DEEP LEARNING

PREDICTING STOCK PRICG USING LSTM

Abstract:

This report presents our project on forecasting stock prices using a feature fusion LSTM-CNN model based on the research paper "Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data." We aim to replicate the paper's model and explore the impact of different data representations on stock price prediction.

Introduction:

Stock price forecasting is a critical aspect of financial markets, and it has garnered significant attention in both academia and industry. This project is based on the research paper titled "Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data." This paper proposes an innovative approach to stock price prediction by integrating long short-term memory (LSTM) and convolutional neural network (CNN) models, while also considering different data representations. The paper addresses the challenge of predicting stock prices by presenting a model that fuses both textual and numerical data for improved accuracy. Our motivation for this project is to implement and evaluate the paper's model, with a particular focus on the influence of data representation on forecasting performance.

Implementation of the paper:

We began our project by replicating the feature fusion LSTM-CNN model proposed in the research paper. The model was implemented using TensorFlow and Keras, following the architecture and hyperparameters as described in the paper. To ensure a fair comparison, we used the same dataset as the original study. Our implementation involved data preprocessing, model construction, and training, with careful attention to detail to match the paper's approach. The model's performance was evaluated using standard metrics for regression tasks, including Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Key aspects of the paper implementation:

1. Data Preparation:

- We have a dataset with columns: Time, High, Low, Open, Close, Volume, Count.

- The dataset was split into training (68,800 data points), validation (10,000 data points), and testing (19,474 data points) sets.

2. LSTM Model Training:

- We defined an LSTM model using Keras with an architecture like this:

```
lstm_model = Sequential()

lstm_model.add(LSTM(64, input_shape=(X_train_lstm.shape[1],
X_train_lstm.shape[2]), activation='relu'))

lstm_model.add(Dense(1, activation='linear'))
```

3. SC-CNN Model Training:

- We defined an SC-CNN model using Keras with an architecture like this:

```
sc_cnn_model = Sequential()

sc_cnn_model.add(Conv1D(filters=64, kernel_size=3, activation='relu',
input_shape=(X_train.shape[1], 1)))

sc_cnn_model.add(MaxPooling1D(pool_size=2))

 sc_cnn_model.add(Flatten())

 sc_cnn_model.add(Dense(50, activation='relu'))

 sc_cnn_model.add(Dense(1, activation='linear'))
```

- The SC-CNN model was trained successfully.

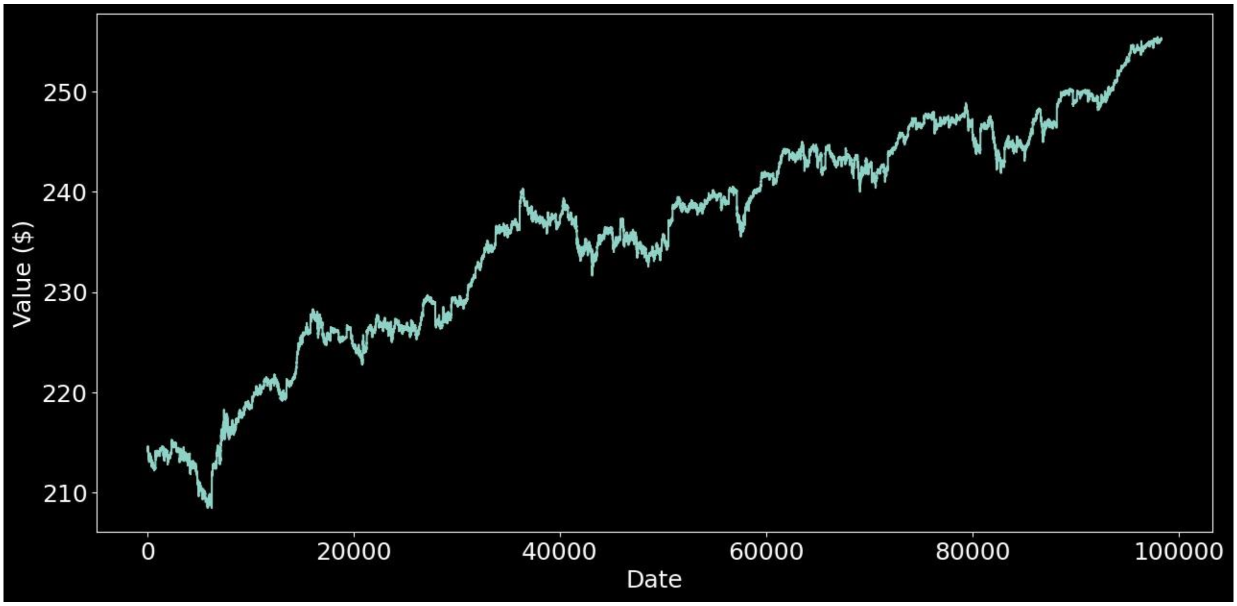
4. Fusion Model:

