Stock Market Prediction

Project Report

INDUSTRIAL TRANNING (ECS599)

BACHELOR OF TECHNOLOGY (CSE: AI ML DL)

PROJECT GUDE:
Mr. Rajendra Prasad Pandey

SUBMITTED BY:
Asim Husain (TCA2259012)

November, 2024

FACULTY OF ENGINEERING & COMPUTING SCIENCES
TEERTHANKER MAHAVEER UNIVERSITY, MORADABAD

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Internshala for providing me with the opportunity to undergo virtual industrial training. Despite the challenges posed by remote learning, the experience I gained during my time at Internshala has been invaluable in enhancing my understanding of Machine Learning

I am particularly grateful to Mr.Rajendra Prasad Pandey for their guidance, support, and patience throughout my training period. Their expertise and mentorship, even in a virtual environment, have been instrumental in my professional development.

I would also like to thank all the employees of Internshala who were involved in my training. Their willingness to share their knowledge and experiences through virtual platforms has been greatly appreciated. The collaborative environment allowed me to learn from diverse perspectives and gain practical insights. This experience has significantly boosted my confidence in applying my skills. I am deeply appreciative of the invaluable lessons and support I received during this training.

Name: Asim Husain
Date:

DECLARATION

We hereby declare that this Project Report titled **Stock Market Prediction** submitted by us and approved by our project guide, the College of Computing Sciences and Information Technology (CCSIT), Teerthanker Mahaveer University, Moradabad, is a bonafide work undertaken by us and it is not submitted to any other University or Institution for the award of any degree diploma / certificate or published any time before.

	Name	**Signature**
Student Name:	Asim Husain	
Project Guide: (External)		
Project Guide: (Internal)	Dr. Rajendra Prasad Pandey	

Brief About the Company

Internshala: Empowering Students and Young Professionals

Internshala is a premier online platform based in India, dedicated to bridging the gap between students and the professional world. Founded in 2010 by Sarvesh Agrawal, an IIT Madras alumnus, Internshala has grown exponentially to become a trusted resource for students and fresh graduates seeking internships, training, and career guidance. With its headquarters in Gurugram, India, and a team of 282 employees, Internshala is well-positioned to cater to the diverse needs of its users.

Mission and Vision

Internshala's mission is to enhance employability among the youth by providing meaningful internships and skill-building opportunities. The platform envisions a future where every student has access to practical experience and professional development, essential for a successful career.

Core Services

1. **Internship Opportunities:** Internshala offers a vast array of internship listings across various domains like engineering, management, media, design, and more. These internships provide hands-on experience with reputable companies and organizations, allowing students to apply their academic knowledge in real-world settings.
2. **Online Training Programs:** Recognizing the need for continuous learning, Internshala provides a wide range of online training programs. These courses cover topics such as programming, digital marketing, data science, soft skills, and more. Designed by industry experts, these courses ensure that students acquire the latest skills and knowledge to stay competitive in the job market.
3. **Career Services:** Internshala offers comprehensive career development resources, including resume building, interview preparation, and career counseling. These services equip students with the necessary tools to navigate the job market effectively and land their dream jobs.

Impact and Reach

Internshala has made a significant impact on the educational landscape in India. With over 8 million students and 100,000+ companies registered on the platform, it has facilitated countless internships and job placements. The platform's user-friendly interface and robust matching algorithms ensure that students find internships that align with their career goals and aspirations.

Internshala's commitment to quality and transparency has earned it the trust of both students and employers. The platform's extensive network includes partnerships with educational institutions and industry leaders, further enhancing its credibility and reach.

Corporate Social Responsibility

Internshala is dedicated to giving back to the community through various initiatives. The company's "Internshala Trainings Young Achiever Scholarship" offers free online training to underprivileged students, empowering them to build successful careers. Additionally, Internshala frequently conducts webinars, workshops, and campaigns aimed at promoting skill development and career readiness among the youth. These efforts ensure that more students and have access to quality education and career opportunities.

Innovation and Growth

Innovation is at the heart of Internshala's operations. The platform continually evolves to meet the changing needs of the job market. By leveraging the latest technology, Internshala provides personalized recommendations and seamless user experiences. Its commitment to continuous improvement ensures it remains a leader in the internship and training space.

Financial and Market Presence

Internshala generates an annual revenue of ₹29.1 crore (approximately $3.63 million) as of March 31, 2023. This substantial revenue stream reflects the platform's widespread influence and success in connecting students with valuable career opportunities. The financial stability of Internshala allows it to invest in further technological advancements and expand its services, ensuring that it remains a top choice for students and employers alike.

Conclusion

In conclusion, Internshala is a transformative force in the realm of education and career development in India. By connecting students with valuable internships and providing essential skill-building opportunities, Internshala is paving the way for a brighter, more skilled workforce. As it continues to expand and innovate, Internshala is well-positioned to remain a key player in shaping the future of work for young professionals. Its commitment to quality, innovation, and social responsibility has set a benchmark in the industry, making it a trusted and reliable platform for both students and employers.

Table Of Content

1.Project Title

Stock Market Prediction: A Machine Learning-Based Prediction Model.

2.Problem Statement

Without tools for predicting stock market trends, investors encounter numerous challenges. Making informed decisions becomes difficult as they often rely on speculation, news snippets, or outdated advice, leading to risky choices and potential financial setbacks. The stock market's volatility means that without predictive capabilities, individuals must navigate frequent fluctuations with minimal guidance. This unpredictability complicates the process of deciding when to buy or sell stocks, often resulting in emotional decisions like panic-selling during downturns or purchasing stocks at their peak prices.

Moreover, the absence of predictive tools forces investors to spend considerable time and effort analysing market data independently, which can be overwhelming, especially for those without a robust financial background. This not only increases the chance of errors but also makes investing a daunting task for beginners. Investors also miss out on maximizing their profits as they cannot accurately forecast which stocks will perform well.

In summary, the lack of stock market prediction tools makes managing investments a stressful and uncertain endeavour. It hinders individuals' ability to grow their wealth and achieve financial stability, often leading to hesitation and a reluctance to invest, thus missing out on the potential benefits that a well-informed investment strategy could offer

3.Problem Description

A Stock Market Prediction model is an advanced analytical tool designed to forecast future stock prices by analysing historical data, market trends, and various influencing factors. These models utilize techniques such as machine learning algorithms, statistical methods, and artificial intelligence to process large data volumes and identify patterns indicating future price movements.

The primary use of these models is to help investors make more informed decisions about buying, selling, or holding stocks.

The significance of Stock Market Prediction models is immense. They reduce investment uncertainty and risk by offering data-driven insights and forecasts, enabling investors to make decisions based on objective analysis rather than emotions. Accurate predictions lead to improved investment returns, better portfolio management, and the ability to anticipate market trends.

