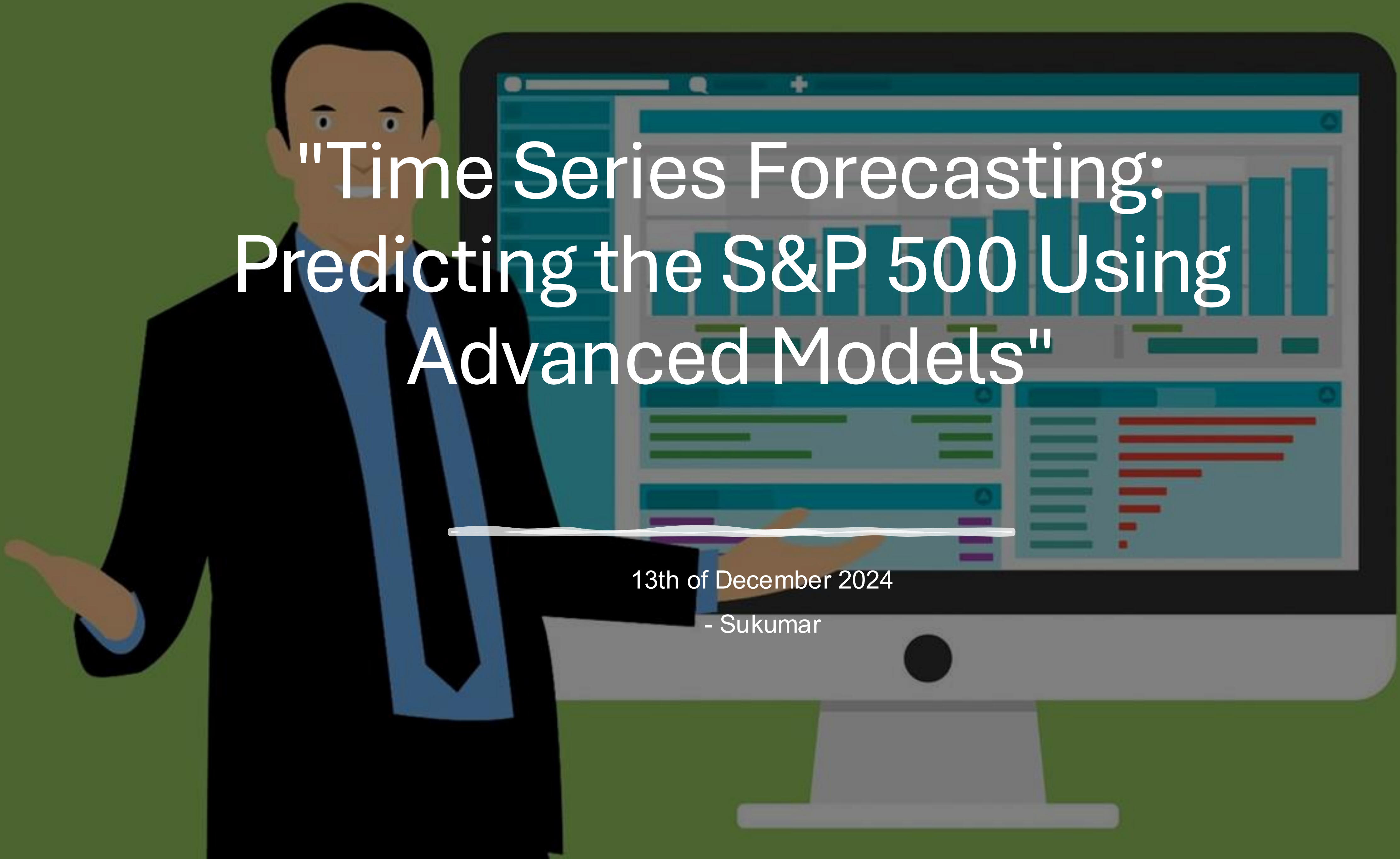


"Time Series Forecasting: Predicting the S&P 500 Using Advanced Models"

