StockMind.ai

Detection Of Stock Market Manipulation Using Market Structure

Team-9:
Adarsh Kumar Singh
Jaideep Singh Garlyal
Karnik Kanojia
Debasmita Paul
Aadya Goel

Visit Our Website
stockmind.ai.com

Problem Statement

The stock market is a complex system that involves a large number of participants, including investors, traders, and market makers. Often, these participants manipulate the market to make a profit at the expense of others.
Therefore, detecting stock market manipulation has become a critical issue for both investors and regulators.

As a result, the purpose of this research project is to investigate how market structure analysis can be used to detect stock market manipulation.
As we address this issue, we hope to provide valuable insights that can help investors and regulators mitigate the risks associated with stock market manipulation.

What is market manipulation?

Market manipulation is the act of using unethical or fraudulent means to interfere with the free and fair operation of the market. It can take many forms, and one way it can occur is through manipulating market structure.

What are the ways to detect manipulation?

- **Sentiment analysis**

  Deep learning algorithms can be trained to analyze large 4
  amounts of social media data, news articles, and other sources of information to detect changes in market sentiment that could be indicative of manipulation.

- **Trading patterns**

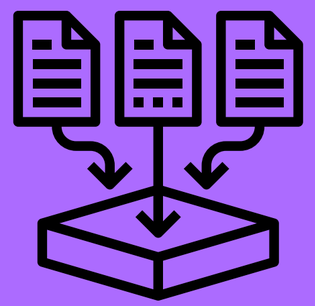
  Deep learning algorithms can be trained to recognize patterns in trading data that may be indicative of manipulation, such as wash trades or spoofing.

- **Market structure**

  Deep learning algorithms can be used to analyze market structure data, such as order book data and trade flow data, to detect changes in market dynamics that could be indicative of manipulation.

- **Compliance**

  Deep learning algorithms can be used to assist with compliance and surveillance efforts by identifying unusual trading activity and flagging it for further investigation.



Data Collection

Currently, we are manually collecting data from the official website of the Indian stock exchanges (NSE & BSE).
It is important for us to find both manipulated and non-manipulated data so that we can train our model more effectively.
After training our model on the data collected by us, we will also test our model with the yfinance (Yahoo finance) python library.
The library contains all the financial data we need to deploy our web app.

Initial Findings

While searching for methods to detect data manipulation we came across a statistical law known as Benford's law.
We observed that this law is capable of detecting the probability of data manipulation.
It is essentially a visual representation of a simple logic discussed in the following slide.
Therefore, we tested it and visualized it against stock market data for a company, which we have attached.

About our dataset

The has been collected from the Bombay Stock Exchange and the National Stock Exchange of India.

The data contains 12 parameters and they are:

1. Open Price
2. High Price
3. Low Price
4. Close Price
5. WAP
6. No. of Shares
7. No. of Trades
8. Total Turnover
9. Deliverable Quantity
10. % Deli. Qty to Traded Qty
11. Spread High-Low
12. Spread Close-Open

What is Benford's law?

Statistically, Benford's law is known as the first-digit law, which states that in many naturally occurring datasets, the first digit of a number is usually small rather than large.

In particular, the law predicts that the first digit will appear about 30% of the time, while the digit 9 will appear less than 5% of the time.

Manipulation detection using Benford's law

If a dataset is significantly different from the expected distribution of the first digits, this may be an indication that it has been altered or fabricated.

For example, if a company's financial statements show an unusually high proportion of numbers starting with the digit 9, it could be a red flag for fraudulent activity.

