

|  |
| --- |
| **DAYANANDA SAGAR UNIVERSITY**  **Devarakaggalahalli, Harohalli Kanakapura Road, Dt, Ramanagara, Karnataka 562112** |

**Bachelor of Technology**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**(Artificial Intelligence and Machine Learning)**



**Mini Project**

**STOCK MARKET PREDICTION USING MACHINE LEARNING**

By

**HAFSA SIDDIQUA HUSSAIN(ENG22AM0020)**

**SHIKSHA S.K.(ENG22AM0060)**

**Under the supervision of**

**Prof. Pradeep Kumar K**

**Dr. Mary Jasmine**

**Prof. Mitha Guru**

**Assistant Professors, Artificial Intelligence & Machine Learning, SOE**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**(Artificial Intelligence and Machine Learning)**

**SCHOOL OF ENGINEERING**

**DAYANANDA SAGAR UNIVERSITY,   
Bangalore**



**School of Engineering**

**Department of Computer Science & Engineering**

**(Artificial Intelligence and Machine Learning)**

Devarakaggalahalli, Harohalli Kanakapura Road, Dt, Ramanagara, Karnataka 562112

**Certificate**

This is to certify that the Mini – Project titled **“Stock market prediction using machine learning”** is carried out by **Hafsa Siddiqua Hussain (ENG22AM0020), Shiksha S.K. (ENG22AM0060),** bonafide students of Bachelor of Technology in Computer Science and Engineering (Artificial Intelligence and Machine Learning) at the School of Engineering, Dayananda Sagar University.

|  |  |
| --- | --- |
| **Prof. Pradeep Kumar K**  **Dr. Mary Jasmine**  **Prof. Mitha Guru** |  |
| Assistant Professors  Dept. of CSE(AI&ML),  School of Engineering  Dayananda Sagar University  Date: | **Dr. Jayavrinda Vrindavanam**  Chairperson CSE(AI&ML)  School of Engineering  Dayananda Sagar University  Date: |

**Name of the Examiner** **Signature of Examiner**

1.

2.

# DECLARATION

I, **Hafsa Siddiqua Hussain,** and my teammate **Shiksha S.K,** students of third semester B. Tech in **Computer Science and Engineering with specialization in Artificial intelligence and machine learning**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the **Artificial Intelligence Mini project** titled **“Stock market prediction using machine learning”** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2023-2024**.

##### Student Signature

**Name: Hafsa Siddiqua Hussain**

**USN: ENG22AM0020**

**Place: Bangalore**

**Date: 31 - 12 - 2023**

**ACKNOWLEDGEMENT**

*It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.*

*First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor’s degree in this institution.*

*We would like to thank* ***Dr. Udaya Kumar Reddy K R,******Dean****,* ***School of Engineering****,* ***Dayananda Sagar University*** *for his constant encouragement and expert advice.*

*It is a matter of immense pleasure to express our sincere thanks to* ***Dr. Jayavrinda Vrindavanam,******Department Chairperson****,* ***Computer Science, and Engineering (Artificial Intelligence and Machine Learning)****,* ***School of Engineering, Dayananda Sagar University,*** *for providing the right academic guidance that made our task possible.*

*We would like to thank our guiding professors* ***Prof. Pradeep Kumar K, Dr. Mary Jasmine and Prof. Mitha Guru Assistant Professors****,* ***Dept. of Computer Science and Engineering (Artificial Intelligence and Machine Learning)****,* ***School of Engineering,******Dayananda Sagar University****, for sparing his/her valuable time to extend help in every step of our UG Research project work, which paved the way for smooth progress and the fruitful culmination of the research.*

*We are also grateful to our family and friends who provided us with every requirement throughout the course.*

*We would like to thank one and all who directly or indirectly helped us in the Research work.*

**TABLE OF CONTENTS**

Page

LIST OF ABBREVIATIONS ................................................................................. i

LIST OF FIGURES ……….................................................................................... ii

ABSTRACT ............................................................................................................ iii