13th of December 2024

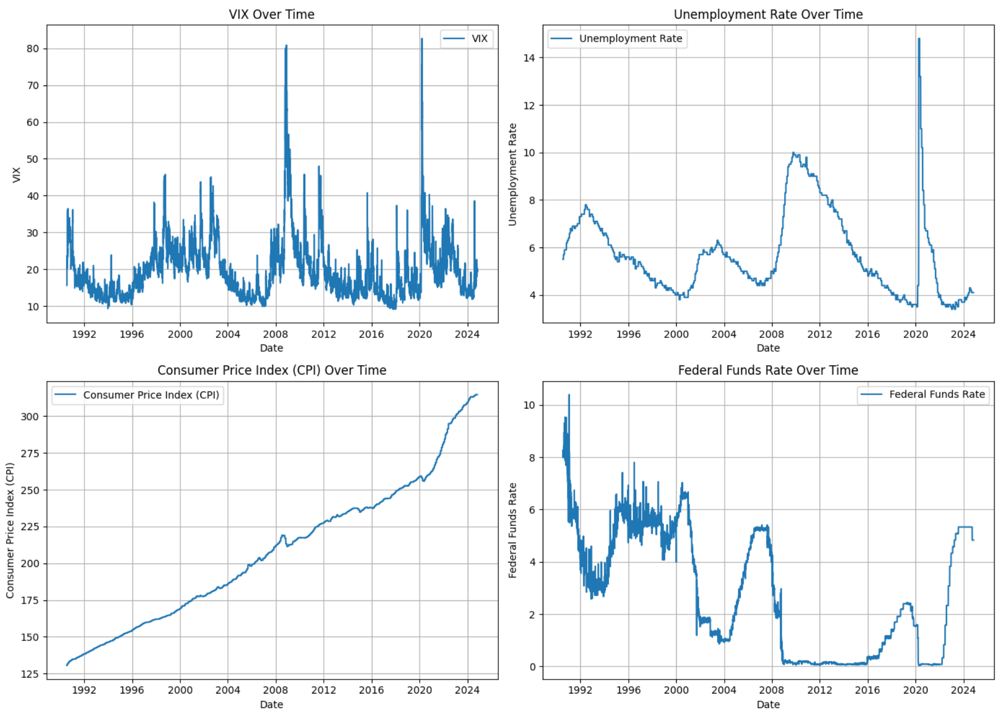
- Sukumar

Introduction to S&P 500 Forecasting

- Forecasting financial data is essential for **investors** and **market analysts**.
- The **S&P 500** serves as a key benchmark for U.S. stock market performance.
- **Accurate predictions** can improve investment decisions, risk management, and market analysis.
- **Goal of This Analysis:**
- Predict future values of the **S&P 500 index** using **time series analysis**.
- Utilize **economic indicators** to improve forecast accuracy.

This Photo by Unknown Author is licensed under CC BY-NC

Data Overview



•The dataset includes historical data from the S&P 500 Index, economic indicators, and other financial metrics.

•Key economic indicators include CPI (Consumer Price Index), Unemployment Rate, VIX (Volatility Index), and RSI (Relative Strength Index).

•**Time Frame:**

•The dataset spans from 1990 to 2024.

•The data frequency is daily/weekly/monthly, depending on the source.

•**Key Features:**

•Key columns in the dataset include: Date, S&P500_Close, VIX, CPI, Uninucleate, RSI, Bearish Score, and 200_MA.

•The dataset includes both price-related and macroeconomic variables.

•**Structure of the Dataset:**

•Each row represents a daily observation with multiple financial and economic features.

