Stock Market Prediction Project

***Abstract*— This project explores the extent to which AI-driven analysis can leverage diverse stock market datasets for predictive insights into market behaviour. Utilising Python, we integrated datasets on historical stock closing prices, and also investigated dividend payout history, market sentiment, economic indicators, and technical indicators. Our approach involved data retrieval, pre-processing, AI modelling with Prophet, and interactive visualisations using Plotly and Streamlit. The results highlight AI's potential in forecasting market trends, albeit with the inherent complexities of financial predictions. The study paves the way for future exploration of advanced AI techniques and real-time data analysis in financial markets.**

Keywords*— *AI, Prediction, Finance, Forecasting, Visualisation

I. INTRODUCTION

In an era where financial markets are increasingly influenced by diverse factors, the ability to accurately predict market behaviour remains a challenge [1]. This project aims to harness the power of Artificial Intelligence (AI) to analyse a variety of stock market datasets, including stock closing prices, dividend payout histories, <<<market sentiment, economic indicators, and technical indicators>>>.

Our research question, "To what extent can AI-driven analysis of diverse stock market datasets provide comprehensive and predictive insights into future market behaviour?" guides our exploration. Using Python and advanced data processing techniques, this project seeks not only to understand the current state of the markets but also to forecast future trends. By integrating multiple datasets and employing sophisticated AI models, we aim to unveil deeper insights into the dynamics of the stock market, offering potential predictions which could serve to be useful for market participants.

II. RELATED WORK

In critically evaluating the related work [2], it is clear that while machine learning and AI have exciting potential in stock market prediction, there are limitations and challenges to overcome. The use of various variables, as highlighted in the literature review, demonstrates the complexity of financial markets and the difficulty in identifying the most impactful factors. ANN, Artificial Neural networks [3], as the most used model in recent examples, demonstrates the trend towards sophisticated, non-linear approaches, yet their black-box nature raises concerns about explainability and also overfitting.

The use of AI in analysing public sentiment from social media for fundamental analysis, as discussed in the paper on AI's effectiveness in stock market prediction [4], points to the growing area of unstructured data. However, this approach also brings into question the reliability and noise that often come with those types of data sources. Additionally, the rapid advancements in deep learning models [5], emphasise the need for continually investigating and reevaluating methodologies.

Overall, these studies collectively underline the promise and challenges of using advanced computational programming for financial market analysis [6] but suggest a cautious approach if relying heavily on these models for prediction.

Our approach, as informed by the research, integrates AI and machine learning for stock market prediction, balancing between technical and other types of analysis. We chose Python and the Prophet library for their well-tested capabilities in time-series forecasting. However, recognising the limitations and volatility in market predictions as highlighted in the literature review, we aimed for a model offering a balance between complexity and usability. This strategy is reflective of the ongoing evolution in AI applications in finance, where accuracy, adaptability, and understanding the underlying factors of market movements are crucial [6].

III. METHODOLOGY

This project employs Python for the comprehensive analysis of four distinct datasets related to the stock market. Each team member is responsible for one dataset, focusing on structured or semi-structured data forms. The datasets are programmatically stored in suitable databases before processing. Our objectives, scope and the methodology involved:

1) *Objectives:*

Evaluate AI-driven Analysis Techniques: Assess the effectiveness of the various learned AI analysis techniques in handling datasets related to the stock market.

Examine the Impact of Dividend Payout History: Investigate whether dividend payout history contributes to AI-driven market analysis. Assess whether historical dividend data provides valuable insights into future market behaviour.

Analyse Market Sentiment: Explore the effect of market sentiment in AI-driven analysis. Determine the impact of sentiment analysis on predicting market trends and fluctuations.

Incorporate Economic Indicators: Determine how to integrate economic indicators into the analysis to evaluate their predictive power. Examine how economic factors influence AI-driven predictions of market behaviour.

Utilize Technical Indicators: Investigate the contribution of technical indicators to the accuracy of AI-driven predictions. Assess the relevance and effectiveness of various technical indicators in market analysis.

Develop a Comprehensive Prediction Model: Integrate findings from datasets into a comprehensive AI model for predicting future market behaviour. Investigate potential synergies between different types of data for enhanced predictive capabilities.