In addition to aiding individual investors, financial institutions, hedge funds, and trading firms also use these models to optimize their trading algorithms and strategies, potentially yielding higher returns.

Overall, these models contribute to more efficient and effective investment strategies, helping investors achieve financial goals while managing risk more effectively. As technology advances, the capabilities of Stock Market Prediction models will continue to grow, further revolutionizing our understanding and interaction with financial markets.

3.1. Scope of the Work

The scope of this project involves collecting and preprocessing historical stock market data, performing exploratory data analysis (EDA), and implementing machine learning models like Linear Regression, Decision Trees, and LSTM for time series forecasting. The project focuses on optimizing model performance through hyperparameter tuning and evaluating results using metrics such as RMSE and MAE. Additionally, it aims to explore various feature engineering techniques to enhance predictive accuracy, providing actionable insights for investors and highlighting areas for future improvements.

3.2. Project Modules

1. Data Collection

- Fetch historical stock price data from APIs like Yahoo Finance, Alpha Vantage, or Quandl.
- Integrate other relevant data sources like news sentiment or macroeconomic indicators.

2. Data Preprocessing

- Clean data by handling missing values and outliers.
- Normalize and scale features to prepare for modelling.
- Feature engineering to create additional predictive features.

3. Exploratory Data Analysis (EDA)

- Visualize data trends, correlations, and patterns.
- Analyse seasonality, volatility, and impact of market events.

4. Model Development

- Implement various models such as Linear Regression, Decision Trees, ARIMA, and LSTM.
- Develop technical indicator-based models for enhanced predictions.

5. Model Training and Evaluation

- Split data into training, validation, and testing sets.
- Train models and evaluate their performance using like RMSE, MSE

6. Hyperparameter Tuning

- Optimize model parameters using techniques like Grid Search, Random Search, or Bayesian Optimization.

7. Prediction and Analysis

- Generate future Stock Market Predictions.
- Compare results across different models and analyse their accuracy.

8. Visualization and Reporting

- Create dashboards or visual reports to present model predictions and insights.
- Provide interactive plots to explore predictions vs. actual performance.

9. Deployment Module

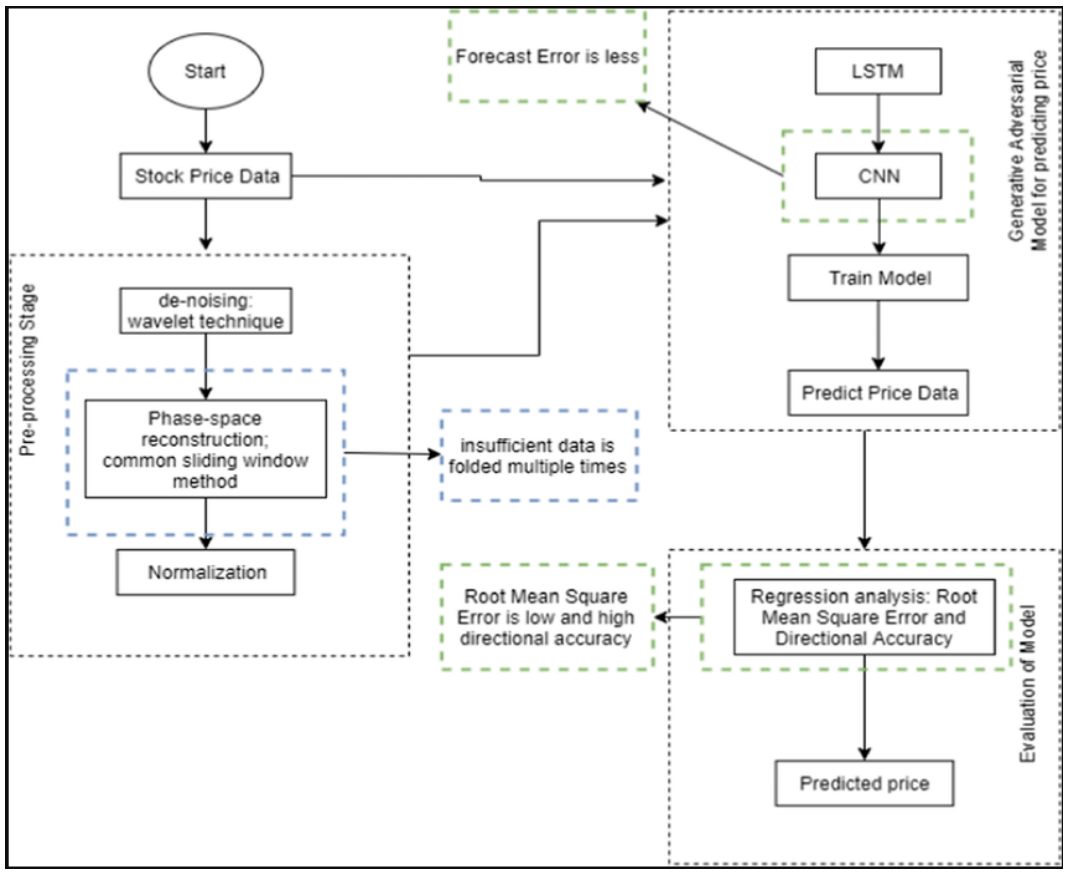
- Deploy the prediction model in a web app or dashboard for real-time predictions.

10. Conclusion and Future Work

- Summarize findings, discuss limitations and suggest areas for future improvements

3.3. Context Diagram

Level 1



4. Implementation Methodology

1. Planning and Requirement Analysis

- Define Objectives: Clearly define the objectives of the Stock Market Prediction, such as predicting future prices, forecasting trends, or making buy/sell recommendations.
- Stakeholder Requirements: Gather requirements from stakeholders, including investors, financial analysts, or academic needs for the project.
- Feasibility Study: Assess technical, operational, and financial feasibility. Determine data availability and the computational resources needed for machine learning models.

2. System Design

- **Architecture Design:** Design the architecture of the system, including data flow, model selection, and integration points. This involves specifying how data will be collected, processed, and fed into predictive models.
- **Module Design:** Outline key modules such as data collection, preprocessing, model training, evaluation, and visualization.
- **Technology Stack:** Choose the technology stack, including programming languages (Python, R), libraries (TensorFlow, scikit-learn), and data storage solutions (SQL, NoSQL databases).

3. Implementation (Coding)

- **Data Collection:** Implement scripts to fetch historical stock data using APIs like Yahoo Finance or Alpha Vantage.
- **Preprocessing Pipeline:** Develop data cleaning, normalization, and feature engineering pipelines to prepare data for modelling.
- **Model Development:** Code various machine learning and deep learning models such as LSTM, ARIMA, or XGBoost.
- **Hyperparameter Tuning:** Implement tuning mechanisms using Grid Search or Random Search to optimize model performance.

