DSC 550 TERM PROJECT MILESTONE 3 - TOOLS FOR INVESTOR'S - S&P 500

ANBUSELVAN MAHALINGAM

BELLEVUE UNIVERSITY

DSC550 DATA MINING

BRETT WERNER

03/02/2024

S&P 500 INDEX DATA ANALYSIS FOR INVESTORS

Introduce the Problem

In the realm of financial markets, accurate forecasting of indices like the S&P 500, DOW 30 and a few other indices are cornerstones for investment strategies. The ability to predict market movements, even with a slight edge over market averages, can significantly impact investment returns. Given the S&P 500's role as a barometer for the overall U.S. stock market, developing predictive models for its performance is both a challenge and a necessity for investors. The same can be expanded for any other index or stock for investor need.

Justification and the Importance

The importance of solving this problem lies in the potential for investors to optimize their portfolios, minimize risk, and identify opportunities for growth. Accurate forecasts can guide investment decisions, from timing market entry and exit to allocating resources among different asset classes. In a broader sense, enhancing forecasting models contributes to the stability and efficiency of financial markets.

Pitch to Stakeholders

To stakeholders, by leveraging the market index and stocks using advanced forecasting models, we can gain actionable insights into future market movements. This model's integration into our investment strategy framework can enhance decision-making processes, improve risk management, and potentially lead to superior returns. The adoption of such predictive tools is not merely an upgrade, it's a transformation towards data-driven finance.

Data Source Explanation

The data for this project was sourced from reputable financial databases that provide historical market data, including prices, volumes, and other relevant market indicators for the S&P 500. This

historical data forms the backbone of our analytical approach, ensuring that our models are built on comprehensive and accurate datasets.

For this project, we focus on the S&P 500 and Dow Jones Industrial Average (Dow 30) indices. The dataset encompasses 5 years of historical data. Reference: Yahoofinance.com for obtaining these indices historical data.

S&P 500: https://finance.yahoo.com/quote/%5EGSPC/history?p=%255EGSPC

The Standard and Poor's 500, or simply the S&P 500,is a stock market index tracking the stock performance of 500 of the largest companies listed on stock exchanges in the United States. It is one of the most commonly followed equity indices and includes approximately 80% of the total market capitalization of U.S. public companies.

Dow 30: https://finance.yahoo.com/quote/%5EDJI/history?p=%255EDJI

The Dow Jones Industrial Average, Dow Jones, or simply the Dow, is a stock market index of 30 prominent companies listed on stock exchanges in the United States.

Exploratory Data Analysis

During the EDA phase, we focused on identifying patterns, trends, and outliers in the S&P 500 historical data.

Comprehensive statistical overview of the dataset:

S&P 500 the average closing price is around 3778, and the mean volume is approximately 4.31 billion, reflecting the average trading activity.

Dow 30 the average closing price is approximately 31000, and the mean volume is around 343 million, suggesting lower trading volumes compared to the S&P 500

	Open_sp500	High_sp500	Low_sp500	Close_sp500	Adj Close_sp500	Volume_sp500	Open_dow30	High_dow30	Low_dow30	Close_dow30	Adj Close_dow30	Volume_
count	1258.000000	1258.000000	1258.000000	1258.000000	1258.000000	1.258000e+03	1258.000000	1258.000000	1258.000000	1258.000000	1258.000000	1.25800
mean	3777.426413	3800.220382	3753.356557	3778.343116	3778.343116	4.306184e+09	30994.985198	31177.771859	30803.684907	31000.664278	31000.664278	3.43304
std	630.107259	631.949095	628.910413	630.461755	630.461755	1.064672e+09	3871.336152	3863.303453	3885.919279	3875.314886	3875.314886	1.04982
min	2290.709961	2300.729980	2191.860107	2237.399902	2237.399902	1.296530e+09	19028.359375	19121.009766	18213.650391	18591.929688	18591.929688	8.61500
25%	3152.619995	3182.530029	3137.015014	3154.027405	3154.027405	3.662092e+09	27248.515625	27375.132812	27147.099610	27276.270508	27276.270508	2.79337
50%	3917.295044	3949.779908	3891.505005	3918.785034	3918.785034	4.048580e+09	32176.530274	32486.354493	31897.425782	32197.125000	32197.125000	3.22580
75%	4329.079956	4361.334961	4291.397705	4328.559814	4328.559814	4.687198e+09	34192.227539	34410.108399	33981.481446	34206.047852	34206.047852	3.81985
max	4804.509766	4842.069824	4785.870117	4839.810059	4839.810059	9.976520e+09	37818.050781	37933.730469	37650.980469	37863.800781	37863.800781	9.15990