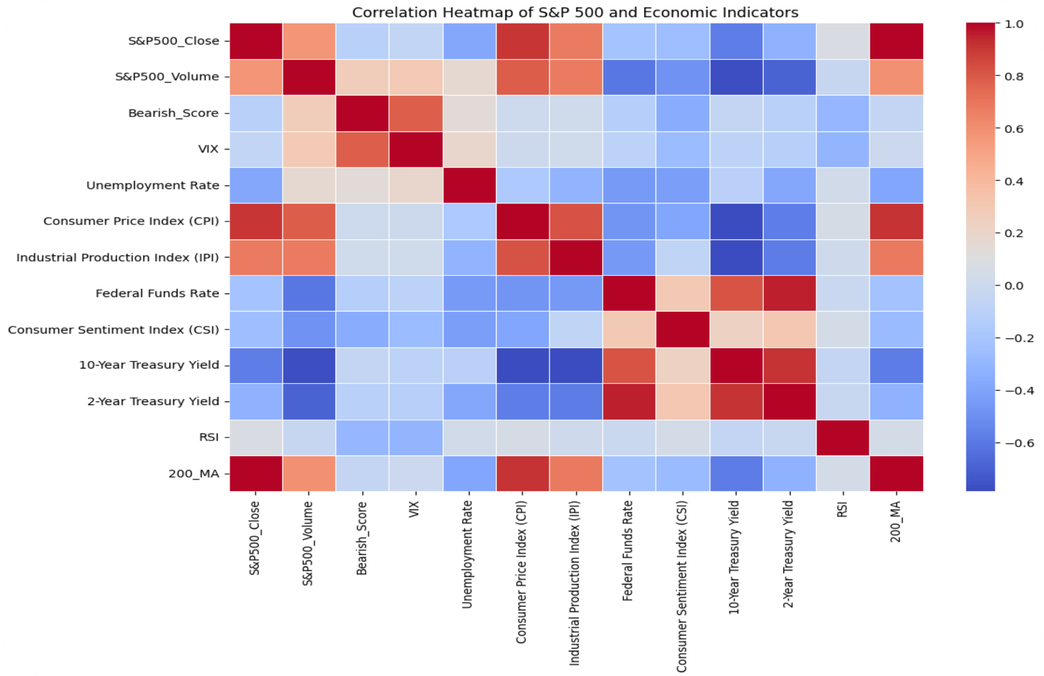
Exploratory Data Analysis (EDA)

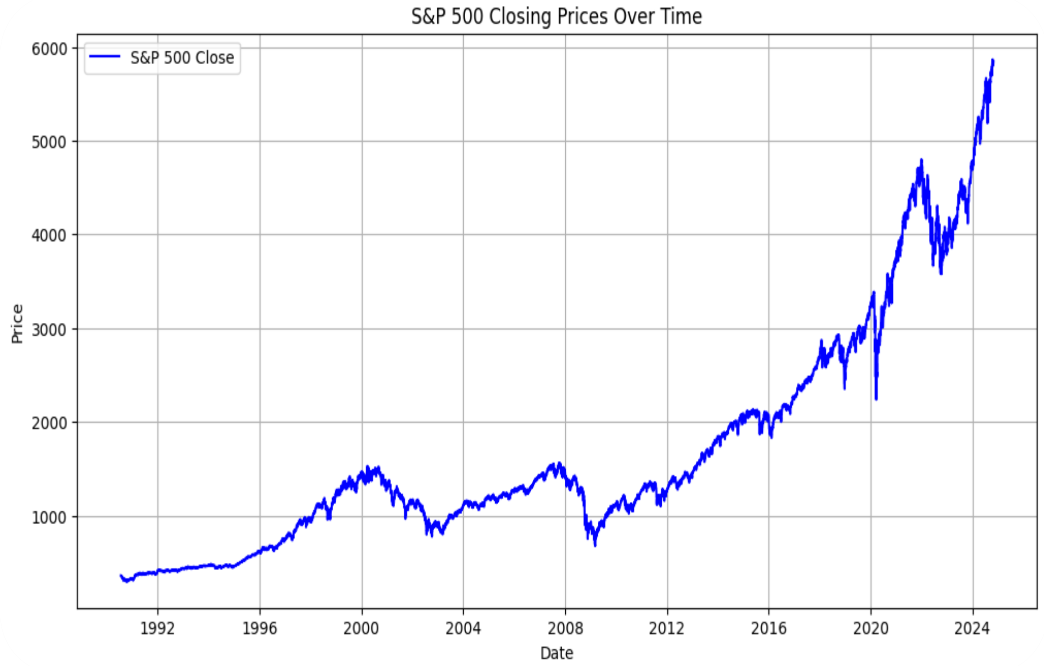
Strong positive correlation between the S&P 500 and CPI, suggesting inflation is a key driver of market growth.

