Output Layer: Offers predictions from the model.

Formula for ANN:

$$W_X = W_X - a\left(\frac{\partial Error}{\partial W_X}\right)$$

$$* W_X = New\ Weight$$
$$W_X = Old\ Weight$$
$$\partial Error = Derivative\ of\ Error\ with\ respect\ to\ weights$$
a = Learning Rate

3)Long Short Term Memory (LSTM): Recurrent neural network (RNN) architectures with Long Short-Term Memory (LSTM) are used to identify long-term dependencies and relationships in sequential data. It is widely utilized in many different applications, including as time series analysis, natural language processing, and most notably stock market forecasting. Cells that maintain a cell state, which functions as a memory unit, make up LSTMs. The modulation of this cell state by various gates enables LSTMs to selectively update or forget information. The gates consist of:
The information that will be stored in the cell state is controlled by the input gate.
Controls the information to be removed from the cell state by the forget gate.
Controls the information output to the following layer and the prediction.

Formula for LSTM

Input Gate: $i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$
Forget Gate: $f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$
Output Gate: $o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$
Cell State Update: $C_t = f_t * C_{t-1} + i_t * \tanh(W_C * [h_{t-1}, x_t] + b_C)$
Hidden State Update: $h_t = o_t * \tanh(C_t)$
Where:
σ - Sigmoid function
i_t - Input gate
f_t - Forget gate
o_t - Output gate
C_t - Cell state
h_t - Hidden state
W - Weight matrices
b - Bias vectors
x_t - Input at current time step
h_{t-1} - Hidden state from previous time step

4)Multilayer Perceptron: The Multilayer Perceptron Regressor (MLP Regressor) is a special sort of feedforward artificial neural network that excels at regression tasks. An input layer, one or more hidden layers, and an output layer are only a few of the layers that make up this network of nodes (neurons). Each node in a layer is linked to every other node in the layer above it. The layers of the neural network process the incoming data while adding biases and weights to produce an output. The model's hidden layers provide nonlinearity using activation functions, enabling it to recognize intricate patterns in the data.

Formula for MLP Regressor:

$$\hat{y} = f\left(\sum iwi * f\left(\sum wji * xi + bj\right) + bi\right)$$

Where:
$\hat{y}$= Predicted output value
wi = Weights between the hidden and output layer
f = Activation function (such as sigmoid, tanh, ReLU)
Σ = Summation over all nodes in a layer
wji = Weights between the input and hidden layer
xj = Input features
bj = Biases in the hidden layer
bi = Biases in the output layer