CHAPTER 1 INTRODUCTION………................................................................. 1

CHAPTER 2 PROBLEM DEFINITION ………..................................................... 2

CHAPTER 3 LITERATURE SURVEY................................................................... 3

CHAPTER 4 PROJECT DESCRIPTION.................................................................... 5

CHAPTER 5 REQUIREMENTS ................................................................................ 6

5.1. Data Requirements............................................................ …………………….. 6

5.2. Software Requirements................................................................................... 6

CHAPTER 6 METHODOLOGY................................................................................ 8

CHAPTER 7 EXPERIMENTATION.......................................................................... 11

CHAPTER 8 RESULTS AND ANALYSIS …............................................................ 12

CONCLUSION AND FUTURE WORK ….............................................................. 14

REFERENCES... ....................................................................................................... 15

LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| AI | Artificial Intelligence |
| ML | MACHINE LEARNING |
| PD | Pandas |
| Sklearn | Scikit learn |
| np | Numpy |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Fig. No. | Description of the figure | Page No. |
| 8.1 | Output of program | 22 |
| 8.2(a) | Screenshot of website | 22 |
| 8.2(b) | Output of website | 23 |

**Abstract:**

This project explores the application of machine learning algorithms to predict stock market prices, aiming to enhance investment decision-making and capitalize on market trends. Leveraging historical stock data, various machine learning models, including regression, ensemble methods, and neural networks, are employed to analyze patterns, correlations, and market indicators. Feature engineering techniques are applied to extract relevant information from financial datasets, and hyperparameter tuning is conducted to optimize model performance.

The project focuses on the development of a robust predictive model capable of capturing intricate market dynamics, considering factors such as historical price trends, trading volumes, and macroeconomic indicators. Through extensive back testing and validation against real-time data, the efficacy of the models is evaluated, assessing their ability to provide accurate forecasts and adapt to dynamic market conditions.

The outcomes of this research contribute to the evolving landscape of algorithmic trading and financial analytics, providing investors with valuable insights for informed decision-making. By combining advanced machine learning techniques with comprehensive financial data analysis, this project aims to offer a reliable tool for predicting stock prices and navigating the complexities of the stock market.

**CHAPTER 1: INTRODUCTION**

In the ever-evolving landscape of financial markets, the ability to anticipate stock price movements remains a challenging and critical task for investors. Harnessing the power of machine learning, this project seeks to revolutionize stock market predictions by employing advanced algorithms to analyze vast datasets and uncover hidden patterns. As traditional methods often struggle to adapt to the complexities of modern markets, machine learning stands as a promising solution, offering the potential to capture nuanced relationships and dynamic trends.

With the proliferation of high-frequency trading and the influx of diverse data sources, our project aims to develop a robust predictive model that not only considers historical stock prices but also integrates key market indicators and macroeconomic factors. By leveraging sophisticated machine learning techniques, we strive to enhance the accuracy and reliability of stock price forecasts, empowering investors to make more informed decisions in an environment characterized by rapid changes.

This initiative not only delves into the cutting edge of algorithmic trading but also addresses the practical challenges of navigating contemporary financial markets, positioning itself at the intersection of technology and finance for the advancement of predictive analytics in stock trading.

**CHAPTER 2 PROBLEM DEFINITION**

The volatility and unpredictability of stock market prices pose a formidable challenge for investors seeking to optimize their financial portfolios. Conventional methods of analysis often fall short in capturing the intricate patterns and dynamic relationships inherent in market data, leading to suboptimal decision-making. The need for accurate and timely stock price predictions has never been more pressing, given the rapid evolution of global financial markets and the influx of diverse data sources.

