A

Mini-Project Report on

“Market Trend Analysis using Machine Learning”

Submitted in partial fulfilment of the requirements for the degree of

BACHELOR OF ENGINEERING

By-

Sidra Khan (22106028)

Saif Khan (22106072)

Tulsi Dubey (22106055)

IN

Computer Science & Engineering (Artificial Intelligence & Machine Learning)

Under the guidance of

Prof. Shraddha Dalvi

Department of Computer Science & Machine Learning

(Artificial Intelligence & Machine Learning)

A.P. Shah Institute of Technology

G.B. Road, Kasarvadavali, Thane (W) – 400615

University of Mumbai 2024 – 2025

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "**MARKET TREND ANALYSIS USING MACHINE LEARNING"** is a bonafide work of Sidra Khan (22106028), Saif Khan (22106072), Tulsi Dubey (22106055) (submitted to the University of Mumbai in partial fulfilment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

_____	_____
Prof. Shraddha Dalvi	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A.P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini Project Report entitled "**MARKET TREND ANALYSIS USING MACHINE LEARNING**" by **Sidra Khan (22106028), Saif Khan (22106072) and Tulsi Dubey (22106055)** is approved for the degree of ***Bachelor of Engineering in Computer Science & Engineering in Computer Science & Engineering (Artificial Intelligence & Machine Learning) 2024 – 2025.***

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane

Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Sidra Khan	Saif Khan	Tulsi Dubey
(22106048)	(22106072)	(22106055)

ABSTRACT

The stock market is one of the most critical pillars supporting a nation's economy, providing companies with a platform to raise significant capital through Initial Public Offerings (IPOs). For investors, IPOs offer an opportunity to purchase newly issued stocks and either become shareholders, benefiting from dividends, or traders, aiming to profit from market fluctuations. A trader's ability to predict future stock price trends can yield substantial financial gains. However, the stock market is notoriously volatile, with prices subject to immediate changes due to various factors such as political developments, company performance, and unforeseen global events. This unpredictability presents significant challenges for investors trying to forecast price movements accurately.

Traditionally, investors have relied on fundamental and technical analyses to inform their trading decisions. Fundamental analysis involves evaluating a company's performance by reviewing quarterly earnings reports and monitoring critical news events to avoid high-risk stocks. However, in recent years, the volume and speed at which financial information is disseminated have overwhelmed investors' ability to process the data effectively. The exponential increase in news outlets, combined with the rapid release of corporate and market reports, has made it nearly impossible for investors to assess all available information manually before making a stock purchase or sale decision.

This project, titled **"Market Trend Analysis Using Machine Learning,"** aims to address this challenge by developing an automated decision support system that leverages machine learning algorithms to predict short-term stock market trends. The system is designed to analyze a company's historical stock data, financial reports, and real-time news updates to provide accurate forecasts that help investors make informed decisions. A key component of this system is the integration of natural language processing (NLP) and sentiment analysis. Since market sentiment plays a crucial role in influencing stock prices, understanding the sentiment behind news articles whether positive, neutral, or negative can provide valuable insights into future price movements.

INDEX

CHAPTER 1
INTRODUCTION

1. INTRODUCTION

The stock market is one of the most important components forming a country's economy. Through IPO – Initial Public Offering, a company is able to raise a substantial amount of money to expand businesses. It is a great opportunity for investors to buy a brand-new stock and become either a stockholder who gets extra benefit from dividends from the firm's shareholder bonus program or a trader who trades stock in the stock market. If the stock trader predicted the stock price trends correctly, he would gain enormous profits. However, the stock market is volatile , daily news events such as developing political situations, the company's performance and other unexpected events affect stock prices immediately in a positive or negative way.

As the result, it is impossible to predict the stock prices and their directions (increase, decrease) accurately, instead investor is only to forecast the upcoming short-term trends. He usually evaluates a company's performance before making the decision to buy stock. The evaluation includes analyzing a company quarterly earnings report and paying attention to the important news to avoid buying overrated or high-risk stocks. However, both the speed of release and the number of daily news outlets have skyrocketed over the last few years which overwhelm investor's ability to thoroughly assess such a huge volume of data. As the result, an auto mated decision support system is essential as it automatically evaluates and shows the prediction for the upcoming stock trends. For example, if the price of the potential stock was predicted to be ‘‘going up’’ tomorrow, investors could either sell the stocks they held at a higher price or wait for the price drops and buy more. Thus, which algorithm is more effective and how to analyze the financial news to increase profits have drawn much interest from the research community.

