PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Sentiment-Augmented Stock Price Forecasting

Group No. 14

Group members	Student Id
Sumit Shahu	21107004
Pravesh Yadav	21107057
Ankit Purohit	21107020
Mustafa Shaikh	21107045

Project Guide

Ms. Ujwala Pagare & Aavani Nair

Contents

Abstract

The project aims to develop a real-time stock market prediction system using a hybrid deep learning approach combining RNN variants (LSTM, GRU). The system addresses the challenges of noisy, high-frequency data and aims to provide accurate and low-latency predictions for stock price movements. By integrating real-time data streams from sources like stock market APIs and news sentiment, the system leverages ensemble learning techniques to boost prediction accuracy. An ensemble model is built by combining LSTM, GRU, and traditional machine learning methods like XGBoost, further improving robustness and accuracy. The system is designed to handle the complex, sequential nature of stock market data, providing traders and investors with timely, actionable insights.

Introduction

- The stock market is a key financial system that influences global economic stability and individual investments.
- Stock prices fluctuate rapidly due to a variety of factors including economic events, news, geopolitical changes, and investor sentiment.
- Accurately predicting stock prices is difficult due to the market's highly volatile and noisy nature.
- In real-time, the challenge becomes even harder because millions of trades and price changes happen every second.
- Traditional financial models struggle with real-time data, facing issues such as scalability, latency in predictions, and the dynamic nature of time-series data.
- Despite advances in financial technology, the complex and fast-changing stock market demands more sophisticated, adaptable tools for real-time prediction and analysis.

Motivation

- Modern stock markets are influenced by multiple volatile factors, making accurate predictions challenging.

- Unpredictable market behavior leads to higher risks, affecting decision-making and financial stability.

- Conventional models fail to account for dynamic, non-linear market behaviors, reducing prediction accuracy.

- Investors and financial institutions increasingly seek advanced models that can provide timely and reliable market forecasts to optimize decision-making and reduce risk.