<u>Trend analysis of closing prices</u>


Closing Prices of S&P 500 and Dow 30
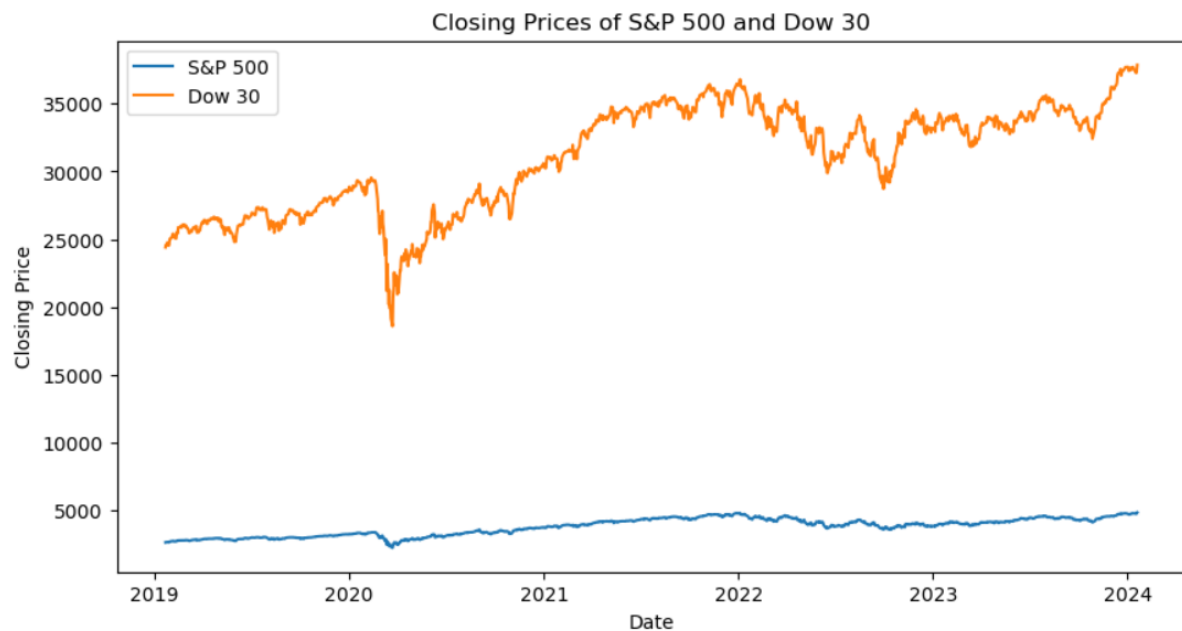
The long-term trend for both indices S&P 500 and DOW 30 has been positive over the last five years. This information can be crucial for long-term investors, suggesting that despite periods of decline, the overall trajectory has been one of growth.

There was a slight down trend and volatility around the year 2020, and both the indices picked up later and gained considerable growth. The 2020 down trend could be the reason of political reason, or pandemic period reason.

The trends between the two indices seem to move in a correlated fashion, rising and falling together, which suggests potential shared market drivers.