Previous studies on this topic were divided into two main approaches: the technical analysis and the fundamental analysis. In the technical analysis, mathematics has been widely used to analyze historical stock price patterns and predict stock prices in the near future. Researchers have applied many algorithms such as multiple kernel learning, deep learning, stepwise regression analysis, etc. Although they achieved good results, it is impossible to predict the stock prices accurately by using only the historical prices because unexpected events can affect the stock prices immediately.

After that, machine learning algorithms are implemented to learn the connection between word patterns and stock prices movements. Although bag of words based approaches have achieved high accuracy, they ignored one crucial element of the directional predictions, which is the sentiment of the article. As shown in Fig. 2.1, one important stage is that the published news articles are interpreted by investors and converted into sentiments (positive, negative); the investors then decide whether to sell/hold/buy stocks based on the sentiment. Market prices aggregate the actions of each investor and them in the price trends. Therefore, combining the sentiment analysis and natural language processing would become more effective.

CHAPTER 2
LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 HISTORY

Here is a comprehensive literature history on market trend analysis, with a focus on the development leading up to and including the paper titled "Deep Learning Approach for ShortTerm Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network".

Market trend analysis has evolved considerably over time, starting with traditional statistical methods and gradually incorporating advanced machine learning and deep learning techniques. Early approaches in market trend analysis relied heavily on statistical models such as the Autoregressive Integrated Moving Average (ARIMA) model, developed by Box and Jenkins in 1976. ARIMA models were foundational in forecasting time-series data by capturing temporal patterns through autoregressive and moving average components. Another significant statistical method was Exponential Smoothing State Space Models (ETS), extensively discussed by Hyndman and Athanasopoulos in 2018. ETS models extend simple exponential smoothing by incorporating error, trend, and seasonality, offering a robust approach to capturing various time-series components.

Technical analysis also played a crucial role in market forecasting, utilizing historical price data and trading volumes to predict future price movements. Techniques such as moving averages and the Relative Strength Index (RSI), introduced by Wilder in 1978, became standard tools for identifying market trends and trading signals. Moving averages smooth out price data to highlight trends, while RSI measures the speed and change of price movements to identify overbought or oversold conditions.

The advent of machine learning brought new methods to market trend analysis, enhancing the predictive capabilities beyond traditional statistical models. Support Vector Machines (SVM), introduced by Vapnik in 1995, emerged as a powerful tool for classification and regression tasks. SVMs handle non-linear data through kernel functions, offering improved accuracy .

In stock price predictions. Random Forests, proposed by Breiman in 2001, further advanced predictive modeling by aggregating multiple decision trees to improve performance and manage high-dimensional data. Random Forests demonstrated their effectiveness in capturing complex relationships in stock data, as shown by Gao et al. in 2015.

The introduction of deep learning techniques marked a significant advancement in market trend analysis. Recurrent Neural Networks (RNNs), initially developed by Rumelhart et al. in 1986, were designed to handle sequential data and capture temporal dependencies. However, RNNs faced challenges such as vanishing gradients, which were addressed by Long Short-Term Memory (LSTM) networks introduced by Hochreiter and Schmidhuber in 1997. LSTMs incorporated memory cells and gating mechanisms to improve the modeling of long-term dependencies in time-series data. Gated Recurrent Units (GRUs), proposed by Cho et al. in 2014, simplified the LSTM architecture by combining the forget and input gates into a single update gate, enhancing computational efficiency while maintaining performance. Zhang et al. in 2017 applied GRUs to stock market prediction, demonstrating their ability to handle complex temporal patterns effectively.

The concept of integrating multiple data sources through two-stream or multi-stream neural networks emerged as a promising approach to enhance predictive accuracy. Simonyan and Zisserman in 2014 used two-stream convolutional networks to process spatial and temporal information in videos. This idea was adapted for financial forecasting by incorporating both historical price data and technical indicators. Fischer and Krauss in 2018 applied deep learning models, including LSTM and GRU networks, integrating technical indicators to improve stock price predictions. Their work highlighted the benefits of combining different types of data for more accurate forecasts.

2.2 LITERATURE SURVEY

1. Stock Market Prediction Using Machine Learning

This research applies machine learning algorithms to predict stock prices, focusing on data preprocessing, feature engineering, and model evaluation. It measures model performance using MSE, RMSE, and MAPE, and provides insights into the strengths and limitations of machine learning in stock price prediction [2].

