Anomaly Detection in Stock Price Time-Series Using an LSTM Autoencoder

Submitted by : Gorthy Hari Krishna Sai
Roll No: 22EC01011

Supervised by Dr. Siddhartha S. Borkotoky
at IIT Bhubaneswar

ABSTRACT

Financial time-series such as stock prices show non-linear and fluctuating patterns influenced by several market and economic factors. Detecting unusual price movements is useful in identifying market instability and supporting trading decisions. In this work, an LSTM-based Autoencoder is trained on normal closing price sequences of Google (GOOG) stock. The model learns typical temporal behavior and reconstructs it. When abnormal variations occur, the reconstruction error increases, which is used to mark anomalies. This method is fully unsupervised and does not require labeled anomaly examples, making it practical for real-time market monitoring.

INTRODUCTION

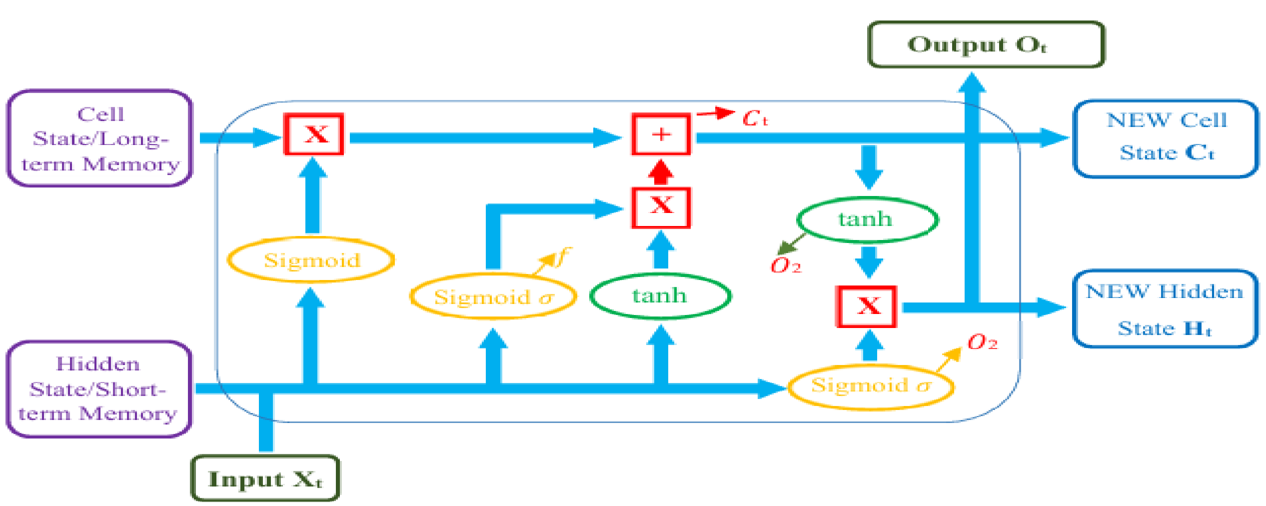
Stock price variations depend on investor activity, global events, and market conditions, leading to complex behavior that traditional statistical models often fail to capture. LSTM networks are effective for modeling sequential data due to their ability to preserve historical dependencies. When combined with an autoencoder framework, the model learns normal patterns during training and identifies deviations based on reconstruction quality. This makes the LSTM autoencoder suitable for anomaly detection in time series of stock prices, where abnormal changes may indicate volatility spikes or unusual trading activity

OBJECTIVES

- Collect and preprocess the **GOOGLE stock price dataset** for time-series analysis.
- Design and train an **LSTM Autoencoder** capable of learning normal closing price patterns.
- Compute the **reconstruction error** and define a suitable **threshold** to detect anomalous price movements.
- **Visualize** the detected anomalies along the price timeline for better interpretation and analysis.
- **Evaluate** the effectiveness of the LSTM Autoencoder in real-world **financial anomaly detection** scenarios.