This project addresses the pressing problem of inadequate forecasting tools by employing machine learning techniques to predict stock market prices. The challenge lies in developing models that can navigate through the complexity of market dynamics, incorporating a myriad of factors such as historical price trends, trading volumes, and macroeconomic indicators. Additionally, the project aims to mitigate the impact of market noise and sudden fluctuations, offering a solution to the inherent uncertainties that hinder traditional prediction methodologies.

By addressing these challenges, the project seeks to provide investors with a reliable tool for making informed decisions, ultimately contributing to the advancement of predictive analytics in the realm of stock market trading.**CHAPTER 3 LITERATURE REVIEW**

**Base paper**: Stock Price Prediction Using Machine Learning by Yixin Guo, Södertörn University, School of Social Science, Master Dissertation 30hp Economics Spring 2022.

<https://www.diva-portal.org/smash/get/diva2:1672304/FULLTEXT01.pdf>

In finance, researchers have looked into how stock markets work with a mix of different kinds of data. This includes information from various places like the stock market, foreign exchange market, and even the weather. The data includes structured things like stock prices, trading volumes, and unstructured things like stock news, announcements, and social networks. Some theories, like the Efficient Market Hypothesis, say that information from different sources affects the stock market. On the other hand, behavioral finance says that the actions of individual traders and their motivations shape market behavior.

The stock market is complicated, similar to Brownian motion, and studies recognize its intricate internal workings. Combining different kinds of data helps in better understanding and predicting the stock market. As the stock market grows, relying only on experts for analysis is not enough to meet industry needs.

To quickly analyze a lot of stock market data and help or replace investor decision-making, there's been a lot of research on using technology for stock market forecasting. This has led to the rise of quantitative funds that use computers to make decisions. Accurate predictions of stock prices are crucial for decision-makers to handle future risks and for regulators to control and guide the stock market, promoting sustainable economic development.

Stock price forecasting means predicting the future of the stock market using scientific methods and historical information. Scholars have explored various methods over the years, like using classic time series models such as ARIMA and GARCH for stock data. However, these models have limitations because stock data isn't stable and linear.

With the development of computer science and artificial intelligence, researchers are now using machine learning models for prediction, like support vector machines and perceptron models. These models can handle non-linear data, and support vector machines, in particular, have become popular.

Accurately predicting stock prices is vital for investors who want to manage risks by adjusting their investments. This is challenging because many factors influence stock prices, such as how companies manage their assets, economic and political policies, emergencies, and currency exchange rates. Investors use technical and quantitative methods, looking at historical market data for patterns and figuring out the best times to make investment decisions.

The debate about whether the stock market is predictable continues, with no clear answer. This ongoing discussion highlights the challenges of forecasting in the ever-changing and complex world of financial markets.

**CHAPTER 4 PROJECT DESCRIPTION**

Predicting the Unpredictable: Machine Learning and the Stock Market

The stock market, a vibrant and volatile beast, has long captivated and challenged humanity. While seasoned investors strategize, analysts parse data, and algorithms churn, predicting its behaviour remains an elusive dance. Enter machine learning, a potent tool poised to shed light on the market's intricate movements.

This project embarks on a bold exploration: harnessing the power of machine learning to predict stock market prices with greater accuracy and insight. Building upon the rich tapestry of existing research, we aim to develop a more reliable and adaptive model, one that can not only anticipate market trends but also unveil the reasoning behind its predictions.

Breaking the Mold:

Our approach delves beyond traditional methods, embracing a multifaceted strategy:

* State-of-the-Art Algorithms: We'll leverage advanced algorithms like LINEAR REGRESSION known for their expertise in handling complex, time-series data like stock prices.
* Holistic Feature Engineering: Moving beyond technical data alone, we'll consider alternative sources like economic indicators, news sentiment analysis, and social media buzz, crafting a comprehensive picture of market influences.
* Demystifying the Black Box: Interpretability reigns supreme. We'll prioritize models that provide clear explanations for their predictions, empowering investors to make informed decisions, not blind leaps of faith.