4. Testing

- **Unit Testing:** Test individual components like data preprocessing functions, model training, and evaluation modules to ensure they work as expected.
- **Integration Testing:** Test the integrated system to verify that data flows correctly through preprocessing, modelling, and evaluation steps.
- **Validation Testing:** Validate model predictions against a validation dataset to ensure accuracy and reliability.

5. Deployment

- **Deployment Strategy:** Decide on deployment methods (local deployment, cloud deployment on AWS, GCP, etc.).
- **Real-time Prediction Setup:** If needed, set up real-time prediction capabilities using APIs or web interfaces (Flask, Django).
- **Monitoring:** Implement monitoring to track model performance over time and detect any drifts or anomalies.

6. Maintenance

- **Model Updates:** Regularly update models with new data to keep predictions accurate.
- **Performance Monitoring:** Continuously monitor performance metrics to identify when retraining is necessary.
- **Bug Fixes and Enhancements:** Address any bugs or improve features based on feedback.

7. Documentation

- **User Documentation:** Provide clear documentation for end-users explaining how to interpret model results.
- **Technical Documentation:** Create technical documents detailing data sources, preprocessing steps, model algorithms, and code structure for developers.

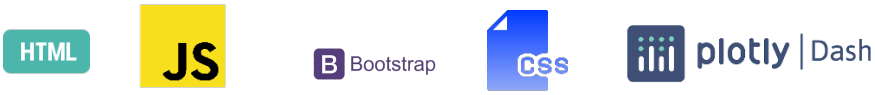
8. Evaluation and Feedback

- **Review Results:** Evaluate overall project performance, comparing the predicted results with actual outcomes.
- **Stakeholder Feedback:** Gather feedback from users and stakeholders to understand the effectiveness and areas for improvement.
- **Iterative Improvements:** Use feedback for continuous improvement of the system.

5. Technology to be Used

5.1. Software Platform

a) Front-End



b) Backend-End

Python 3.12.6 (backend programming, data processing, and ML model development.)

APIs

- Yahoo Finance API (fetching stock market data)

Libraries

- Pandas (for data manipulation and preprocessing)
- NumPy (for numerical computations and array handling)
- Streamlit (for creating the interactive web application)
- Keras (for loading and using the machine learning model)
- Datetime (for handling dates)
- Sklearn.Preprocessing (MinMaxScaler for scaling the data)
- Matplotlib (for data visualization and plotting)
- TensorFlow (TensorFlow is a powerful lib for machine learning and neural networking)

5.2. Hardware Platform

- Windows
- Linux
- MacOS

5.3. Tools

Latest Versions

- **Git** (For managing code changes and version control)
- **GitHub** (For hosting code repositories and collaboration)
- **Jupyter Lab** (For interactive development and experimentation with data analysis and model training)
- **PyCharm / VS Code** (IDEs used for coding, debugging, and testing backend code)

6. Advantages of the Project

- **Informed Decisions:** Utilizes data-driven insights for better investment choices, reducing reliance on guesswork.
- **Risk Management:** Helps identify and mitigate investment risks with timely predictions.
- **Accuracy:** Machine learning enhances prediction accuracy by analysing large datasets.
- **Time Efficiency:** Automates data analysis, saving investors valuable time.
- **User-Friendly:** Accessible interface via Streamlit for all users, regardless of technical expertise.
- **Customizable Insights:** Tailors analyses to specific stocks and parameters, aligning with individual goals.
- **Continuous Improvement:** Model adapts and improves with new data.
- **Visualization:** Provides clear visual representations of trends and predictions.
- **Comprehensive Analysis:** Incorporates various factors for a holistic market view.
- **Scalability:** Flexible and adaptable to different markets and investment needs

7. Assumptions

- **Historical Data Relevance:** Past stock prices and patterns can help predict future price movements.
- **Market Efficiency:** The market is semi-efficient, allowing patterns to be detected before they fully impact stock prices.
- **Stationarity:** The time series data is stationary or can be transformed to be stationary for accurate model predictions.
- **Model Accuracy:** The quality and recency of data are crucial for accurate predictions.
- **External Factors Considered:** The model reasonably accounts for external factors like economic indicators and company performance.

8. Future Scope and Further Enhancement of the Project

- **Integration of Sentiment Analysis**: Use news and social media data to improve prediction accuracy.
- **Advanced Machine Learning Models**: Explore newer models like Transformers and reinforcement learning.
- **Real-Time Data Processing**: Implement live data streaming for real-time predictions.
- **Scalability to Multi-Asset Classes**: Extend predictions to cryptocurrencies, commodities, and forex.
- **Cloud Deployment**: Deploy models on cloud platforms for better scalability and accessibility.
- **Interactive Dashboards**: Develop user-friendly dashboards for better visualization and interpretation.
- **Algorithmic Trading**: Integrate predictions with automated trading systems for immediate execution

9.Project Repository Location

Sn	Project Artifacts (Softcopy)	Location (Mention lab ID, Folder Location etc)	Verified By Project Guide	Verified By Lab Incharge
1.	Project Synopsis Report (Final Version)			
2.	Project Progress updates			
3.	Project Requirement specifications			
4.	Project Report (Final Version)			
5.	Test Repository			
6.	Any Other Document, Give Details			

10. Definitions of Acronyms and Abbreviations

1. **AI - Artificial Intelligence:** The simulation of human intelligence processes by machines, especially computer systems.
2. **API - Application Programming Interface:** A set of protocols and tools for building software and applications, allowing different systems to communicate.
3. **CSS - Cascading Style Sheets:** A stylesheet language used for describing the presentation of a document written in HTML or XML.
4. **HTML -** The standard language for creating web pages and web applications.
5. **HTTP - Hypertext Transfer Protocol:** The foundation of any data exchange on the web, and it is a protocol used for transmitting hypertext requests and information.
6. **HTTPS - Hypertext Transfer Protocol Secure:** An extension of HTTP. It is used for secure communication over a computer network.
7. **ML - Machine Learning:** A type of artificial intelligence that allows software applications to become more accurate in predicting outcomes without being explicitly programmed to do so.
8. **NLP - Natural Language Processing:** A field of AI that gives machines the ability to read, understand, and derive meaning from human languages.
9. **SQL - Structured Query Language:** A standard programming language used to manage relational databases and perform various operations on the data in them.
10. **UI - User Interface:** The means by which a user and a computer system interact, specifically the use of input devices and software.

11. Conclusion

In conclusion, the stock market prediction project utilizes advanced machine learning models, effective data processing, and interactive web applications to provide valuable insights for investors. By leveraging historical data and predictive analytics, the project aims to enhance decision-making, mitigate risks, and improve investment outcomes. The integration of various tools and techniques, such as Keras, TensorFlow, Streamlit, and yfinance, ensures a robust and user-friendly platform. Overall, this project serves as a powerful tool for investors to navigate the complexities of the stock market and achieve better financial stability and growth.

