Isabela State University Linker:
Journal of Engineering, Computing, and Technology
Volume 1, Issue 1
ISSN (Print): 3082-3676 ISSN (Online): 3082-3684
DOI: https://doi.org/10.65141/ject.v1i1.n4

Forecasting Gasoline Retail Prices Using Predictive Analytics

Joseph T. Banatao[1], Vince Aldrin F. Panganiban[2], Igancio F. Datul[3], Amy Lyn M. Maddalora[4]
Isabela State University, Cabagan, Isabela, Philippines[1,2,3,4]
josephbanatao14@gmail.com; amy.m.maddalora@isu.edu.ph

RESEARCH ARTICLE INFORMATION

Received: July 18, 2023
Reviewed: May 21, 2024
Accepted: May 29, 2024
Published: June 29, 2024

ABSTRACT

Forecasting gasoline and diesel retail prices is crucial for consumers, businesses, and policymakers due to the significant economic impact of fuel costs. This study aimed to address the challenge of predicting these prices using predictive analytics. It sought to identify weekly price trends, develop accurate forecasting models for various types of gasoline and diesel, and create an application to integrate these models. Data from Kaggle, encompassing 1,362 instances with attributes (date, regular, midgrade, premium, and diesel) over 26 years (1995-2021), were analyzed. The findings indicated inconsistent prices across gasoline types and a recent sharp decline in retail prices. Using the ARIMA (1,0,0) model, the study developed a forecasting model with minimal differences between actual and forecasted prices. The five-week forecast revealed a steady price increase, with a low RMSE of 0.22 to 0.26, indicating excellent model accuracy. Additionally, an application with a dashboard for visualizing weekly gasoline and diesel price trends was developed. This study successfully provides a model for forecasting retail gasoline prices, helping consumers anticipate price increases.

Keywords: *ARIMA Model, gasoline retail prices, forecasting, predictive analytics, time series analysis*

Introduction

Gasoline is one of the essential needs of consumers, especially those whose livelihood depends on passenger vehicles. Different gasoline stations have different prices, which is why forecasting gasoline retail prices helps consumers be aware of how gasoline prices increase and understand the reasons behind these increases. To an ordinary consumer, a sudden increase in gasoline prices can significantly impact their expenses, potentially disrupting their budget plans. This is especially critical for workers earning minimum wage or below, particularly in rural areas. Unexpected changes in oil prices have introduced another source of concern globally, as such changes have become more pronounced in recent years (Iyke, 2019). Hence, predicting gasoline prices may help consumers plan their budgets more effectively.

Forecasting natural gas prices has become an essential tool for stakeholders in the natural gas market, enabling better decision-making, and risk management, reducing the gap between demand and supply, and optimizing resource usage based on accurate predictions. For instance, the study by Zhang *et al.* (2019) evaluated various natural gas price forecasting models to enhance future prediction accuracy. Oil prices are characterized by their unpredictability, influenced by a multitude of factors. Numerous methods exist for analyzing and forecasting retail gasoline prices, with time series analysis being one of the traditional models employed for this purpose. Time series analysis can learn from historical data relationships and trends, aiding in making data-driven predictions or decisions.

In this light, this study was proposed to forecast retail gasoline prices in the Philippines, aiding establishment owners and consumers in strategic and budgetary planning. Generally, the study aimed to forecast gasoline and diesel retail prices by analyzing data from the past 26 years. Specifically, it aimed to determine the trend of weekly retail prices for each type of gasoline, develop a model to forecast retail gasoline prices and evaluate its accuracy, and develop an application that integrates the developed model.

This study investigated the prediction of retail gasoline prices by analyzing a 26-year dataset obtained from Kaggle. Employing time series analysis, this study aimed to develop a reliable model for forecasting future gasoline prices and assess using the RMSE (Root Mean Square Error) of the ARIMA (Autoregressive Integrated Moving Average) Model. The dataset encompassed prices for regular, midgrade, premium, and diesel gasoline which were expressed in US dollars per gallon. The classification of gasoline types was accomplished through interviews with retailers in Cabagan, Isabela such as Flying V and Jet Fuel. Notably, the study's forecasting scope was limited to weekly changes, excluding daily fluctuations. Additionally, predictions were solely based on existing data, offering insights into historical trends rather than speculative future projections.