Volume Trend Analysis against Date:
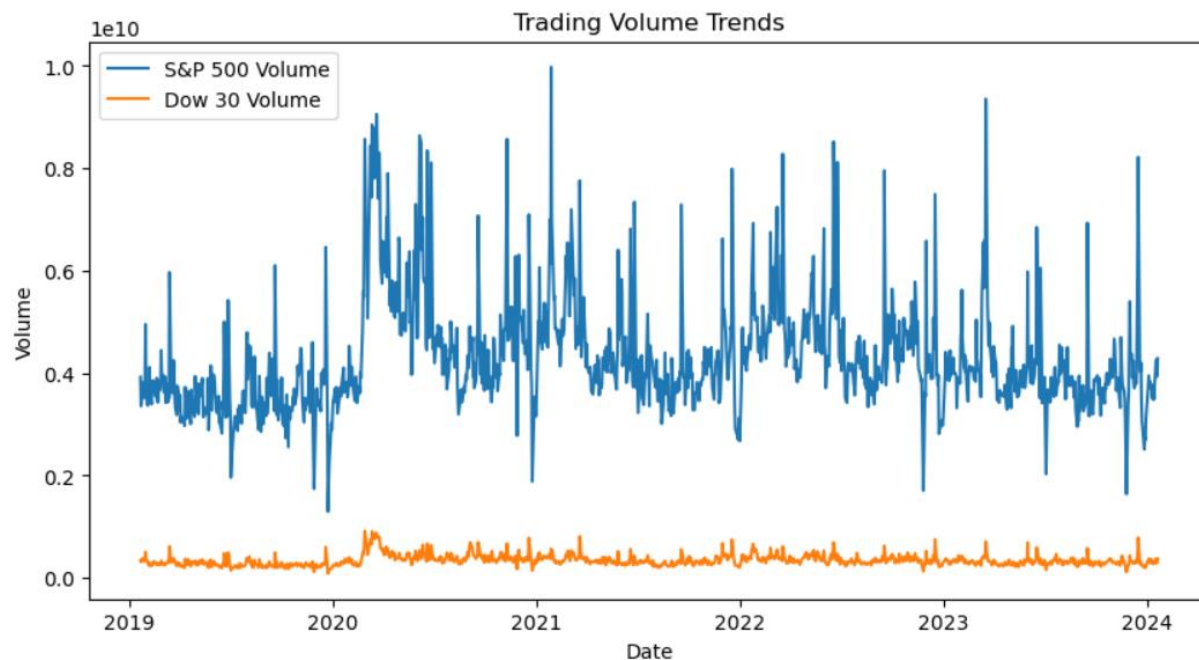
The S&P 500 shows significantly higher trading volumes compared to the Dow 30.

The S&P 500's volume displays spikes volatility in volume this depicts the investor sentiments about buy ad sell. The trading volume for the Dow 30 appears relatively consistent with fewer spikes, which may indicate less frequent trading or lower liquidity in comparison to the S&P 500.

Around 2020 both the indices had some down trend in volume, this can be related with the price trend visualization.

Data preparation and EDA

We have created the following new data elements based on the closing price.

50-day moving average - a key indicator used by traders and investors to understand and react to market trends

Short term weekly and monthly mean closing price - 7-day moving average and 30-day moving average respectively.