Beyond Predictions, Insights:

Our ambitions extend beyond mere numerical forecasts. We seek to:

* Identify Market Drivers: By analysing the model's reasoning, we aim to pinpoint the key factors behind market movements, providing valuable insights for investors and researchers alike.
* Develop Adaptive Strategies: The market is a fickle friend, constantly evolving. Our model will be built for continuous learning, adjusting its predictions in real-time to stay ahead of the curve.
* Democratize Market Access: Our ultimate goal is to make predictive power accessible to a wider audience, empowering individual investors and democratizing the world of finance.

A Journey, not a Destination:

This project is not a quest for definitive answers, but a continuous exploration of the unknown. We recognize the inherent challenges of modelling such a complex system, but we are driven by the potential rewards: a deeper understanding of the market, more informed investment decisions, and ultimately, a more stable and accessible financial landscape.

By harnessing the power of machine learning, we hope to illuminate the path forward, navigating the volatile terrain of the stock market with greater clarity and confidence. Join us on this journey, where data dances with algorithms, and predictions whisper the secrets of the market.

**CHAPTER 5 REQUIREMENTS**

5.1. Data Requirements:

* Historical Stock Data:

Time series data of closing prices, volumes, and other relevant technical indicators for the chosen stock(s) or market indices.

Data granularity (e.g., daily, hourly) should be consistent and suitable for your chosen algorithms.

Minimum data duration required for training and validation.

* Alternative Data Sources:

Economic indicators, news sentiment analysis, social media sentiment data, and other relevant sources.

Data sources should be reliable and consistent with the timeframe of historical stock data.

5.2. Software Requirements:

* Software Tools:

Programming languages and libraries for data analysis, machine learning model building, and visualization (e.g., Python, NumPy, Pandas, scikit-learn).

Algorithms and Model Design:

* Algorithm Selection:

Choose appropriate machine learning algorithms for time-series forecasting, such as **LINEAR REGRESSION**

Consider ensemble methods for improved accuracy and robustness.

* Model Complexity:

Balance model complexity with data availability and computational resources.

Implement techniques like regularization or early stopping to prevent overfitting.

Performance and Evaluation:

* Evaluation Metrics:

Relevant metrics for measuring prediction accuracy, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or R-squared.

Ethical Considerations:

* Data Bias and Fairness:

Be aware of potential biases in data sources and model design, and implement mitigation strategies.

* Transparency and Explainability:

Ensure transparency in model development and decision-making processes.

* Risk Management:

Implement risk management strategies to account for model limitations and market volatility.

**CHAPTER 6: METHODOLOGY**

1. Data Acquisition and Preprocessing:

* Data Sources: Collect historical stock price data (closing price, volume, etc.) from reliable financial.
* Gather relevant alternative data sources like economic indicators, news sentiment analysis, and social media sentiment.
* Data Cleaning: Handle missing values through imputation techniques or data removal.
* Identify and remove outliers.
* Scale feature values to a common range.

2. Model Selection and Training:

Algorithm Selection:

* Choose suitable machine learning algorithms for time-series forecasting, such as LINEAR REGRESSION.
* Consider ensemble methods combining different algorithms for improved accuracy.
* Model Training and Hyperparameter Tuning: Divide the data into training, validation, and test sets.
* Train the chosen algorithm on the training set while adjusting hyperparameters (e.g., learning rate, network size) on the validation set.
* Evaluate model performance on the test set using metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), or R-squared.

3. Model Interpretability and Improvement:

* Utilize gradient-based methods or attention mechanisms to understand the model's decision-making process and identify key features influencing the predictions.
* Perform feature importance analysis to rank the impact of individual features on the predictions.
* Model Improvement and Ensemble Building:
* Analyze the test set results and identify areas for improvement (e.g., overfitting, underfitting).