Benford's law visualization

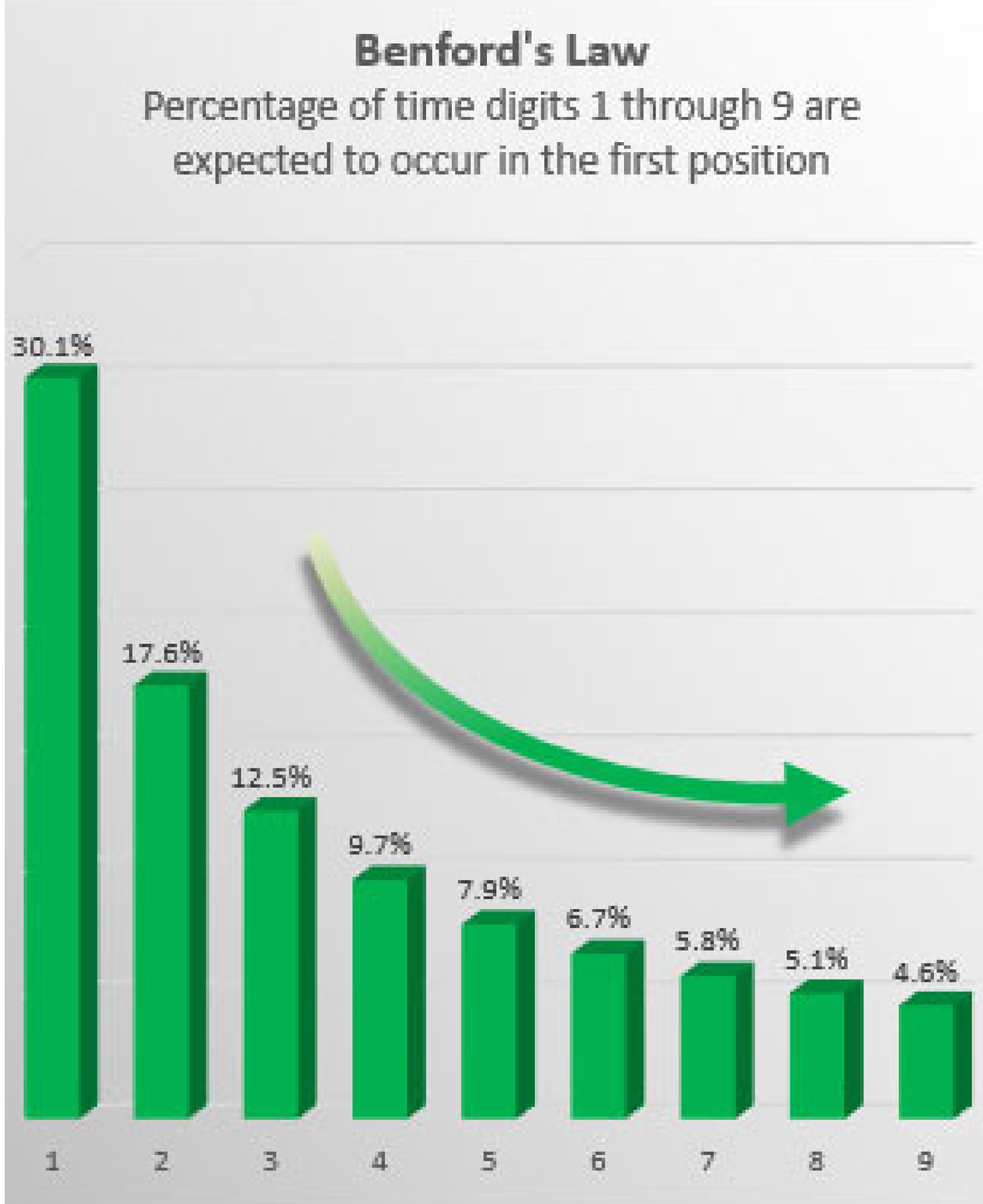
This graph is a visual representation of Benford's law. In this test, we will compare the visualization of Adani Enterprise's stock market data against this graph.