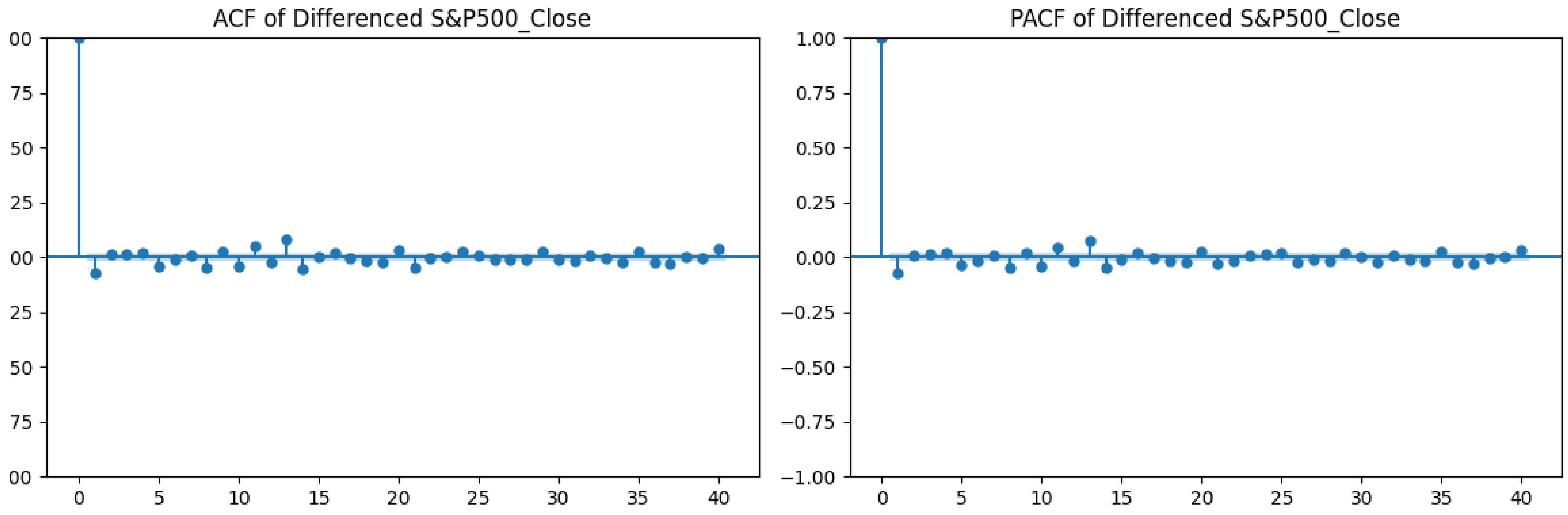
Moderate negative correlation between the S&P 500 and Unemployment Rate, indicating that high unemployment can lead to market downturns.

The S&P 500 index shows a general upward trend over time, with sharp drops during economic crises like 2008 and 2020.

Major market fluctuations are linked to broader economic events such as financial crises, inflation, and market shocks.