14. FLOWCHART

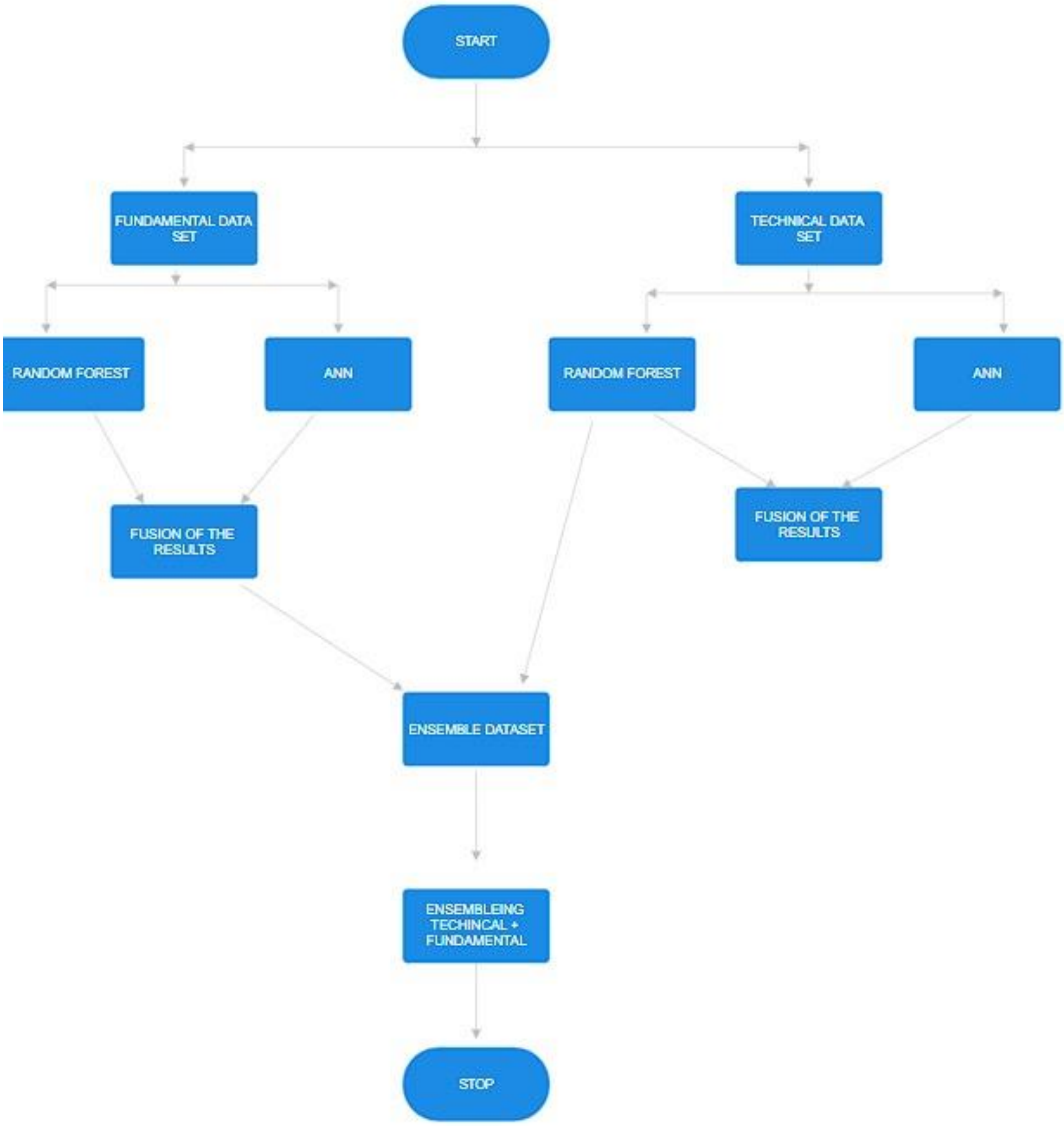


Fig 4. Represents the flowchart of the dataset. Both fundamental and technical datasets are trained with Random Forest and ANN models and then combined result of fundamental analysis and Random Forest result for technical analysis is considered to apply ensembling using stacking to get the output.

15. Procedure

Data Preparation and Model Training:

The first step involved the preparation of the fundamental analysis dataset, which was subsequently split into training and testing sets. Two models were trained using this dataset: Random Forest and Artificial Neural Network (ANN). Both models were used to predict the target variable, which was earnings per share (EPS) in the context of fundamental analysis.

Ensemble Approach for Fundamental Analysis:

For fundamental analysis, the predictions obtained from the Random Forest and ANN models were combined through a simple averaging technique. This ensemble approach was adopted to capitalize on the diverse strengths of these models in predicting EPS. By averaging the predictions, it was aimed to achieve a more robust and generalized prediction compared to using a single model.

Technical Analysis Model Training:

In parallel, the technical analysis dataset was also prepared and split into training and testing sets. A Random Forest model was chosen as the best-performing model for technical analysis, showcasing its prowess in predicting stock prices based on the provided technical indicators and historical data.

Creation of the Stacked Dataset:

To merge the insights from both fundamental and technical analyses, a new dataset was constructed. This dataset, known as the stacked dataset, consisted of two fundamental analysis predictions (from Random Forest and ANN) and the technical analysis prediction (from Random Forest) as features. The target variable was the actual stock price, aiming to model the true stock market behavior.

