The foreign exchange market is one of the most dynamic and highly liquid markets globally, with trillions of dollars traded every day. Among the many currency pairs, the exchange rate between the US Dollar (USD) and the Chinese Yuan (CNY) is of particular significance due to its essential role in global trade, international investment, and geopolitical economics. The USD/CNY pair is also crucial for businesses, policymakers, and individuals who are exposed to fluctuations in exchange rates due to their international operations or investments.

Given that I am planning to go to Shanghai for my required study abroad at NYU, I found it particularly relevant to model the USD/CNY exchange rate to gain a deeper understanding of its dynamics and fluctuations. In this analysis, I set out to forecast the future movements of the USD/CNY exchange rate using various forecasting techniques, with an emphasis on time series models and machine learning algorithms. The primary models explored include the AutoRegressive model, Exponential Smoothing, and Support Vector Regression.

The data used for this analysis was sourced from Yahoo Finance, which provided daily high exchange rate data for the USD/CNY pair. The data spans approximately six months for simpler models and one full year for more complex models that require a longer historical context for training and validation. By focusing on the high exchange rate, I captured the peak values on each day, which is useful for observing significant shifts or spikes in the exchange rate.

The primary goal was to model and predict the future movement of the exchange rate based on past data. To start, I visualized the data to observe any inherent patterns or trends. This step revealed periods of significant volatility. At the same time, there were also stable periods where the exchange rate exhibited more predictable, steady movements. This mixture of volatility and stability in the data presented a unique challenge for forecasting models, as they needed to account for both predictable trends and unexpected shocks.

AutoARIMA is widely regarded as a robust and efficient model for time series forecasting due to its ability to automatically identify the optimal parameters for autoregressive (AR), moving average (MA), and differencing (I) components. This makes it particularly useful for modeling data that exhibits clear trends and seasonality. In theory, AutoARIMA is well-suited for datasets that are stationary or can be made stationary with appropriate differencing, and it can automatically handle many of the complexities of time series data such as seasonality, noise, and non-stationarity. For a time series like the USD/CNY exchange rate, which often exhibits both trend and volatility, AutoARIMA should be able to identify and adjust to these characteristics effectively. However, in my analysis, I encountered several challenges when attempting to apply AutoARIMA to this data.

The primary issue I faced was related to the frequency problem inherent in the dataset. The data for the USD/CNY exchange rate was collected on a daily basis, but AutoARIMA requires a predefined frequency to handle the time series appropriately. Initially, I tried to resolve this by using pandas' asfreq() function, which is intended to adjust the frequency of a time series. However, this caused several missing values due to the misalignment of dates between the data points. I attempted various strategies to handle the missing data, such as using the dropna() function to remove null values and the ffill() method to forward-fill missing values, but these methods led to issues where the model still failed to converge properly.

Additionally, another complication arose from the differencing process, which is often necessary to make a time series stationary. Differencing removes trends and seasonality, which is essential for building reliable forecasting models. However, the differencing process exacerbated the missing data issue, as it introduced new null values that could not be resolved even with the methods mentioned above. As a result, the model struggled to identify the underlying pattern and failed to produce accurate forecasts. Given these difficulties, I opted to switch from AutoARIMA to a more straightforward Autoregressive (AR) model, which doesn’t require as much parameter tuning or sophisticated preprocessing.

After switching from AutoARIMA to the Autoregressive model, I found that the simpler AR approach performed better in this scenario. While the predictions generated by this model were not perfect, they were relatively consistent with the overall trend of the exchange rate. The AR model helped to capture the general direction of the exchange rate, which was crucial for long-term forecasting. Although it couldn't handle the high volatility of the data, it gave a reliable estimate of whether the next data point would be higher or lower than the previous one. In volatile periods, the AR model was not able to adjust quickly enough, which reflected its limitation in capturing sudden shifts, but for the general trend, it provided decent forecasts. This made the AR model a suitable choice given its simplicity and effectiveness at modeling linear relationships.

Following the AutoARIMA attempts, I next applied the Holt-Winters Exponential Smoothing model. This model is well-suited for time series data that exhibits both trend and seasonality, which makes it an ideal candidate for financial data like the USD/CNY exchange rate. The Holt-Winters method uses exponential smoothing, which applies different weights to past observations based on how recent they are, thus emphasizing the most recent values while giving less weight to older values. This makes the model effective at capturing changes in trends and seasonality, which are common in financial time series data.