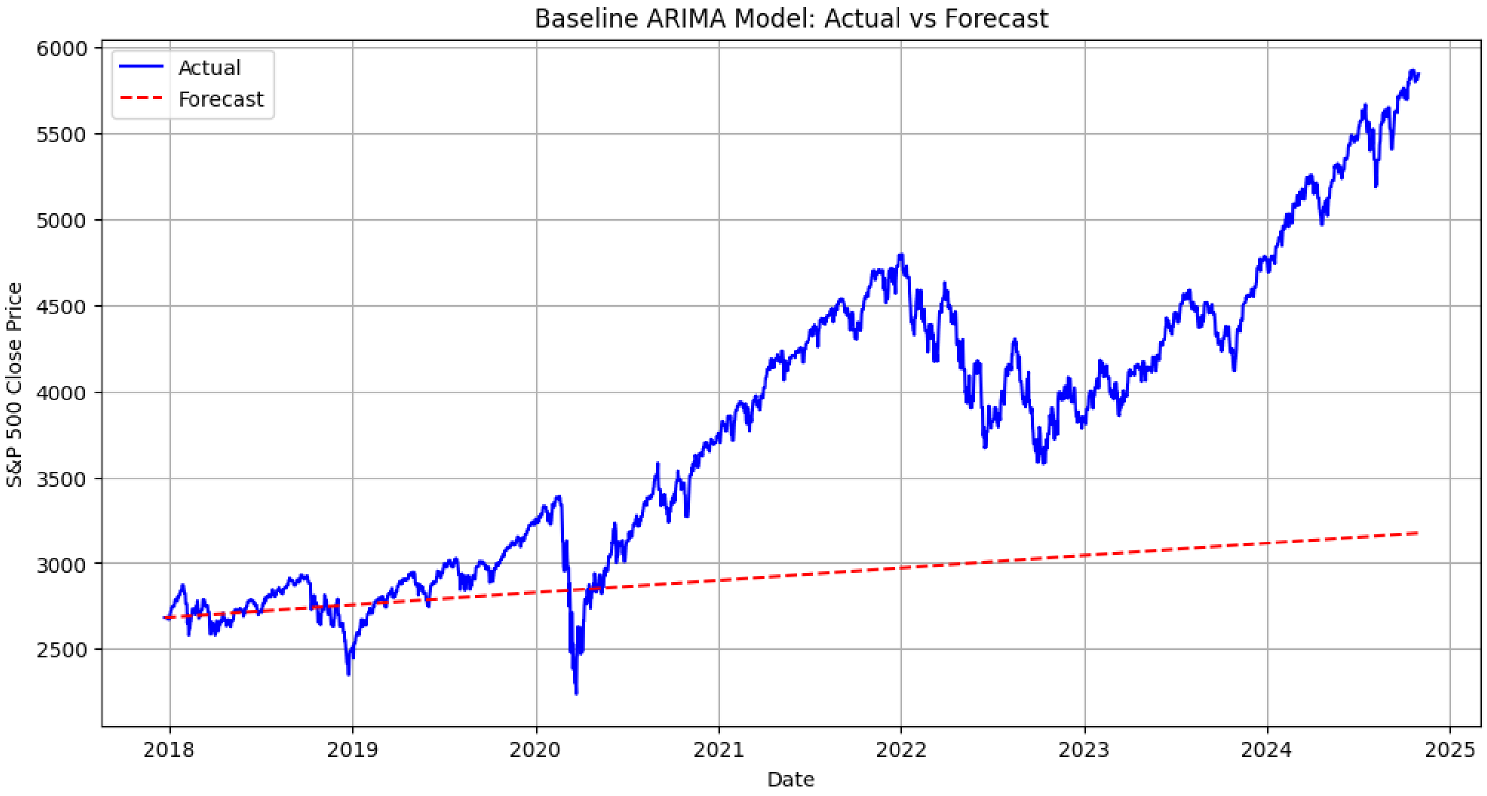






Baseline Model: ARIMA

- "ARIMA (Auto-Regressive Integrated Moving Average) is a widely used model for forecasting time series data.
- **It combines three components: AR (Auto-Regressive):** Relates the current value to previous values (lags).
- **I (Integrated):** Involves differencing the data to make it stationary.
- **MA (Moving Average):** Models the relationship between the residual errors.
- For ARIMA to work effectively, the data must be stationary, meaning its statistical properties (mean, variance) do not change over time. If the data is not stationary, differencing is applied to remove trends. Stationarity is vital because ARIMA models rely on the assumption that past values help predict future ones.

