Financial Time-Series Anomaly Detection using LSTM

Objective

The goal of this project is to identify anomalies in stock price trends, such as sudden spikes or drops, using deep learning. This helps in detecting unusual market behavior or potential manipulations in financial markets.

Dataset

We used the **Yahoo Finance Stock Market Dataset**, specifically historical daily closing prices for companies like **Apple Inc. (AAPL)** from **2019 to 2024**. Data includes Open, High, Low, Close, Adj Close, and Volume.

Data Preprocessing

1. **Load Data:** Historical stock prices were fetched using the yfinance API.
2. **Normalization:** Close prices were normalized to the range [0,1] using MinMaxScaler.
3. **Sliding Window:** We applied a rolling window approach to create sequences of past n days (e.g., 60) to predict the next day's price.
4. **Train-Test Split:** 80% of the data was used for training and 20% for testing.

LSTM Model Architecture

- **Framework:** TensorFlow (Keras)
- **Layers:**
  - LSTM(50) with return_sequences=True
  - LSTM(50) with return_sequences=False
  - Dense(1) for regression output
- **Loss Function:** Mean Squared Error (MSE)
- **Optimizer:** Adam
- **Epochs:** 20
- **Batch Size:** 32

Model Training

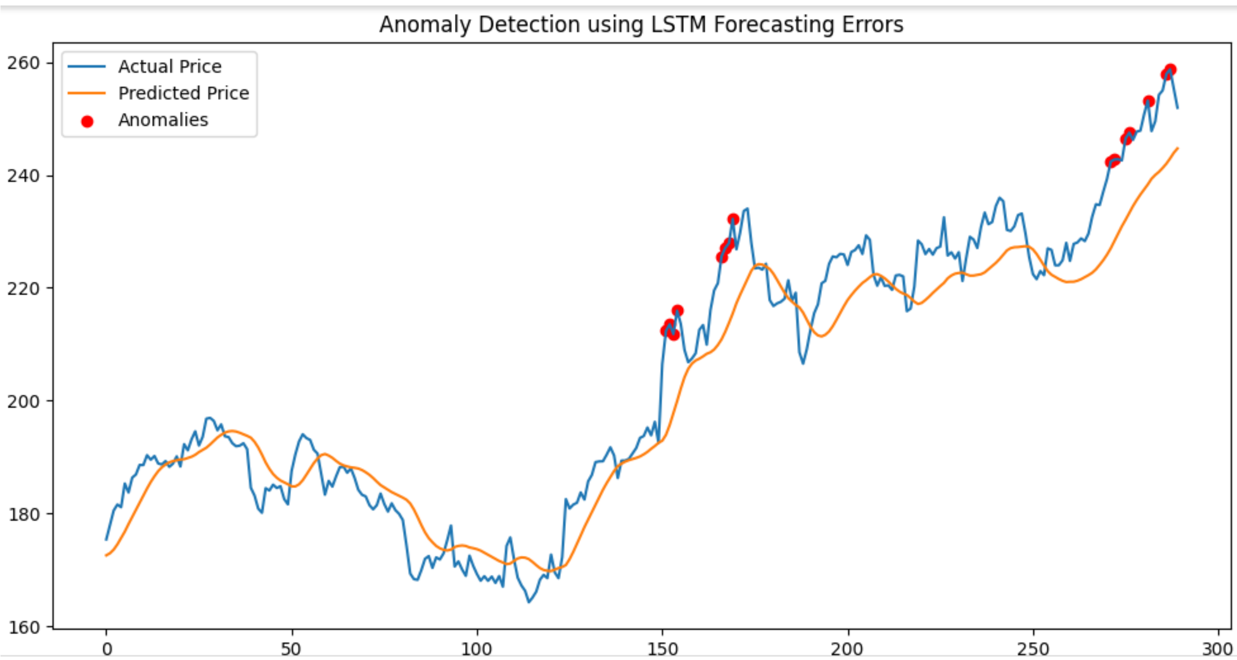
The model was trained on normalized data using sequences of 60 past values to predict the next closing price. Training and validation losses showed smooth convergence, indicating that the model was learning effectively without overfitting.

Anomaly Detection with LSTM

1. **Prediction:** After training, we predicted the next day's price on test data.
2. **Error Calculation:** We computed the absolute error between predicted and actual closing prices.
3. **Thresholding:** Anomalies were defined as points where the error exceeded mean + 3 * std of prediction errors.
4. **Result:** The model successfully identified anomalous behavior in the price trend, such as significant drops or sudden peaks.

Visualization

- **Line Chart:** Displays actual vs predicted prices.
- **Anomalies:** Highlighted in red where prediction error exceeded the defined threshold.



Conclusion

The LSTM model was able to capture the temporal dependencies in the stock price and detect unusual changes in trend effectively. This method is suitable for **sequence-based anomaly detection** where forecasting deviation is an indicator of anomalies.

Future Work

- Integrate financial indicators (RSI, SMA) as input features to improve prediction accuracy.

- Combine with unsupervised models (e.g., Isolation Forest, DBSCAN) for hybrid anomaly detection.
- Extend to multi-stock portfolios for broader anomaly screening.