Logn tern moving average - 90 days (Quaterly) mean closing price
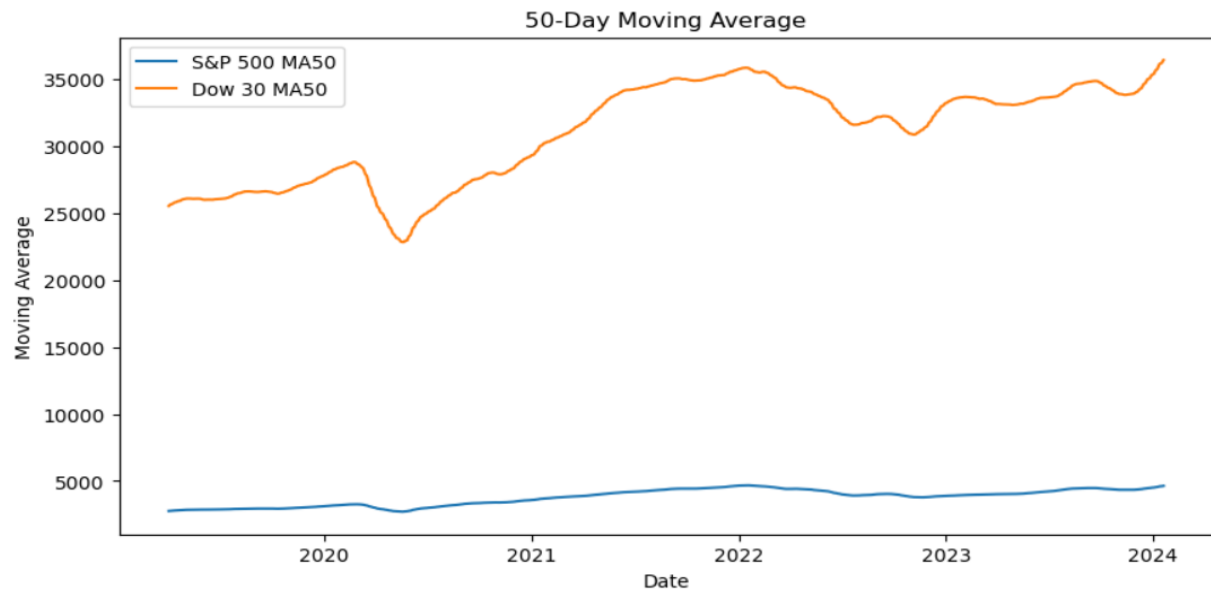
The MA50 of the Dow 30 appears to be at a higher level compared to the S&P 500 same as the closing price trend analysis.
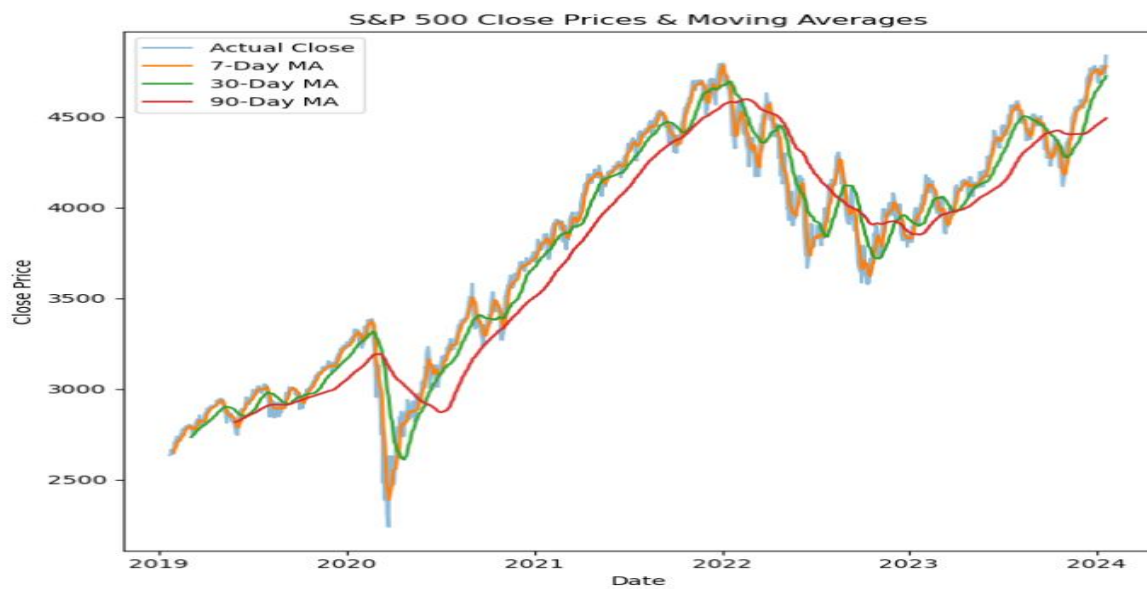
Here both the S&P 500 and the Dow 30 almost follows the same pattern over the past 5 years, There was a sharp dip around 2020 this could be the reason of COVID period. For the long term both the index outperformed the market and how's the sing of bull.

The actual closing price - Daily

It is the most volatile line on the chart, with all the daily market movements reflected. It shows real-time market   sentiment and is sensitive to daily market news, economic reports, and other events. There was a sharp dip around 2020, this could be the reason for COVID period and market down trend.

The 7-Day Moving Average - short-term trend line

It helps to identify the short-term trend and momentum. If the 7-day MA is trending upward and crosses above the longer-term MAs, it is a Bull sign, If the 7-day MA is trending downward and crosses below the longer-term MAs it is a Bear sign. Currently for the year 2024, we can see the Bullish trend which started from the year 2023 mid approximately.

The 30-Day Moving Average - medium-term trend line

The 30-day MA medium term trend direction can be established when comparing the actual closing prices to the 30-day MA. If the actual closing price stays above the 30-day MA, it is a Bull sign.If the actual closing price stays below the 30-day MA, it is a Bear sign.