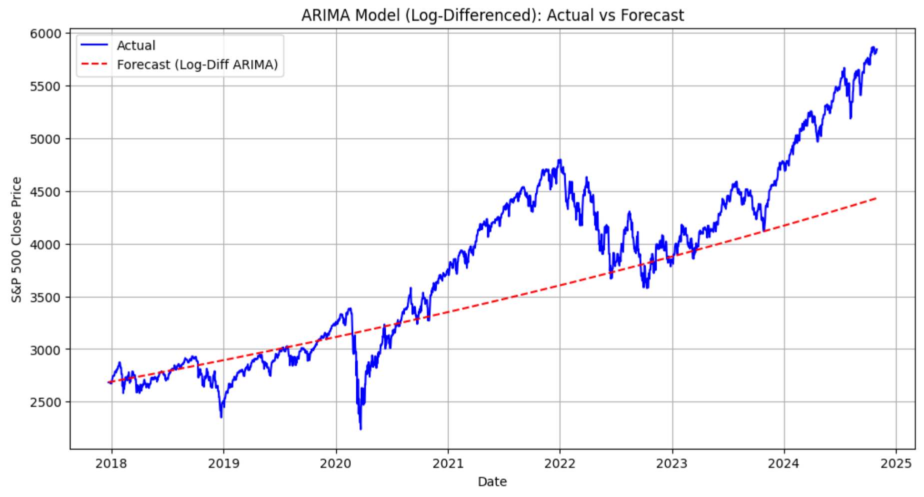


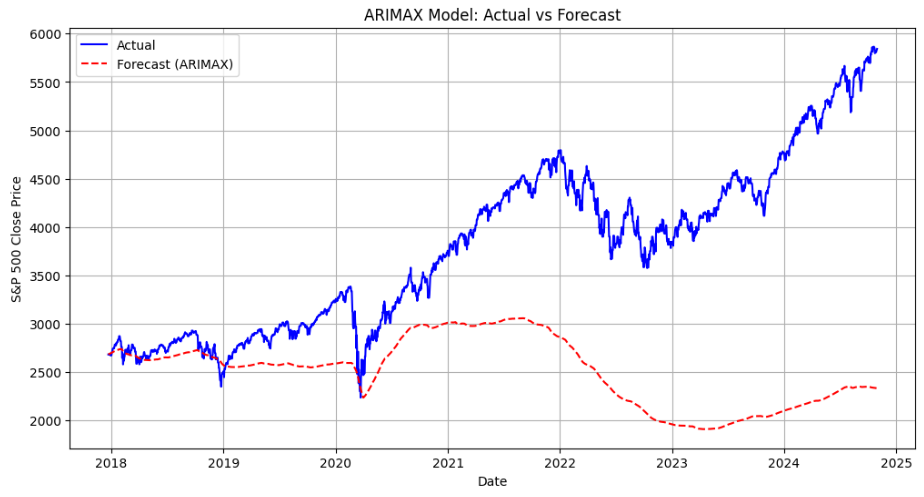
Baseline Model: ARIMA

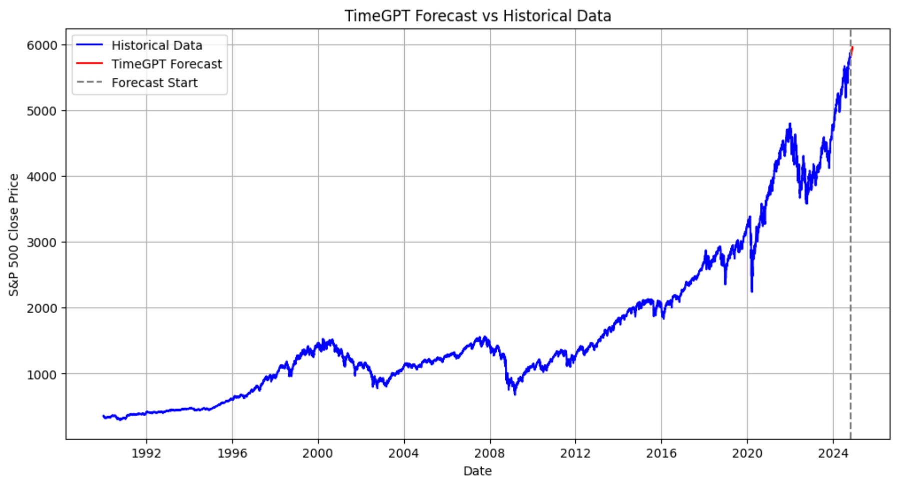
- The plot shows the ARIMA model's forecasted values (blue) compared to the actual observed values (orange). This demonstrates the model's accuracy and performance over time.