Methods

This study utilized descriptive research and a developmental research design. A descriptive research design is useful in this study to learn from historical data and understand how it might influence future outcomes. The chance of future events based on historical data was determined using a predictive method. The goal was to provide the best forecast of what will happen in the future, rather than only knowing what has happened. A developmental research design was also utilized in this study since an application was developed to integrate the forecasting model for gasoline retail prices.

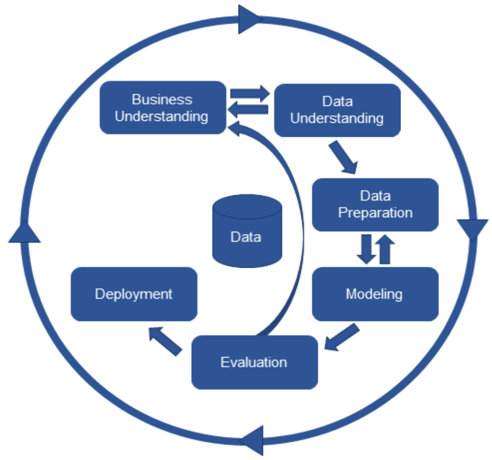

Figure 1. *CRISP-DM Model*

The CRISP-DM (Cross-Industry Standard Process for Data Mining) process model was adapted in this study, which offers a structured approach to project planning for data mining. It consists of six logical steps, from business understanding to modelling and deployment. CRISP-DM is the most widely used data analytics method that can be adapted to specific business needs (O'Hara *et al.*, 2023). Hence, this framework is the most suitable for this study since all the process steps had been undertaken to achieve the research objectives.

Business Understanding

Gasoline prices impact the economic status of countries, particularly the Philippines. Higher gas prices mean not just less money in consumers' pockets but also higher costs for businesses, some or all of which will be passed on to customers later. Individual consumers have a tendency to reduce their fuel purchase expenses. Forecasting gasoline retail prices helps consumers be aware of how gasoline prices increase and what the reasons are for the increase. This study may help both the owner and the consumer make better strategic and budgeting decisions. In order to establish strategies and regulations to mitigate the consequences of such oscillations, predictive models that can reliably and precisely predict such unforeseen events are required.

Data Understanding

In this study, Kaggle was used to obtain raw data in the global market from the internet. The independent variable of this dataset was the weekly data, and the dependent variable was the retail prices for each type of gasoline.

Data Preparation

The data used in this study consisted of 1,362 records and 14 attributes, namely date, A1, A2, A3, R1, R2, R3, M1, M2, M3, P1, P2, P3, and D, for the past 26 years, from 1995 to 2021.

Data cleaning and preprocessing were done manually through the use of MS Excel. The built-in function COUNTIF and conditional formatting were used to find missing and duplicate values. It was found that the dataset had no missing values or duplicates. However, unnecessary attributes (other types of gasoline) had been removed and renamed. Based on an interview with certain gasoline retail stores in Cabagan, Isabela, the types of gasoline identified in the dataset were those regularly purchased by customers in the Philippines. The price of the gasoline was converted from gallon to liter using MS Excel with the formula: gallon divided by 3.785. After data preprocessing, five attributes were finally obtained, as shown in Table 1.

Table 1. Attributes of the Gasoline and Diesel Retail Price Dataset

Name	Type	Description
date	date	weekly changes in gasoline retail prices.
regular	numeric	regular conventional retail prices (R2)
midgrade	numeric	midgrade conventional retail prices (M2)
premium	numeric	premium conventional retail prices (P2)
diesel	numeric	diesel retail prices (D)

Modelling

This study used time series analysis for forecasting gasoline and diesel retail prices. A time series is a group of data points that have been enumerated, graphed, or otherwise arranged according to time. Time series is most frequently defined as sequences taken at successive, equally spaced moments in time. Time series analysis is the use of a model to forecast future values based on values that have been observed.

A non-seasonal ARIMA model is classified as an ARIMA (p, d, q) model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.