12. Reference

Data Camp

- https://www.datacamp.com/courses/machine-learning-with-python

Internshala

- https://trainings.internshala.com/machine-learning

YouTube:

- https://youtu.be/OXwZtlcTiuk
- https://shorturl.at/xLCOT

Google:

- https://shorturl.at/NapFT
- https://www.geeksforgeeks.org/libraries-in-python/

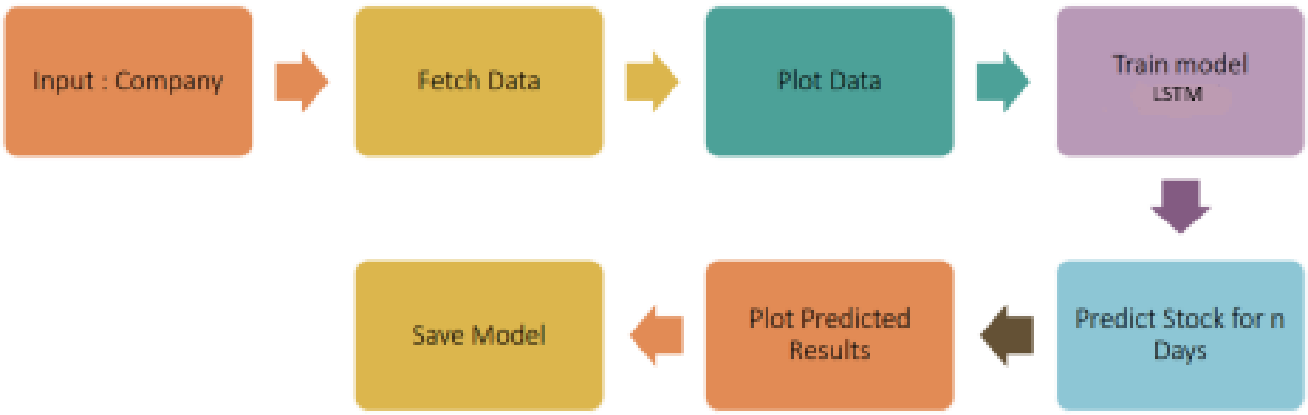
GitHub:

- https://rb.gy/oqchkf
- https://github.com/alisonmitchell/Stock-Prediction.git

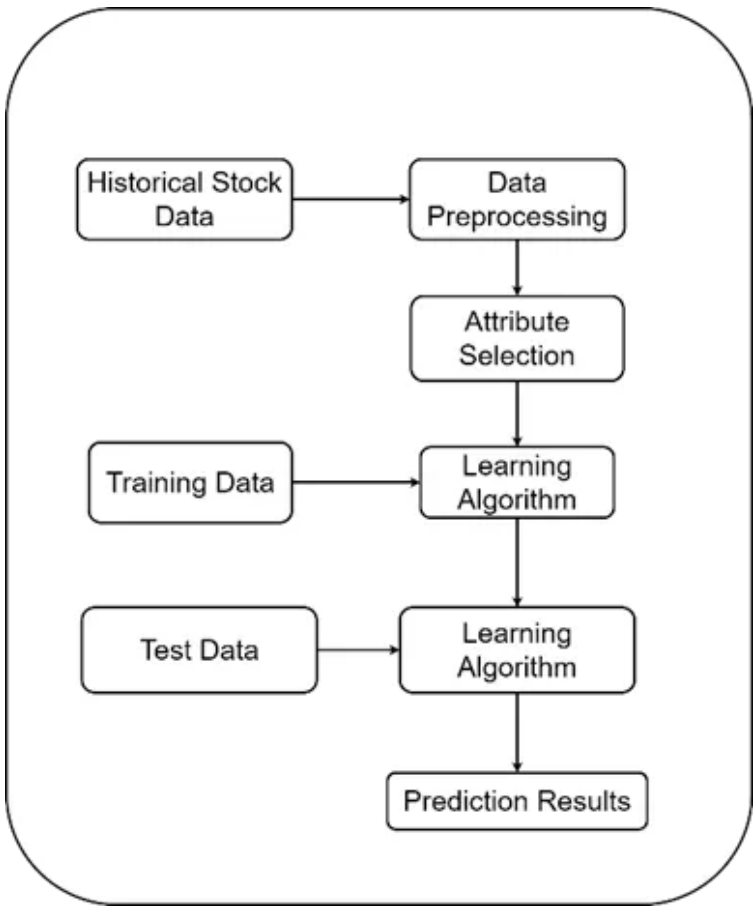
https://github.com/Rajat-dhyani/Stock-Price-Predictor.git

Data Flow Diagram (DFD)