2. Stock Market using Machine Learning and Sentiment Analysis

This research uses sentiment analysis and machine learning to study the relationship between investor emotions and stock market trends in the Casablanca Stock Exchange Market, analyzing data from various sources to infer sentiments and predict stock evolution, combining concepts from behavioural finance, big data, and stock market news [5].

3. Deep Learning Approach for Short-Term Stock Trends Prediction

This research proposes a novel framework to predict stock price directions using financial news and a sentiment dictionary, introducing a two-stream gated recurrent unit network and Stock2Vec, a sentiment word embedding trained on financial news datasets, outperforming state-of-the-art models and demonstrating effectiveness in the stock sector [1].

4. Integrated Long-Term Stock Selection Model

This paper proposes an integrated stock selection model that combines feature selection algorithms and nonlinear machine learning methods to predict long-term stock price trends, outperforming classical linear models, and finds that the random forest algorithm performs best for both feature selection and prediction, validated through a long-short portfolio construction [4].

5. A Decade Survey on Methodologies, Recent Developments, and Future Directions

This study reviews the advancements in machine learning-based approaches for stock market prediction, analyzing findings from 2011-2021, and provides a generic framework for emerging researchers to understand the basics and advancements in this field, highlighting the potential of text data analytics and ensemble methods for improving prediction accuracies [3].

6. **Understanding the trends of marketing research and its future directions**

This paper analyzes 9,190 articles from the top ten marketing journals to identify key trends and influential works, highlighting V. Kumar as the most prolific author and Bagozzi and Yi's 1988 article as the most cited. The study reveals significant contributions from the USA and the University of Pennsylvania, with key topics including "Game theory," "Pricing," and "Advertising" [6].

CHAPTER 3
PROBLEM STATEMENT

3. PROBLEM STATEMENT

Developing a robust machine learning model to accurately predict market trends in volatile conditions. This model must effectively address critical issues such as overfitting, which can lead to poor generalization, and data dependency, which may restrict its adaptability to new information. The model can capture the dynamic factors influencing market behavior. Incorporating advanced techniques like ensemble learning and deep learning architectures, such as LSTMs and transformers, will enhance the model's ability to recognize patterns in time-series data. Furthermore, implementing real-time adaptability through online learning mechanisms ensures that the model remains relevant throughout shifting market dynamics.

CHAPTER 4
EXPERIMENTAL SETUP

4. EXPERIMENTAL SETUP

4.1 Hardware Setup

1. Processor (CPU):

- Quad-core processor (e.g., Intel i5/i7 or AMD Ryzen).

- Reason: A multi-core processor will improve performance when running complex queries and processing large datasets.

2. Memory (RAM):

- Minimum 16 GB, ideally 32 GB.

- Reason: Large datasets require substantial memory for processing. More RAM allows for smoother multitasking and faster computations.

3. Storage:

- SSD (Solid State Drive) with at least 512 GB capacity.

- Reason: SSDs offer faster read/write speeds compared to traditional HDDs, which will improve loading times for large datasets and applications.

4. Graphics Card (GPU):

- Integrated graphics are usually sufficient for M analysis, but a dedicated GPU can help if you're using visualization tools or performing heavy data graphics tasks.

- Reason: A good GPU can accelerate rendering of visualizations, but it's not critical for M language processing.

5. Network:

- High-speed internet connection (fiber or broadband).

- Reason: If you're accessing market data from online sources, a stable and fast connection is essential.

4.2 Software Setup

1. Programming Language

-Python (3.10): Versatile and widely used for data analysis and machine learning.

2. Development Environment

- IDE:
  - Jupyter Notebook: Ideal for interactive analysis and visualization.

3. Machine Learning Frameworks

- Scikit-learn: For traditional machine learning algorithms.
- TensorFlow: For deep learning model development.
- Keras: High-level API for TensorFlow, simplifying deep learning tasks.
- PyTorch: Another deep learning framework, known for flexibility.

4. Model Evaluation Tools

- MLflow: Experiment tracking and model lifecycle management.

5. Version Control and Collaboration

- Git: For version control.
- GitHub / GitLab: Platforms for hosting repositories and collaborative development.