ARIMA (1,0,0) is a first-order autoregressive model if the series is stationary and autocorrelated and can be predicted as a multiple of its own previous value, plus a constant. This is an "ARIMA (1,0,0) + constant" model. If the mean of Y is zero, then the constant term would not be included. If the slope coefficient is positive and less than 1 in magnitude (it must be less than 1 in magnitude if Y is stationary), the model describes mean-reverting behavior in which the next period's value should be predicted to be times as far away from the mean as this period's value. If the slope of the coefficient is negative, it predicts mean-reverting behavior with an alternation of signs, i.e., it also predicts that Y will be below the mean next period if it is above the mean this period.

Rapid Miner was used for modeling and forecasting because it is easy to use and suitable for the data.

Evaluation

In the evaluation of the forecasting model, the Root Mean Square Error (RMSE) was used to determine the accuracy of the model. It is a common method for calculating a model's error in predicting quantitative data. One of the two basic performance measurements for a regression model is the Root Mean Square Error (RMSE). It calculates the typical difference between values that a model predicts and actual values. It provides an estimate of the model's accuracy, or how well it predicts the desired result. The lower the RMSE, the better the model and its predictions.

Deployment

The rapid application development approach was used in the development of the dashboard for forecasting gasoline retail prices. Visual Basic.NET was used to develop an application that would allow gasoline store owners as well as consumers to view and visualize retail gasoline prices. One feature of the application consisted of a dashboard that visualizes the weekly trend of gasoline prices. The forecasting model was also integrated into the application, which may help the gasoline store owner make informed decisions on whether or not to stock additional fuels. MySQL served as the database for the system, and it was also used to manage the dataset.

Ethical Considerations

The researchers ensured the responsible and ethical conduct of their research on forecasting gasoline and diesel retail prices.

Results and Discussion

This section presents the findings, analysis, interpretation, and discussion of the gathered data to forecast gasoline retail prices through analysis of the obtained data for the past 26 years.

Weekly Trends of Each Type of Gasoline Retail Prices

The weekly patterns and trends of gasoline and diesel retail prices from 1995 to 2021 are depicted in Figures 2-5.

The weekly trend of retail prices for regular gasoline is highlighted in Figure 2. The results indicate that the highest peak of retail price for regular gasoline occurred on July 7, 2008, reaching 1.07 USD. However, the price sharply declined in 2009. In contrast, the lowest point was recorded on February 15, 1999, with a price of 0.237 USD. Furthermore, as depicted in the figure, from 2012 to 2015, there was no stable price for regular gasoline, as prices continued to fluctuate.

Figure 2. *Weekly Trend of Regular Gasoline Price*

The weekly trend of midgrade gasoline retail prices is shown in Figure 3. It reveals that the highest peak was recorded on July 07, 2008, with a retail price of 1.096 USD, while the lowest point was on February 15, 1999, with a retail price of 0.262 USD.

Figure 3. *Weekly Trend of Midgrade Gasoline Price*

The weekly trend of premium gasoline retail prices is revealed in Figure 4. The results indicate that the highest peak, with a price of 1.131 USD, was recorded on July 7, 2008. Conversely, the lowest point, at 0.286 USD, occurred on February 15, 1999

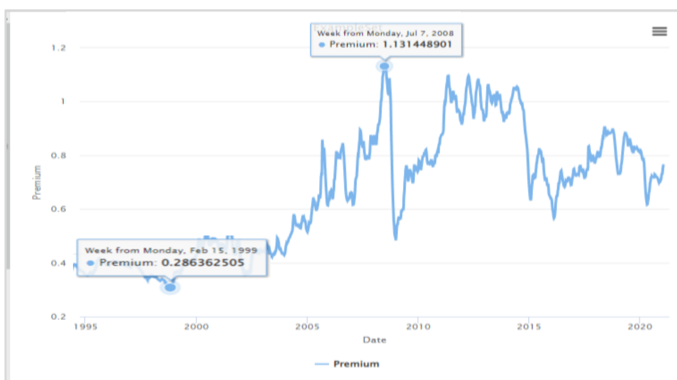

Figure 4. *Weekly Trend of Premium Gasoline Price*

The weekly trend of diesel retail prices is depicted in Figure 5. The findings indicate that the highest peak, at 1.249 USD, occurred on July 07, 2008, while the lowest point, at 0.253 USD, was recorded on February 15, 1999. Similar to other types of gasoline, there was an abrupt decrease in retail prices in both 2009 and 2016.