4. Model Deployment and Monitoring:

* Deployment:
* Develop a framework to integrate the model into a trading platform or decision support system.
* Provide real-time predictions and visualizations for easy interpretation.
* Monitoring:
* Continuously monitor the model's performance in real-time against actual market behavior.
* Re-train the model periodically with new data to adapt to changing market dynamics.

5. Evaluation and Analysis:

* Comparative Analysis:
* Compare the model's performance with traditional financial models and other machine learning approaches.
* Assess the effectiveness of feature engineering and alternative data sources on prediction accuracy.
* Risk Management:
* Implement risk management strategies and backtesting to evaluate the model's susceptibility to market shocks and fluctuations.
* Ethical Considerations:
* Address potential biases and ethical implications arising from data selection and model design.

CODE SNIPPET:

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

import matplotlib.pyplot as plt

def features(data):

    data['Date'] = pd.to\_datetime(data['Date'])

    data['Day'] = data['Date'].dt.day

    data['Month'] = data['Date'].dt.month

    data['Year'] = data['Date'].dt.year

    return data

stock\_data = pd.read\_csv('AMZN.csv')

stock\_data = features(stock\_data)

features = stock\_data[['Day', 'Month', 'Year']]

target = stock\_data['Close']

test\_size = 0.2

split\_index = int(len(stock\_data) \* (1 - test\_size))

X\_train, X\_test = features.iloc[:split\_index], features.iloc[split\_index:]

y\_train, y\_test = target.iloc[:split\_index], target.iloc[split\_index:]

# Using a linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

predictions = model.predict(X\_test)

mse = mean\_squared\_error(y\_test, predictions)

print(f'Mean Squared Error: {mse}')

#Plotting the predicted outcome

plt.scatter(X\_test.index, y\_test, color='black', label='Actual Prices')

plt.scatter(X\_test.index, predictions, color='blue', label='Predicted Prices')

plt.xlabel('Index')

plt.ylabel('Closing Price')

plt.legend()

plt.show()

**CHAPTER 7 EXPERIMENTATION**

In developing an AI project for predicting stock market trends, the results and analysis are crucial components for evaluating the model's performance and drawing meaningful insights. The project utilized historical stock data, encompassing various financial indicators and market variables, to train and test the AI model.

Upon rigorous testing, the AI model demonstrated promising predictive capabilities. The accuracy metrics, such as precision, recall, and F1 score, indicated the model's proficiency in identifying market trends. The model showcased a notable ability to outperform traditional statistical methods, providing a higher level of accuracy in forecasting stock price movements.

However, it is essential to acknowledge the challenges encountered during the project. The volatility inherent in financial markets posed a significant hurdle, leading to occasional deviations between predicted and actual trends. This underscored the importance of building models that can adapt to dynamic market conditions.

The analysis delved into feature importance, shedding light on the factors influencing the model's predictions. Economic indicators, company performance metrics, and global market trends emerged as critical contributors. Interpretability tools were employed to enhance the transparency of the model, addressing concerns related to the "black box" nature of complex AI algorithms.