The 90-Day Moving Average - longer-term trend line

The 90-day MA - long-term trend direction can be established when comparing the actual closing prices to the 90-day MA. If the actual closing price stays above the 90-day MA, it is a Bull sign. If the actual closing price stays below the 90-day MA, it is a Bear sign.

Based on investors' financial plan and their goal investors can follow the short-, medium- and long-term trend analysis and plan their investment options.

Here a similar activity can be performed for any stock or index and identify their trend over the period of time same as above.

Model and evaluation

Considering the goal of predicting stock index/price based on historical data we choose the time series analysis ARIMA model for its simple computational power need. Here we consider S&P 500 stock index historical data and apply ARIMA model fit with optimized p, d, q parameter values to forecast the future values.

We have identified the best ARIMA model parameter p, d, q as (2, 2, 3) with the minimum best AIC: 13117.71 value. We have applied the ARIMA model fit with the best parameter and predicted the next 5 days stock market prediction for the index price.

Model Validation

Predicted 5 days index price are given below.

```
Last Business day from training dataset: 2024-01-19 00:00:00
Last index from training dataset: 1257
Forecasted Values of next 5 business days from last business day 2024-01-22 to 2024-01-26
1258    4843.351933
1259    4843.208616
1260    4846.742907
1261    4847.066569
1262    4849.783500
```

And the Actual index price for the days,

```
Date: 2024-01-22 00:00:00, Actual Closing Price: 4850.430176
Date: 2024-01-23 00:00:00, Actual Closing Price: 4864.600098
Date: 2024-01-24 00:00:00, Actual Closing Price: 4868.549805
Date: 2024-01-25 00:00:00, Actual Closing Price: 4894.160156
Date: 2024-01-26 00:00:00, Actual Closing Price: 4890.970215
```

The Mean Squared Error (MSE) between the forecasted values and the actual closing prices for the specified dates is approximately 979.48, and the Root Mean Squared Error (RMSE) is approximately 31.30.

The ARIMA(2, 2, 3) model showed a reasonable ability to forecast future stock index closing prices over a short horizon. The forecasts were relatively close to the actual prices, with deviations within an expected range given market volatility.

The model's performance, as indicated by an RMSE of about 31.30, suggests that while the forecasts were not perfect, they captured the general direction and magnitude of price movements to a useful extent.

Forecast the index price for next 252 business/trading days(approximately one trading year)
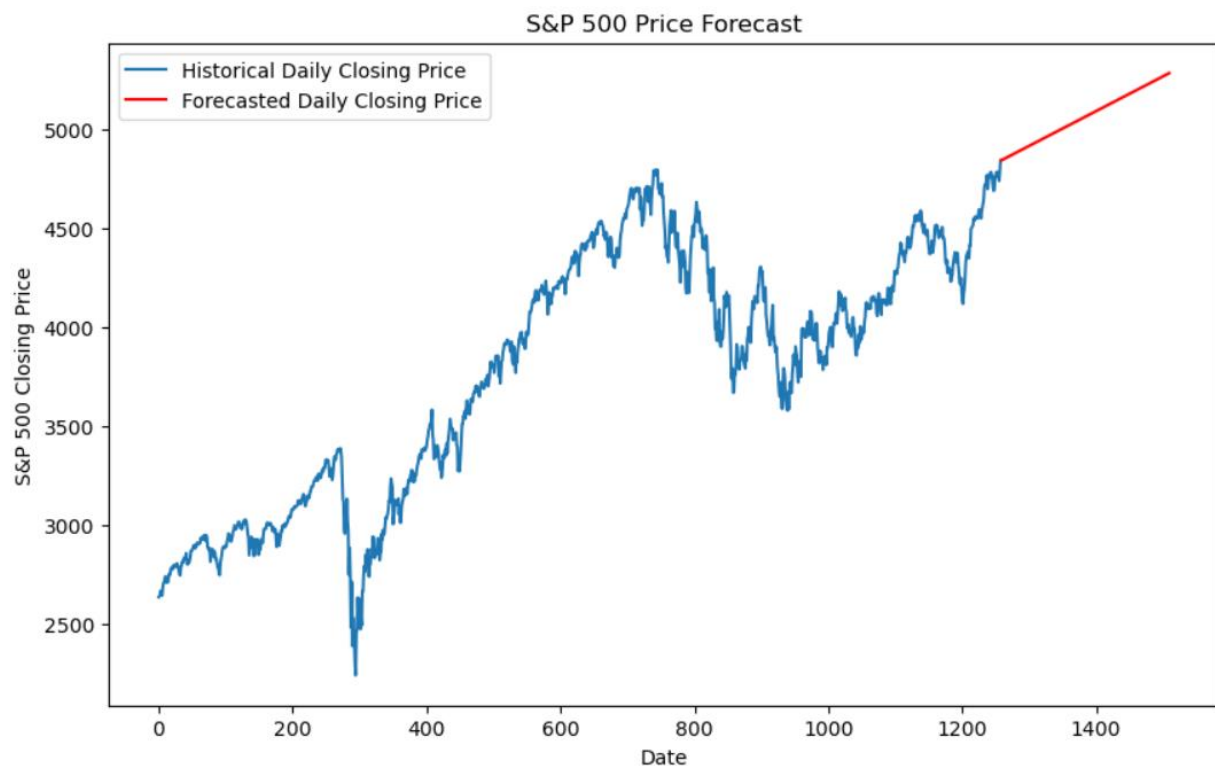
Conclusion