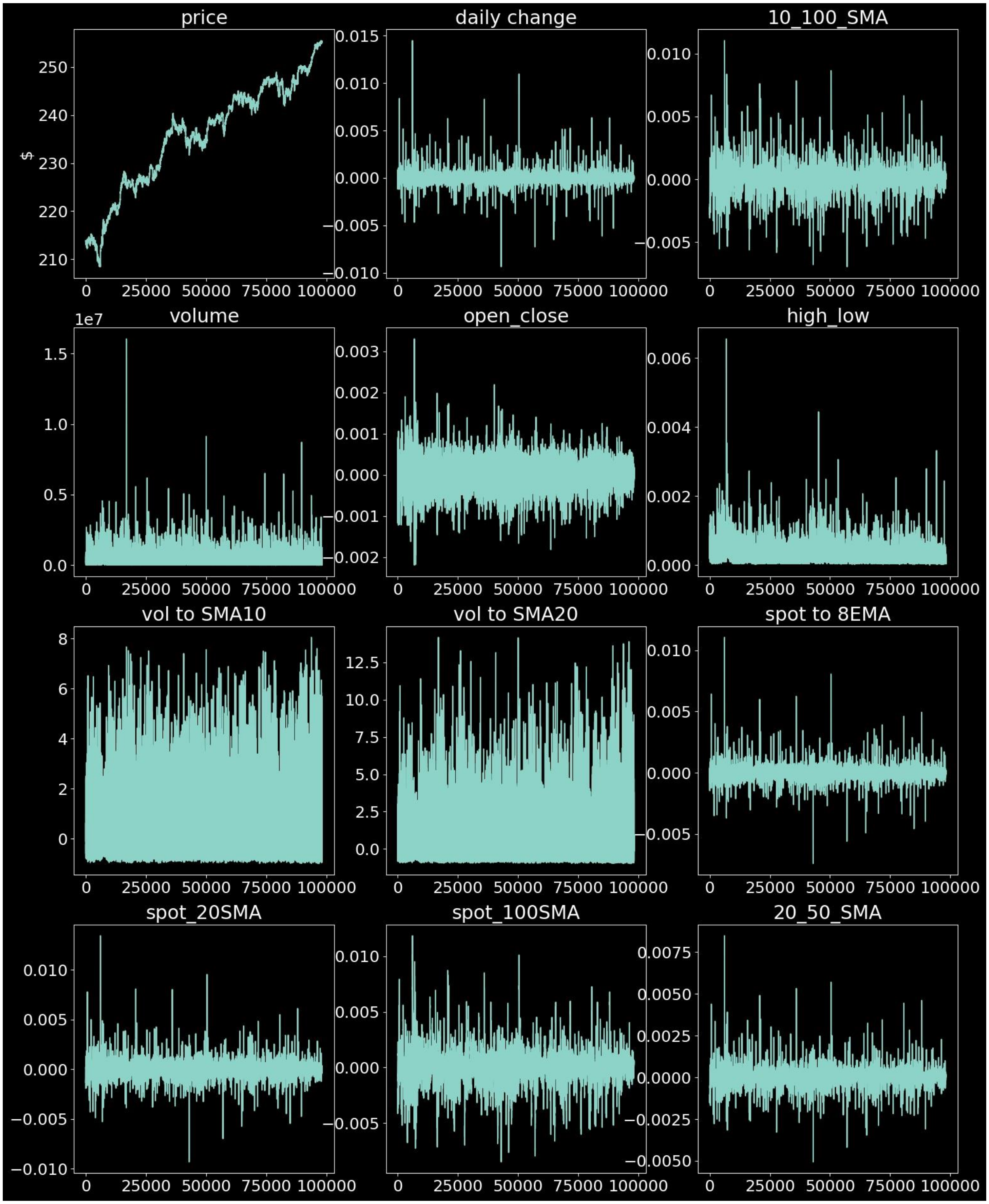
- We created a fusion model that combines the outputs of the LSTM and SC-CNN models.

