STOCK ANALYSIS + PREDICTION
(USING - DTW& LSTM)

Data Exploration and System Management Using

Artificial Intelligence

Hiran Christa Bel Bommella

BTU Cottbus-Senftenberg

Cottbus, Germany

February 09,2024

1. Abstract:

This project aims at Dynamic Time Warping (DTW) for stock price pattern recognition, analysis, and Long Short-Term Memory (LSTM) neural networks for prediction. DTW is a robust similarity measure to identify identical historical price patterns. By normalizing and down sampling price data, DTW efficiently compares current price movements with past patterns, assisting in pattern recognition. The DTW and LSTM offer a comprehensive approach to understanding stock price dynamics and forecasting future trends, empowering investors with actionable insights for informed decision-making.

2. Project Goals:

1. Utilize Dynamic Time Warping (DTW) to identify similar historical price patterns in stock data.
2. Enhance understanding of stock price dynamics through efficient comparison and recognition of historical patterns using DTW.
3. Provide actionable insights to investors for informed decision-making in the stock market based on LSTM predictions.

Introduction:

In the dynamic and complex world of financial markets, identifying past pricing patterns and effectively forecasting future trends are crucial for investors trying to make correct choices. Traditional technical analysis methods rely on primitive pattern identification techniques, which may miss small but substantial connections in price movements over time. To address these problems, this study proposes an idea that uses Dynamic Time Warping (DTW) and Long Short-Term Memory (LSTM) neural networks to analyze and improve the accuracy of stock price predictions.

Dynamic Time Warping (DTW) Analysis:

An algorithm for measuring similarity between two temporal sequences, which may vary in speed. Dynamic Time Warping (DTW) is a powerful metric for similarity commonly used in time series research, notably for identifying patterns over time. DTW, unlike traditional distance metrics, can accept time series of varied durations and periodic shifts, making it ideal for analyzing stock price movements with inconsistencies and nonlinearities. By finding similar historical price patterns using DTW, this initiative hopes to provide significant insights into the fundamental dynamics of stock values to understand the stock market. It finds minimal cost path through the cost matrix. It shows two things firstly, how similar are two points. Secondly, which points correspond to one another?

Long Short-Term Memory (LSTM) Neural Networks:

Long Short-Term Memory (LSTM) neural networks, on the other hand, specialize in identifying historical relationships in sequential data, making them ideal for time series prediction use cases. In this project, LSTM networks are used to model the historical dynamics of stock price movements, leveraging the network's ability to capture long-term dependencies and learn complex patterns from historical data. LSTM is particularly useful in analyzing stock market data because it can handle data with multiple input and output timesteps. For example, a company's stock price may be influenced by various factors such as economic indicators, market trends, and company-specific news.

3. Material and methods

3.1. Data

In this project, Apple Inc.'s (AAPL) stock values spanning from December 1, 1990 to December 29, 2023, is used. Yahoo Finance is the source, accessible through the convenient yfinance Python package.

yfinance is a Python package that makes it easy to retrieve historical market data from Yahoo Finance. yfinance permits to access the historical stock data for various financial tools, including equities, ETFs (Exchange-Traded Funds), indices, and cryptocurrencies.

The link to access the historical stock price data is Yahoo Finance.

	Open	High	Low	Close	Adj Close	Volume
Date						
1990-12-03	0.332589	0.341518	0.330357	0.340402	0.272950	165401600
1990-12-04	0.334821	0.345982	0.334821	0.343750	0.275634	152152000
1990-12-05	0.343750	0.359375	0.338170	0.358259	0.287268	218388800
1990-12-06	0.368304	0.372768	0.361607	0.368304	0.295323	532246400
1990-12-07	0.366071	0.381696	0.366071	0.379464	0.304271	329660800
...	...	...	...	...	...	...
2023-12-22	195.179993	195.410004	192.970001	193.600006	193.600006	37122800
2023-12-26	193.610001	193.889999	192.830002	193.050003	193.050003	28919300
2023-12-27	192.490005	193.500000	191.089996	193.149994	193.149994	48087700
2023-12-28	194.139999	194.660004	193.169998	193.580002	193.580002	34049900
2023-12-29	193.899994	194.399994	191.729996	192.529999	192.529999	42628800