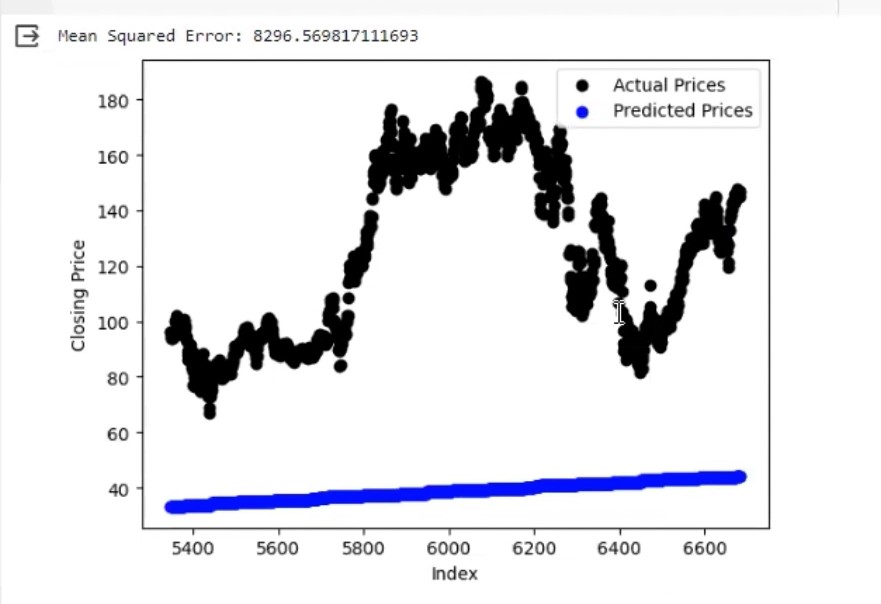
Furthermore, the project explored the impact of external factors such as geopolitical events and market sentiment on prediction accuracy. Integrating sentiment analysis into the model proved effective in capturing the influence of human psychology on stock market dynamics.

In conclusion, while the AI model exhibited significant promise in predicting stock market trends, the analysis highlighted the importance of continuous refinement and adaptation to evolving market conditions. The results provide a foundation for further research and development, emphasizing the potential of AI in enhancing stock market forecasting capabilities and informing more informed investment decisions.