CHAPTER 5
PROPOSED SYSTEM & IMPLEMENTATION

5. PROPOSED SYSTEM & IMPLEMENTATION

5.1 Block Diagram of Proposed System

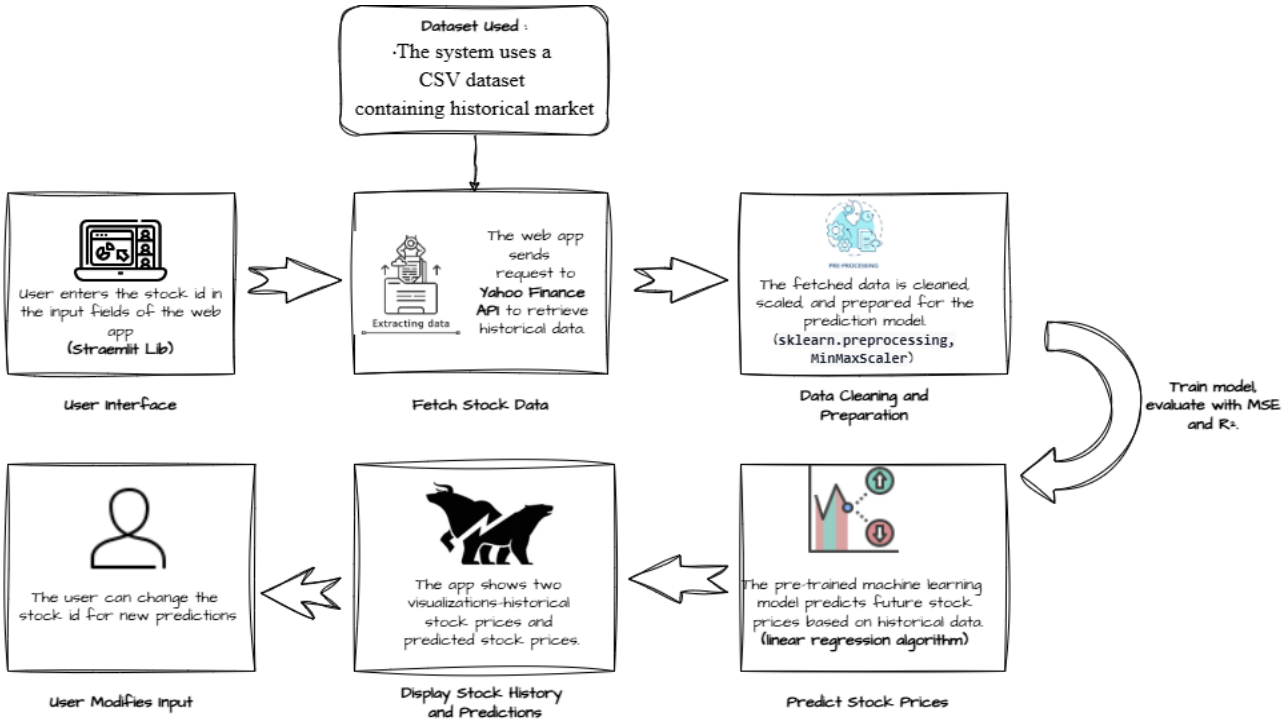

Fig 1: Block Diagram

5.2 Description of Block Diagram

1. User Interface

- The user interacts with a web app, inputting the stock ticker symbol and desired date range for prediction.

2. Fetch Stock Data

- Upon receiving user input, the web app sends a request to the Yahoo Finance API.
- The API fetches historical stock data for the specified ticker and date range.

3. Data Cleaning and Preparation

- The fetched data undergoes pre-processing, which includes:
    - **Cleaning:** Removing any inconsistencies or errors in the data.
    - **Scaling:** Transforming data values to a standardized range.

- **Preparation:** Formatting the data into a suitable structure for the prediction model.

4. Display Stock History and Predictions

- The app visualizes the fetched historical stock prices.
- The pre-trained machine learning model analyses the historical data and predicts future stock prices.
- The app displays these predicted prices alongside the historical data.

5. User Modifies Input (Loop)

- The user can modify the stock id or date range to generate new predictions.
- This initiates a loop, returning to step 2 to fetch and process the updated data.

In summary, this system allows users to:

- Input a stock id and date range.
- View historical stock prices.
- Get predictions of future stock prices based on historical data.