While the Holt-Winters model did a decent job of capturing the underlying trend in the USD/CNY exchange rate, it ultimately underperformed due to its inability to account for sudden and significant fluctuations or shocks in the data. One of the key shortcomings of the Holt-Winters model is its reliance on smooth trends. As a result, it has trouble responding to sharp and abrupt changes in the data, which are typical in the foreign exchange market. For example, during August of the analysis period, there was a significant and unexpected drop in the USD/CNY exchange rate. This kind of abrupt shift caused the Holt-Winters model to produce predictions that were consistently lower than the actual values, as it failed to recognize the sudden volatility and adjust accordingly.

After encountering limitations with AutoARIMA and Holt-Winters, I turned to Support Vector Regression (SVR), a machine learning model that is not inherently designed for time series forecasting but can still handle complex relationships in data. SVR works by mapping the original data into a higher-dimensional space using a kernel function and then fitting a regression model to predict the target variable. This approach allows SVR to capture non-linear relationships between features, making it especially useful for time series data that may have complex interactions or dependencies that are not easily captured by traditional statistical models like ARIMA or Holt-Winters.

SVR, with its flexibility, provided significantly better results compared to the other models, especially in capturing the underlying trend of the exchange rate. The main strength of SVR lies in its ability to map the data into higher-dimensional spaces, which helps capture complex and non-linear dependencies. For financial data such as the USD/CNY exchange rate, this is a significant advantage, as exchange rates are affected by a multitude of factors, including geopolitical events, economic indicators, and market sentiment. These factors often exhibit non-linear relationships that simpler models struggle to capture. SVR was able to learn these patterns, resulting in more accurate predictions of the direction and magnitude of future exchange rate movements. Additionally, the ability to fine-tune hyperparameters such as the kernel type, regularization parameter, and gamma made it a highly adaptable model.

The performance of each model was evaluated using various metrics, including the Mean Absolute Percentage Error for time series models and the Mean Squared Error for the SVR model. The AutoArima struggled to converge and did not produce reliable forecasts. Its performance was hampered by issues with missing values and the frequency of the data. The simplified AutoRegressive model I used in its place was able to be somewhat accurate in its predictions of the general trend of the exchange rate.

The Holt-Winters model performed better than AutoARIMA but still had difficulty in capturing sharp fluctuations, especially during volatile periods. It was able to predict the overall trend to some extent, but its inability to react to sudden shocks made it less reliable in fast-moving markets. After replacing the AutoArima with the AutoRegressive model, the Holt-Winters model was the worst of all the ones I tried.

On the other hand, SVR emerged as the best model in terms of predictive accuracy. The model's ability to handle non-linear relationships allowed it to capture the general trend of the exchange rate more effectively than the other models. The MSE was relatively low, indicating that the model provided reliable forecasts.

While SVR outperformed the other models, there are still some potential areas for improvement. One such approach could involve combining the strengths of different models to create a hybrid forecasting system. For example, the Autoregressive (AR) model could be used to capture the overall trend in the data, while SVR could be applied to capture more complex non-linear patterns. By combining these models, we could leverage the simplicity and robustness of the AR model for long-term trend prediction while using SVR to handle more intricate, short-term fluctuations and non-linear relationships

In a hybrid forecasting system, the strengths of both AR and SVR could complement each other in a way that improves accuracy and flexibility. For example, the AR model is well-suited to capture the long-term trend and smooth patterns in the data, while SVR can focus on predicting short-term fluctuations and reacting to sudden volatility. A combined approach would allow us to build a more robust system capable of forecasting the USD/CNY exchange rate with higher precision, as it would account for both linear and non-linear factors. The combination of both models could lead to an overall performance improvement and a more adaptable forecasting framework.

In conclusion, the analysis of the USD/CNY exchange rate using various forecasting models highlighted the strengths and weaknesses of each approach. The AutoRegressive model and Holt-Winters provided some useful insights but struggled with issues related to missing values and volatility. On the other hand, SVR performed significantly better by capturing the non-linear relationships in the data.

For practical purposes, combining the AutoRegressive model with SVR is likely to yield the most reliable results. By leveraging the strengths of both models, it is possible to create a robust forecasting system that can more accurately predict whether the USD/CNY exchange rate will increase or decrease. One possible way to combine the AR model and SVR would be to use the AR model for longer-term trend forecasting, capturing the broad direction of the exchange rate over weeks or months, and then applying the SVR model to predict more immediate movements, such as daily fluctuations or reactions to news events. Another potential approach could involve using AR to identify significant turning points or trend reversals, while SVR fine-tunes predictions by incorporating the latest market information and adjusting to sudden changes in the data. By merging the strengths of both models, this hybrid forecasting system would not only be more accurate but also more adaptable, able to handle both the linear and non-linear aspects of currency movements. Ultimately, such a hybrid model could offer a comprehensive and powerful solution for forecasting the USD/CNY exchange rate with greater precision and reliability.