AUTOENCODERS

- Autoencoders provide an efficient way to learn **compact feature representations** by compressing data into a latent space and then reconstructing it back.
- They operate in an **unsupervised** manner, requiring no labeled data, and learn to capture the most important patterns present in the input.
- The **encoder** extracts meaningful features from the input, while the **decoder** attempts to reconstruct the original data from the latent representation.
- The difference between the input and the reconstruction (called **reconstruction error**) guides the learning process.
- Autoencoders are particularly useful for **anomaly detection**, since they learn only the characteristics of **normal data**. When an unusual pattern occurs, the reconstruction error increases, making anomalies easy to identify.



$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f),\ \ i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o),\ \ h_t = o_t \odot \tanh(c_t)$$

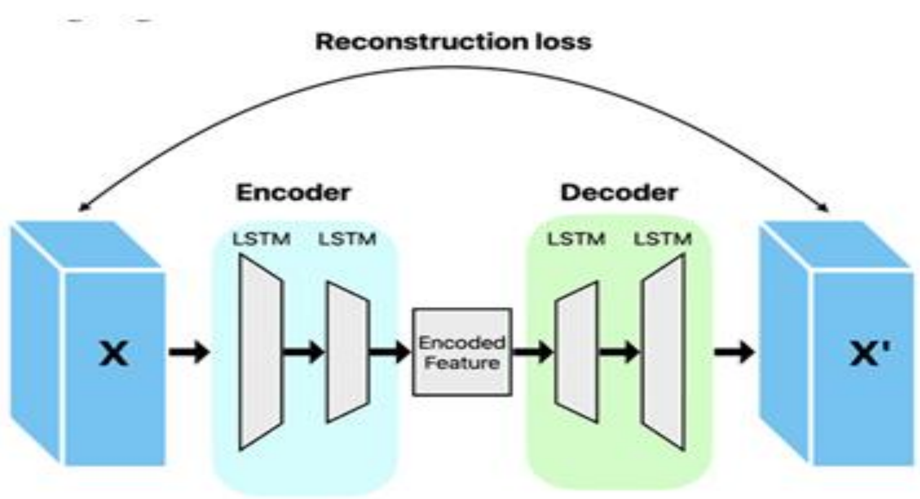
Fig 1. LSTM Architecture

LSTM NEURAL NETWORK

- **LSTM (Long Short-Term Memory)** networks are a type of recurrent neural network designed to handle sequential data.
- They capture **long-range temporal dependencies** using memory cells and gating mechanisms.
- The gates allow the network to **retain important information** while discarding irrelevant history.
- LSTMs help overcome the **vanishing gradient problem** seen in standard RNNs.
- Widely used in **time-series forecasting, speech processing, and financial data modeling**, where sequence context matters.

LSTM AUTOENCODER

- An LSTM Autoencoder combines **sequence modeling** with **representation learning** by encoding a time-series into a compressed latent vector and then reconstructing it.
- It is trained on **normal patterns only**, enabling it to model typical behavior in the data. During inference, **unusual or abnormal patterns** reconstruct poorly, resulting in a **higher reconstruction error**, which is used to detect anomalies.
- This makes LSTM Autoencoders particularly powerful for **financial anomaly detection and sensor fault monitoring**, where labels are scarce but normal behavior is abundant.



LSTM Auto Encoder

DATA PREPROCESSING

- The GOOG stock dataset is first sorted and indexed by date to maintain temporal order.
- The *Close* price values are selected and normalized using **StandardScaler** to ensure stable training and reduce scale bias.
- Missing or noisy values are checked and handled (not required here as the dataset is clean).
- The dataset is then split into **training (80%)** and **testing (20%)** segments for evaluation.

LSTM AUTOENCODER MODEL

- The **Encoder** compresses the input sequence into a compact **latent representation**, capturing essential trend patterns.
- The **Decoder** reconstructs the original sequence from this latent space using an LSTM-based generator.
- Architecture Used:

– LSTM(128) → Dropout(0.2) → RepeatVector(30) → LSTM(128, return_sequences=True) → TimeDistributed(Dense(1))

- The model learns by minimizing **Mean Squared Reconstruction Error** between input and output.