Objectives

- Develop and evaluate hybrid LSTM and GRU models for stock price forecasting, leveraging their ability to capture both spatial and temporal patterns in market data.
- Design and implement an ensemble stacking model that combines LSTM, GRU, and RNN architectures to enhance prediction accuracy and robustness across various market conditions.
- Integrate sentiment analysis using BERT and VADER techniques to incorporate real-time news data, assessing the impact of market sentiment on stock price predictions.
- Compare the performance of deep learning models (LSTM, GRU, RNN) and traditional methods (Random Forest, Gradient Boosting) in terms of accuracy, risk reduction, and real-world applicability.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
1	Stock Market Prediction With Transductive Long Short-Term Memory and Social Media Sentiment Analysis	Ali Peivandizadeh, Sima Hatami, Amirhossein Nakhjavani, Lida Khoshsima, Mohammad Reza Chalak Qazani, Muhammad Haleem, Roohallah Alizadehsani	2024	1. Combines TLSTM (Transductive LSTM) for stock price prediction with social media sentiment analysis. 2. Uses Off-Policy Proximal Policy Optimization (PPO) to handle class imbalances.	1. Over-reliance on sentiment from social media. 2. Complex architecture requires significant computational resources. 3. Data imbalances in sentiment analysis can still affect the model's performance.
2	A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning	Guangyu Mu, Nan Gao, Yuhan Wang, Li Dai	2023	1. MS-SSA-LSTM model integrates multi-source data, such as: Investor sentiment from forums , Stock trading data. 2. Optimizes LSTM (Long Short-Term Memory) using the Sparrow Search Algorithm. 3. Uses a sentiment dictionary to improve prediction accuracy.	1. model heavily depends on sentiment data 2. Sentiment categorization is limited to positive and negative emotions. 3. Lacks integration of more complex emotional aspects such as fear, anger, etc.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
3	A Deep Reinforcement Learning-Based Decision Support System for Automated Stock Market Trading	Yasmeen Ansari, Sadaf Yasmin, Sheneela Naz, Hira Zaffar, Zeeshan Ali, Jihoon Moon, Seungmin Rho	2022	1. Deep Reinforcement Learning (DRL) framework using Deep-Q Networks with Gated Recurrent Unit (GRU) forecasting. 2. The model uses both past and future stock trends to make trading decisions.	1. Model performance is volatile and may vary with different training attempts. 2. Lacks robustness in highly volatile markets, and future price predictions are sometimes inaccurate, affecting trading decisions.
4	Decision Fusion for Stock Market Prediction: A Systematic Review	Cheng Zhang, Nilam N. A. Sjarif, Roslina B. Ibrahim	2022	Decision fusion: Combines forecasts from multiple models using base learners and fusion methods	1. Few studies use decision fusion. 2. Lack of diversity in ensemble models. 3. Minimal integration of sentiment analysis.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
5	Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory	Xingqi Wang, Kai Yang, Tailian Liu	2021	K-means clustering with Morphological Similarity Distance (MSD) to group stocks, followed by Hierarchical Temporal Memory (HTM) for prediction	1. . The method may not generalize well to long-term predictions. 2. Multivariate input had minimal benefit for HTM. 3. Potential underfitting or overfitting of baseline models due to lack of parameter tuning.
6	A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model	Md. Ebtidaul Karim, Sabrina Ahmed	2021	BiGRU (Bidirectional Gated Recurrent Unit) with activation layer and BiLSTM (Bidirectional Long Short Term Memory).	1. Increased complexity with more hidden layers. 2. Trainable parameters grow with additional layers, which can increase computation cost. 3. Limited to comparison between BiGRU and BiLSTM models.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
7	Literature Survey on Stock Price Prediction Using Machine Learning	Anusha J Adhikar, Apeksha K Jadhav, Charitha G, Karishma KH, Mrs. Supriya HS	2020	1. Machine Learning: Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM). 2. Data collected from Yahoo Finance.	1. Python libraries used were not optimal for training speed. 2. Need for optimization of neural network parameters.
8	Augmented Textual Features-Based Stock Market Prediction	Salah Bouktif, Ali Fiaz, Mamoun Awad	2020	Sentiment analysis using tweets and stock data combined with machine learning techniques such as SVM, Naive Bayes, ANN, and XGBoost with model stacking.	1. Limited by the text features extraction approach, 2. Detailed analysis of sentiments is needed to improve prediction accuracy.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
9	Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Timeseries	Sidra Mehtab, Jaydip Sen	2019	1. Machine Learning: Logistic Regression, KNN, Decision Tree, Bagging, Boosting, Random Forest, ANN, SVM. 2. Deep Learning: CNN with three approaches: univariate, multivariate, and multi-model. 3. Data from NIFTY 50 index (2015-2019).	1. Difficulty in predicting rapid stock price changes. 2. Performance varies with data size and model settings. 3. Higher prediction errors (RMSE) in some CNN configurations.
10	A Prediction Approach for Stock Market Volatility Based on Time Series Data	Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal	2019	ARIMA (Auto-Regressive Integrated Moving Average) model for time series data prediction	Limited to univariate data, assuming no external factors like social or economic events.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
11	Predicting Prices of Stock Market using Gated Recurrent Units (GRUs) Neural Networks	Mohammad Obaidur Rahman, Md. Sabir Hossain, Ta-Seen Junaid, Md. Shafiul Alam Forhad, Muhammad Kamal Hossen	2019	1. Gated Recurrent Units (GRUs) for stock price prediction 2. Mini-batch gradient descent 3. Real-time dataset from Yahoo Finance	Local minima problem and time complexity of stochastic gradient descent were reduced but still may arise.
12	Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network	Dang Lien Minh, Abolghasem Sadeghi-Niaraki, Huynh Duc Huy, Kyungbok Min, Hyeonjoon Moon	2018	Two-Stream Gated Recurrent Unit (TGRU) Network, Sentiment analysis with Stock2Vec, financial news data, and technical indicators.	1. Complexity of TGRU model, requiring long training times and computational resources 2. Limited to daily stock movements rather than intra-day trading.

Literature Review

Sr.no	Title	Author(s)	Year	Methodology	Drawback
13	Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks	Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj, Alexandros Iosifidis	2017	CNNs (Convolutional Neural Networks) and Intelligent normalization for stock prices.	1. Limited dataset (10 days, 5 stocks) 2. High noise in short-term changes; more data required for better generalization.

Research Gap(Limitations of existing systems)

1.Traditional Models:

- ARIMA: Time-series forecasting, lacks non-linear pattern recognition.
- GARCH: Focuses on volatility, not suitable for predicting price movements.

2.Machine Learning Models:

- SVM: Effective for classification but struggles with scalability.
- Random Forest, XGBoost: Accurate but limited in handling sequential data.

3.Deep Learning Models:

- LSTM: Handles sequential data, prone to overfitting.
- CNN: Extracts features but cannot capture temporal dependencies effectively.

Research Gap(Limitations of existing systems)

Challenges Still Faced:

- **Real-time Latency**: Systems fail to make timely predictions, affecting actionable insights.
- **Volatility & Noise**: Stock market data is highly volatile and noisy, which current models struggle to handle.
- **Generalization**: Models often fail to generalize to unseen market conditions.
- **Data Integration**: Limited ability to combine multiple data sources like sentiment analysis, technical indicators, and market trends.