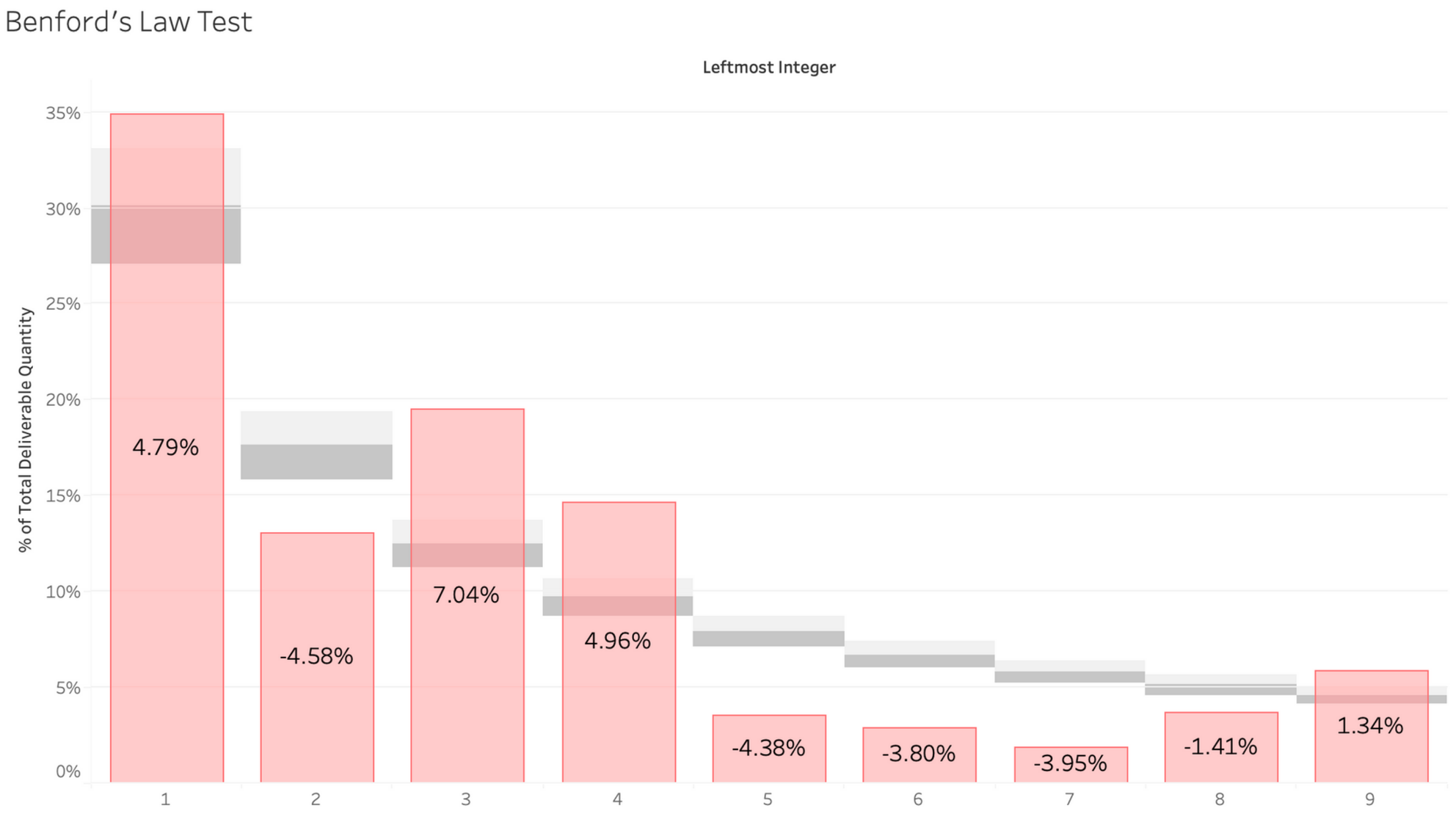
The dataset was collected from the official website of the BSE.

We have taken only two parameters from the data and those are the total no. of shares vs the total deliverable quantity.