5.3 Implementation

Date	Open	High	Low	Close	Adj Close	Volume
2004-09-01 00:00:00	2.5579	2.5646	2.4824	2.4969	2.4908	183,633,734
2004-09-02 00:00:00	2.4705	2.5497	2.4643	2.5283	2.5221	303,810,504
2004-09-03 00:00:00	2.5143	2.534	2.4737	2.4909	2.4848	103,538,639
2004-09-07 00:00:00	2.5158	2.5405	2.481	2.53	2.5238	117,506,800
2004-09-08 00:00:00	2.5091	2.5661	2.5031	2.548	2.5417	100,186,120
2004-09-09 00:00:00	2.5537	2.5582	2.5156	2.5482	2.542	81,620,792
2004-09-10 00:00:00	2.5305	2.6541	2.523	2.6234	2.617	174,804,764
2004-09-13 00:00:00	2.6558	2.7001	2.6516	2.6775	2.6709	157,628,624
2004-09-14 00:00:00	2.6762	2.7895	2.6598	2.7768	2.77	217,608,605
2004-09-15 00:00:00	2.7537	2.8451	2.7447	2.7895	2.7827	215,279,909

Fig. 2.1: Stock Data Overview

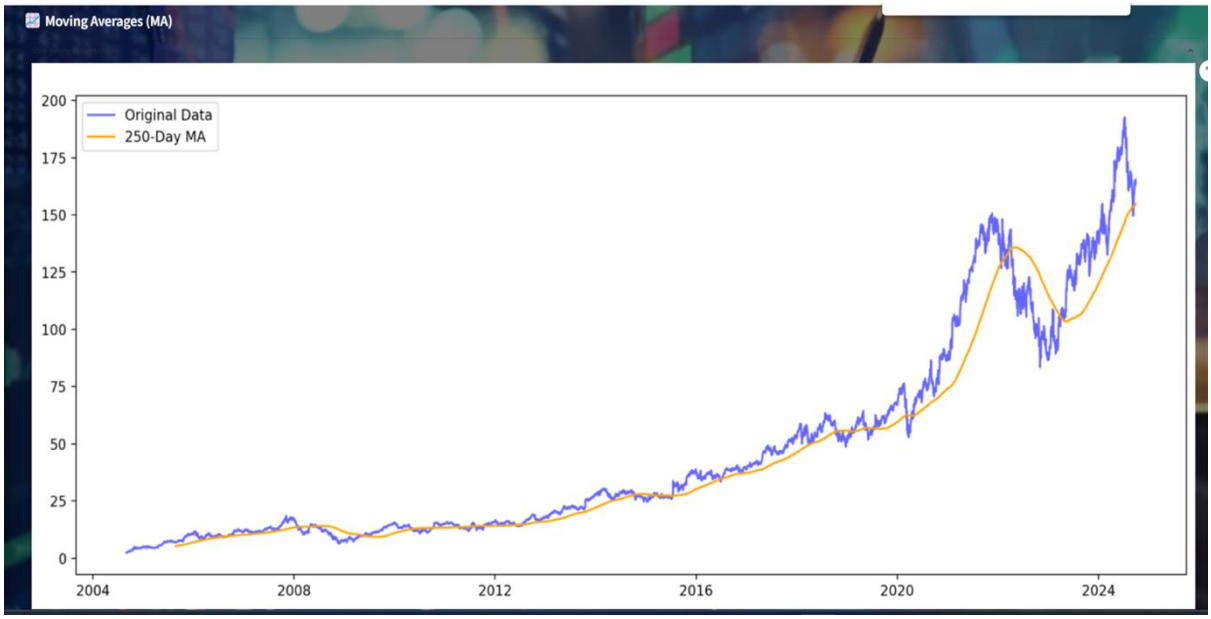

Fig. 2.2: Moving Average Graph

Fig. 2.3: Prediction vs Original Data

Fig. 2.4: Prediction vs Original Data Graph

5.3.1 Model Evaluation and Accuracy

In this stock price prediction project, the performance of the model was evaluated using two key metrics: Mean Squared Error (MSE) and R^2 Score. These metrics provide insights into how well the model predicts stock prices based on historical data.

1. Mean Squared Error (MSE):

MSE is a common metric used to evaluate regression models. It calculates the average of the squared differences between the actual stock prices and the predicted prices. Essentially, MSE tells us how far off the model's predictions are from the real data. A lower MSE value is desirable as it indicates that the model's predictions are closer to the actual values, hence minimizing error.