8332 rows × 6 columns

Fig 1. AAPL stock dataset

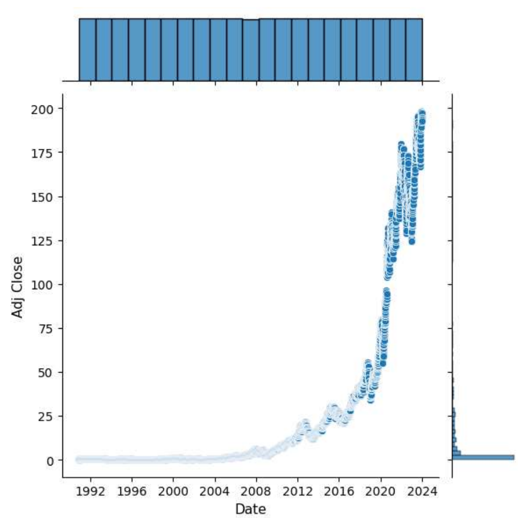
3.2. Methods

1. **Python Libraries:** For data retrieval, manipulation, visualization, and technical analysis, libraries used are yfinance, seaborn, numpy, pandas, matplotlib, tensorflow and ta.

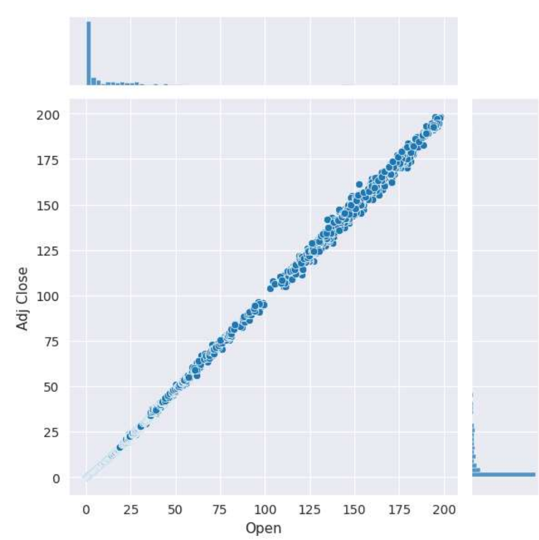
2. **Data Retrieval and Exploration**:
    1. Utilizing the yfinance library to download historical stock price data.
    2. Exploring the data to understand its structure, including attributes such as date, open price, high price, low price, close price, adjusted close price, and trading volume.

3. **Data Preprocessing**:
    1. Handling missing values, if any, ensuring data cleanliness.
    2. Normalizing the data to a common scale to facilitate analysis.
    3. The closing price data is extracted and transformed into percentage change using pct_change(), and then NaN values are dropped.

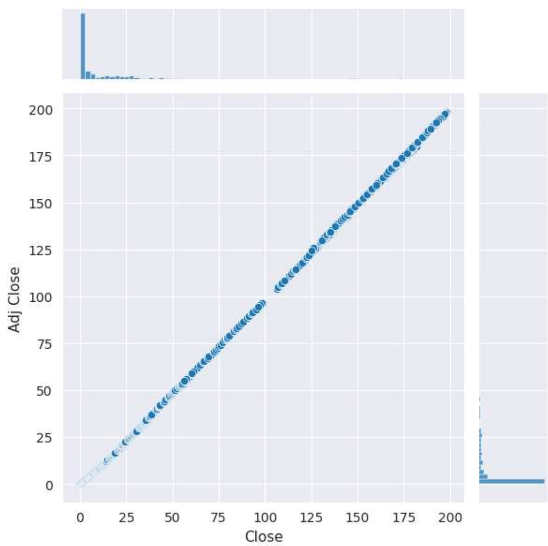
The below graphs are created to visualize the relationship between Adj Close (the final closing value) and all other the columns in the data set.