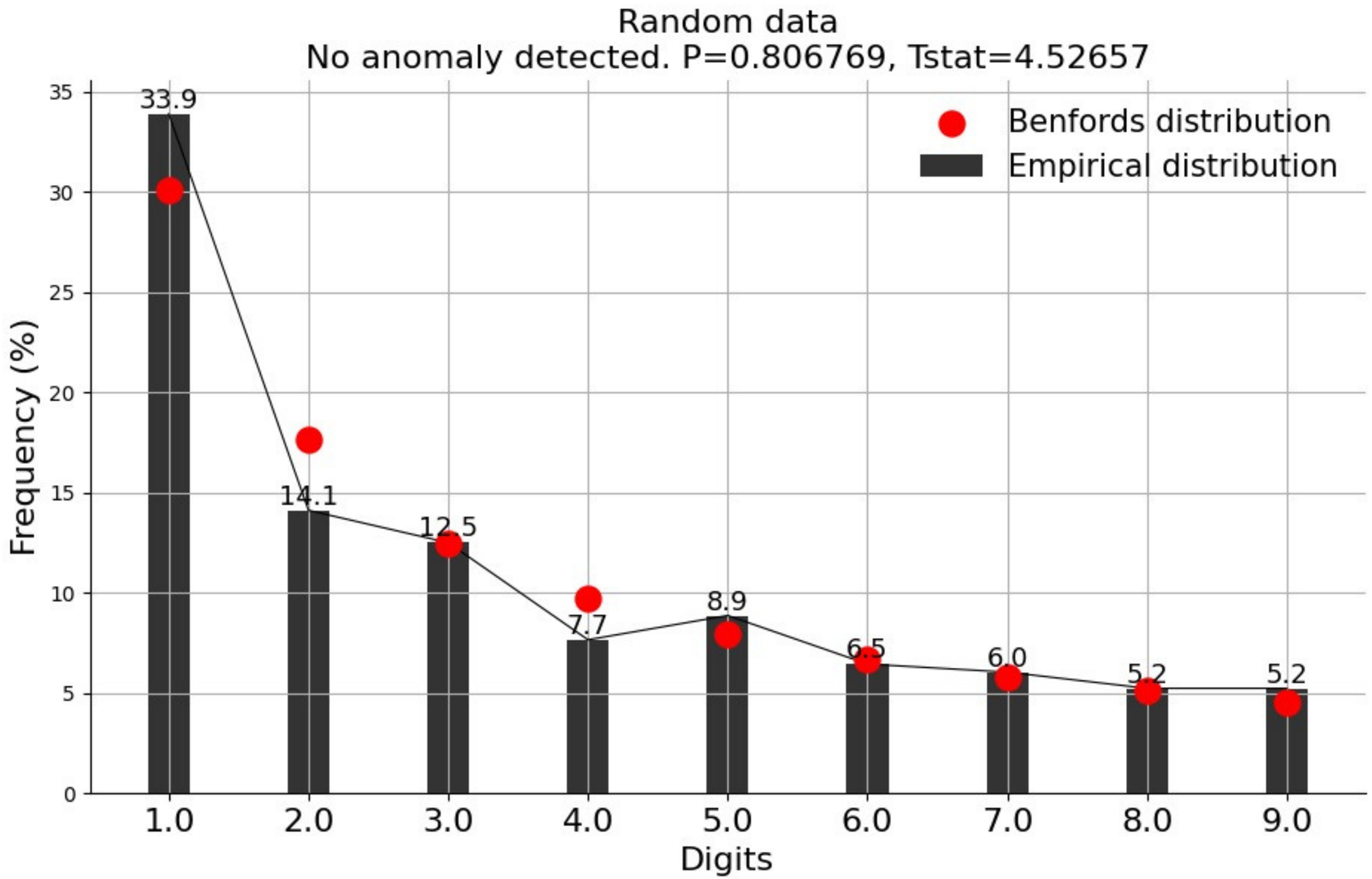
The next few slides contain the comparison between both graphs.

The visualizations were created on the tableau. and plotly.