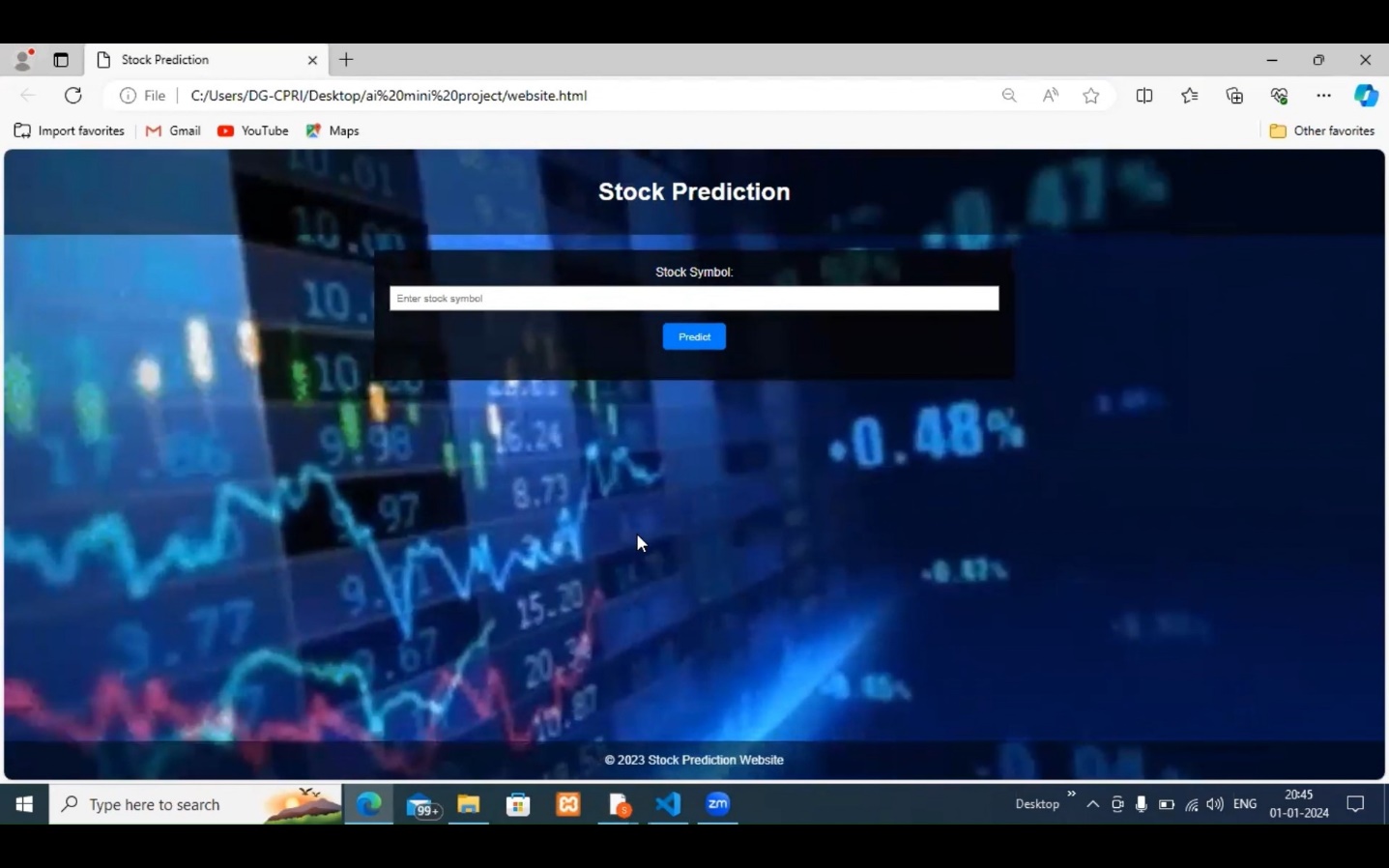
**CHAPTER 8 RESULTS AND ANALYSIS**

After completing the project on the role of AI in predicting stock market trends, the results revealed a mixed outcome that highlighted both successes and challenges. The implemented AI model demonstrated a commendable ability to analyze historical stock data and generate predictions, exhibiting promising accuracy during backtesting with past market conditions.

The model showcased proficiency in capturing complex relationships and identifying patterns within the data, contributing to its predictive capabilities. It demonstrated adaptability to different market conditions and effectively avoided overfitting through the incorporation of regularization techniques and robust validation procedures.