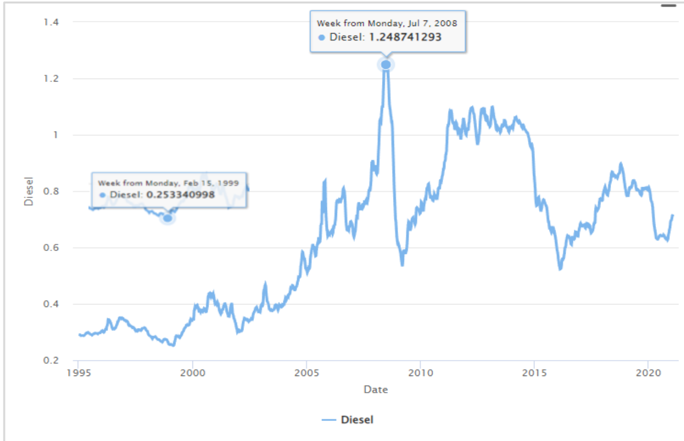

Figure 5. *Weekly Trend of Diesel Retail Price*

Forecasting Model for Gasoline and Diesel Retail Prices

The generated forecasting models for retail prices of gasoline and diesel are displayed in Figures 6 to 9. These models provide price forecasts for a period of 24 weeks (6 months), offering a comprehensive and detailed perspective on the trends and patterns in the data.

The forecasting model for regular gasoline is depicted in Figure 6. The forecasted price of regular gasoline ranges from 0.6099 USD to 0.6214 USD over a period of five weeks. As illustrated in the figure, the results show a gradual increase in gasoline prices over time.

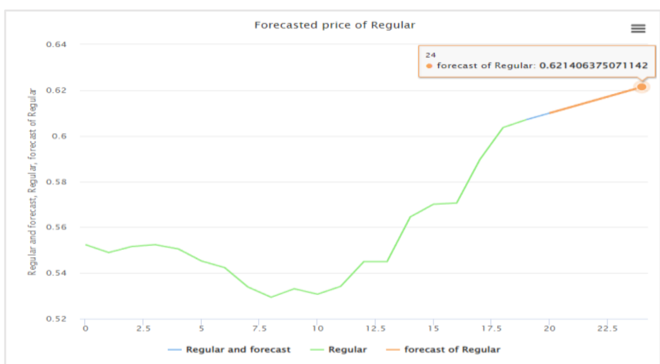

Figure 6. *Forecasting Model of Regular Gasoline*

The forecasting model for midgrade gasoline is revealed in Figure 7. The study's results indicate that the forecasted price ranges from 0.6930 USD to 0.7022 USD, suggesting a slight increase in the price of midgrade gasoline over the five weeks.

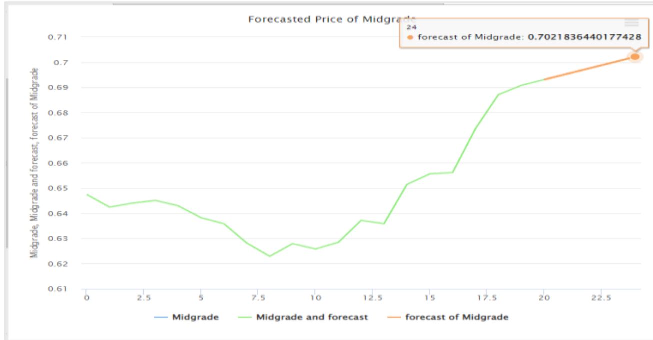

Figure 7. *Forecasting Model of Midgrade Gasoline*

In the forecasting model for premium gasoline depicted in Figure 8, findings reveal that the forecasted price starts at 0.7684 USD and continues to increase up to 0.7777 USD. Thus, from week 1 to week 5, there is a little but constant increase in the price of premium gasoline.

