Full Report: Financial Time-Series Anomaly Detection

1. Introduction

Anomaly detection in financial time-series is crucial for identifying unusual patterns that may indicate fraud, market shifts, or systemic issues. This project applies to a deep learning approach using LSTM Autoencoders to detect such anomalies.

2. Dataset Preprocessing

The preprocessing pipeline involved the following:

2.1 Data Collection

- Historical stock data was downloaded using the Yahoo Finance API via yfinance.

2.2 Cleaning

- Null values were dropped to maintain data integrity.
- The datetime index was set to align with the temporal nature of the dataset.

2.3 Feature Engineering

- Financial indicators such as Moving Average (MA), Exponential Moving Average (EMA), Relative Strength Index (RSI), Bollinger Bands, and MACD were calculated to enhance the dataset's predictive capability.

2.4 Normalization

- Min-Max scaling was applied to compress the features into a 0–1 range, helping the LSTM model converge more efficiently.

2.5 Windowing

- Sequential time windows (TIME_STEPS = 30) were constructed to train the model on temporal trends, ensuring each sample represented 30 consecutive days.

3. Model Selection and Rationale

3.1 Model Chosen: LSTM Autoencoder

Why LSTM?

- LSTM (Long Short-Term Memory) networks are designed to handle time-series data and retain long-term dependencies, crucial for financial patterns.

Why Autoencoder?

- Autoencoders are effective for unsupervised anomaly detection. They learn to reconstruct input sequences. Poor reconstruction implies anomaly.

3.2 Architecture

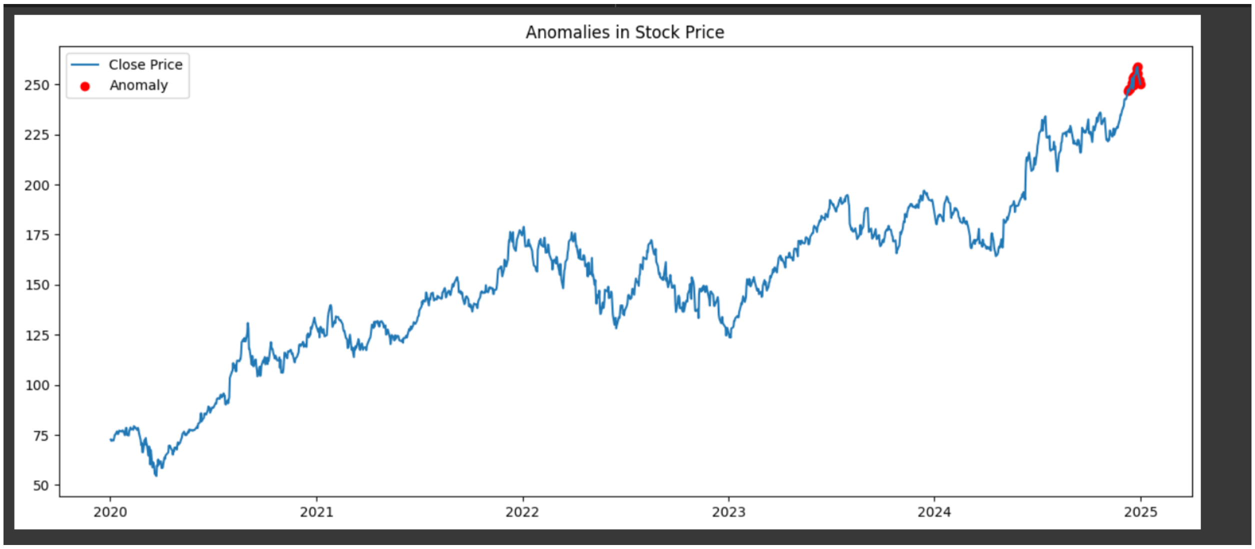
- Encoder: LSTM layers compress the input.
- Decoder: LSTM layers attempt to reconstruct the input.
- Dropout: Used for regularization.
- Loss Function: Mean Squared Error (MSE) between original and reconstructed input.

4. Challenges Faced and Solutions

Challenge	Solution
Temporal data handling	Used LSTM layers capable of modeling time-dependent relationships.
Lack of labeled anomalies	Used unsupervised learning based on reconstruction error.
Determining anomaly threshold	Defined threshold using statistical heuristics: mean + 3*std of error.
Model overfitting	Applied Dropout and early stopping.
Feature complexity	Created technical indicators to improve model input richness.

5. Results and Visualizations

5.3 Anomaly Detection Visualization



Anomalies were visually marked on the stock price plot.

- **Red markers**: Represent detected anomalies.
- **Observation**: Anomalies tend to cluster around market volatility periods, suggesting the model detects significant behavioral shifts.

6. Interpretation

- The LSTM Autoencoder successfully modeled normal behavior in stock price sequences.
- Reconstruction errors reliably highlighted anomalies without requiring labeled data.
- Technical indicators enhanced model performance by providing richer features.

7. Conclusion

The approach demonstrated that:

- LSTM Autoencoders are effective for unsupervised anomaly detection in financial time-series.
- Careful preprocessing, including feature engineering and normalization, is essential.

- Anomaly detection can provide early warning signs of unusual market activity.