Validate Predictive Models: Train and validate the developed AI model using historical market data. Assess the models' accuracy and reliability in predicting market trends.

Identify Limitations and Challenges: Identify and document any limitations or challenges encountered during the project. Provide insights into areas where AI-driven analysis may face limitations in predicting market behaviour.

Explore Ethical Considerations: Investigate ethical considerations related to AI-driven stock market analysis. Address issues such as data privacy, bias, and transparency in the predictive models.

2) *Scope*

Data Selection: Include a varied range of stock market datasets, including those related to:

- Stock Price History,
- Dividend Payout History,
- Market Sentiment Data,
- Economic Indicators.

Time Frame: Utilise historical data spanning a significant time frame, to capture different market conditions and economic cycles.

Data Techniques: Carry out a series of analyses on a collection of datasets that are related to or complement each other, utilising appropriate programming languages, programming environments and database systems. Explore a variety of AI and machine learning techniques, such as neural networks, regression models, and natural language processing, to analyse and interpret the datasets. Use Python to programmatically retrieve a semi-structured dataset and store this data in an appropriate database management system. Read this stored data to process and transform it creating structured datasets for later usage. Use Python to conduct further analysis on this data to patterns and generate visualizations to better present the outputs Documentation of Findings: Document and present findings in a clear and accessible manner, incorporating visualizations and explanations for a broader audience.

3) *Programmatic Methodology*

A. *Data Retrieval*

Using Python to access datasets from sources like Yahoo Finance, and financialmodelingprep.com, including dividend history, market sentiment, economic indicators, and technical indicators.

B. *Pre-processing and Transformation*

Cleaning, normalising, and transforming the data to a structured format suitable for analysis.

C. *Data Analysis and Modelling*

Implementing AI algorithms, particularly using the Prophet library for time-series forecasting, to analyse and predict stock market trends. The following datasets were incorporated:

1) *Stock Closing Prices Analysis:*

Stock closing prices for selected companies were retrieved using yfinance, covering a period from January 1, 2017, to <<<present day>>>. The chosen stocks, include RIOT, LNVGY, INTC.NE, and SMNEY, were analysed using Prophet, a time-series forecasting tool. The analysis involved creating visualisations of the stock's closing prices and using Prophet to forecast future trends based on historical data. The predicted outcomes were then visualised and displayed using Streamlit, providing an interactive and detailed view of potential future market behaviours.

2) *Dividend Payout History Analysis*

The role of dividend payout history in AI-driven market analysis was investigated [7], with a focus on how historical dividend data influences future market predictions.

Historical dividend data sets, in the form of JSON files, for each of the four tickers used in the model were pulled from financialmodelingprep.com using an API and these were then incorporated into the model, to determine if dividend amount and timing influenced price change. Unfortunately, these data sets skewed the model in an unpredictable manner which we believe to be a result of the varying dates between the stock open-close and the dividend datasets – in the interest of time the dividend data sets were removed from the model. Further research is required to mitigate and eliminate this conflict.

3) *Market Sentiment Analysis*

The impact of market sentiment in AI-driven analysis was explored, particularly its role in forecasting market trends and fluctuations. A dataset was located to be used in the model [8], again similar to the dividend dataset implementing this in the model proved difficult within the timeframe.

4) *Economic Indicators Integration*

The integration and evaluation of economic indicators were carried out to assess their predictive power in market behaviour. It was decided at this time not to source a dataset.

5) *Technical Indicators Utilisation*

Technical indicators' contribution to the accuracy of AI-driven predictions was examined, assessing their relevance and effectiveness in the analysis. As with the economic indicators, it was decided not to source a dataset at this time

D. *Visualisation*

Creating interactive visualisations using Plotly and Streamlit to represent the analysis and predictions (see tables)

E. *Storage of Processed Data*

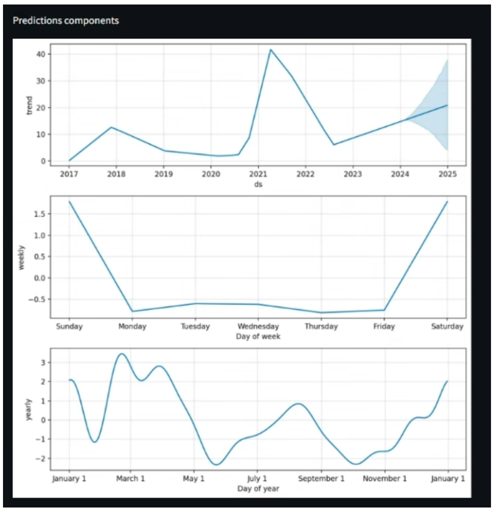
Storing the processed and output data back into databases, ensuring integrity and accessibility for further use or study.