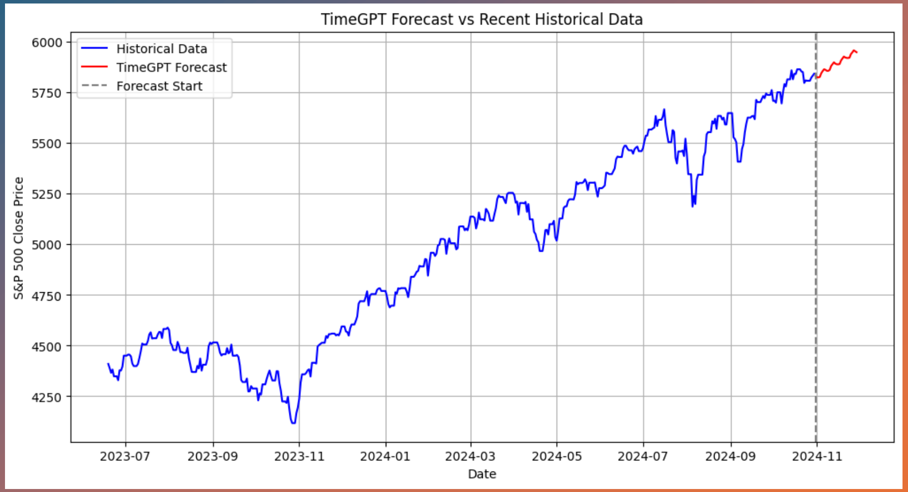
Enhancing ARIMA with Economic Indicators: ARIMAX/SARIMAX

1. **ARIMAX Model**:
   1. **ARIMAX (Auto-Regressive Integrated Moving Average with Exogenous Variables)** extends the ARIMA model by incorporating **external factors** (exogenous variables) that influence the time series.
   2. These **exogenous variables** can include economic indicators such as:
      1. **CPI (Consumer Price Index)**: Helps forecast future price movements.
      2. **Unemployment Rate**: Can predict economic trends impacting financial markets.
   3. By adding these variables, ARIMAX improves **forecasting accuracy** because it accounts for influences beyond the time series' own past values.
2. **SARIMAX Model**:
   1. **SARIMAX (Seasonal ARIMAX)** extends ARIMAX to handle **seasonality** in time series data.
   2. Useful for time series data that exhibit **seasonal patterns** (e.g., quarterly sales, monthly temperature).
   3. Adds **seasonal components** to ARIMAX, making it valuable for forecasting financial data that experiences regular cycles or seasonality (e.g., financial market cycles).
3. **Why Use ARIMAX/SARIMAX?**
   1. Both models provide better forecasting performance by including external factors (ARIMAX) or accounting for seasonality (SARIMAX).









Advanced Modeling: TimeGPT

Benefits of TimeGPT:

- **Accurate long-term forecasting**: It captures both the trend and seasonality in time series data more effectively.
- **Anomaly detection:** TimeGPT is particularly powerful at identifying and forecasting outliers or unusual trends that other models might miss.
- **Superior to ARIMA:** It goes beyond simple linear relationships to offer better predictions for volatile or highly complex datasets like the S&P 500.

Why TimeGPT for Time Series Forecasting?

- **Complexity Handling**: TimeGPT performs well with complex, volatile, and seasonally fluctuating data like stock market trends.
- **Forecasting Power**: TimeGPT outperforms traditional models in predicting future values with high precision.

Advanced Modeling: LSTM (Long Short-Term Memory)

Why Use LSTM for Time Series Forecasting?