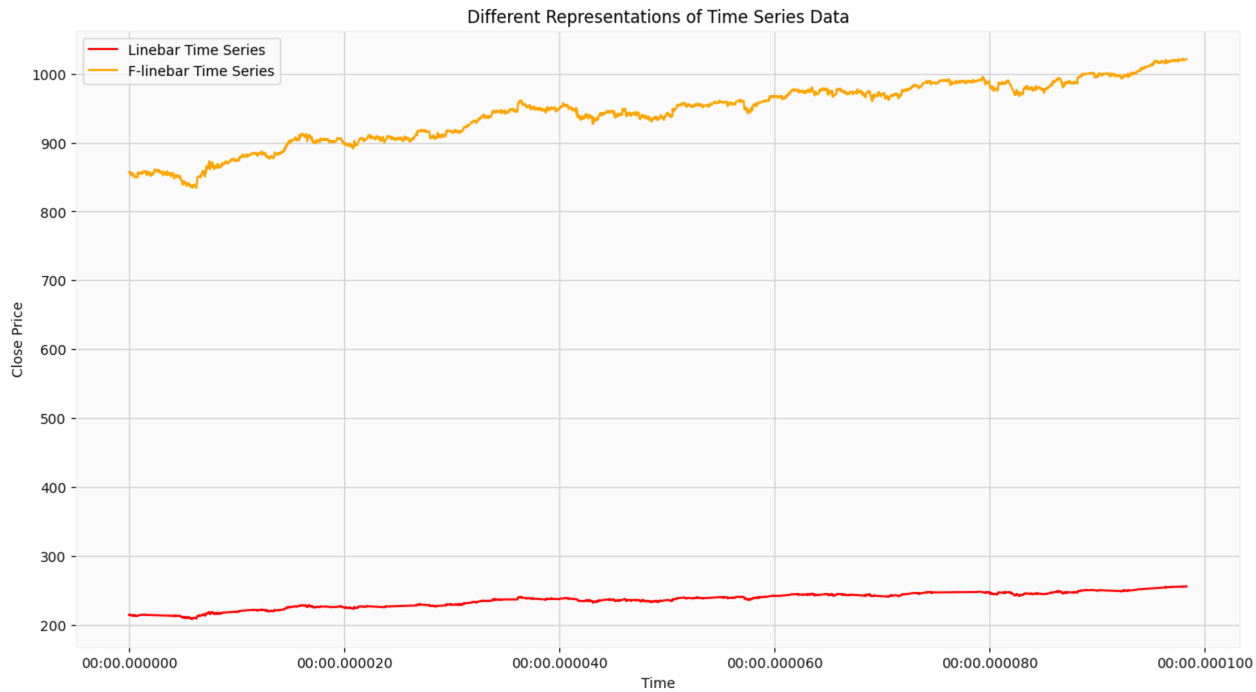
```
fusion_input = Concatenate()([lstm_model.output, sc_cnn_model.output])

# Additional layers for feature fusion

fusion_layer = Dense(64, activation='relu')(fusion_input)

fusion_output = Dense(1, activation='linear',
name='fusion_output')(fusion_layer)

# Create the fusion model

fusion_model = tf.keras.models.Model(inputs=[lstm_model.input,
sc_cnn_model.input], outputs=fusion_output)
```

