# Time Series Forecasting for the Automotive Industry: Evaluating Traditional and Deep Learning

By: Steven Sullivan, Robert Lignowski, Uditi Shah, and Ahmad Javed

*Abstract*

The purpose of this project is to create and implement a stock price prediction model for several major automotive companies Ford, Toyota, Volkswagen, and Tesla. A time series forecasting model was created by analyzing historical stock price data from 2020 to 2024, as well as using predictive models via Arima, Prophet, XGBoost, and LSTM. The effectiveness of the models was tested by using evaluation methods such as RMSE and MAPE. Our analysis found that XGBoost had the most accurate predictions based on the lack of seasonality and/or patterns present in the historical data.

*Background*

The current state of the automotive industry has become an area of interest due to recent political and economic developments. Factors such as global trade, government policies, and political influence have shown to be potential indicators and predictors of how the market is doing currently as well as how it may perform in the future. Given how unpredictable the current market is, this project focuses on how easily market prices can be predicted and what are the major influences on the market itself.

*Methods*

We used historical and current stock price data from 2020 – 2025 via Investing.com as the primary data source. Stock prices data was gathered for all four car companies (Ford, Toyota, Volkswagen, and Tesla) and timelines were synced to ensure the same data points were used for each prediction model. Data was also cleaned and verified to make sure all date formats and column names were the same for all datasets, as well as checking for any missing values.

We used four algorithms to create our time series forecast models; ARIMA, Prophet, XGBoost, and LTSM to see which method was the best at predicting stock prices. Each model was selected to test different methods used in time series forecasting, and all models were evaluated using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error)

*Results*

Across all companies, we found little to no seasonality or patterns in stock prices over a five-year period. While all companies did show a spike in stock price in late 2021/ early 2022 there were very little overall trends across companies. XGBoost gave the most accurate predictions, as the model resulted in the lowest RMSE and MAPE scores. LSTM models also worked well, followed by Prophet and then ARIMA. The low seasonality and non-linearity were the major factors in determining which models would perform better than others, with XGBoost and LSTM having shown to be more equipped to handle irregular data as opposed to more traditional methods such as ARIMA and Prophet.

*Conclusion*

While overall economic and global factors can result in a change of stock prices (i.e. Executive Order 14037 in late 2021), we determined that stock prices are a result of multiple national, international, and market dependent factors. In terms of predictive modeling for stock prices, it is best to choose models that fit that data present to get the most accurate predictions.

In the case of automotive stock prices, due to the low seasonality and high volatility of the stock prices, models such as XGBoost and LSTM are the most effective as they can be engineered to better handle this type of data as compared to more traditional models. In the future, we can look to explore hybrid models and conduct more in-depth research on broader macroeconomic features and how to integrate them into our predictive models.

**GitHub Repository Link**

https://github.com/AhmJav01/DSCI-592-Group-8

## Introduction

While the automotive industry has always been a competitive and evolving market, recent global, economic and political developments have made many question what the future of the industry is going to look like, and how much influence do new policy shifts affect the state of the market. Tesla's CEO Elon Musk's involvement with the current Administration has made people wonder if this partnership will lead to a drastic shift in sales and stock price not just only amongst electric vehicles, but also amongst domestic and international car companies as well. Along with other factors such as tariffs, inflation and a shift towards electric vehicles, some wonder if it is possible to make sense of the current market state and if we can make predictions of what the future may bring.

This project focuses on creating a time series forecasting model to predict stock prices for four major car companies, each representing a different section of the market; Ford (American), Toyota (Asian), Volkswagen (European) and Tesla (Electric). This project will use historical and current stock price data ranging from 2020 to 2025 to train and test several algorithms to create a prediction model. These models include traditional time series models such as ARIMA and Prophet, as well as more engineered learning models such as XGBoost and LSTM. Our aim is to implement each model amongst the car companies and evaluate the using methods such as RMSE and MAPE.

By looking at company, region, and sector specific parts of the market, we hope to determine which model type is the best at predicting automotive stock prices as well as determine what influences are most consequential on the stock market. We hope that by creating multiple models using data from several car companies, we can lay the foundation for a highly accurate predictive system that can be used to help determine what the future of the market will look like.

## Data Pre-Processing and Acquisition

For our stock prediction analysis, we collected historical stock price data for four major automotive companies: Tesla, Ford, Toyota, and Volkswagen. The data was sourced from Investing.com under the "General - Historical Data" section. We selected these companies to capture both traditional and electric vehicle market trends. The data was downloaded manually in CSV format for each stock, ensuring access to reliable and up-to-date financial information.

Once acquired, the datasets underwent thorough preprocessing to ensure consistency and quality. First, we standardized the date formats and converted all relevant columns to appropriate numeric types. This included cleaning the ‘Volume’ column, where values with suffixes like ‘K’ and ‘M’ were converted to actual numeric values, and removing percentage signs from the ‘Change %’ column. The ‘Date’ column was also converted to datetime format to allow for accurate time-series analysis.