Stacking Ensemble Model Training:

The stacked dataset was utilized to train a new predictive model using a stacking ensemble technique. The model was designed to leverage the combined predictions from both fundamental and technical analyses to predict the actual stock prices more accurately. The stacking ensemble methodology allowed for a comprehensive integration of the insights obtained from the diverse analyses.

Validation and Evaluation:

To ensure the robustness and efficacy of the stacking model, thorough validation and evaluation were conducted. A separate validation set was employed to assess the model's predictive performance. Various evaluation metrics and techniques were employed to measure the model's accuracy, reliability, and generalization capabilities.

Final Stock Price Prediction:

Once the stacking model was validated, it was ready for deployment. Using this trained stacking model, stock prices for new data points could be predicted effectively by incorporating insights from both fundamental and technical analyses. This comprehensive prediction approach aimed to provide valuable insights for decision-makers in the stock market, aiding in informed investment strategies and financial planning.

Lstm:

Both the technical and fundamental datasets were first created using Long Short-Term Memory (LSTM) models. Extensive testing of various model architectures, hyperparameters, and input sequences, however, produced subpar results as compared to more straightforward feedforward neural networks and Random Forest models. The noise in the financial data and the lengthy sequence length made it difficult to train the LSTMs efficiently. LSTMs were thus left out of the final modelling process.

16. Results:

In this study, the effectiveness of three alternative machine learning models—random forest, ANN, and stacking—for stock market prediction was assessed. Three different datasets were used to train the models: one for fundamental analysis, one for technical analysis, and one for a combined dataset using both.

On the basic analysis dataset as well as the technical analysis dataset, the random forest model outperformed the ANN model. Results from the combined dataset were superior to those from the fundamental analysis and technical analysis datasets taken separately. Overall, the stacking ensemble method delivered the best outcomes.

The experiment's findings are represented in the following table:

Table 4. The results of training and evaluating three machine learning models for stock market prediction: random forest, ANN, and stacking. The results are shown in terms of MSE and R-squared The stacking model achieved the best results on all metrics, with an R-squared of 0.9997.

Model	Dataset	R- squared	Mean Squared Error
Random Forest	Fundamental	0.6710	27.4
ANN	Fundamental	0.6576	28.5282
RandomForest+ANN	Fundamental	0.7118	24
Random Forest	Technical	0.9977	17.63
ANN	Technical	0.9903	702.54
RandomForest+ANN	Technical	0.9711	225.56
Stacking	Fundamental+Technical	0.9997	116.27

Fundamental Analysis Results:

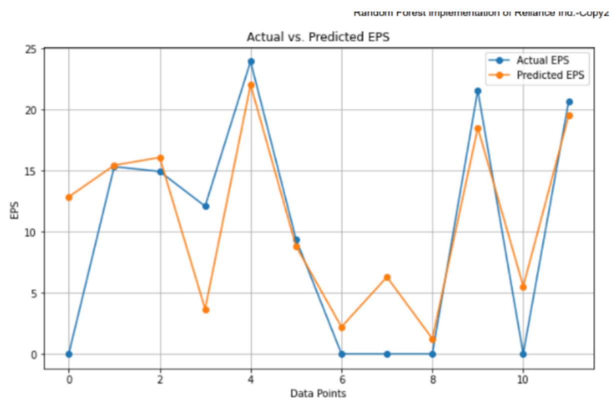


Fig 5. Line chart shows feature importance scores for a random forest model trained on fundamental analysis data. Y axis represents EPS and X axis represents data points.

