Work Report: Improvement in Price Prediction Model Accuracy through Data Segmentation and Training Optimization

*Prepared by [**Nahid Shahab**] on [**2025-02-23**]*

"Here is the link to my report: Google Colab Link."

https://colab.research.google.com/drive/1S-Z-76JgkymIzIRLlknZFHaPEaD-46R7#scrollTo=bVFjKfR9JjnQ

And "Here is the link to my report :GitHub link

https://github.com/NShahab/Uniswap-Decentralized-Marketplace

Introduction

In this new version of the price prediction model, the prediction accuracy has been improved. This enhancement is a result of dividing the data into three separate segments (80% training, 10% validation, and 10% test). The test data was kept completely separate from the training and validation datasets, leading to a more realistic evaluation of the model's performance.

Main Changes and Results

1. **Better Data Segmentation:**
   - The data was divided into three separate segments.
   - The test segment comprises 10% of the data, which was not used during training.
2. **Increased Prediction Accuracy:**
   - LSTM and GRU models were tested on the new data.
   - Evaluation metrics (MSE, RMSE, MAE, R^2) indicate improved model accuracy.
3. **Optimization of the Training Process:**
   - The batch size was increased from 32 to 64 to speed up model training.
   - EarlyStopping and ReduceLROnPlateau were used to control the training process.

Model Evaluation Results

Comparison of LSTM and GRU models across different timeframes:

Timeframe	Model	MSE	RMSE	MAE	R^2	Next Predicted Price
1 Hour	LSTM	150141207.75	12253.21	12125.00	0.19	82363.88
1 Hour	GRU	234924588.89	15327.25	14242.03	-0.26	70421.97
4 Hours	LSTM	15179946.32	3896.15	3299.43	0.92	98510.84
4 Hours	GRU	52491576.04	7245.11	6999.79	0.72	85675.01
1 Day	LSTM	21560799.64	4643.36	3539.02	0.90	99707.49
1 Day	GRU	15759063.84	3969.77	2937.40	0.93	93681.94

It is observed that the LSTM model outperforms the GRU model across all timeframes, especially in shorter timeframes.

Implementation of the New Model

```python
Copy
# Code implementation is included in this section
# Optimization of the model training process

# Function to train the model
def train_model(model, X_train, y_train, X_val, y_val, epochs=100, batch_size=64):
    early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)

    reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=5, min_lr=1e-6, verbose=1)

    model_checkpoint = ModelCheckpoint(
        'best_model.keras', monitor='val_loss', save_best_only=True, verbose=1)

    history = model.fit(
        X_train, y_train,
        validation_data=(X_val, y_val),
        epochs=epochs,
        batch_size=batch_size,
        callbacks=[early_stopping, reduce_lr, model_checkpoint],
        verbose=1
    )
    return model, history
```

Link to Colab Notebook

Comparison of Evaluation Metrics and Next Predicted Prices for Different Timeframes

1 Hour:

- **Model: LSTM**
  - MSE: 150141207.75
  - RMSE: 12253.21
  - MAE: 12125.00
  - R²: 0.19
  - Next Predicted Price: 82363.88
- **Model: GRU**
  - MSE: 234924588.89
  - RMSE: 15327.25
  - MAE: 14242.03
  - R²: -0.26
  - Next Predicted Price: 70421.97

4 Hours:

- **Model: LSTM**
  - MSE: 15179946.32
  - RMSE: 3896.15
  - MAE: 3299.43
  - R²: 0.92
  - Next Predicted Price: 98510.84
- **Model: GRU**
  - MSE: 52491576.04
  - RMSE: 7245.11
  - MAE: 6999.79
  - R²: 0.72
  - Next Predicted Price: 85675.01

1 Day:

- **Model: LSTM**
  - MSE: 21560799.64
  - RMSE: 4643.36
  - MAE: 3539.02
  - R²: 0.90
  - Next Predicted Price: 99707.49
- **Model: GRU**
  - MSE: 15759063.84
  - RMSE: 3969.77
  - MAE: 2937.40
  - R²: 0.93
  - Next Predicted Price: 93681.94

Interpretation of Results

1. Interpretation of Evaluation Metrics

- **MSE (Mean Squared Error):**
  - Lower values are better.
  - This metric shows the average squared error. The lower the value, the better the model's performance.
- **RMSE (Root Mean Squared Error):**
  - Lower values are better.
  - This metric is the square root of the average squared error and has the same unit as the original data, making it easier to interpret.
- **MAE (Mean Absolute Error):**
  - Lower values are better.
  - This metric shows the average absolute error and is less sensitive to large errors.
- **R² (R-Squared):**
  - Values closer to 1 are better.

o This metric indicates how much variance in the data is explained by the model. It ranges between 0 and 1:
  - 1: The model perfectly explains the data.
  - 0: The model explains none of the data.
  - Negative: The model performs worse than a simple model (e.g., the mean).