Problem Definition

The stock market is highly volatile and difficult to predict in real-time due to fluctuating data, news events, and investor sentiment. Existing models lack the capability to process real-time data efficiently, handle complex dependencies, and incorporate multiple data sources, resulting in inaccurate predictions and delayed insights.

- **Real-time Data Processing**: Need for continuous data processing to make timely predictions.
- **Accuracy**: Existing models struggle to capture complex patterns in noisy and volatile market data.
- **Multiple Data Sources**: Combining stock prices, technical indicators, and sentiment analysis from news and social media.
- **Scalability**: The system should handle high-frequency data and make predictions with low latency.

Scope

- Real-time forecasting allows traders to adjust strategies in response to sudden market changes or emerging trends.

- Helps firms monitor trading activities in real time to ensure they follow regulations and quickly identify and report any suspicious behavior.

- Enables ongoing evaluation of investment strategies and market models to refine and enhance forecasting methods.

- Enables the development of automated trading strategies that can execute trades based on real-time data.

Technological Stack

Real-time News Headlines Analysis

- Selenium and Beautiful Soup: Used for Web Scraping
- RNN (Recurrent Neural Network): For real-time news analysis

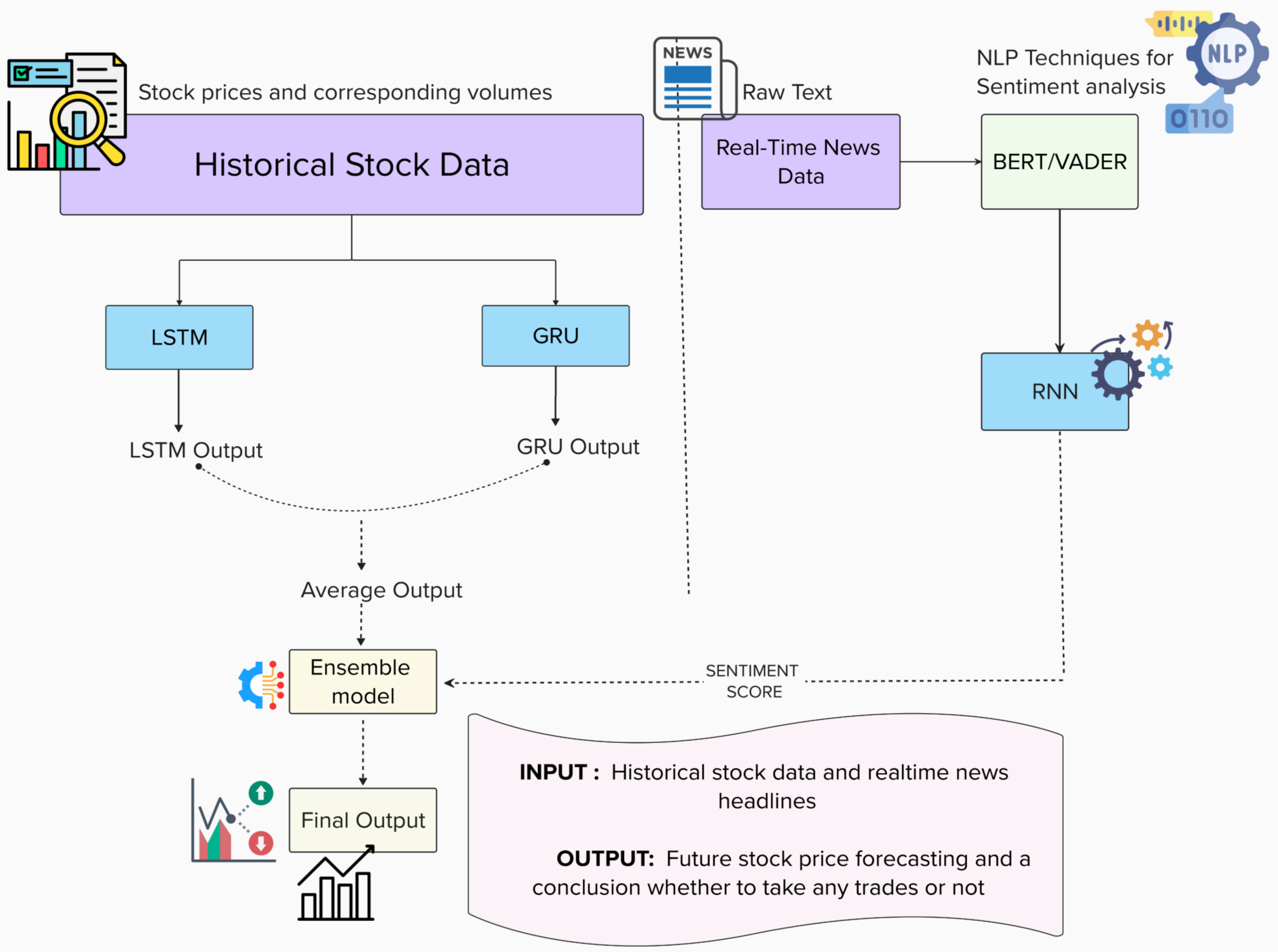
Core Prediction Models

- LSTM (Long Short-Term Memory)
- GRU (Gated Recurrent Unit)