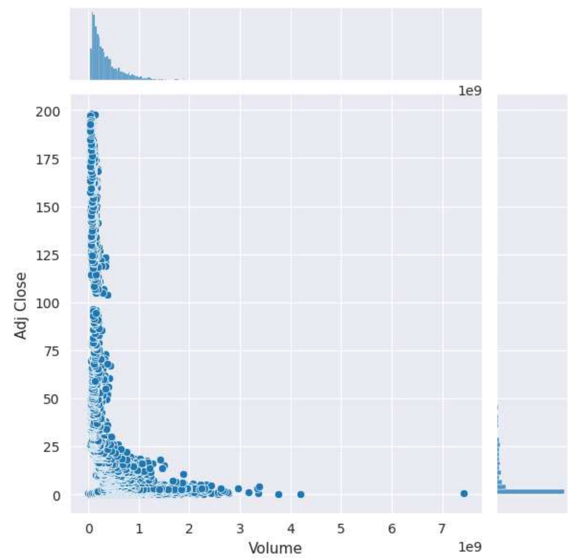
2.1 Date v/s Adj Close



2.2 Open v/s Adj Close



2.3 Close v/s Adj Close



2.4 Volume v/s Adj Close

4. Finding Similar Patterns using DTW

The methods that were created to find the patterns are:

- *Normalize Method:* Normalizes a time series data.
- *dtw_distance Method:* Calculates the DTW distance between two current window and past window time series. Fastdtw predefined function is used to make the process faster
- *find_most_similar_pattern:* Finds the most similar pattern in historical price data to a given window of days.

  There is a variable "min_distance" that stores the smallest distance between the past window and current window.

  It compares each calculated distance with the minimal distances found so far, and are recorded in min_distance. If the calculated distance is less than any of the previously determined minimum distances, that entry in min_distances is updated with the new distance and the current start index.

Visualization:

- Visualizing the most similar patterns identified by DTW in the historical stock price data.
- Aligning and reindexing these patterns to gain insights on past trends.
- Plotting various graphs to understand the relationship between different features and the adjusted close price.

4. LSTM model for Stock Prediction

- *Data preprocessing:*

  The dataset is divided into two sets: training and testing. Training takes up 80% of the data.

  The data is then scaled with MinMaxScaler to normalise it between 0 and 1.
- *Training the model:*

  The training data is created by employing a window of 60 preceding prices to generate input sequences (x_train) and their matching output (y_train).
- *Model Training:*

  The prepared training data is used to define and train an LSTM model.
- Then the results are plotted on graph.

5. Results

5.1. Scenario for Study

The project focused on Stock Analysis + Prediction using DTW and LSTM, the study aims to analyze historical stock price data to identify patterns and trends using dynamic time warping (DTW) techniques. The primary objective is to understand the underlying patterns in the historical data of a selected stock (e.g., Apple Inc. - AAPL) and assess their relevance for predicting future price movements. The study will involve collecting historical stock price data from reliable sources like Yahoo Finance and applying DTW to detect similar patterns in the data. By examining the identified patterns, the study seeks to gain insights into the stock's historical performance and its potential implications for future investment strategies. The project will focus solely on utilizing DTW for pattern recognition, with LSTM models reserved for future predictive analysis. Through this research, valuable insights into the behavior of the selected stock and the effectiveness of DTW for pattern analysis in financial data will be gained, contributing to the broader understanding of stock market dynamics.

5.2. Results obtained

dynamic time warping (DTW) analysis Results

By using dynamic time warping (DTW) analysis, similar patterns are recognised in the historical stock price data for apple (APPL)

Patterns were identified using different window sizes, including 15, 20, and 30 days, to capture varying trends in the data.

The graphs also give the month and year the similar has identified.