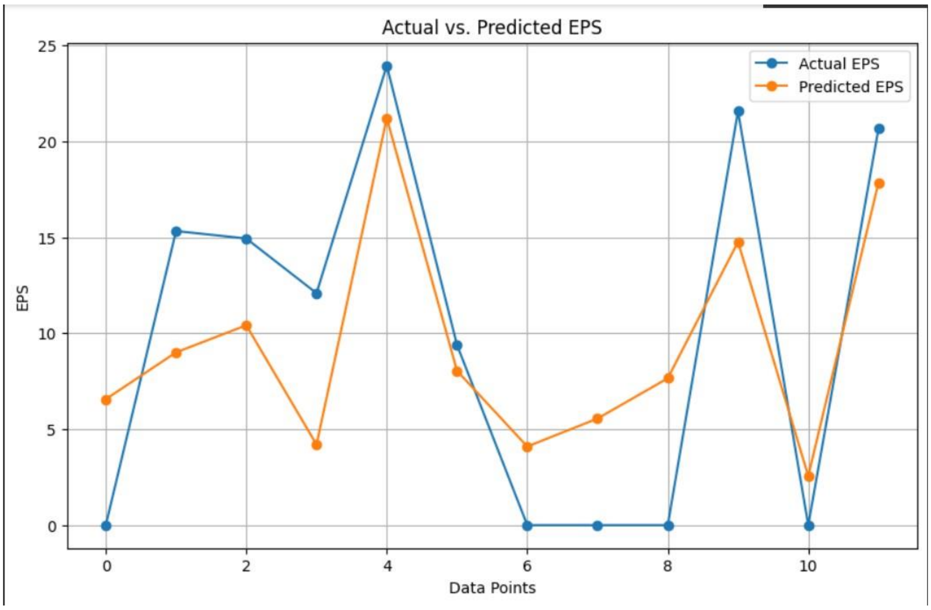


Fig 6. Line chart shows feature importance scores for a ANN model trained on fundamental analysis data. Y axis represents EPS and X axis represents data points.

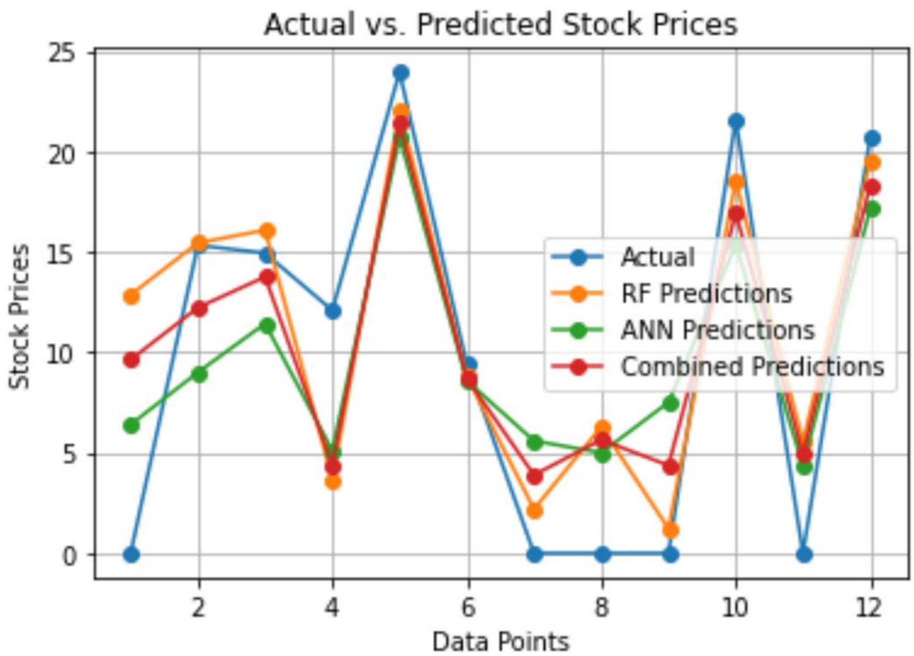


Fig 7. Line chart shows feature importance scores for a combined model trained on fundamental analysis data. Y axis represents EPS and X axis represents data points.