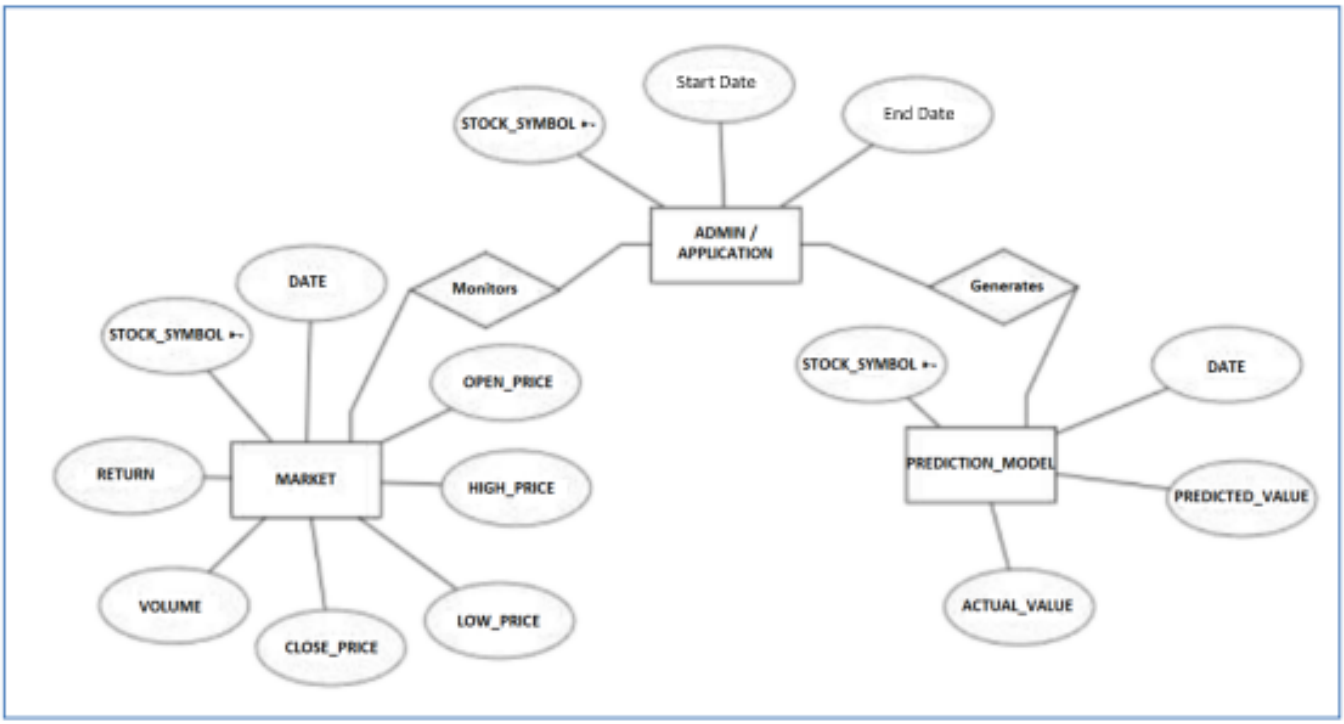
DFD – 1



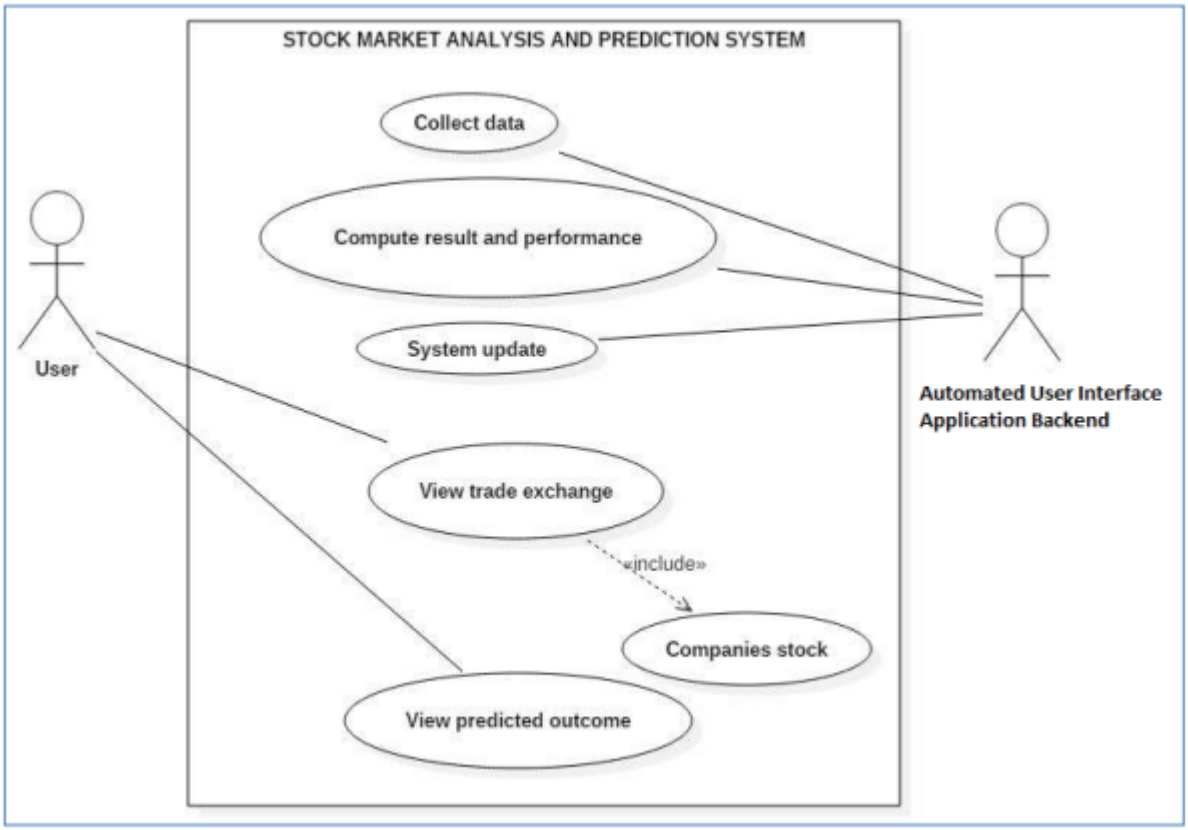
DFD – 2



Entity Relationship Diagram (ERD)



Use Case Diagram (UCD)



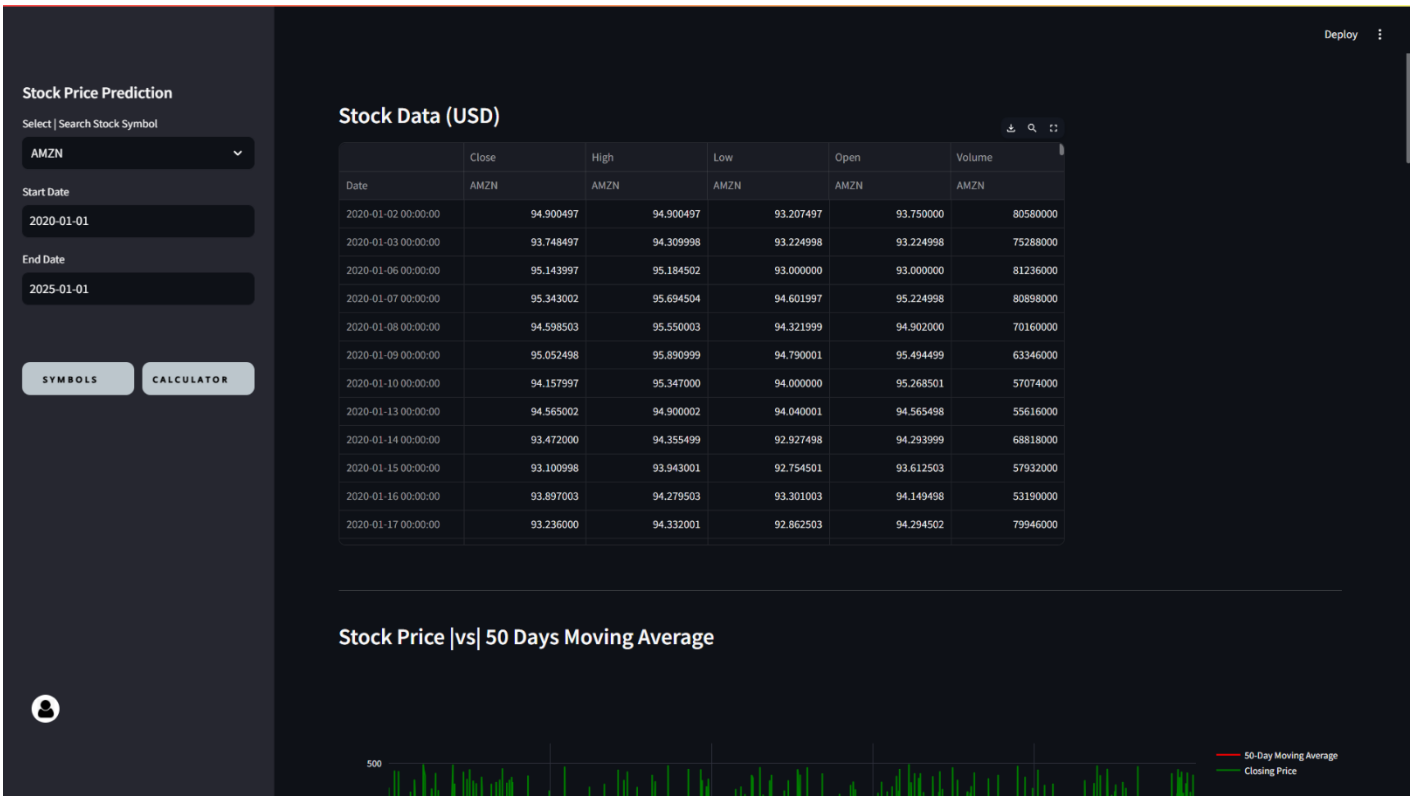
Data Dictionary (DD)