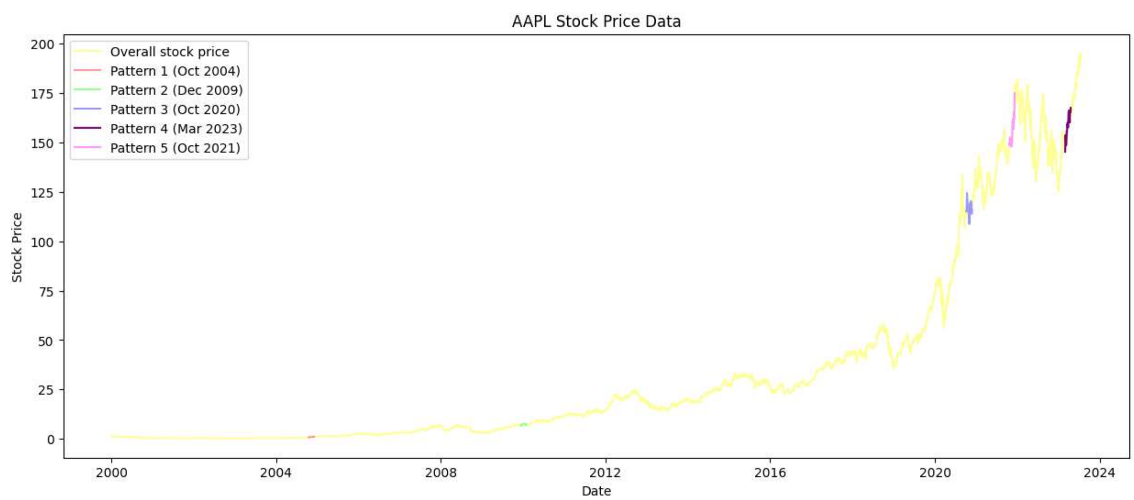


Fig 3.1 The graph visualise the 4 patterns similar to current pattern in past years (window size - 15 days)

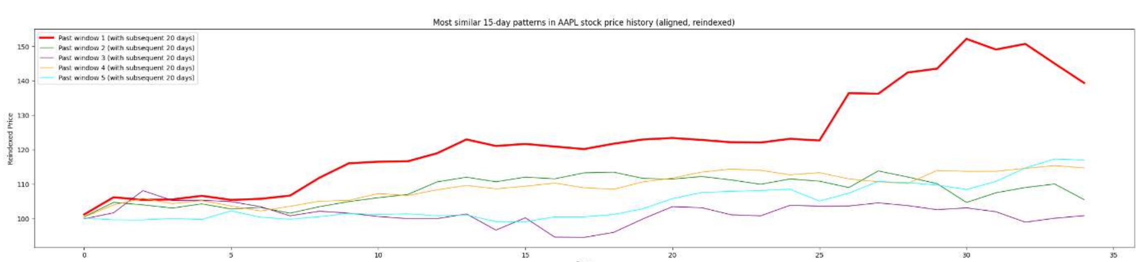


Fig 3.2 The graph visualise the 4 patterns similar to current pattern in past years (where pattern1 corresponds to past window 1)

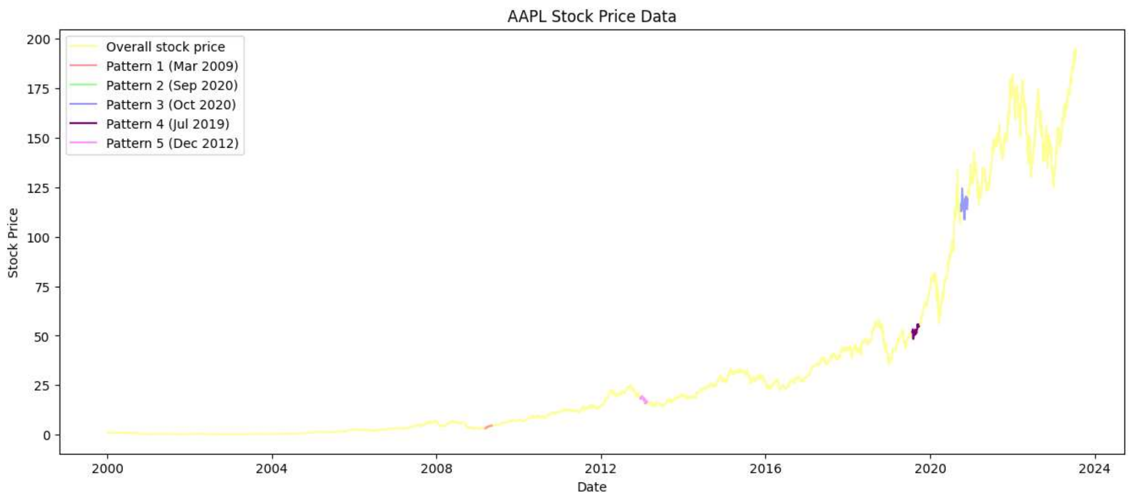


Fig 4.1 The graph visualise the 4 patterns similar to current pattern in past years (window size - 20 days)