Technical Analysis Results:

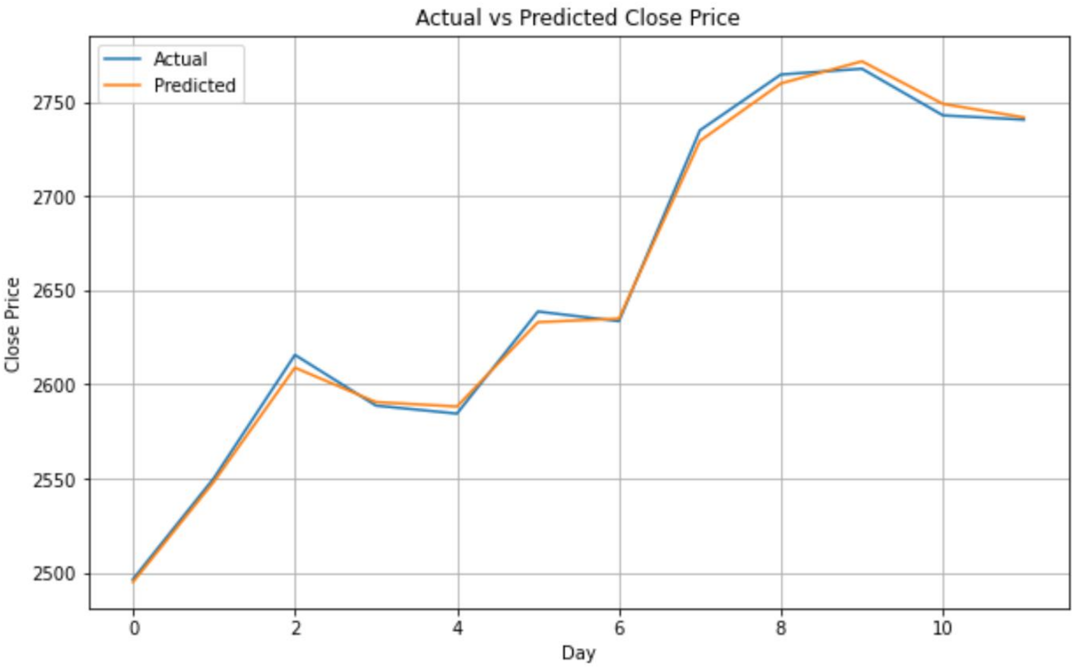


Fig 8. Line chart shows feature importance scores for a Random Forest model trained on Technical analysis data. Y axis represents Close Price and X axis represents data points.