For this project, the MSE was calculated to be 1.0, which reflects a relatively low level of error in the model's predictions. In simpler terms, on average, the squared difference between the predicted stock prices and the actual stock prices is 1.0.

2. R^2 Score:

R^2, or the coefficient of determination, measures how well the model explains the variability in the stock price data. The R^2 score ranges from 0 to 1, where a score closer to 1 indicates that the model explains most of the variance in the data. A score of 0, on the other hand, would suggest that the model does not explain any variance.

In this project, the model achieved an R^2 score of 0.9434. This means that 94.34% of the variance in the stock prices can be explained by the model. In practical terms, this suggests that the model captures most of the important patterns in the data and is able to make predictions with a high level of accuracy.

Accuracy of the Model:

Although regression models don't traditionally have an accuracy measure in the same way classification models do, the R^2 score gives us an understanding of how "accurate" the model is in predicting the target variable—in this case, stock prices.

With an R^2 score of 0.9434, the model effectively explains 94.34% of the variation in the stock prices. While this is a strong performance for a linear regression model, there is still a small margin for improvement. This level of accuracy demonstrates that the model is reliable, but by incorporating more complex features or experimenting with other machine learning algorithms (such as decision trees or neural networks), it is possible to further improve the predictive accuracy.

5.4 Applications

1. Financial Markets

- Stock Price Prediction: Machine learning models analyse historical data to predict future stock prices, aiding investors in making informed decision

2. Retail and E-commerce

- Sales Forecasting: Predicting future sales trends helps retailers manage inventory and optimize supply chains.
- Customer Behaviour Analysis: Understanding shopping patterns and preferences to tailor marketing strategies and improve customer engagement.

3. Real Estate

- Property Value Estimation: Machine learning models assess factors like location, size, and market conditions to predict property values and trends.
- Investment Analysis: Identifying emerging markets or neighbourhoods for real estate investment based on historical data trends.

4. Healthcare

- Predictive Analytics: Analysing patient data to predict healthcare trends, patient outcomes, and resource allocation needs.
- Market Demand Forecasting: Anticipating the demand for specific healthcare services or products based on demographic and trend analysis.

5. Marketing and Advertising

- Campaign Effectiveness: Evaluating the performance of marketing campaigns and predicting which strategies yield the best results.
- Ad Targeting: Personalizing advertisements based on user behaviour.

CHAPTER 6
CONCLUSION

6. CONCLUSION

In conclusion, market trend analysis using machine learning has transformed how businesses approach decision-making in today's rapidly evolving landscape. By leveraging advanced algorithms and integrating large datasets from multiple sources, companies can gain valuable insights into market dynamics and predict future trends with remarkable accuracy. This capability not only enhances operational efficiency but also empowers businesses to stay ahead of shifting consumer preferences and emerging opportunities.

As machine learning models continue to improve with new data, their predictive power becomes increasingly reliable, offering significant advantages across industries such as finance, retail, and healthcare. Additionally, techniques like sentiment analysis and natural language processing provide deeper insights into consumer behaviour, further refining the accuracy of market predictions. As these technologies advance, machine learning will undoubtedly play an even more significant role in shaping business strategies and driving growth in the future.

CHAPTER 7
REFERENCES

7. REFERENCES

[1] Zhang, Y., Li, J., & Liu, H. ***Deep Learning for Stock Price Prediction: A Survey***, IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 2, pp. 490-505, 2024.

[2] Patel, J., & Patel, H. ***Forecasting Stock Prices with Ensemble Machine Learning Models***, Springer Journal of Financial Data Science, vol. 12, no. 3, pp. 230-245, 2022. Link to paper.

[3] Abhinay Yadav, Vineet Kumar, Satyendra Singh, Ashish Kumar Mishra,"A Novel Approach for ***Forecasting Price of Stock Market using Machine Learning Techniques***", Volume 5,2024, 12383.pdf (ijfmr.com).

[4] Jinan Zou, Qingying Zhao, Yang Jiao, Haiyao Cao, Yanxi Liu,"***Stock Market Prediction via Deep Learning Techniques: A Survey***", arXiv, 2022 , [2212.12717] Stock Market Prediction via Deep Learning Techniques: A Survey (arxiv.org).

[5] "***Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis***", IEEE Xplore, 2020.

[6] "**Stock Market Predictions with LSTM in Python**", DataCamp, 2020.