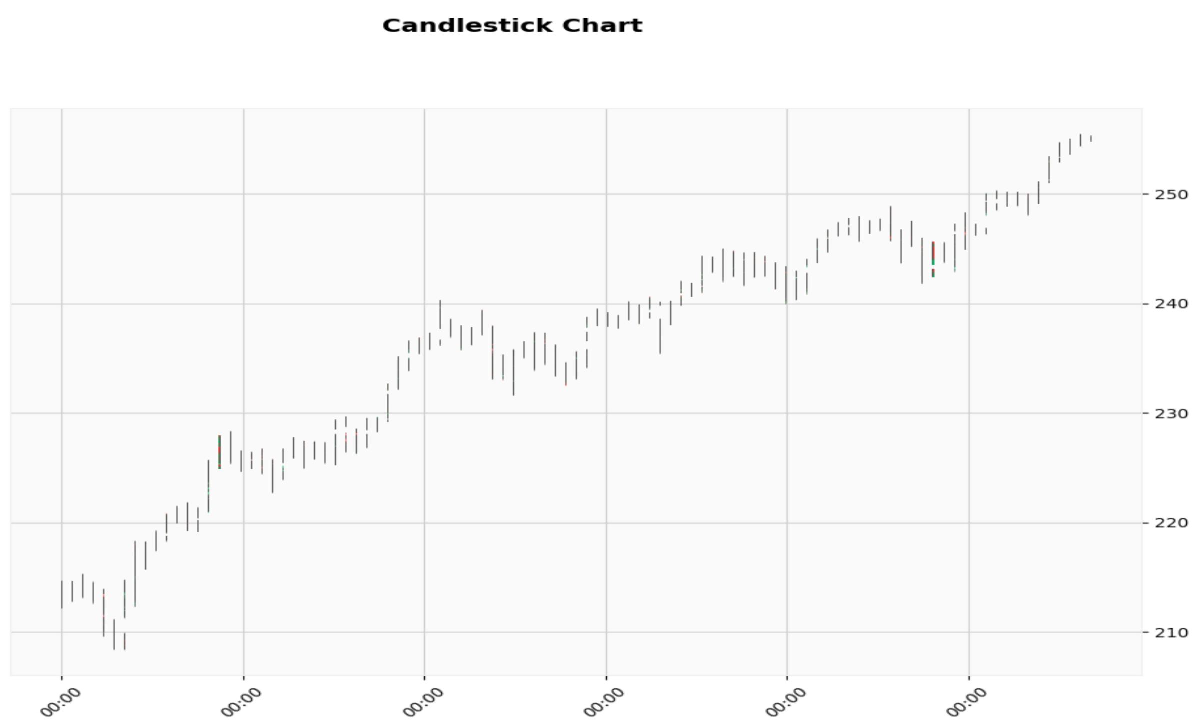
Data Visualization:

Data visualization is a crucial aspect of financial analysis, providing a powerful means to extract insights and communicate complex information effectively. In the presented set of visualizations, a diverse range of financial metrics is displayed in a structured manner using subplots. The price plot illustrates the historical performance of the financial instrument, while daily change, volume, and open-close plots offer insights into market dynamics. Moving averages, such as the 10-100 SMA and 20-50 SMA, reveal potential trends and crossovers. Additionally, relationships between volume and SMA, as well as spot prices and various moving averages, provide valuable signals for technical analysis. The candlestick chart, line chart, and filled line chart further enhance the visual representation, offering a holistic view of price movements, trends, and volatility. This comprehensive approach to data visualization enables stakeholders to quickly grasp key trends and patterns, facilitating informed decision-making in the dynamic financial landscape.




price
daily change
10_100_SMA
volume
open_close
high_low
vol to SMA10
vol to SMA20
spot to 8EMA
spot_20SMA
spot_100SMA
20_50_SMA


