INTRODUCTION

The stock market involves a large number of buyers and sellers. A rising stock market often aligns with a progressing economy, and a decline in the market may indicate a poorly performing one. A stock index is a collection of stocks that are bundled together to replicate an economy, market, or a sector. The index we have chosen for this analysis is the S&P 500. The S&P 500 is a market-capitalization-weighted stock market index that tracks the stock performance of 500 of the largest U.S. public companies. It is one of the most commonly followed equity indices and is considered by many experts as the best representation of the US stock market.

Stock and financial markets tend to be unpredictable and even illogical, making it hard to find reliable patterns and predict the future direction of a particular stock. Such problems can be dealt with using Machine Learning and Deep Learning models. Deep learning excels at handling complex structures and extracting relationships that further increase the accuracy of the generated results. However, stock prices are not randomly generated values; instead, they can be treated as a discrete-time series model based on well-defined numerical data items collected at successive points at regular intervals of time. Thus, transforming the time series using ARIMA is a better approach than forecasting directly, as it provides more reliable results. In fact, to date, the ARIMA model has remained the most widely used time series model for forecasting stock market series.

For our project, we plan to analyze the S&P 500 dataset from 2019 to 2022 to understand the world events that shaped the US stock market during this time frame. We also delve deep into the details of the dataset within individual sectors to better understand the impact of specific sectors and companies within the index. Ultimately, we perform a time-series forecast on a particular company’s stock to comprehend its future within the index.

ABSTRACT

Our primary goal in analyzing and predicting the S&P 500 Financial Index is to gain a deeper understanding of the global factors influencing the Index's current form and how these factors will impact the stocks of companies moving forward. We aim to conduct an in-depth analysis of how this index functions and comprehend the various components that drive the market. Analyzing a cumulative dataset and making inferences based on it would not provide a decisive representation of the whole picture. Hence, we intend to create individual datasets for each of the 11 sectors in the S&P 500 Index: Communication Services, Consumer Discretionary, Consumer Staples, Energy, Financials, Healthcare, Industrials, Information Technology, Materials, Real Estate, and Utilities. Subsequently, we analyze these sectors to understand the companies that impact the market the most and are most affected by world events. We also perform a separate analysis on companies within a particular sector to gain a deeper understanding of their role in shaping the Index. Finally, we draw conclusions from these analyses and conduct Time Series Forecasting on a company's data to predict its future.

PROBLEM FORMULATION

Our problem statement arises from the current nature of the S&P 500 Index. Numerous world events have significantly impacted the financial landscape, making it challenging to predict the Index's future. Given that the Index influences not only companies but also the US Dollar, understanding the events shaping the Index is vital. We plan to use our understanding to apply it to Time Series forecasting, predicting the future of Pfizer’s portfolio within the index. Pfizer's market position is crucial, and its performance affects public perception and future trends.

DATASET SOURCE

The internet is teeming with data about the S&P 500 Index. However, analyzing the entirety of historical data would be impractical. We focus on recent data post-2019, as significant events during this period have stirred concerns among financial experts. Events like the COVID Pandemic, US Elections, Russia-Ukraine War, Crisis in Yemen, and Elon Musk’s Twitter purchase have disrupted markets and made traditional prediction methods challenging. We utilize a dataset from Kaggle, a Data Science repository, containing attributes like Date, Open, High, Close, and Volume.

METHODOLOGY

Our analysis begins by plotting S&P 500 trends. While helpful, these trends alone don't fully represent stock behavior. Candlestick charts, representing High, Low, Open, and Close values, provide better insight. Certain events like COVID and elections are evident in these charts. Sector-wise analysis offers deeper understanding, including line graphs, box plots, and bar plots. These depict trends and disparities among sectors. The Healthcare sector, for example, benefited from increased demand due to COVID.

We forecast using an ARIMA model. Stationarity tests and auto ARIMA help determine the best model. Forecasting Pfizer's data shows its future trend. However, the model's accuracy is tested using the Ljung-Box test.

CONCLUSION

Our analysis of the S&P 500 Index highlights the mutual influence of world events and companies. In-depth sector-wise analysis reveals varying impacts of events on different sectors. Pfizer's future trend is forecasted using ARIMA. This project successfully conducts a Time Series Analysis and Prediction of the S&P 500 Financial Index for 2019-2022.