User Table (USR)

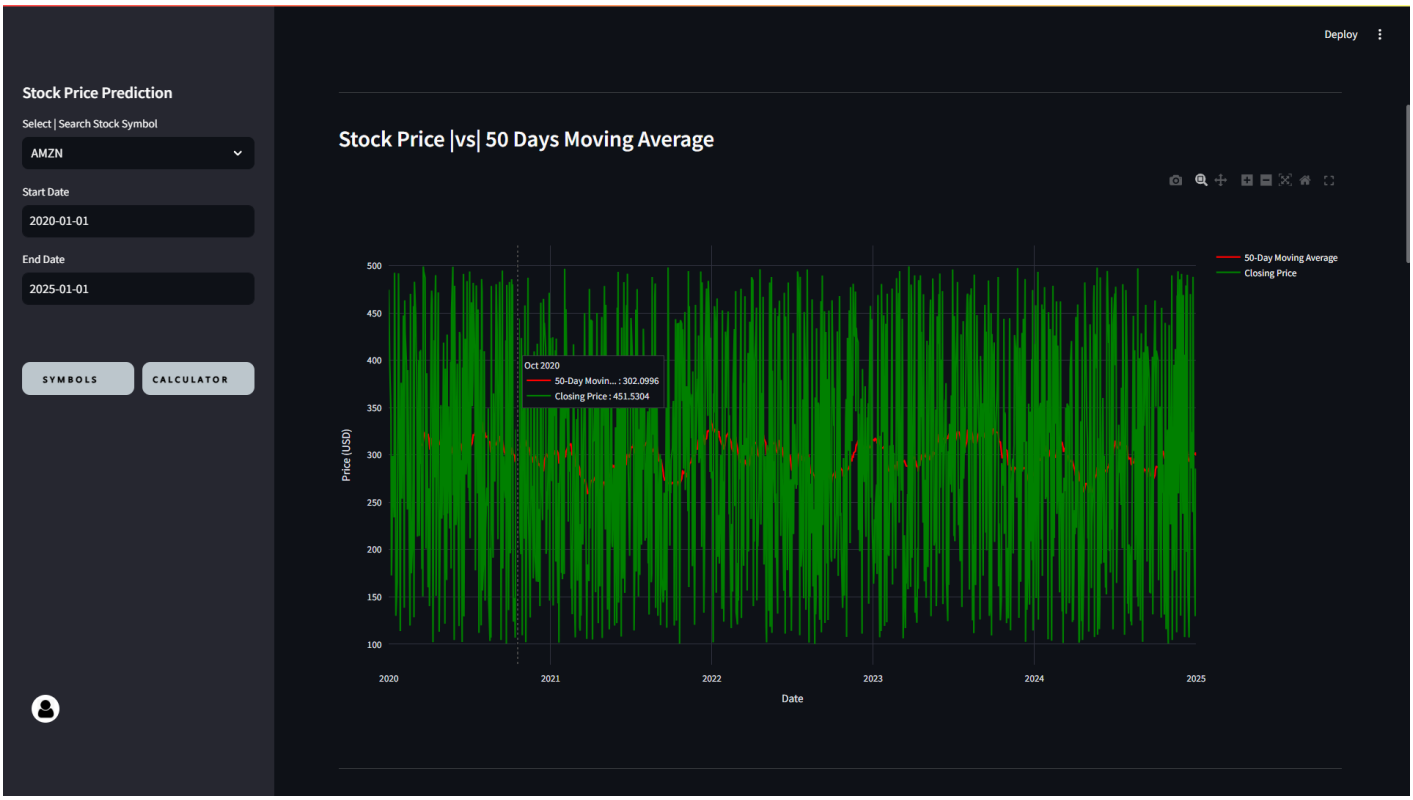
Fields	Data Type	Description
USR-Name	Text	Asim Husain
USR-Contact-No	Number	+91 7310647227
USR-Address	Text	Shahpur Sirpura, Sambhal, UP, India