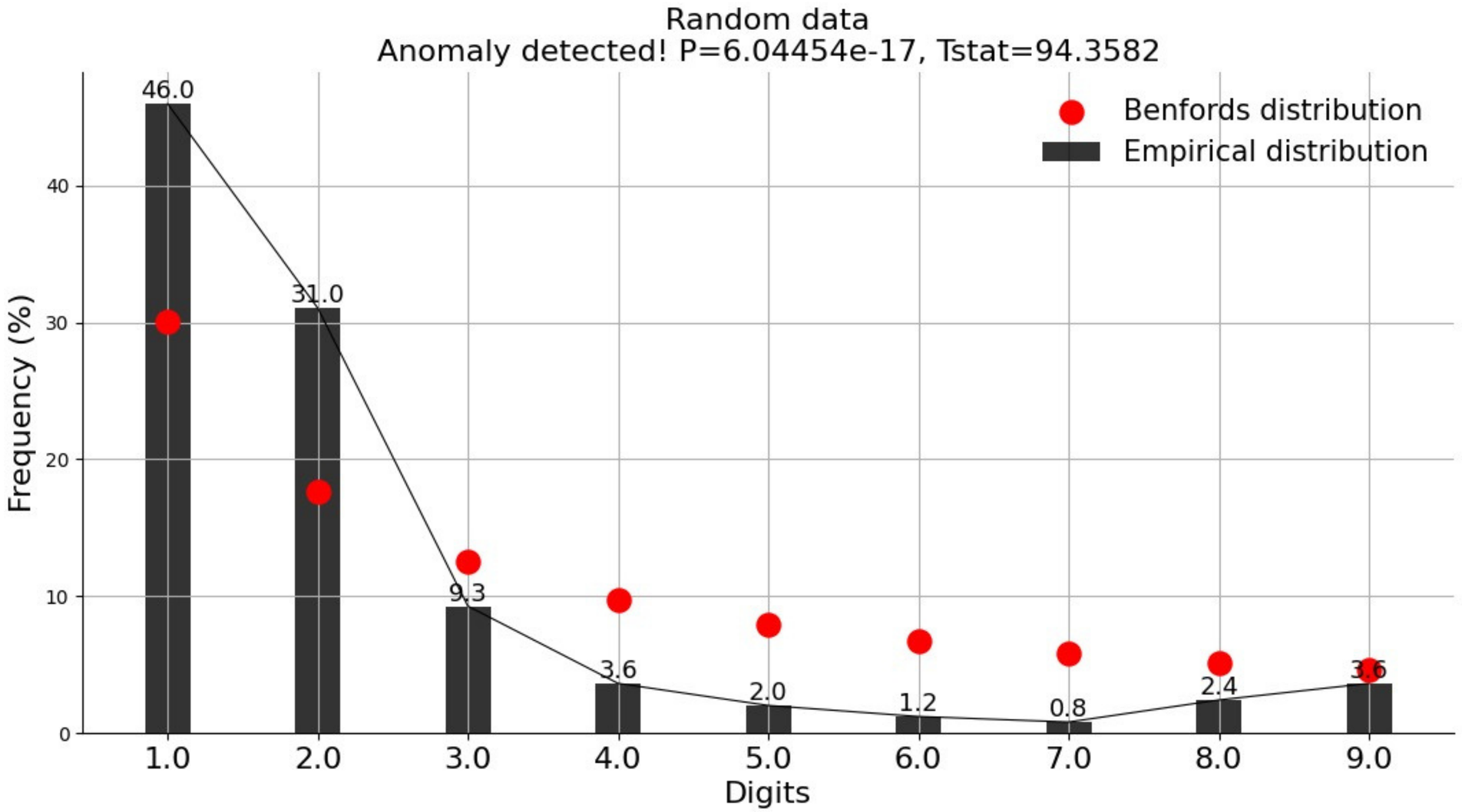




Visual representation of Benford's law against Adani enterprises stock market data of 2022