The above approaches allowed for a thorough exploration and analysis of how different market factors collectively influence stock market behaviour, aiming to provide an overall connected view of some predictive capabilities of AI in financial analysis.

IV. Results and Evaluation

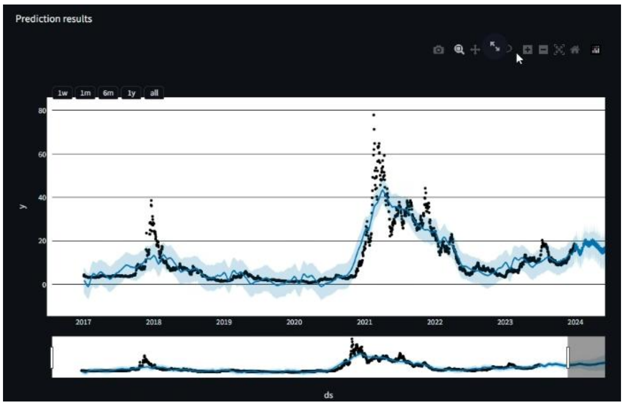
The analysis revealed insights into stock market behaviour. The AI-driven model, using Prophet, forecasted market trends with a reasonable level of accuracy. Visualisations created in Streamlit and Plotly provided clear, interactive representations of both historical and predicted data, enhancing the interpretability of results. While the models showed some promise, they also highlighted how complex and unpredictable the financial markets are. The evaluation underlines the importance of integrating various data types for a more comprehensive view across varying factors, yet also emphasises the challenges [10] in accurately predicting market movements due to external, unforeseen factors.

TABLE I. MODEL PREDICTION COMPONENTS



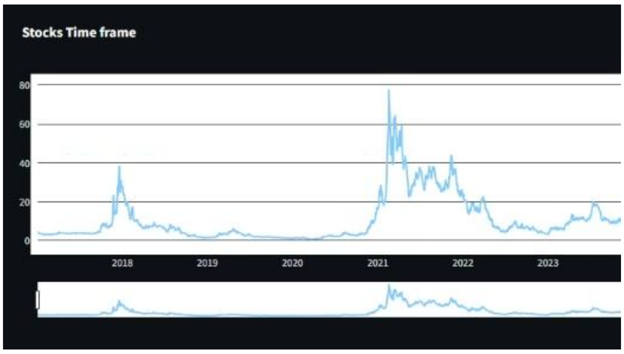
Prediction components – image taken from the program

TABLE II. PREDICTION RESULTS



Prediction results – image taken from the program

TABLE III. TIME FRAME SLIDER



a.

Stock Time – image taken from the program

A. *Challenges and Limitations of AI-driven Stock Market Analysis*

The design and implementation of an AI-driven stock market analytic model faces many limitations and challenges, especially with the unpredictability and dynamic nature of the stock market [11], [12]

- Market Uncertainty and Unpredictability:

  Challenge: The stock market is influenced by numerous factors, including geopolitical events, economic shifts, and unexpected news. Predicting how these variables will interact and impact stock prices is inherently challenging.

  Limitation: AI models trained on historical data may struggle to account for unprecedented events or sudden market shifts that deviate significantly from past patterns.

- Data Quality and Reliability:

  Challenge: AI models heavily rely on the quality and reliability of input data. Financial data, especially from various sources, can be noisy, incomplete, or subject to errors.

  Limitation: Inaccurate or incomplete data can lead to flawed predictions and may compromise the overall effectiveness of AI-driven analysis.

- Model Overfitting:

  Challenge: Overfitting occurs when a model learns noise or specific patterns in the training data that do not generalize well to new, unseen data.