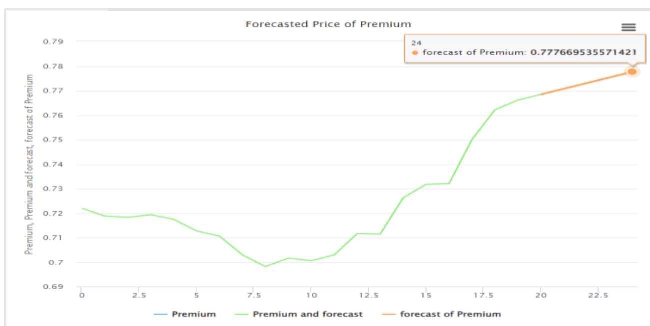

Figure 8. *Forecasting Model of Premium Gasoline*

In the forecasting model for diesel presented in Figure 9, the price trajectory unfolds. It initiates at 0.7216 USD in week 1, gradually climbing to 0.7379 USD by week 5.

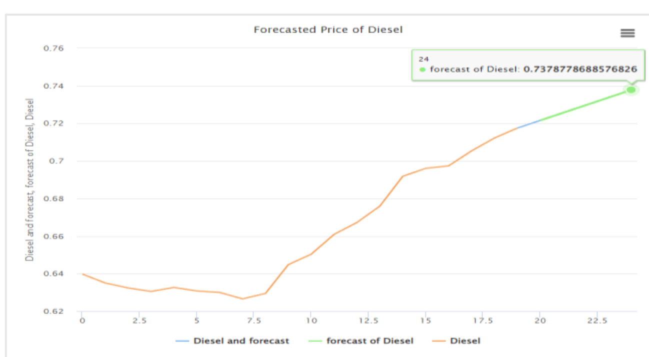

Figure 9. *Forecasting Model of Diesel*

Accuracy of the Forecasting Model

In the ARIMA model, RMSE serves as a crucial metric for evaluating the accuracy of each forecasting model. The findings indicate that all forecasted values yielded RMSE scores below 0.50. Such consistently low RMSE values strongly indicate outstanding accuracy across all forecasting models, underscoring the remarkable precision in their predictions. Consequently, these results affirm the success of the study in developing a robust forecasting model tailored for predicting gasoline retail prices.

Table 2. Performance of the ARIMA Models

ARIMA Model	AR Coefficient	MA Coefficient	RMSE
	Regular		
(1, 0, 0)	0.9999		0.22
(1, 0, 1)	1.0896	0.1768	0.74
(1, 1, 1)	0.9799	-1.5327	0.52
	Midgrade		
(1, 0, 0)	0.9999		0.22
(1, 0, 1)	1.0864	0.1791	0.61
(1, 1, 1)	0.8946	-1.7416	0.78
	Premium		
(1, 0, 0)	0.9999		0.23
(1, 0, 1)	1.0773	0.2969	0.69
(1, 1, 1)	0.9515	-1.7964	0.68
	Diesel		
(1, 0, 0)	0.9999		0.26
(1, 0, 1)	1.0632	0.3655	0.59
(1, 1, 1)	0.8736	-1.3280	0.89

The findings from Table 2 suggest that the ARIMA (1,0,0) model outperforms both the ARIMA (1,0,1) and ARIMA (1,1,1) models. This conclusion is based on the AR and MA coefficient, and RMSE values of the ARIMA models.

The ARIMA (1,0,0) model showcases a substantial AR coefficient of 0.9999, which suggests a positive association between current and past gasoline prices. Furthermore, the notably low RMSE values (0.22, 0.23, and 0.26) indicate a good fit of the model to the data.

In contrast, within the ARIMA (1,0,1) model, the AR coefficients (ranging from 1.0632 to 1.0896) remain elevated, signifying a strong positive correlation. Similarly, its MA coefficients (ranging from 0.1768 to 0.3655) are positive. However, the elevated RMSEs (ranging from 0.59 to 0.74) suggest a somewhat inferior fit compared to the previous model.

On the other hand, the AR coefficients of ARIMA (1,1,1) model (ranging from 0.8736 to 0.9799) are comparatively lower. Notably, its MA coefficient takes on a negative value (ranging from -1.527 to -1.3280). Moreover, the RMSEs (ranging from 0.52 to 0.89) are the highest among the three models, indicating a suboptimal fit relative to the others.