SEQUENCE CREATION

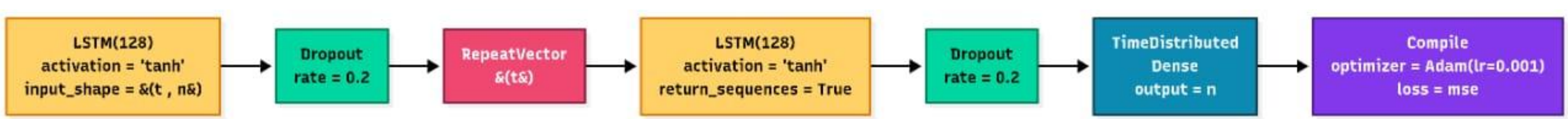
- LSTM models require sequential input, so the normalized price data is converted into overlapping time windows.
- Each input sample contains **30 consecutive price points**, representing short-term movement patterns.
- This allows the network to learn **temporal dependencies** rather than isolated values.

ANOMALY DETECTION METRIC

- After training on normal patterns, each sequence's **reconstruction error** is computed:
- A **threshold** is derived from training errors:

$\theta = \mu_E + 2\sigma_E$ = 95^{th} percentile of Training reconstruction loss.

- If **RE > threshold**, the point is flagged as an **anomaly**.
- This works because **normal patterns reconstruct well**, while **abnormal patterns reconstruct poorly**.