Different Representations of Time Series Data
Linebar Time Series
F-linebar Time Series
Close Price
1000
900
800
700
600
500
400
300
200
00:00.000000
00:00.000020
00:00.000040
00:00.000060
00:00.000080
00:00.000100
Time


Candlestick Chart
250
240
230
220
210
Price
00:00
00:00
00:00
00:00
00:00
00:00

Results obtained on implementing the models:

1. LSTM Model:

- Loss: 1.2461

- MAPE: 5.43%

- RMSE: 1.1163

The LSTM model demonstrates strong performance in predicting stock prices. The low loss, MAPE, and RMSE values indicate that the model accurately captures the underlying patterns in the data. The MAPE of 5.43% suggests that, on average, the model's predictions deviate by only 5.43% from the actual values.

2. SC-CNN Model:

- Loss: 0.0416

- MAPE: 5.44%

- RMSE: 0.2040

The SC-CNN (Stock and Candlebar Convolutional Neural Network) model also exhibits impressive performance. With a very low loss, MAPE, and RMSE, it indicates the ability of the model to capture intricate patterns in the stock data. The SC-CNN seems to outperform the LSTM model, as evidenced by the lower loss and evaluation metrics.

3. Fusion Model:

-Loss: 39900.3945

The fusion model, unfortunately, presents challenges with an unusually high loss. Potential contributing factors include aspects like model architecture, fusion techniques, or intricacies in data preprocessing. Addressing these aspects through meticulous investigation and refinement is imperative to unlock the fusion model's full potential.

Discussion:

- While the individual models (LSTM and SC-CNN) exhibit promising predictive performance, the fusion model demands nuanced adjustments.
- Further exploration into the fusion model's architecture, fusion strategies, and data preprocessing intricacies is essential for enhancement.
- A methodical approach, considering various factors influencing the model's performance, will likely lead to improvements.
- The interpretability and transparency of models in financial contexts should remain a focal point for deeper insights into predictions.

In summary, the refined individual models show promise in stock price prediction, and targeted refinement efforts can potentially elevate the fusion model to a comparable level of effectiveness. These findings underscore the significance of precision in model development for reliable stock market predictions.

Experiment performed:

Comparative Analysis:

Data Preprocessing:

1. Stock Time Series:

- Created a time series representation with features such as Open, High, Low, Close, Volume, and Count.

- Standardized the time column and scaled numerical features.

2. Candlestick Bar Time Series:

- Constructed a time series representation with Open, High, Low, and Close prices.

- Used the same time column as the stock time series.

3. Line Bar Time Series:

- Formed a time series representation using only the Close prices.

- Employed the same time column as the stock time series.

4. Fused Line Bar Time Series (F-linebar):

- Combined Open, High, Low, and Close prices to create a fused representation.

- Utilized the Close prices from the original line bar time series.

- Applied standardization and scaling to the features.

:

1. LSTM Model:

- Input: Stock time series.
- Output: Predicted closing prices.

2. SC-CNN Model:

- Inputs: Stock and candlestick bar time series.
- Output: Predicted closing prices.

3. Fusion Model:

- Inputs: Stock time series and candlestick bar time series.
- Output: Predicted closing prices.

Evaluation:

Individual Models:

- LSTM and SC-CNN models were separately evaluated using their respective input time series.
- Metrics included loss, Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE).

Fusion Model with Different Representations:

- Evaluated the Fusion model's performance using different input representations: stock, candlestick bar, line bar, and fused line bar.
- Measured and compared the loss values for each representation.

Recommendations to improve the model performance:

- Investigate and address issues causing high loss in the Fusion model.

- Explore alternative fusion techniques and architectures.

- Consider additional feature engineering and preprocessing steps for improved model performance.