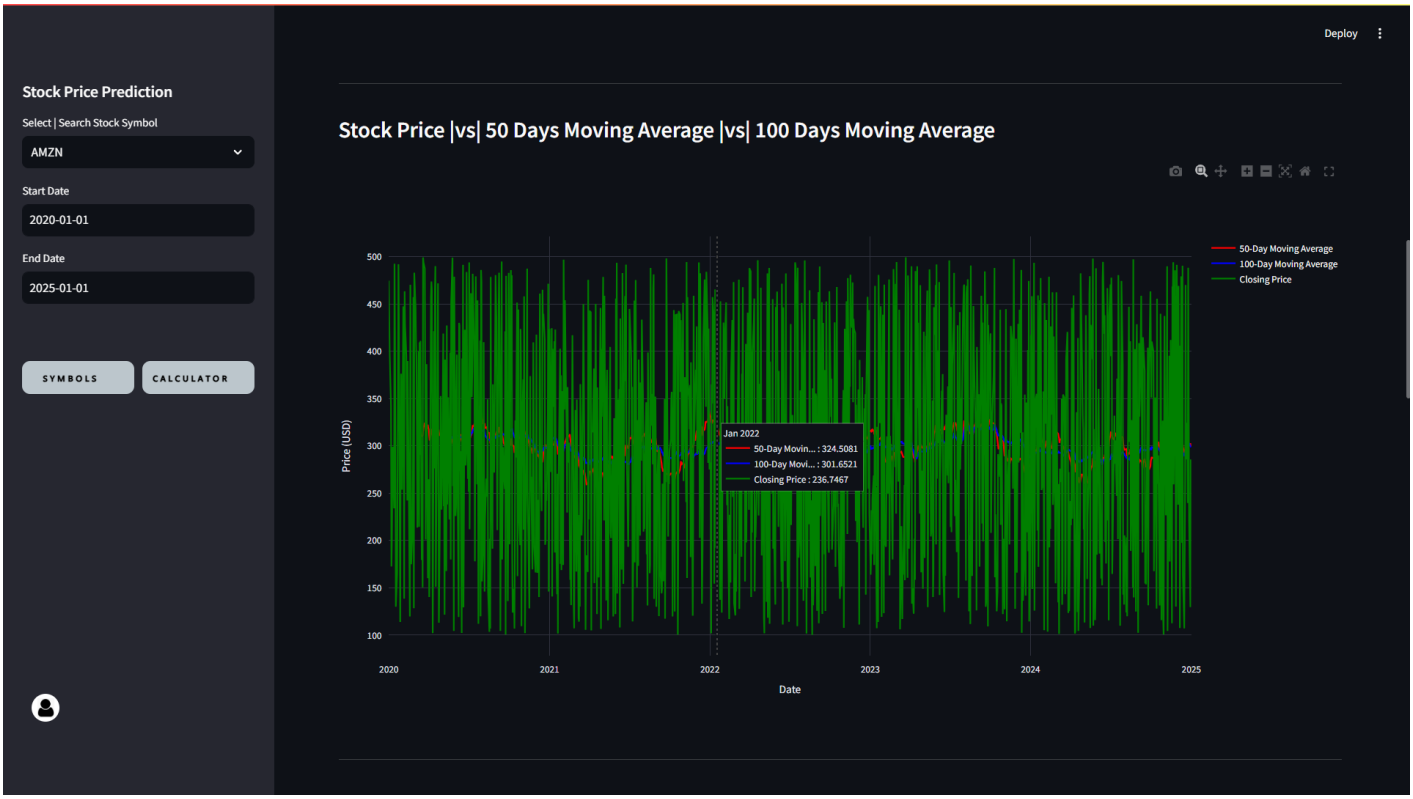
Screenshots

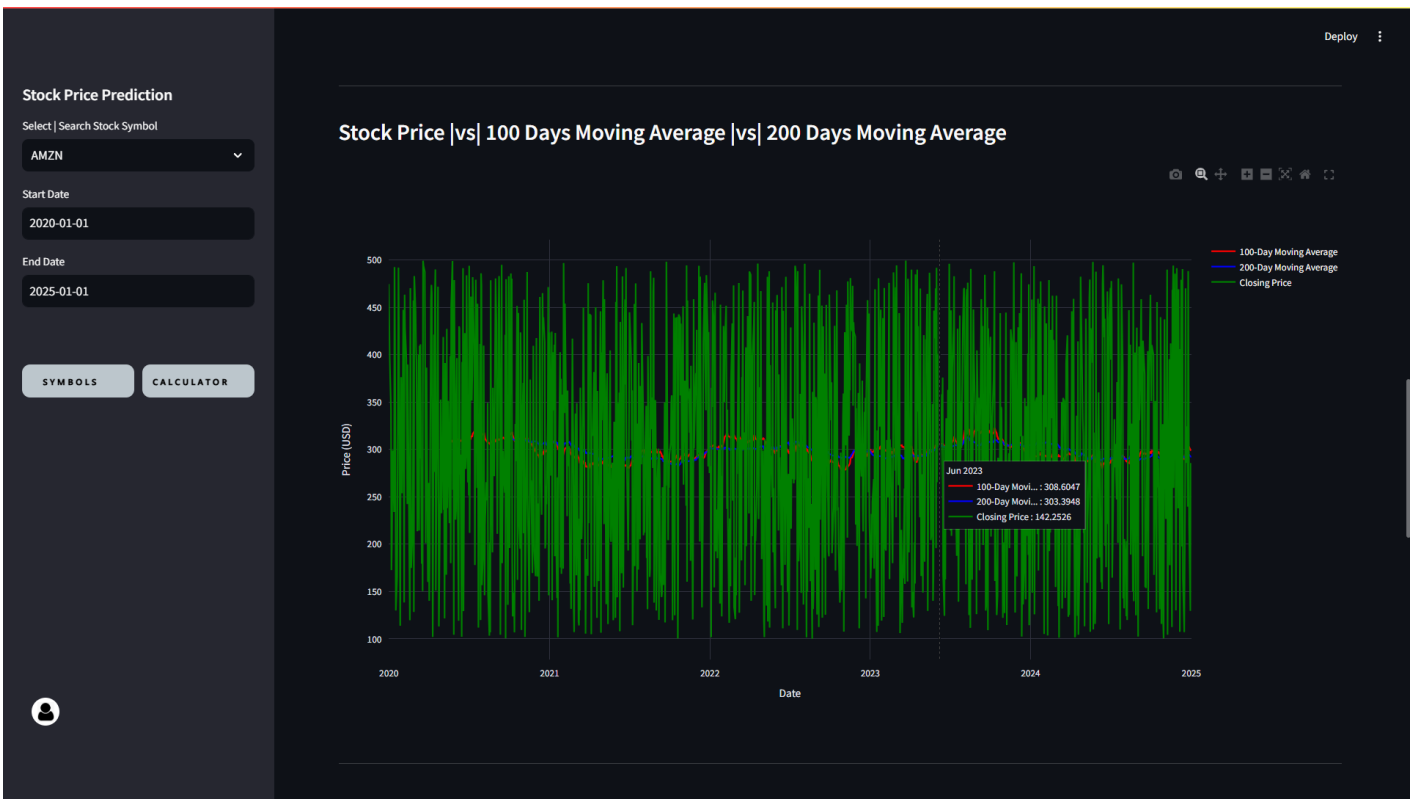
Amazon (AMZN)