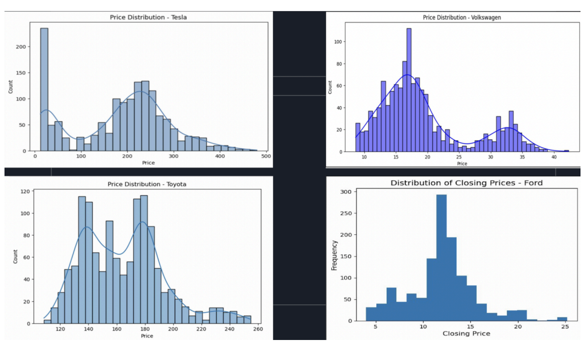
We addressed missing values by filling them with the median of the respective column, which helped preserve the data’s integrity without introducing significant bias. Outliers were identified using box plots and handled accordingly. Furthermore, we removed unnecessary columns, unified column headers, and ensured all datasets were aligned in terms of time frame and structure. This preprocessing phase ensured that the data was clean, consistent, and ready for visualization, analysis, and predictive modeling.

## EDA and Visualisation

We performed exploratory data analysis on historical stock data for Tesla, Ford, Toyota, and Volkswagen to uncover patterns, trends, and data quality issues. Before analysis, the datasets were cleaned by handling missing values and standardizing formats. All 'Date' columns were converted to datetime format, and numerical columns like 'Volume' and 'Change %' were transformed to remove non-numeric characters. After preprocessing, we confirmed there were no missing values across any of the datasets, allowing for consistent analysis.

To understand the distribution of data, we used histograms and time series plots. Tesla exhibited the most volatility, with sharp fluctuations in stock prices and frequent outliers, as shown in its box plots. In contrast, Toyota displayed the most stability. Volume spikes in Tesla and Ford stocks aligned with key events, suggesting potential external influences on trading activity. These trends were visualized using line charts for stock prices and bar plots for trading volume over time.

Finally, correlation matrices were generated for each dataset to assess relationships between variables. Strong positive correlations were observed among price-related features like 'Price', 'Open', 'High', and 'Low'. Volume and 'Change %' had weaker but still informative relationships with other variables. This analysis helped us identify important features for future modeling and confirmed the quality and structure of our cleaned datasets.




To gain a deeper understanding of the stock data, we employed a range of visualizations. Line graphs were used to illustrate the time series trends of stock prices for Tesla, Ford, Toyota, and Volkswagen. These plots revealed that Tesla had the most dynamic price changes over time, while Toyota maintained more consistent pricing patterns. Such visual cues helped highlight the volatility and stability across different companies.

Histograms were used for univariate analysis to show the distribution of daily stock prices and trading volumes. These visualizations helped us understand how often certain price ranges occurred and identified the central tendencies and spreads within each stock's dataset. Additionally, histograms of daily high prices and trading volumes revealed trading behaviors and investor activity over time.

We also utilized box and whisker plots to detect outliers in the stock prices. Tesla's box plot showed numerous extreme values, indicating frequent sharp market movements, while Toyota and Ford showed fewer outliers, reflecting more stable trends. Finally, correlation heatmaps were created to visualize the relationships between numerical variables like Price, Volume, and Change %. These helped in identifying redundant features and informed our feature selection process for modeling.

## Machine Learning Steps

We applied multiple models to forecast stock prices, starting with ARIMA and progressing to Prophet, XGBoost, LSTM, and a hybrid model for Volkswagen. Data was split using an 80/20 train-test ratio. For LSTM, a 30-day input window was used to structure the data for supervised learning.

ARIMA served as the baseline, with parameters selected using auto_arima. It captured basic trends but struggled with non-linear patterns. Prophet was configured with yearly seasonality and performed moderately, but not as well as other models.

XGBoost used lag features and rolling statistics, converting time-series data into a regression format. It was the best individual model overall. For Volkswagen, it achieved an RMSE of 1.4951 and a MAPE of 0.0999.

LSTM was trained after scaling and reshaping data into 3D arrays. It captured temporal patterns well, performing strongly on Volkswagen with an RMSE of 0.5467 and MAPE of 0.0364.

A hybrid model combining SARIMA and LSTM was implemented only for Volkswagen. The two models' outputs were averaged, resulting in the best overall performance: RMSE of 0.5388 and MAPE of 0.0360.

**Feature Engineering and Analysis**

For XGBoost, we created lag features (lag_1 to lag_7) and rolling statistics (mean and std over 3-day windows). This transformed the time series into a tabular regression format.

For LSTM, we normalized the data and used a 30-day input sequence, reshaped into 3D arrays for training.