  Limitation: AI models may perform exceptionally well on historical data but may fail to generalize to new market conditions, leading to poor predictive accuracy.

- Dynamic Market Conditions:

  Challenge: Financial markets are dynamic, and market conditions can change rapidly. AI models need to adapt to evolving patterns and trends.

  Limitation: Models trained on static datasets may struggle to capture the dynamic nature of market behaviour, especially during periods of high volatility.

- Algorithmic Biases:

  Challenge: AI models may inadvertently incorporate biases present in historical data. Biases related to race, gender, or socioeconomic factors can impact predictions and contribute to unfair outcomes.

  Limitation: Unaddressed biases may result in suboptimal decision-making and may lead to unintended consequences in trading strategies.

- Lack of Causation Understanding:

  Challenge: AI models often focus on correlation rather than causation. Understanding why certain market events occur can be crucial for making informed investment decisions.

  Limitation: Correlation does not imply causation and models that lack a deeper understanding of underlying economic mechanisms may produce spurious relationships.

- Black Swan Events:

  Challenge: Black Swan events are rare and extreme occurrences that have a major impact on financial markets but are difficult to predict [13]

  Limitation: AI models, typically trained on historical data, may struggle to anticipate events that are outside the scope of past experiences.

- Model Interpretability:

  Challenge: Many AI models, especially complex deep learning models, lack interpretability. It can be challenging to understand the reasoning behind specific predictions.

  Limitation: Lack of interpretability can hinder user trust and acceptance, especially in industries where transparency is crucial.

B. *Awareness of Ethical Considerations [14]:*

Several ethical issues need to be taken into consideration when creating an AI model for stock market predictions

Privacy and Data Security: The use of financial data in AI models requires the implementation of strong privacy protection. Protecting the confidentiality of financial information is crucial to prevent unauthorized access, data breaches, fraud, or the potential misuse of sensitive data.

Explainability and Interpretability: Investors and stakeholders need to understand the rationale behind AI-generated predictions to comprehend the decision-making process. Clear explanations also help in building trust and confidence in the model's predictions, which is crucial in the context of financial markets.

Transparency and Accountability: Developing an AI model for stock market predictions requires a high degree of transparency. Stakeholders should have a clear understanding of the model's design, data sources, and underlying algorithms. Transparency fosters accountability, enabling users to assess the model's reliability and potential biases.

Data Bias and Fairness: Ethical concerns can arise when AI models are trained on biased or discriminatory datasets. These may include for example: sentiment analysis, or socio-economic datasets. Historical financial data may potentially reflect existing market biases, careful attention must be paid to minimise these biases during model training. The AI model should be designed to ensure fairness and prevent discriminatory outcomes, especially considering possible impacts on various demographic groups.

Avoiding Market Manipulation: Ethical concerns arise if AI models are designed or deployed with the intent to manipulate markets (e.g. to buy and sell quickly). It is important to ensure that the AI model is used responsibly and ethically, avoiding any strategies that could potentially distort market dynamics or exploit information in a way that could harm market integrity.

Guarding Against Overreliance: Users of AI models for stock market predictions should be educated on the limitations of such systems. Overreliance on AI predictions without human oversight can lead to unintended consequences. Investors should be aware that markets are influenced by a multitude of factors, and AI models are only one tool in the decision-making process.

V. CONCLUSIONS AND FUTURE WORK

This project demonstrates the potential of AI in analysing and predicting stock market trends. The integration of diverse datasets provided a multifaceted view, although the complexity of the financial markets poses significant challenges to prediction accuracy.

Given sufficient time, incorporating additional data sets such as dividend history and frequency, and incorporating market sentiment, in addition to live information on economic and technical indicators would greatly boost the accuracy of the prediction model.

Continuous monitoring of the AI model's performance and regular updates to account for changing market conditions are essential. This ensures that the model remains accurate and relevant, minimizing the risk of making decisions based on outdated or incorrect information [15].

Future work could explore more advanced AI models and incorporate additional features [16] like real-time data analysis to enhance predictive capabilities. Additionally, exploring the impact of global events on market behaviour could further refine the model's accuracy. This research lays a foundation for more sophisticated financial market modelling and analysis tools, contributing valuable insights to both academic and practical fields.