As a result, the ARIMA (1,0,0) model emerged as the most optimal model for forecasting gasoline prices, demonstrating superior performance across the evaluated metrics. Additionally, in a study by He (2023), the ARIMA (1,0,0) model outperformed the ARIMA (2,1,2), exhibiting a higher RMSE of 1.103 compared to the ARIMA (1,0,0) model.

Developed Application for Forecasting Gasoline Retail Prices

The dashboard of the application, shown in Figure 10, consists of different gasoline types, including diesel, and visualizes the weekly trend for a specific week. It illustrates the trend of how gasoline prices fluctuate over time.

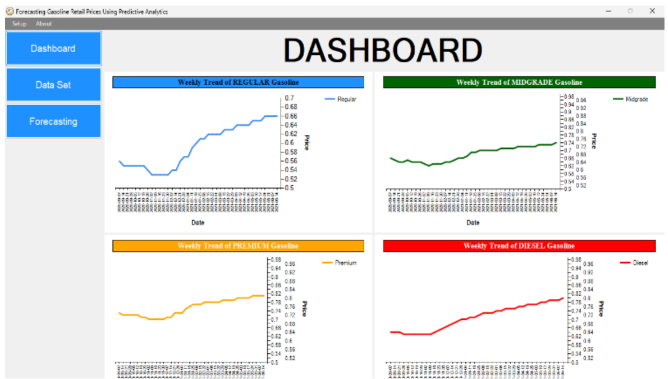

Figure 10. *Dashboard*

The integration of the forecasting model is illustrated in Figure 11. In this section, users can view the forecasted prices for three types of gasoline–regular, midgrade, and premium–as well as diesel, based on the inputted actual prices. Both actual and predicted prices are displayed here. The color coding for the lines is as follows: orange represents regular gasoline, red indicates midgrade, blue denotes premium, and gray stands for diesel.

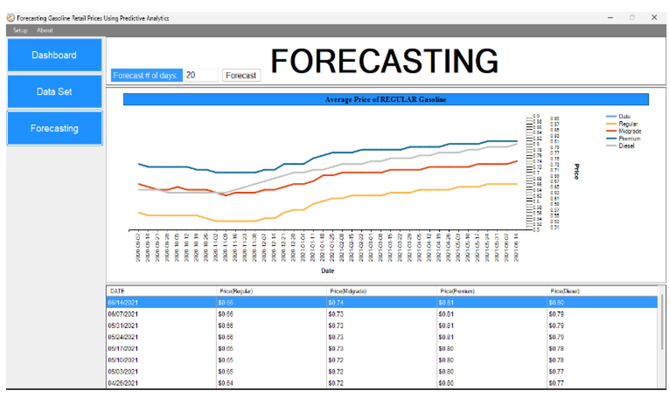

Figure 11. *Integrated Forecasting Model*

Conclusion and Future Works

Given the variation in prices across different gasoline stations, forecasting gasoline retail prices is essential for consumers to understand the factors influencing price fluctuations. In pursuit of this objective, the study sought to forecast gasoline retail prices through an analysis of data spanning 26 years.

Since gasoline prices have been rising and falling over the last few years, there is no consistent price for each type of gasoline. It can be concluded that there has been an abrupt decrease in retail prices in the recent years. In addition, the application of time series analysis employing the ARIMA model effectively produced forecasts for retail prices across various gasoline types. The study discerned minimal variance between the forecasted and actual retail prices. Notably, the forecast for gasoline prices spanned five weeks, unveiling a consistent upward trend in prices over this period.

Furthermore, the forecasting models yielded remarkably low RMSE values, indicating a high level of accuracy for each model. Consequently, these models demonstrated exceptional predictive capabilities, affirming the success of this study in constructing reliable frameworks for forecasting retail gasoline prices. The developed application seamlessly integrated the forecasting model for retail gasoline prices, offering a user-friendly dashboard interface. This dashboard provides insightful visualizations of the weekly trends across four distinct gasoline types, enhancing accessibility and facilitating informed decision-making.