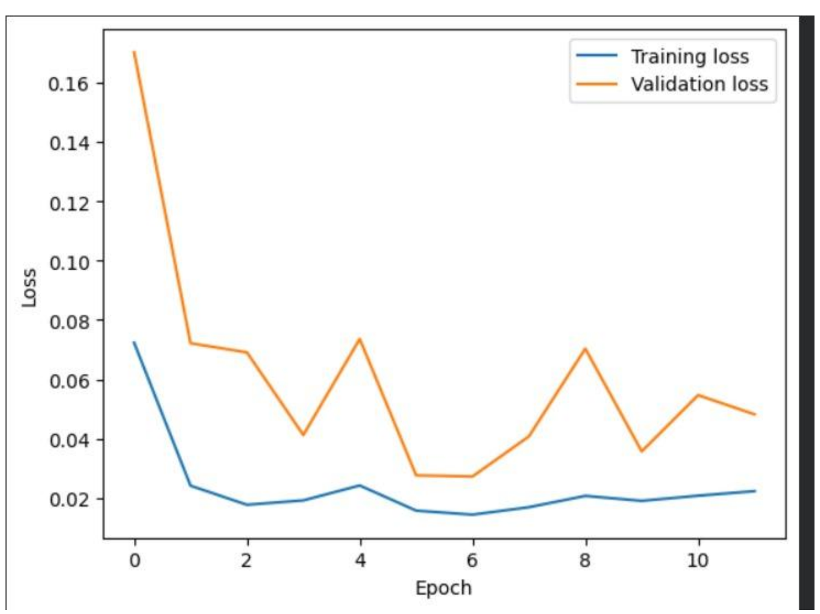
LSTM Autoencoder Model

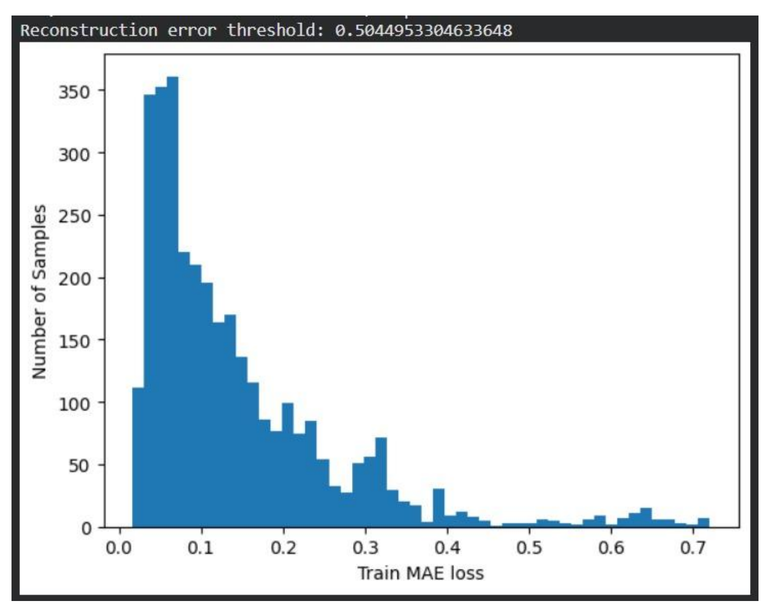
RESULTS

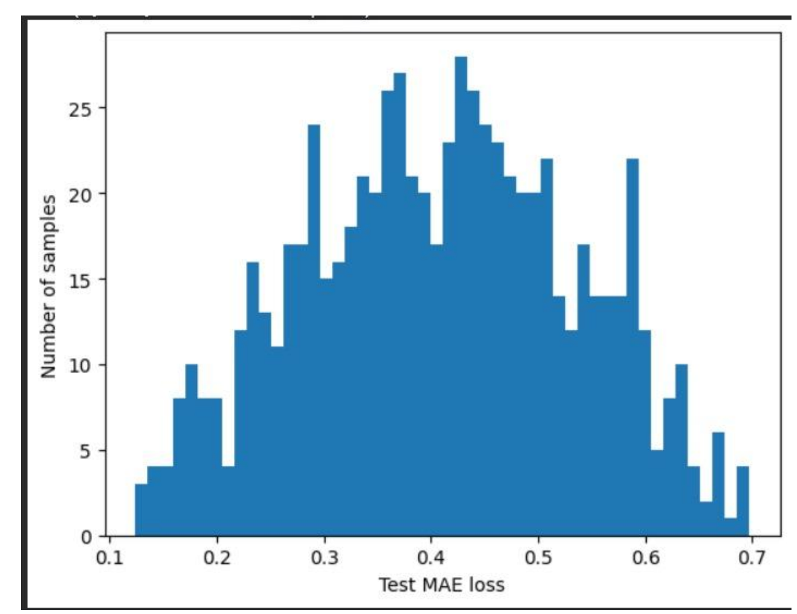
- • The training and validation loss decreased smoothly, indicating stable learning of normal patterns.
  • The reconstruction error on normal sequences remained low, confirming that the model captured typical price behavior.
  • A threshold was calculated based on training reconstruction error distribution.
  • Sequences with sudden price spikes or volatility changes produced **high reconstruction error**, crossing the threshold.
  • Detected anomalies aligned with **rapid market shifts and trend reversals**, validating the model's effectiveness.

DISCUSSIONS

• The LSTM Autoencoder successfully learned typical price movement patterns from historical GOOG data.
• Normal sequences were reconstructed accurately, while unusual price shifts resulted in higher reconstruction error.
• This shows that the model is **effective for unsupervised anomaly detection**, where labeled anomalies are unavailable.
• However, the model relies on a **fixed threshold**, which may be sensitive to varying market volatility.
• Overall, the approach works well for detecting **abrupt market changes**, but performance can be improved using multivariate inputs and adaptive thresholding.