The ARIMA(2, 2, 3) model showed a reasonable ability to forecast future stock index closing prices over a short horizon. The forecasts were relatively close to the actual prices, with deviations within an expected range given market volatility.

The model's performance, as indicated by an RMSE of about 31.30, suggests that while the forecasts were not perfect, they captured the general direction and magnitude of price movements to a useful extent and there is a need for improvement in the model.

Assumptions

This project assumed that all available information was already reflected in the market price, and we were using only the historical data for understanding the data and identifying the relationship. This assumption might not account for all external influences on market volatility and index movements, we need to consider all other market influencing variables too.

Limitations and Implications

Here the analysis was performed based on historical data, but in the real scenario we need to consider additional market impacting variables (macroeconomic factors or sector-specific performances). The following are some additional variables that could have been considered for the analysis.

- ♦ Interest Rates: Changes in interest rates can affect the cost of borrowing, consumer spending, and investment levels.

- ♦ Inflation Rates: Inflation impacts purchasing power and can influence central bank policies, affecting stock prices.

- ♦ Gross Domestic Product (GDP): Economic growth indicators such as GDP growth rate reflect the overall health of the economy and can influence market sentiment.

- ♦ Unemployment Rates: High unemployment can signal economic distress, while low rates may indicate economic strength, impacting consumer spending and corporate profits.

- ♦ Election Results: Political changes can affect market regulation, economic policies, and investor confidence.

- Trade Policies and Tariffs: Trade agreements or disputes can impact industry sectors differently, affecting international trade and company earnings.

- Geopolitical Tensions: Conflicts, wars, or tensions between countries can create market volatility and uncertainty.

- Investor Sentiment: General market mood can drive market movements independent of fundamental indicators.

- Market Trends and Momentum: Trends, whether based on technical analysis or market momentum, can influence investor behavior and stock prices.

Challenges

The analysis acknowledged potential oversights, such as not fully incorporating macroeconomic factors or sector-specific performances, which could have provided more comprehensive insights into the indices' movements, I need to find a way to include all the market influencing factors and conduct the data exploration and proceed for further analytical process.

Is this model ready to be deployed

For some extent the model and its prediction results can be used for investors with our implications advising market investments are subjected to risk and investors should agree and proceed using the model.

Recommendations

In order to fulfil and improve the model's accuracy in prediction we should consider including all the market influencing variables (macroeconomic factors or sector-specific performances), one of a simple solution would be adding one or more derived variables from those macroeconomic factors or sector-specific performance to define a market greedy index considering the economy and political situations, industry rank based on current market need, growth index, current value scores for the

current price. All together collect the historical data to it the model should perform better than the current model.

References

Yahoo finance , For obtaining the histoical data . [URL](URL)

Zacks investment research , For obtaining knowledge about stocks/index and its performance. [URL](URL)