All group members used the same feature engineering process for XGBoost and LSTM. For Volkswagen, additional seasonal modeling was done using SARIMA, which was combined with LSTM in the hybrid model. This hybrid approach improved accuracy by leveraging both statistical and deep learning strengths.

**Final Model Results**

This project evaluated multiple forecasting models to predict the stock prices of four major automotive companies: Volkswagen, Toyota, Tesla, and Ford. The models tested included ARIMA, Prophet, LSTM, and XGBoost, each applied to historical price data to evaluate their ability to capture stock trends and volatility.

Among all models, XGBoost delivered the most consistent and accurate performance across Volkswagen, Toyota, and Ford. Its ability to model complex, non-linear relationships and integrate structured external features made it highly effective in capturing stock price behavior influenced by real-world events. XGBoost was particularly strong in handling noise, reacting to short-term fluctuations, and maintaining stability across different time periods.

For Tesla, which has historically shown more volatile price movements, LSTM outperformed other models. Its strength in sequence modeling and memory-based architecture allowed it to capture Tesla's rapid changes in trend more effectively than traditional approaches.

Across the board, Prophet underperformed, often failing to adapt to the unpredictable shifts in stock prices due to its rigid assumptions about seasonality and trend. ARIMA showed reasonable performance but lacked the flexibility of machine learning and deep learning models, particularly in the presence of sharp discontinuities or external shocks.

A key insight from the model comparisons is that stock prices in the automotive sector are heavily influenced by factors such as earnings reports, CEO announcements, macroeconomic events, and news sentiment. XGBoost's superior performance in this regard can be attributed to its capability to accommodate such influences through feature engineering and flexible boosting techniques.

In conclusion, while LSTM showed strength in modeling Tesla's volatility, XGBoost emerged as the most robust and generalizable model overall, making it the final recommended model for stock price prediction in this study.

**Lessons Learned & Future Work**

Over the course of this project, we learned some very important lessons that we will be able to apply in our future career paths. The importance of data preprocessing was a major one, as we had to handle missing values, sync timelines, and change data types in order for us to be able to analyze and interpret our dataset. We also got the chance to gain experience with other techniques and models that we learned about but did not always get a chance to put into practice in earlier classes.

We got a good look at time-series forecasting, and learned how to implement it more effectively, and how to make allowances for its high sensitivity to seasonal trends. We also gained experience with multiple different machine learning models, eventually coming to the conclusion that XGBoost and LSTM were the best ones to be applied to non-seasonal/irregular data compared to more traditional models. We would recommend future research into hybrid modeling techniques and the inclusion of broader economic variables to further improve forecasting accuracy and interpretability.

**Summary**

This project was a unique opportunity to further our understanding of data science and machine learning, providing a comprehensive comparison of traditional and modern approaches to time series forecasting. By comparing traditional statistical methods (ARIMA, Prophet) with machine learning (XGBoost) and deep learning (LSTM) models to assess prediction accuracy, we were able to determine the best approach to predict the fluctuations and influences on stock prices, and to develop an approach for feature engineering and deep learning. We were also able to examine which models applied best to which situation and which stock, discovering that we could not use a one-size-fits-all approach to our analysis. Ultimately, it is obvious that there is no universal model best suited for all stock prediction tasks, but for the automotive sector—characterized by low seasonality and high responsiveness to external factors—machine learning and deep learning models like XGBoost and LSTM are effective and accurate. These models are flexible, adaptable, and capable of capturing complex patterns, making them valuable tools in financial forecasting.

**Reference**


- Chomiak, T., Hu, B. “Time-series forecasting through recurrent topology.” *Commun Eng* 3, 9 (2024). https://doi.org/10.1038/s44172-023-00142-8

- Dede Tarwidi, Sri Redjeki Pudjaprasetya, Didit Adytia, Mochamad Apri, “An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach,” *MethodsX*, Volume 10, 2023, 102119, ISSN 2215-0161, https://doi.org/10.1016/j.mex.2023.102119.

- Dieckmann, Jonas. “Getting Started Predicting Time Series Data with Facebook Prophet.” *Medium*, TDS Archive, 30 Jan. 2024, medium.com/data-science/getting-started-predicting-time-series-data-with-facebook-prophet-c74ad3040525.

- Kolambe, Milind. (2024). “Forecasting the Future: A Comprehensive Review of Time Series Prediction Techniques.” *Journal of Electrical Systems*. 20. 575-586. 10.52783/jes.1478.

- Pereira da Veiga, C., Pereira da Veiga, C.R., Girotto, F.M. et al. “Implementation of the ARIMA model for prediction of economic variables: evidence from the health sector in Brazil.” *Humanit Soc Sci Commun* 11, 1068 (2024). https://doi.org/10.1057/s41599-024-03023-3