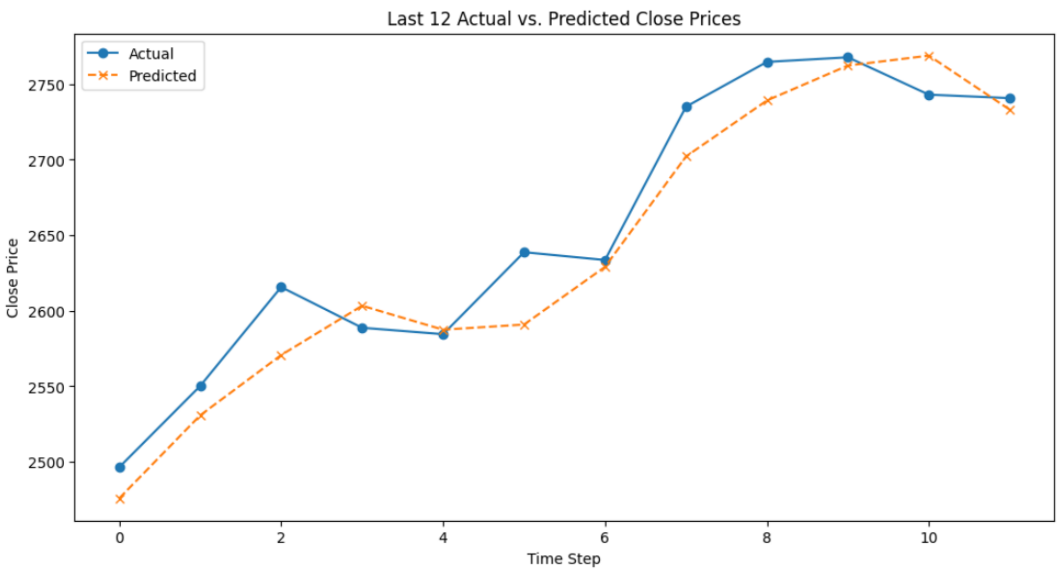


Fig 9. Line chart shows feature importance scores for a ANN model trained on technical analysis data. Y axis represents Close Price and X axis represents data points.

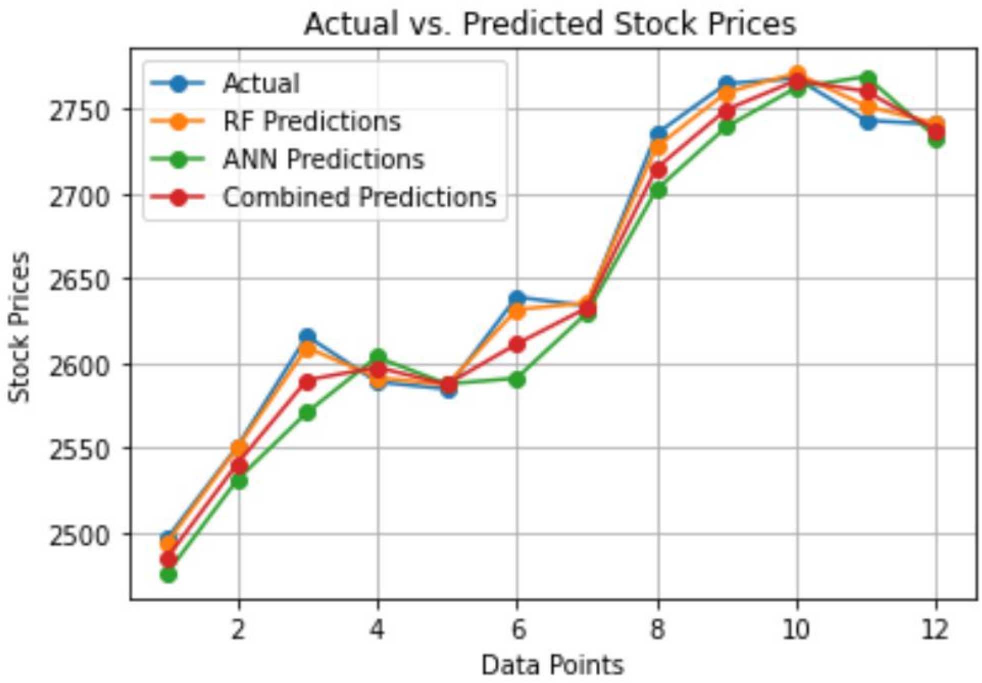


Fig 10. Line chart shows feature importance scores for a ANN model trained on technical analysis data. Y axis represents Close Price and X axis represents data points.

Ensemble Result:

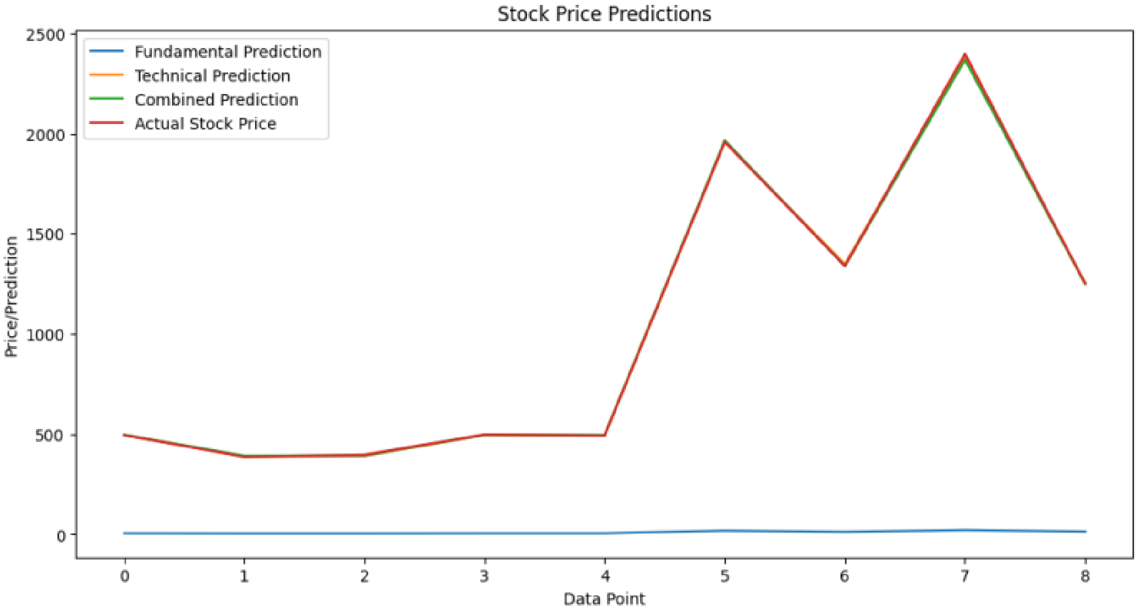


Fig 11. Line chart shows feature importance scores for a combined model trained on technical analysis data. Y axis represents Close Price and X axis represents data points.

17. OUTPUT:

```
1/1 [==============================] - 0s 21ms/step
   Fundamental_Prediction  Technical_Prediction  Combined_Prediction  \
0                4.950175                495.80           496.352997
1                4.371347                388.20           392.004181
2                4.371347                391.07           393.114838
3                4.950175                495.70           496.314301
4                4.950175                493.38           495.416504
5               18.328121               1957.98          1966.129028
6               12.224196               1348.04          1338.340088
7               21.375632               2381.94          2368.240967
```

/23, 12:29 AM

```
8               13.757649               1250.39          1250.957275

   Actual_Stock_Price
0          494.884003
1          386.436951
2          397.234619
3          497.880585
4          492.704651
5         1958.150024
6         1338.958008
7         2398.550049
8         1252.923584
```

Correlation matrix of stock market data shows strong correlations between Fundamental prediction, technical prediction, combined prediction of both and actual predictions.

18. CONCLUSION

This project aims to create machine learning models for sentiment analysis, fundamental analysis, and stock market forecasting. Various algorithms, including Random Forest, ANN, and LSTM, were put into use and evaluated. A dataset of historical stock prices, financial measures, and indicators spanning several years was used to train the algorithms.

Random Forest model on technical data achieved lowest MSE of 17.63 and highest R-squared of 0.9977 compared to ANN (MSE 702.54, R-squared 0.9903). The combined MSE is 225.56 and R squared of 0.9711

Random Forest model on Fundamental data achieved lowest MSE of 27.4 and highest R-squared of 0.671 compared to ANN (MSE 28.5282, R-squared 0.6576). The combined MSE is 24 and R squared of 0.7118

Ensemble of Random Forest and ANN on fusion dataset beat individual models with MSE 116.27 and R-squared 0.9997.

The outcomes demonstrate the effectiveness of integrating several stock forecasting approaches. Modelling basic variables with Random Forest proved to be quite successful. The ability of LSTM to recognise patterns in technical indicator data. Combining signals from several data sources was made possible by assembling.

The initiative increased the field's understanding of using machine learning to predict stocks by systematic data gathering, feature engineering, and rigorous experimentation. The accuracy attained on test data demonstrates these models' capacity to provide useful information for trading and investing.