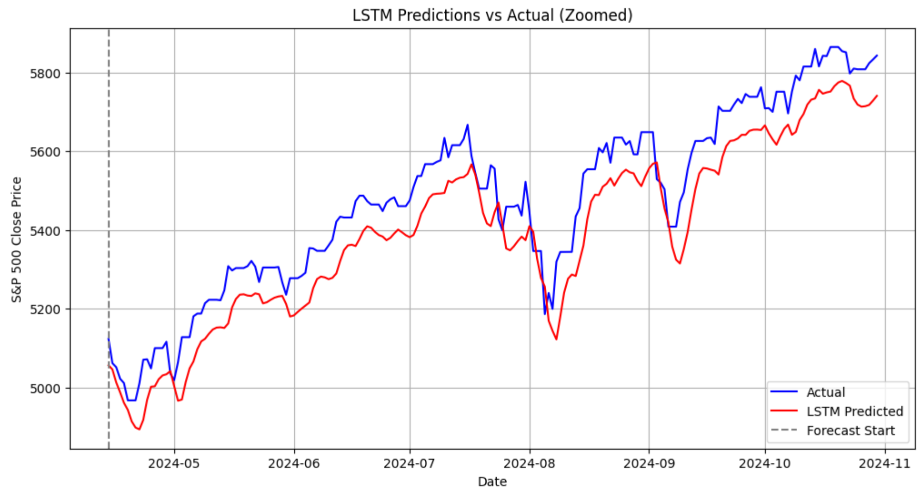
•LSTM performs well when data has **long-term dependencies** (e.g., trends, volatility).

•**Outperforms ARIMA** in capturing complex patterns and shifts in time series data.

Training Process:

•**Training History Plot**: Shows how well the model is fitting the training data over time.

•This plot can help you identify if the model is overfitting (where validation loss starts increasing while training loss decreases) or if it's underfitting (both losses remain high).



Model Comparison

Summarizing Model Performance:

- **ARIMA (Baseline):** Good for basic forecasting, but struggles with volatility and external factors.
- **ARIMAX:** Incorporates exogenous variables (CPI, Unemployment Rate), leading to improved forecasting accuracy.
- **TimeGPT:** Advanced deep learning model, captures complex trends, and has lower errors than ARIMA.
- **LSTM:** A deep learning model excelling at capturing long-term dependencies, with lower MSE/MAE values in certain test sets, making it highly accurate for forecasting recent trends.

Strengths and Weaknesses:

ARIMA: Simpler and less computationally expensive but may not account for external factors.

ARIMAX: Accounts for external variables but can be sensitive to the quality and relevance of input variables.

TimeGPT: Offers excellent forecasting accuracy with complex trends but may require more computational power and tuning.

LSTM: Very accurate for recent data, capturing long-term dependencies, but may be more complex and sensitive to how the data is structured.

Comparative Metrics:

Model	MSE	MAE	RMSE	MAPE (%)
ARIMA (Baseline)	1,297,896	887.28	1,139.25	20.34%
ARIMA (Log-Differenced)	300,007.49	396.75	547.73	9.30%
ARIMAX	2,688,741.41	1,285.08	1,639.74	29.76%
TimeGPT (Overall CV)	-	76.82 (MAE)	104.83(RMSE)	-
LSTM	3,325.86	44.51	*Calculated from MSE*	1.12%

Conclusion

Economic Indicators:

ARIMAX: No significant improvement in S&P 500 forecasting.

Economic factors (e.g., CPI, Unemployment Rate) not strong short-term predictors.

Advanced Methods:

TimeGPT: Technical indicators (e.g., 200-day MA, RSI) more influential than economic factors.

LSTM: Performed well with proper data cleaning and preparation (lower MAE and MAPE).

Research Question Answered:

Economic Indicators: Broad economic data did not improve short-term forecasts.

Technical Indicators: More valuable for short-term forecasting.

Advanced Models: TimeGPT and LSTM offer better predictions and insights than ARIMA.

Insights:

Data Handling & Model Selection: Key for accurate short-term forecasts.

Deep Learning (LSTM): Shows promise when data is well-prepared.