# Stock Bottleneck

Rishit Desai, Parnav Harinathan, Kartik Daswani, Niharka Bhende, Gaurav Jha

March 21, 2024

## 1 Introduction

Neural networks have revolutionized various fields, including finance. Autoencoders, a specific type of neural network architecture, are particularly well-suited for unsupervised learning tasks. This project explores the application of an autoencoder for stock market data analysis. Stock markets are inherently complex, with data exhibiting non-linear relationships and high dimensionality. Autoencoders can potentially address these challenges by learning efficient data representations and identifying underlying patterns.

## 2 Problem Statement

This project focuses on developing an autoencoder model using yfinance to analyze historical stock price data. The model will be trained for anomaly detection, learning to reconstruct normal stock price patterns. Deviations from these reconstructed patterns will be flagged as potential anomalies, indicating unusual market behavior.

## 3 Background Work

Several recent studies have demonstrated the effectiveness of autoencoders for financial data analysis.

Rizvi et al (2020) [1] used autoencoder for detecting stock price manipulation and obtained results show dramatic improvements in the detection performance over existing price manipulation detection techniques.

de Azevedo Takara et al [2] for Anomaly Detection in IBOVESPA Stock Market Index used and compared long short-term autoencoder, bidirectional long short-term autoencoder, and convolutional autoencoder models and concluded that these models perform quite well and can be applied to real stock market data

Other relevant works include studies by Albahli et al (2022) [3] who used autoencoders for dimentionality reduction of stock technical indicators (STIs), resulting in less correlation between the STIs.Then these STIs along with yahoo finance data were used to predict stock market prices.

These studies highlight the versatility of autoencoers in various stock market analysis tasks.

## 4 Proposed System

The proposed system comprises of a simple autoencoder as well as a variational autoencoder architecture for anomaly detection in stock prices.

### 4.1 Input Data

The model will be trained on daily historical opening, closing, high, and low prices of any publicly traded stocks retrieved from Yahoo Finance.

### 4.2 Network Architecture

The autoencoder will consist of an encoder and decoder with symmetrical structures. Each layer will use rectified linear unit (ReLU) activation functions except for the final output layer, which will use a tanh activation function.

### 4.3 Latent Space Dimensionality

The dimension of the encoded representation (latent space) will be a hyperparameter chosen through experimentation using GridSearchCV to balance reconstruction accuracy and capturing significant features.

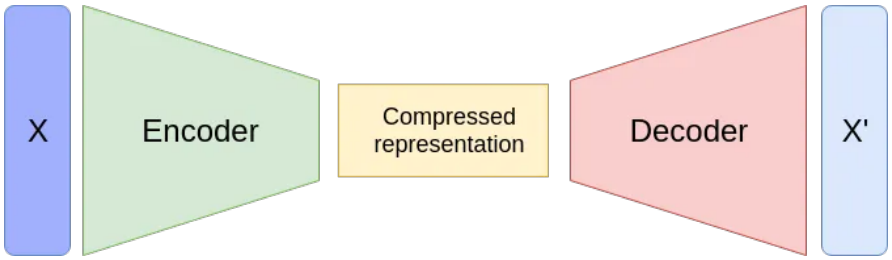


Figure 1: Autoencoder Architecture.

### 4.4 Training Process

The model will be trained using the Adam optimizer and a mean squared error (MSE) loss function to minimize the reconstruction error between the input data and the decoded output. The data will be preprocessed by normalization to ensure features are on the same scale. Training will be conducted for a specified number of epochs with a defined batch size.

## 5 Working/Implementation

The system will be implemented using Python with TensorFlow/Keras libraries. The yfinance library will be used to download historical stock data from Yahoo Finance. We are tine tuning the hyperparameters using GridSearchCV. Data visualization can be achieved using Streamlit. Finally, the application can be deployed on the cloud for anywhere access.

## 6 Results and Discussion

Latent Dimension Optimization: We optimized the number of latent dimensions, the encoded representation size. We found that 16 dimensions yielded the best results when used within the encoder and decoder architecture. Lower dimensions led to underfitting, while higher dimensions caused overfitting and necessitated adjustments to the architecture. Finding the ideal value for this hyperparameter was crucial.

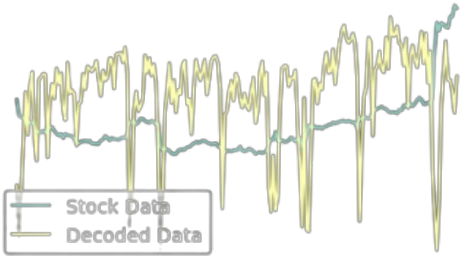


Figure 2: Before Optimization


Figure 3: After Optimization

Optimizer Selection: We evaluated various optimizers, including Adam, AdaGrad, AdaDelta, RMSProp, and SGD. Adam outperformed the others in terms of convergence speed based on the number of epochs and chosen latent dimensions.

# 7 Conclusion and Future Work

This project demonstrates the feasibility of using an autoencoder for anomaly detection in stock market data. The model effectively learns data representations and flags potential anomalies. Future work includes:

- Experimenting with different autoencoder architectures, such as convolutional autoencoder, long short-term autoencoder, to potentially improve anomaly detection accuracy
- Integrating other real-time sources like news for stock price prediction tasks.
- Investigating the application of the proposed approach to other financial datasets.

# References


[1] B. Rizvi, A. Belatreche, A. Bouridane, and K. Mistry, "Stock price manipulation detection based on autoencoder learning of stock trades affinity," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, July 2020.

[2] L. de Azevedo Takara, V. C. Mariani, and L. dos Santos Coelho, "Autoencoder neural network approaches for anomaly detection in ibovespa stock market index,"

[3] S. Albahli, T. Nazir, A. Mehmood, A. Irtaza, A. Alkhalifah, and W. Albattah, "Aei-dnet: A novel densenet model with an autoencoder for the stock market predictions using stock technical indicators," *Electronics*, vol. 11, no. 4, 2022.