2. Interpretation of Results for Each Timeframe

1 Hour:

- **LSTM:**
  o MSE: 150141207.75
  o RMSE: 12253.21
  o MAE: 12125.00
  o R^2: 0.19
  o Next Predicted Price: 82363.88
- **GRU:**
  o MSE: 234924588.89
  o RMSE: 15327.25
  o MAE: 14242.03
  o R^2: -0.26
  o Next Predicted Price: 70421.97

Interpretation:

- The LSTM model performs better than the GRU model in this timeframe.
- The R^2 value for LSTM is positive (0.19) but close to zero, indicating that the model explains only a small portion of the data variance.
- The R^2 value for GRU is negative (-0.26), indicating that this model performs worse than a simple model.
- The RMSE and MAE errors are high for both models, indicating that the predictions are not very accurate.

4 Hours:

- **LSTM:**
  o MSE: 15179946.32
  o RMSE: 3896.15
  o MAE: 3299.43
  o R^2: 0.92
  o Next Predicted Price: 98510.84
- **GRU:**
  o MSE: 52491576.04
  o RMSE: 7245.11

- o MAE: 6999.79
- o R²: 0.72
- o Next Predicted Price: 85675.01

Interpretation:

- The LSTM model performs better than the GRU model in this timeframe.
- The R² value for LSTM (0.92) is very good, indicating that the model explains 92% of the data variance.
- The R² value for GRU (0.72) is also acceptable but weaker compared to LSTM.
- The RMSE and MAE errors for LSTM are significantly lower than for GRU.

1 Day:

- **LSTM:**
    - o MSE: 21560799.64
    - o RMSE: 4643.36
    - o MAE: 3539.02
    - o R²: 0.90
    - o Next Predicted Price: 99707.49
- **GRU:**
    - o MSE: 15759063.84
    - o RMSE: 3969.77
    - o MAE: 2937.40
    - o R²: 0.93
    - o Next Predicted Price: 93681.94

Interpretation:

- In this timeframe, the GRU model performs better than the LSTM model.
- The R² value for GRU (0.93) is very good, indicating that the model explains 93% of the data variance.
- The R² value for LSTM (0.90) is also good but slightly weaker than GRU.
- The RMSE and MAE errors for GRU are lower than for LSTM.

3. Summary

- **1 Hour Timeframe:** The models perform poorly. The data in this timeframe may be noisy, or the models may not have been trained sufficiently.
- **4 Hours Timeframe:** The LSTM model performs better than the GRU model. The predictions are relatively accurate.
- **1 Day Timeframe:** The GRU model performs better than the LSTM model. The predictions are very accurate.

4. Improvement Points

1. **1 Hour Timeframe:**
   - Examine the data to remove noise or errors.
   - Increase the volume of training data or adjust hyperparameters.
2. **4 Hours and 1 Day Timeframes:**
   - Improve model accuracy by tuning hyperparameters or using more data.
3. **Predicted Prices:**
   - Compare predicted prices with actual data to ensure model accuracy.

These improvements indicate significant progress in the accuracy and efficiency of price prediction models.

LSTM Model - 1 Hour Price Prediction

GRU Model - 1 Hour Price Prediction

LSTM Model - 4 Hours Price Prediction

GRU Model - 4 Hours Price Prediction

LSTM Model - 1 Day Price Prediction

GRU Model - 1 Day Price Prediction

MSE Comparison

	LSTM	GRU
1 Hour	150141207.75	234924588.89
4 Hours	15179946.32	52491576.04
1 Day	21560799.64	15759063.84

RMSE Comparison

	LSTM	GRU
1 Hour	12253.21	15327.25
4 Hours	3896.15	7245.11
1 Day	4643.36	3969.77

MAE Comparison

	LSTM	GRU
1 Hour	12125.00	14242.03
4 Hours	3299.43	6999.79
1 Day	3539.02	2937.40

R² Comparison

	LSTM	GRU
1 Hour	0.19	-0.26
4 Hours	0.92	0.72
1 Day	0.90	0.93