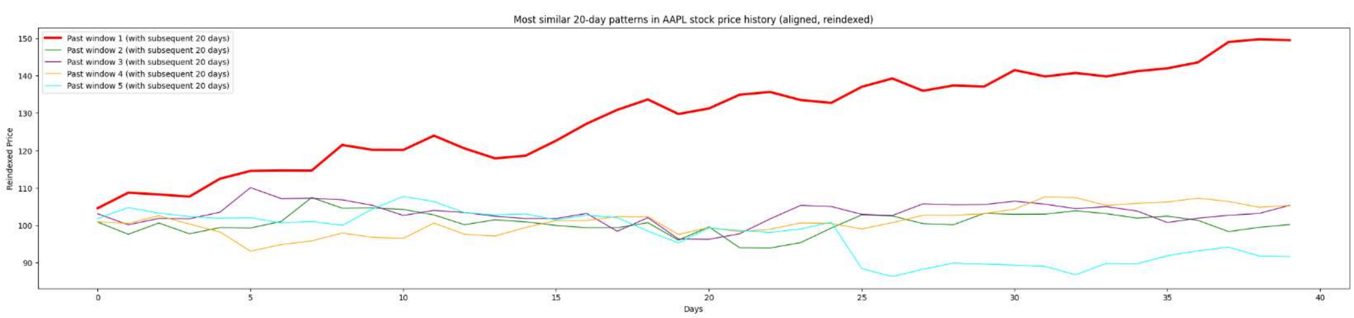


Fig 4.2 The graph visualise the 4 patterns similar to current pattern in past years (where pattern1 corresponds to past window 1)

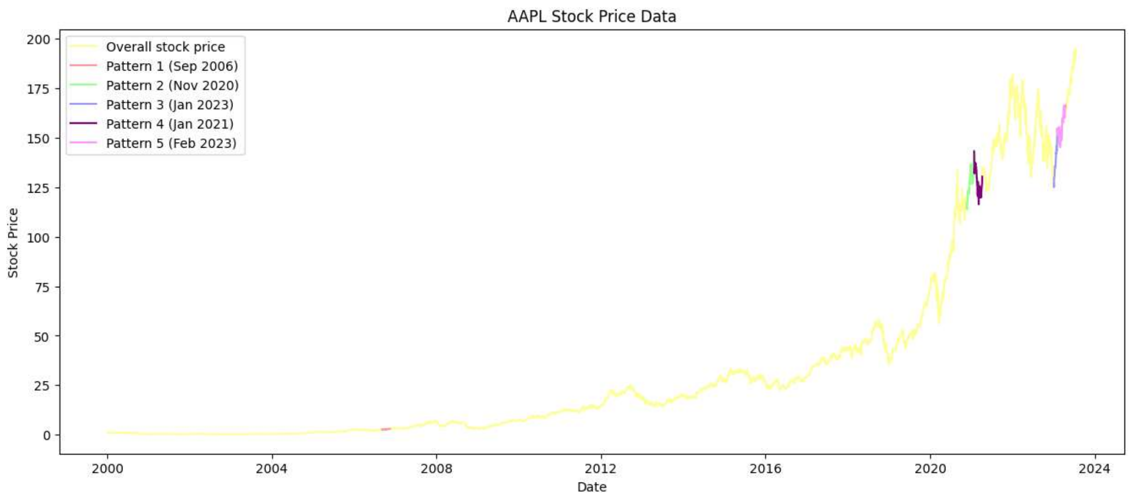


Fig 5.1 The graph visualise the 4 patterns similar to current pattern in past years (window size - 30 days)