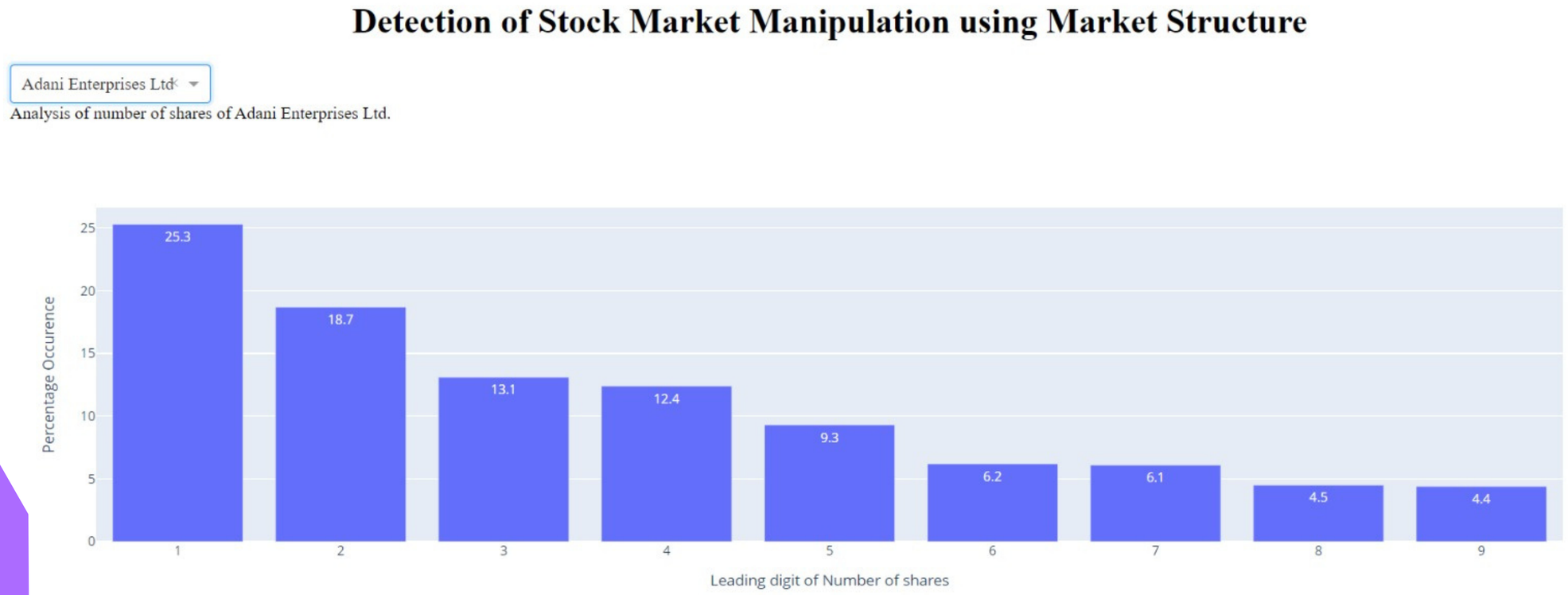


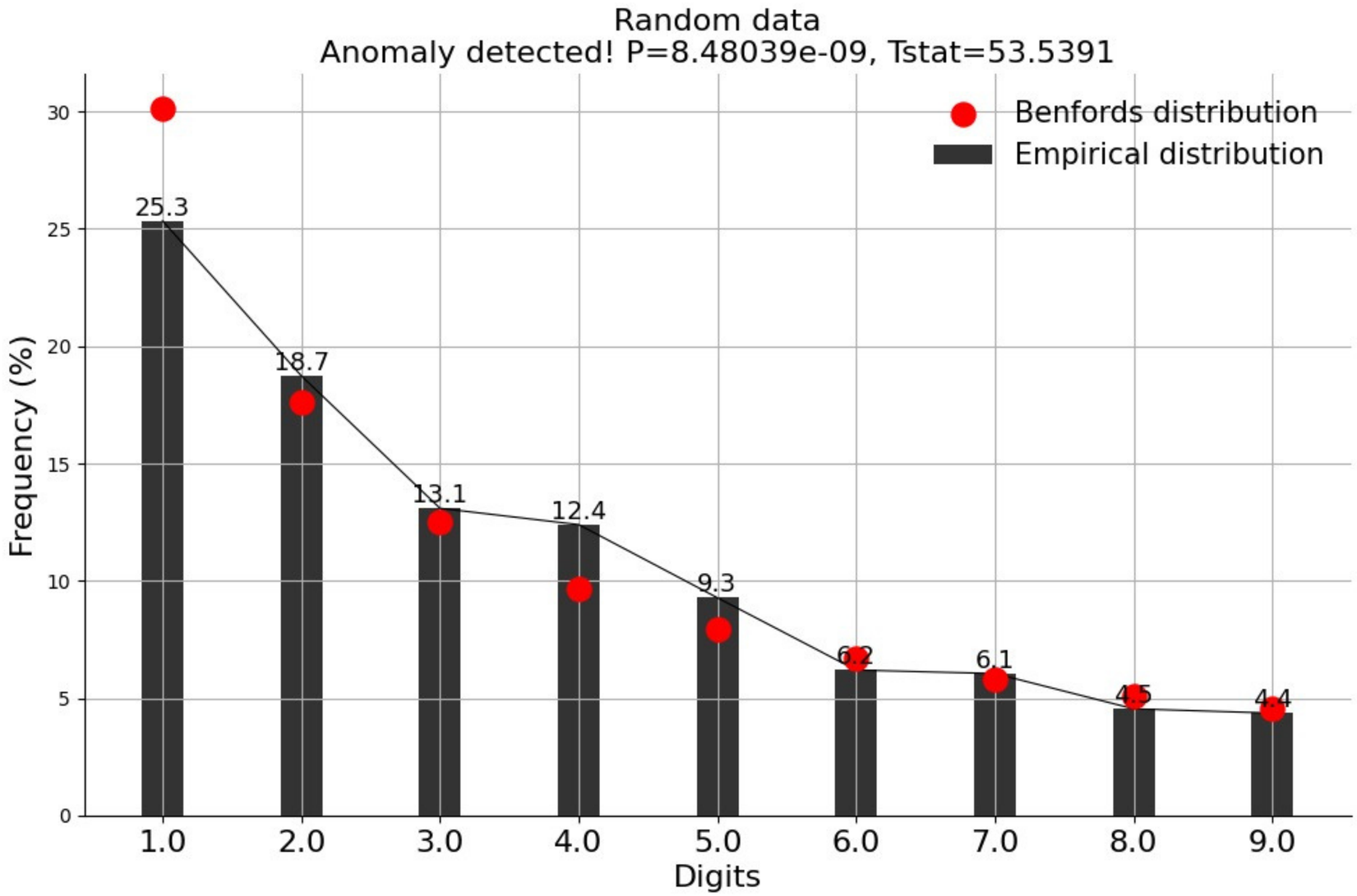
Visual representation of Benford's law against Adani enterprises stock market data of 2022