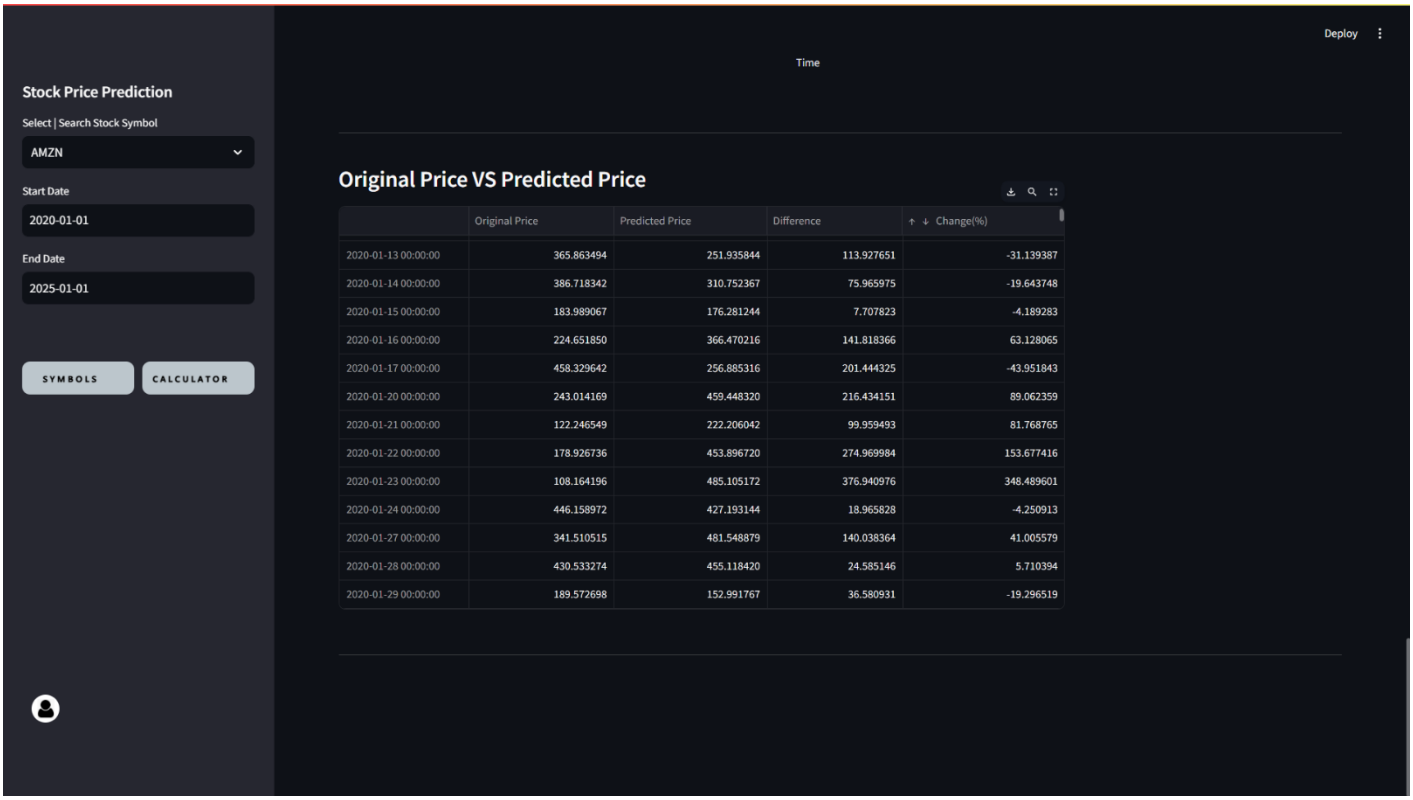


	Close	High	Low	Open	Volume
Date	AMZN	AMZN	AMZN	AMZN	AMZN
2020-01-02 00:00:00	94.900497	94.900497	93.207497	93.750000	80580000
2020-01-03 00:00:00	93.748497	94.309998	93.224998	93.224998	75288000
2020-01-06 00:00:00	95.143997	95.184502	93.000000	93.000000	81236000
2020-01-07 00:00:00	95.343002	95.694504	94.601997	95.224998	80898000
2020-01-08 00:00:00	94.598503	95.550003	94.321999	94.902000	70160000
2020-01-09 00:00:00	95.052498	95.890999	94.790001	95.494499	63346000
2020-01-10 00:00:00	94.157997	95.347000	94.000000	95.268501	57074000
2020-01-13 00:00:00	94.565002	94.900002	94.040001	94.565498	55616000
2020-01-14 00:00:00	93.472000	94.355499	92.927498	94.293999	68818000
2020-01-15 00:00:00	93.100998	93.943001	92.754501	93.612503	57932000
2020-01-16 00:00:00	93.897003	94.279503	93.301003	94.149498	53190000
2020-01-17 00:00:00	93.236000	94.332001	92.862503	94.294502	79946000




Deploy
Stock Price Prediction
Select | Search Stock Symbol
AMZN
Start Date
2020-01-01
End Date
2025-01-01
SYMBOLS
CALCULATOR
Stock Price |vs| 50 Days Moving Average |vs| 100 Days Moving Average
50-Day Moving Average
100-Day Moving Average
Closing Price
Jan 2022
50-Day Movin... : 324.5081
100-Day Movi... : 301.6521
Closing Price : 236.7467
Price (USD)
Date


Deploy
Stock Price Prediction
Select | Search Stock Symbol
AMZN
Start Date
2020-01-01
End Date
2025-01-01
SYMBOLS
CALCULATOR
Stock Price |vs| 100 Days Moving Average |vs| 200 Days Moving Average
100-Day Moving Average
200-Day Moving Average
Closing Price
Jun 2023
100-Day Movi... : 308.6047
200-Day Movi... : 303.3948
Closing Price : 142.2526
Price (USD)
Date



	Original Price	Predicted Price	Difference	Change(%)
2020-01-13 00:00:00	365.863494	251.935844	113.927651	-31.139387
2020-01-14 00:00:00	386.718342	310.752367	75.965975	-19.643748
2020-01-15 00:00:00	183.989067	176.281244	7.707823	-4.189283
2020-01-16 00:00:00	224.651850	366.470216	141.818366	63.128065
2020-01-17 00:00:00	458.329642	256.885316	201.444325	-43.951843
2020-01-20 00:00:00	243.014169	459.448320	216.434151	89.062359
2020-01-21 00:00:00	122.246549	222.206042	99.959493	81.768765
2020-01-22 00:00:00	178.926736	453.896720	274.969984	153.677416
2020-01-23 00:00:00	108.164196	485.105172	376.940976	348.489601
2020-01-24 00:00:00	446.158972	427.193144	18.965828	-4.250913
2020-01-27 00:00:00	341.510515	481.548879	140.038364	41.005579
2020-01-28 00:00:00	430.533274	455.118420	24.585146	5.710394
2020-01-29 00:00:00	189.572698	152.991767	36.580931	-19.296519

INTERNSHALA TRAININGS

Certificate of Training

Asim Husain

has successfully completed a 6-week online training on **Machine Learning**. The training consisted of Introduction to Machine Learning, Data, Introduction to Python, Data Exploration and Pre-processing, Linear Regression, Introduction to Dimensionality Reduction, Logistic Regression, Decision Tree, Ensemble Models, Clustering (Unsupervised Learning), and Machine Learning with AI modules.
Asim scored 100% marks in the final assessment and is a top performer in the training.
We wish Asim all the best for future endeavours.

Sarvesh Agrawal
FOUNDER & CEO, INTERNSHALA

IITM PRAVARTAK CATALYSING INNOVATION

Dr. Shankar Raman
CEO, IITM PRAVARTAK TECHNOLOGIES FOUNDATION

Date of certification: 2024-08-13

Certificate no. : 3c8qenodme0

For certificate authentication, please visit https://trainings.internshala.com/verify_certificate