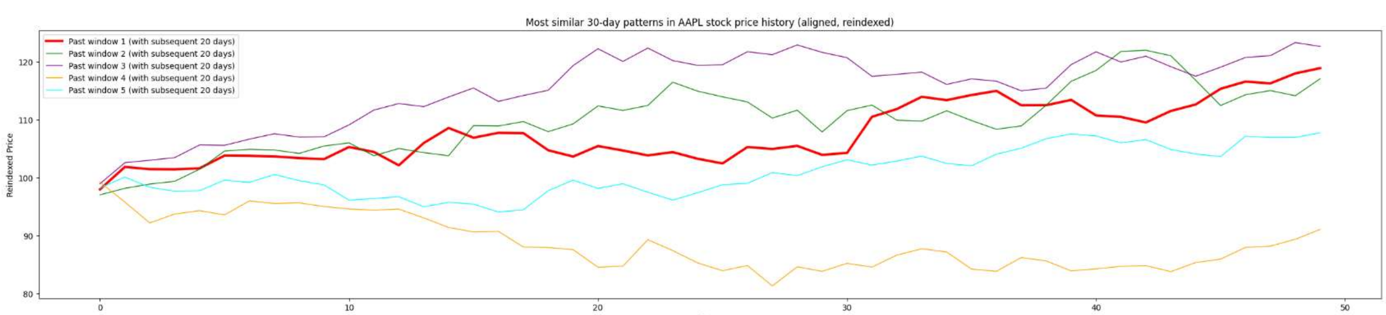


Fig 5.2 The graph visualise the 4 patterns similar to current pattern in past years (where pattern1 corresponds to past window 1)