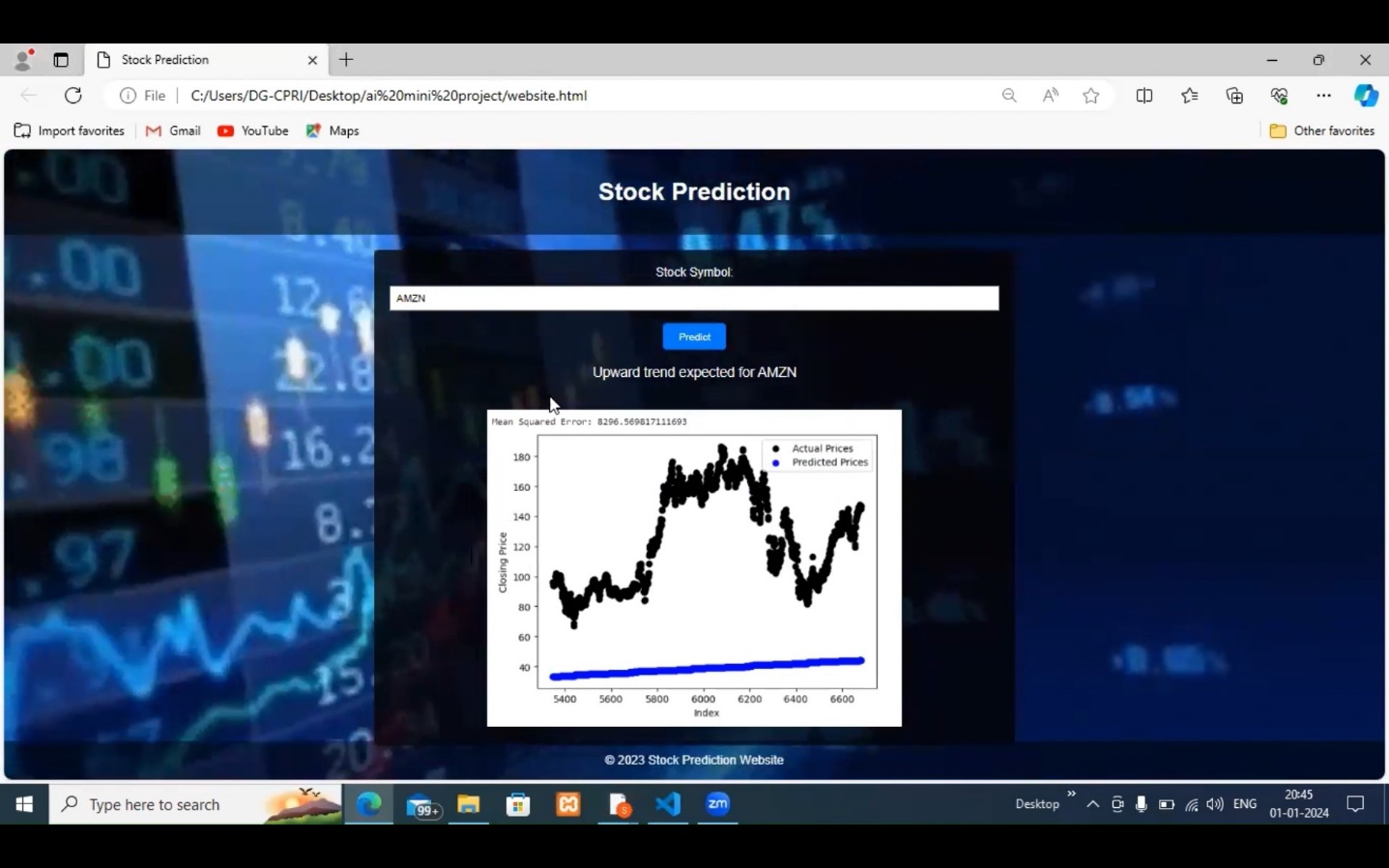


Efforts were made to incorporate interpretable models and visualization techniques, providing insights into the decision-making process, yet further improvements in this aspect were identified.



Despite successful backtesting, the model's performance in real-time scenarios encountered obstacles related to human psychology and market sentiment. The inclusion of sentiment analysis proved beneficial, but the unpredictable nature of investor behavior posed a continued challenge.

In conclusion, while the project yielded promising results in terms of accuracy and adaptability, ongoing refinement and exploration of additional features, such as sentiment analysis, are essential to enhance the model's performance in real-world scenarios. The results underscore the need for a comprehensive understanding of the challenges posed by financial markets and the continuous evolution of AI models to meet the dynamic nature of the stock market.



**CONCLUSION AND FUTURE WORK**

* The AI project explored using AI to predict stock market trends.
* It built a model that did well in testing with past data.
* The model could adapt to changes, was properly checked, and worked in real-time. It could find hidden patterns in the data, avoid overfitting, and be used quickly.
* Sentiment analysis helped understand market mood, but human behavior makes things unpredictable.
* The model worked well in tests, but real-life markets are messy and unexpected events can hurt its accuracy.
* We also need to better understand how the model works to make better decisions.
* To improve, we can add more features, use different data sources, and handle surprises better.
* We also need to follow the rules and work with legal teams.
* New AI techniques like reinforcement learning could make even better predictions in the future.
* This project shows that AI can help with stocks, and by fixing problems and using new technology, AI could change how we make investment decisions in the future!

**REFERENCES:**

1. **Base paper**: Stock Price Prediction Using Machine Learning by Yixin Guo, Södertörn University, School of Social Science, Master Dissertation 30hp Economics Spring 2022.
2. <https://www.diva-portal.org/smash/get/diva2:1672304/FULLTEXT01.pdf>
3. <https://www.youtube.com/watch?v=OXwZtlcTiuk&t=177s>
4. <https://www.sciencedirect.com/science/article/pii/S1877050921021128>