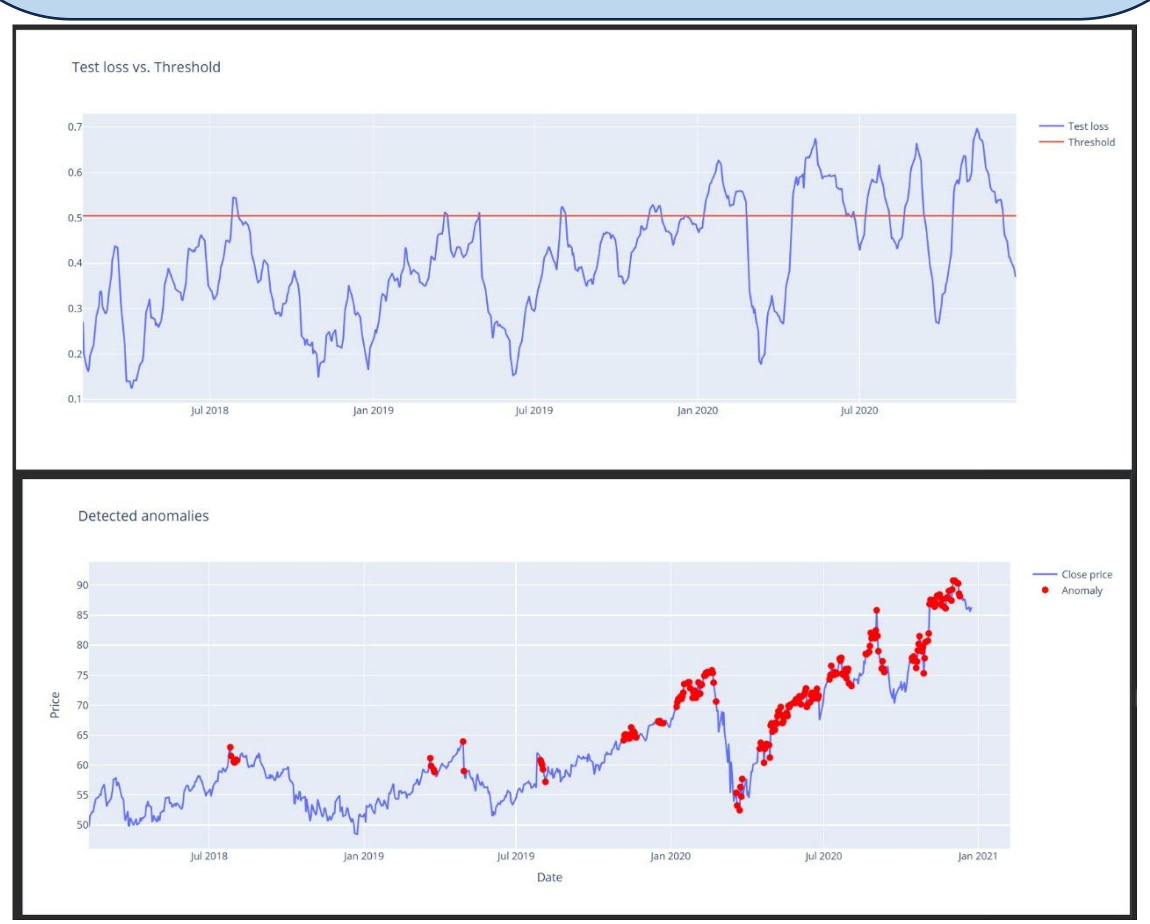




Results and Plots

FUTURE WORK

- Use **multivariate input** (Open, High, Low, Volume, indicators).
  • Apply **Attention-based LSTM** for better feature weighting.
  • Use **adaptive / dynamic thresholds** instead of fixed ones.
  • Integrate anomaly alerts into **trading or risk-monitoring systems**.
  • Add **explainability tools** (e.g., SHAP, saliency maps) to interpret anomalies.



Detected Anomalies, Reconstruction plot of a time step

CONCLUSION

The LSTM Autoencoder model effectively learned the normal behavior of the GOOG stock price through unsupervised reconstruction learning.
• By using reconstruction error and a statistical threshold, the system was able to **identify unusual or abnormal price movements**.
• The results demonstrate that LSTM Autoencoders are well-suited for **time-series anomaly detection**, especially in financial data where labeled anomalies are limited.

REFERENCES

- Lachekhab, F. et al. (2024). *LSTM-Autoencoder Model for Anomaly Detection in Electric Motors.*
- Lee, Y. et al. (2024). *LSTM-Autoencoder Based Anomaly Detection in Wind Turbine Vibration Data.*
- Hochreiter, S. & Schmidhuber, J. (1997). *Long Short-Term Memory.* **Neural Computation**.
- Chollet, F. *Keras Documentation.* https://keras.io
- Yahoo Finance. *GOOG Stock Historical Data.* https://finance.yahoo.com
- Homayouni, H. et al. (2021). *IDEAL: Autocorrelation-Based LSTM Autoencoder Framework.*