Visual representation of Benford's law against Adani enterprises stock market data of 2022

Adani Enterprises





Visual representation of Benford's law against Adani enterprises stock market data of 2022

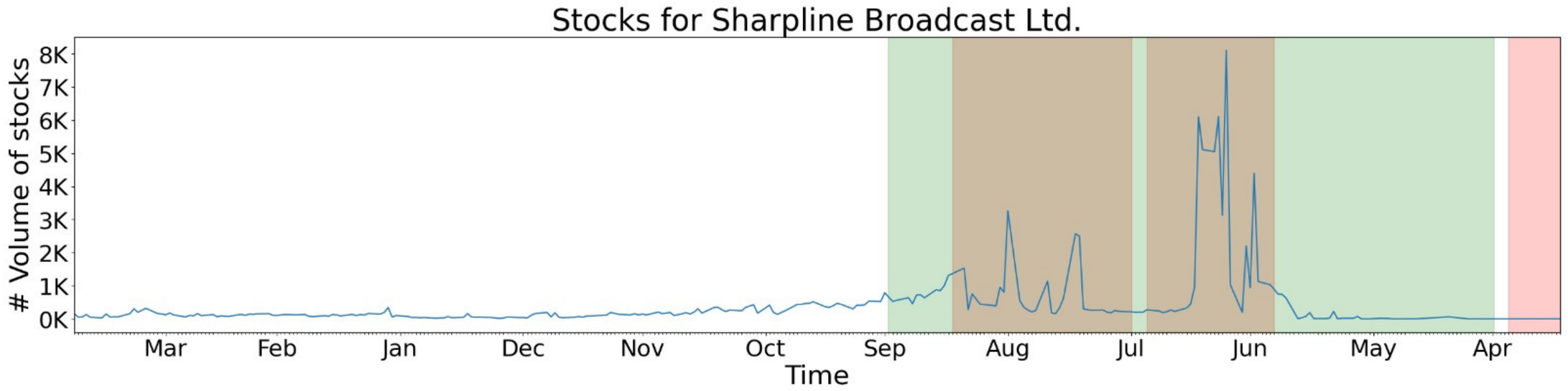
Model-1: LSTM with Autoencoder

An LSTM autoencoder model combines the capabilities of Long Short-Term Memory (LSTM) networks and autoencoders. It consists of an encoder and a decoder, where the encoder compresses input sequences into a lower-dimensional representation.