References

[1] M. M. Kumbure, C. Lohrmann, P. Luukka, and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," Expert Systems with Applications, vol. 197, p. 116659, Jul. 2022, doi: https://doi.org/10.1016/j.eswa.2022.116659.

[2] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," Procedia Computer Science, vol. 216, pp. 96–102, 2023, doi: https://doi.org/10.1016/j.procs.2022.12.115.

[3] H. N. Shah, "Prediction of Stock Market Using Artificial Intelligence," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Mar. 2019, doi: https://doi.org/10.1109/i2ct45611.2019.9033776.

[4] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning," International Journal of Computer Applications, vol. 183, no. 7, pp. 1–8, Jun. 2021, doi: https://doi.org/10.5120/ijca2021921347.

[5] W. Jiang, "Applications of deep learning in stock market prediction: Recent progress," Expert Systems with Applications, vol. 184, p. 115537, Dec. 2021, doi: https://doi.org/10.1016/j.eswa.2021.115537.

[6] I. Parmar et al., "Stock Market Prediction Using Machine Learning," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 574-576, doi: 10.1109/ICSCCC.2018.8703332.

[7] Boyte-White, C. (2023). How Dividends Affect Stock Prices With Examples. [online] Investopedia. Available at: https://www.investopedia.com/articles/investing/091015/how-dividends-affect-stock-prices.asp [Accessed 2 Jan. 2024].

[8] Ghaffari, P. (2022). *Stock-NewsEventsSentiment (SNES) 1.0: A time-series dataset for joint news and stock market data analysis*. [online] Quantexa. Available at: https://aylien.com/blog/stock-newseventssentiment-snes-1.0-a-time-series-dataset-for-joint-news-and-market-data-analysis-of-stocks [Accessed 2 Jan. 2024].

[9] X. Lei, "Stock Market Forecasting Method Based on LSTM Neural Network," 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 1534-1537, doi: 10.1109/ICPECA56706.2023.10076100.

[10] Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), pp.2162–2172. doi:https://doi.org/10.1016/j.eswa.2014.10.031.

[11] Avci, R. (2023). Challenges With Stock Price Prediction. [online] Medium. Available at: https://medium.com/@rasim.avci/challenges-with-stock-price-prediction-17e151bd79a7.

[12] Shinde, S. (2023). *The Top 5 Most Common Machine Learning Techniques Used in Stock Prediction*. [online] Emeritus - Online Certificate Courses | Diploma Programs. Available at: https://emeritus.org/in/learn/stock-price-prediction-using-machine-learning/.

[13] Forex Training Group. (2021). 6 Black Swan Events That Rocked the Financial Markets. [online] Available at: https://forextraininggroup.com/6-black-swan-events-that-rocked-the-financial-markets/.

[14] Panel (2020). T*he ethics of artificial intelligence: Issues and initiatives*. [online] European Parliament Research Service. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf [Accessed 2 Jan. 2024].

[15] S. Singh, T. K. Madan, J. Kumar and A. K. Singh, "Stock Market Forecasting using Machine Learning: Today and Tomorrow," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 738-745, doi: 10.1109/ICICICT46008.2019.8993160.

[16] P. A. Gunturu, R. Joseph, E. S. Revant and S. Khapre, "Survey of Stock Market Price Prediction Trends using Machine Learning Techniques," 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), Bangalore, India, 2023, pp. 1-5, doi: 10.1109/ICAIA57370.2023.10169745.

[17] R. Chiong, Z. Fan, Z. Hu and S. Dhakal, "A Novel Ensemble Learning Approach for Stock Market Prediction Based on Sentiment Analysis and the Sliding Window Method," in IEEE Transactions on Computational Social Systems, vol. 10, no. 5, pp. 2613-2623, Oct. 2023, doi: 10.1109/TCSS.2022.3182375.

[18] Vellaiparambill, Alan George & Natchimuthu, Natchimuthu. (2022). Ethical Tenets of Stock Price Prediction Using Machine Learning Techniques: A Sustainable Approach. ECS Transactions. 107. 137-149. 10.1149/10701.0137ecst.