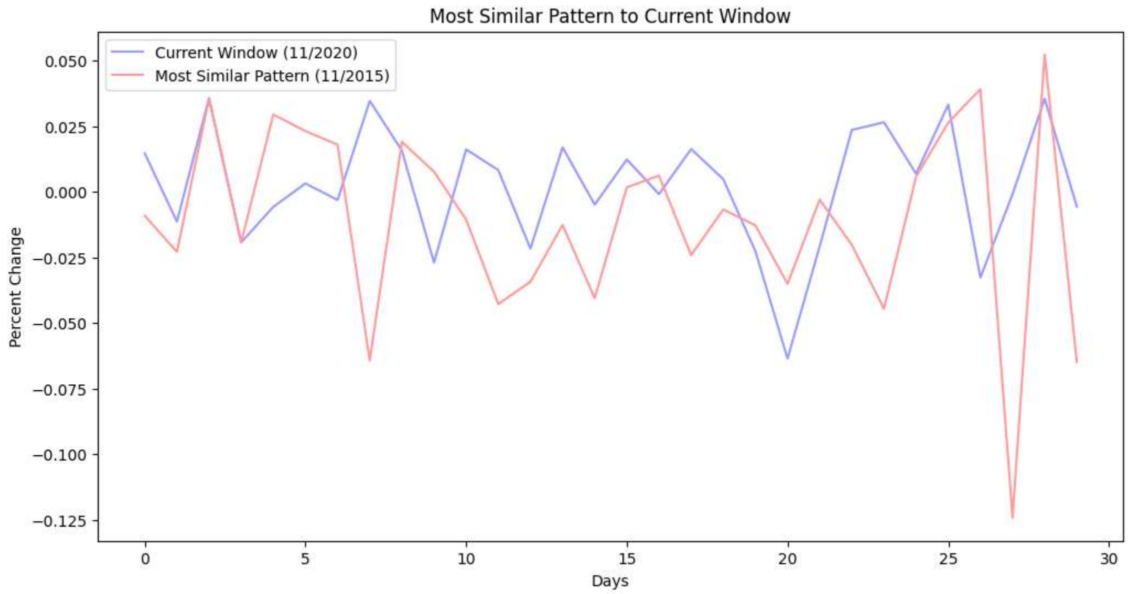


Fig 6 The graph plots similar pattern between two years 2020,2015

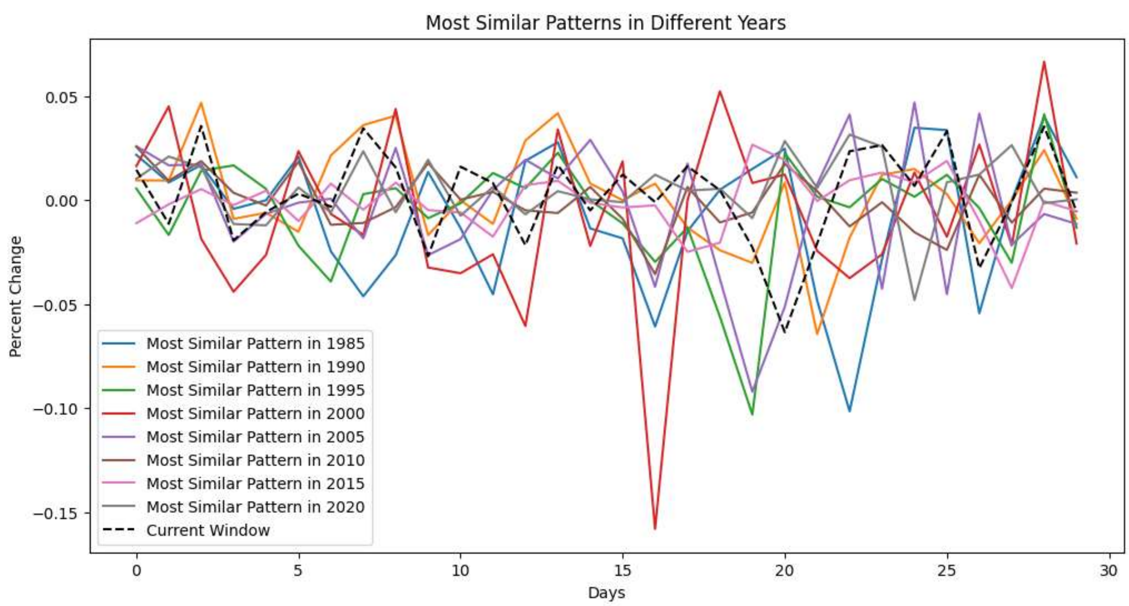


Fig 7 The graph plots similar patterns across different years

LSTM model results for Stock Prediction

From the graph, it can be seen that the predicted price is mostly similar to the actual price.

Fig 8. The original prices and the predicted prices are plotted on a graph to visualize how well the model performs in predicting the stock prices**.**

```
54/54 [==============================] - 2s 28ms/step
3.545571115096271
```

Fig 9. The RMSE (Root Mean Squared Error) score of 3.545571115096271

6. Summary

Certainly, predicting stock market performance is a complex task influenced by various factors. Here are some simple points highlighting the challenges:

1. Dynamic Nature: Stock prices are subject to constant changes
2. Multifactorial Influence: Numerous factors impact stock prices, including economic data, company performance, geopolitical events, and investor sentiment.

3. Historical Patterns Aren't Definitive: While historical data analysis is a common tool, past performance doesn't guarantee future results.
4. Black Swan Events: Unpredictable and rare events.

7. Literature

- S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, India, 2017, pp. 1643-1647, doi: 10.1109/ICACCI.2017.8126078. keywords: {Companies;Time series analysis;Logic gates;Predictive models;Forecasting;Data models;Machine learning;Time series;Stock market;RNN;LSTM;CNN},

- Tian Han, Qinke Peng, Zhibo Zhu, Yiqing Shen, Huijun Huang, Nahiyoon Nabeel Abid,A pattern representation of stock time series based on DTW, Physica A: Statistical Mechanics and its Applications,Volume 550,2020,124161, ISSN 0378-4371.

- Lee, Kunwoo & Jun, Suk & Jeong, & Jae, Suk. (2012). Trading Strategies based on Pattern Recognition in Stock Futures Market using Dynamic Time Warping Algorithm. Journal of Convergence Information Technology. 7. 185-196. 10.4156/jcit.vol7.issue10.22.