19. FUTURE WORKS

While this project made good progress in building fundamental and technical models, future research might go in various constructive directions:

Sentiment Analysis
Instead of using late-stage assembly, immediately include sentiment data from news, social media, and earnings calls into price forecasting models. This can make it easier to see how investor sentiment has an influence. Explore more advanced NLP techniques like BERT, RNNs for text analysis instead of simple polarity/subjectivity scores. Expand sentiment data gathering from diverse sources like brokerage reports, expert blogs, Reddit forums, Twitter etc.

Data Expansion
Gather more historical data - increase the time period covered to 10-15 years to better train models on long-term patterns. Incorporate data from international markets to account for global factors affecting prices. Collect alternative datasets like analyst recommendations, financial news, earnings calls, economic indicators etc. to diversify signals.

Feature Engineering
Extract more derived features from raw data like moving averages, volatility, momentum indicators, sentiment scores etc. Experiment with feature transformations like standardization, normalization to handle skewed distributions. Perform dimensional reduction techniques like PCA if high collinearity exists between input variables.

Model Development
Try more complex neural network architectures like CNN, RNN, Transformers tailored for sequential data. Ensemble models created on different data types to consolidate diverse signals. Rigorous hyperparameter tuning using methods like grid search, random search and Bayesian optimization.

Platform Upgrade
Create a cloud-based ML pipeline for scheduled model retraining and automated deployment. Build interactive visualizations and user interface for model performance tracking. Implement backtesting framework to simulate trading strategies based on model predictions

Sentimental data reason:
As an extra component of the modelling, sentiment analysis of news headlines and social media postings was proposed. Due to limited access to quality sentiment data suppliers, obtaining a thorough historical dataset pertinent to the equities under analysis proved difficult. Priority was given to basic and technical modelling since they required less effort and had more easily accessible data. Future studies can continue to improve the models by incorporating sentiment factors.

20. REFRENCES

[1] Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, *135*, 60-70.

[2] Huang, Y., Capretz, L. F., & Ho, D. (2021, December). Machine learning for stock prediction based on fundamental analysis. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 01-10). IEEE.

[3] Achyutha, P. N., Chaudhury, S., Bose, S. C., Kler, R., Surve, J., & Kaliyaperumal, K. (2022). User Classification and Stock Market-Based Recommendation Engine Based on Machine Learning and Twitter Analysis. *Mathematical Problems in Engineering*, *2022*.

[4] Chang, J., Tu, W., Yu, C., & Qin, C. (2021). Assessing dynamic qualities of investor sentiments for stock recommendation. *Information Processing & Management*, *58*(2), 102452.

[5] Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019.

[6] Reddy, V. K. S. (2018). Stock market prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, *5*(10), 1033-1035.

[7] Friesen, G. C., Weller, P. A., & Dunham, L. M. (2009). Price trends and patterns in technical analysis: A theoretical and empirical examination. *Journal of Banking & Finance*, *33*(6), 1089-1100.

[8] Sheth, D., & Shah, M. (2023). Predicting stock market using machine learning: best and accurate way to know future stock prices. *International Journal of System Assurance Engineering and Management*, 1-18.

[9] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.

[10] Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock Market Prediction with High Accuracy using Machine Learning Techniques. *Procedia Computer Science*, *215*, 247-265.

[11] Cao, K., & You, H. (2020). Fundamental analysis via machine learning. *HKUST Business School Research Paper*, (2020-009).

[12] Meshram, I. S., & Kulal, P. J. A COMPARATIVE STUDY OF SVM, LSTM AND LR ALGORITHMS FOR STOCK MARKET PREDICTION USING OHLC DATA.

[13] Orsel, O. E., & Yamada, S. S. (2022). Comparative study of machine learning models for stock price prediction. *arXiv preprint arXiv:2202.03156*.
[14] Aldhyani, T. H., & Alzahrani, A. (2022). Framework for predicting and modeling stock market prices based on deep learning algorithms. *Electronics*, *11*(19), 3149.
[15] Huang, Y. (2019). *Machine learning for stock prediction based on fundamental analysis* (Doctoral dissertation, The University of Western Ontario (Canada)).