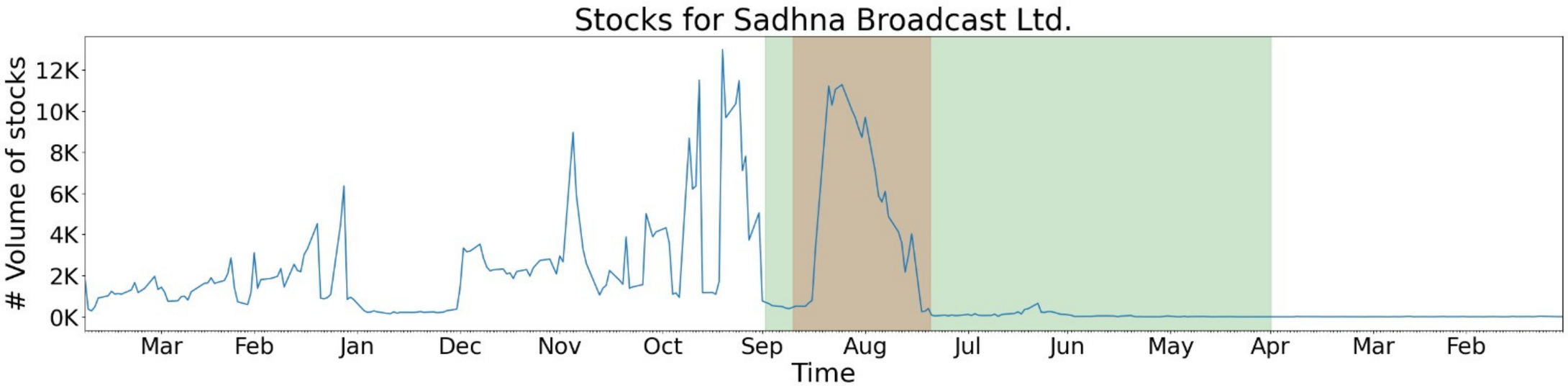
The decoder aims to reconstruct the original input sequence from the encoded representation. Training involves minimizing the reconstruction error, enabling the model to learn a compact representation that captures important features.

LSTM autoencoders are effective for tasks involving sequential or time series data, such as anomaly detection and feature extraction.

Results of Model-1



Results of Model-1



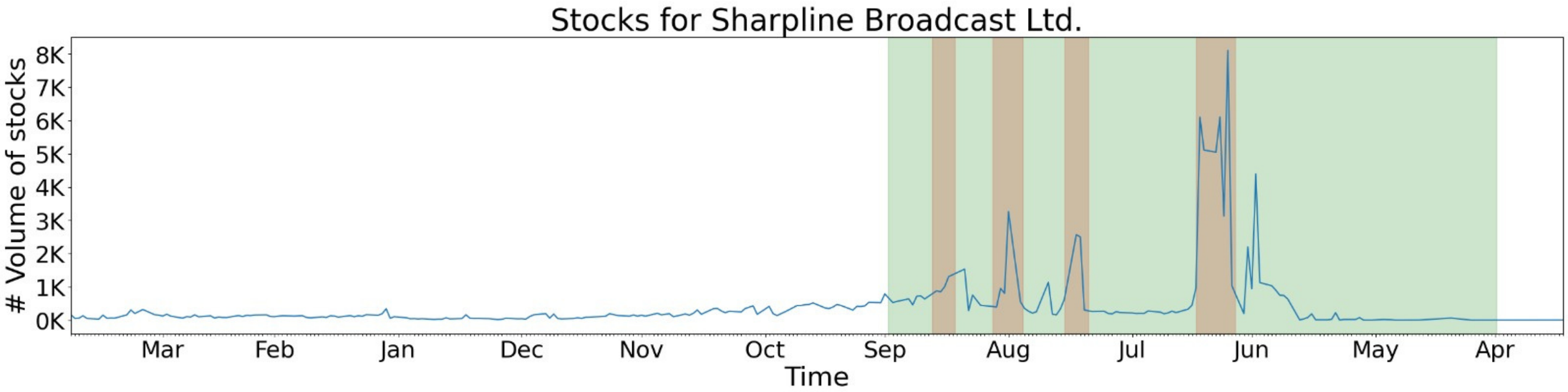
Model-2: TadGAN

TADGAN (Time Series Anomaly Detection with Generative Adversarial Networks) is an algorithm for detecting anomalies in time series data. It utilizes a generative adversarial network (GAN) architecture consisting of a generator and a discriminator.

The generator generates synthetic time series data that resembles real data, while the discriminator learns to distinguish between real and synthetic data.

Anomalies are detected by comparing the real data with the generated data and identifying instances with significant discrepancies. TADGAN has shown promising results in various domains, including finance, where it can be applied to detect anomalies and potential market manipulation.

Results of Model-2



StockMind.ai

Thank You

For Your Attention

Visit Our Website
stockmind.ai.com