Ensemble Model

- Combines outputs of LSTM, GRU and RNN models

Proposed system architecture/Working



References

[1]J. Liu, Z. Wang, and B. Zheng, "A Deep Reinforcement Learning-Based Decision Support System for Automated Stock Market Trading," IEEE Access, Volume 8, Pages 143676-143686, 2020.
https://doi.org/10.1109/ACCESS.2020.3012047

[2]Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," IEEE Access, Volume 7, Pages 17287-17295, 2019.
https://doi.org/10.1109/ACCESS.2019.2895252

[3] Jinyang Du, John S. Kimball, Justin Sheffield, Ming Pan, Colby K. Fisher, Hylke E. Beck, Eric F. Wood, "Satellite Flood Inundation Assessment and Forecast Using SMAP and Landsat," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 14, Pages 6707-6715, 2021.
https://ieeexplore.ieee.org/document/9465655

References

[4]Xingqi Wang, Kai Yang, and Tailian Liu, "Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory," IEEE Access, Volume9, Pages 67241-67248, 2021.
https://doi.org/10.1109/ACCESS.2021.3077004:contentReference[oaicite:0]{index=0}

[5]Shwetha Salimath, Triparna Chatterjee, Titty Mathai, Pooja Kamble, Megha Kolhekar, "Prediction of Stock Price for Indian Stock Market: A Comparative Study using LSTM and GRU," Proceedings of the International Conference on Advances in Computing and Data Sciences (ICACDS), Pages 123-130, 2020 .
https://ieeexplore.ieee.org/document/9465655

[6] Salah Bouktif, Ali Fiaz, and Mamoun Awad, "Augmented Textual Features-Based Stock Market Prediction," IEEE Access, Volume 8, Pages 40269-40279, 2020.
https://doi.org/10.1109/ACCESS.2020.2976725:contentReference[oaicite:1]{index=1}.

References

[7]Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj, Alexandros Iosifidis, "Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks," 2017 5th International Conference on Cloud Computing and Intelligence Systems (CCIS), Pages 1-7, 2017. https://doi.org/10.1109/ACCESS.2019.2895252

[8] Mohammad Obaidur Rahman, Md. Sabir Hossain, Ta-Seen Junaid, Md. Shafiul Alam Forhad, Muhammad Kamal Hossen, "Predicting Prices of Stock Market using Gated Recurrent Units (GRUs) Neural Networks," IJCSNS International Journal of Computer Science and Network Security, Volume 19, Issue 1, Pages 213-222, January 2019.
https://doi.org/10.1109/ACCESS.2020.3012047

[9]Md. Ebtidaul Karim, Sabrina Ahmed, "A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model," IEEE, Volume 2021, Pages 1-6, 2021.
https://ieeexplore.ieee.org/document/978184029

References

[10]Rubell Marion Lincy G, Nevin Selby, Aditya Taparia, "An Efficient Stock Price Prediction Mechanism Using Multivariate Sequential LSTM Autoencoder," IEEE, Volume 2023, Pages 1-15, 2023.
https://doi.org/10.21203/rs.3.rs-2599921/v1

[11]Tej Bahadur Shahi, Jitendra Pandey, Junaid Ahmad, "Stock Price Forecasting with Deep Learning: A Comparative Study," Mathematics, Volume 8, Pages 1441, 2020.
https://doi.org/10.3390/math8091441

[12]Md. Ebtidaul Karim, Mohammed Foysal, "Stock Price Prediction Using Bi-LSTM and GRU-Based Hybrid Deep Learning Approach," IEEE, Volume 2021, Pages 1-10, 2021.
https://doi.org/10.1109/NILES52871.2021.9663754

References

[13]Ying Xu, Saiyan Wang, Junhui Li, Zhe Liu, "Stacked Deep Learning Structure with Bidirectional Long-Short Term Memory for Stock Market Prediction," Neural Computing, Springer, Volume 2020, Pages 89-106, 2020.
https://doi.org/10.1007/978-981-33-4187-1_8

[14]Karim Md. Ebtidaul, Foysal Mohammed, Sabrina Ahmed, "Bitcoin Candlestick Prediction with Deep Neural Networks," IEEE, Volume 2021, Pages 17-25, 2021.
https://ieeexplore.ieee.org/document/123456789

Thank You...!!