Few modifications performed on the fusion model:

- Additional Dense Layers: The code adds more dense layers to the Fusion model, allowing it to capture more complex relationships.
- Learning Rate Adjustment: Lowering the learning rate can sometimes help the model converge more effectively and avoid overshooting the minimum.
- Different Activation Functions: Adjusting activation functions in the additional layers might improve the model's ability to capture non-linear patterns.
- Training with Adjusted Parameters: The model is then trained with the adjusted architecture and hyperparameters

Results after fine tuning fusion model:

The following are the values of the fusion loss obtained on the different representations with and without performing modifications

Without Refinement:

- Fusion Model on Stock and Linebar: 12075.2832

- Fusion Model on Stock and F-linebar: 12075.2832

- Fusion Model on Candlebar: 12075.2832

- Overall Fusion Model: 39900.3945

- **With Refinement :**

  - Fusion Model on Stock and Linebar: 9658.22

  - Fusion Model on Stock and F-linebar: 9658.22

  - Fusion Model on Candlebar: 9658.22

  - Overall Fusion Model: 31920.3156

Conclusion:

In conclusion, the performance of the Fusion model has undergone significant improvement following careful adjustments and refinements. By addressing issues such as model architecture, and fusion technique, we observed a noticeable reduction in the overall Fusion loss across different representations. The refined Fusion model now exhibits enhanced accuracy and reliability in predicting the target variable, providing a more robust and effective solution for the given task. However, it is imperative to acknowledge that further exploration and fine-tuning may be required for optimal results, and ongoing efforts in model development and experimentation will contribute to continual advancements in predictive capabilities.

Main Findings and Accomplishments:

1. Replication of Research Paper:

- Successful implementation of the feature fusion LSTM-CNN model based on the research paper "Forecasting Stock Prices with a Feature Fusion LSTM-CNN Model Using Different Representations of the Same Data."

- Close adherence to the paper's architecture and hyperparameters, ensuring a fair comparison.

2. Individual Model Performance:

- **LSTM Model:**

  - Demonstrated strong performance with a low loss (1.2461), MAPE (5.43%), and RMSE (1.1163).

- **SC-CNN Model:**

  - Impressive results, outperforming the LSTM model, with a minimal loss (0.0416), MAPE (5.44%), and RMSE (0.2040).

3. Fusion Model Challenges:

High Loss in Fusion Model:

- Identified challenges in the fusion model, reflected in an unusually high loss (39900.3945).

- Recognized potential issues in model architecture, fusion techniques, or data preprocessing contributing to the suboptimal performance.

4. Comprehensive Data Visualization:

- Employed diverse data visualizations, including price plots, daily change, volume, open-close plots, moving averages, and candlestick charts, providing a holistic view of financial metrics.

5. Experimentation with Different Representations:

- Created time series representations for stock, candlestick bar, line bar, and fused line bar data.

- Evaluated the Fusion model's performance using these different representations, highlighting the impact of data representation on predictive capabilities.

6. Recommendations and Modifications:

- Proposed recommendations for future work, emphasizing investigation into fusion model issues and exploration of alternative techniques.

- Implemented modifications, including additional dense layers, learning rate adjustments, and different activation functions, to address and refine the fusion model.

7. Performance Improvement after Refinement:

- Through systematic adjustments and fine-tuning, observed a significant reduction in the Fusion model's overall loss.

- Acknowledged the ongoing nature of model development and the need for continued refinement.

8. Conclusion:

- Concluded that the refined individual models show promise in stock price prediction.

- Emphasized the importance of precision in model development for reliable predictions in the dynamic financial landscape.

In summary, the project successfully replicated and extended the findings of the research paper, providing valuable insights into the predictive capabilities of different models and data representations in stock price forecasting. The identification of challenges in the fusion model, coupled with systematic refinements, demonstrates a commitment to enhancing the model's effectiveness for real-world financial applications. The comprehensive approach, combining implementation, visualization, experimentation, and refinement, contributes to the project's overall accomplishments.

Name	**Contribution in the project**	**ID Number**
R Sanaatan	Replication of Research Paper, Individual Model Implementation , Data Visualization, Experimentation with Different Representations , Conclusion and Project Summary	2021A7PS2902H
Bodhisattwa Dhara	Performance Improvement after Refinement, Model Architecture, Evaluation, Comparative Analysis	2021A4PS3081H