With this conclusion, this research recommends then that consumers may use the aforementioned application as it provides valuable insights for informed decision-making in budget planning. Gasoline stations and owners may also use the developed application for forecasting natural gas prices, which is a powerful and essential tool that allows them to make better decisions for managing the potential risk, reducing the gap between demand and supply, and optimizing the usage of resources based on accurate predictions.

On the other hand, future researchers could increase the model's accuracy, and data equality has to be enhanced. Another future effort may investigate other data mining techniques. They may use other algorithms to further improve the model's accuracy and identify major contributing factors class-specifically from the collected data. They may use a specific location as the scope of their study and use other related variables to improve the model and the developed application. Lastly, future research endeavors could explore the potential application of the findings from this study in various fields.

References

[1] Anderson, S. T., Kellogg, R., Sallee, J. M., & Curtin, R. T. (2011). Forecasting gasoline prices using consumer surveys. *American Economic Review, 101*(3), 110–114. https://doi.org/10.1257/aer.101.3.110

[2] Bounteous. (2020, September 15). Forecasting with a time series model using Python: Part one. *Senior Data Scientist.* https://www.bounteous.com/insights/2020/09/15/forecasting-time-series-model-using-python-part-one

[3] Bora, N. (2021). Understanding ARIMA model for machine learning. Capital One Tech. https://www.capitalone.com/tech/machine-learning/understanding-arima-models/

[4] Chandra, et al. (2020). Predictive analytics in big data. *Journal of Information and Computation Science, 43*. https://doi.org/10.4018/978-1-4666-9840-6.ch002.

[5] Edwards, J. (2023). What is predictive analytics? Transforming data into future insights. *CIO.* https://www.cio.com/article/228901/what-is-predictive-analytics-transforming-data-into-future-insights.html

[6] Georgiou, A. D. (2021). An introduction to time series forecasting. *InfoWorld – New Tech Forum.* https://www.infoworld.com/article/3622246/an-introduction-to-time-series-forecasting.html

[7] He, X. J. (2023). Forecasting gasoline price with time series models. *Communications of the IIMA, 21*(1), Article 1. https://doi.org/10.58729/1941-6687.1440

[8] Iyke, B. N. (2019). Real output and oil price uncertainty in an oil-producing country. *Bulletin of Monetary Economics and Banking, 22*(2), 163–176. https://doi.org/10.21098/bemp.v22i2.1095

[9] Lam, W. D. (2014). A survey of predictive analytics in data mining with big data (Master's thesis). Athabasca University.

[10] LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K., & Sun, W. (2016, May 25). Where predictive analytics is having the biggest impact. *Harvard Business Review.* https://hbr.org/2016/05/where-predictive-analytics-is-having-the-biggest-impact

[11] Miguel, C. G. (2021, January 27). What is the difference between regular and premium fuel? *Philkotse.com.* https://philkotse.com/safe-driving/difference-between-regular-premium-gas-9717

[12] Nyangarika, A., Mikhaylov, A., & Richter, U. H. (2018). Oil price factors: Forecasting on the base of modified ARIMA model. *International Journal of Energy Economics and Policy, 9*(1), 149–159. https://doi.org/10.32479/ijeep.6812

[13] Octane Fuel. (2018, January 2). *Shamrock Tire & Auto Repair News-Center.* Retrieved from https://www.shamrocktire.com/About/News-Center/ArticleID/9918/Selecting-the-Right-Octane-Fuel

[14] O'Hara, R., Haylon, L., & Boyle, D. (2023). A data analytics mindset with CRISP-DM. *SF Magazine.* Retrieved from https://sfmagazine.com/articles/2023/february/a-data-analytics-mindset-with-crisp-dm?psso=true

[15] Pierre, S. (2021). A guide for time series forecasting in Python. *WorldQuant Predictive & Built In.* Retrieved from https://builtin.com/data-science/time-series-forecasting-python

[16] Sabu, K. M., & Kumar, T. K. M. (2020). Predictive analytics in agriculture: Forecasting prices of arecanuts in Kerala. *Procedia Computer Science, 171*, 699–708. https://doi.org/10.1016/j.procs.2020.04.076

[17] Shah, G. N. (2015, April 14). 5 reasons VB.NET is better than C#. *MigrateTo.net.* https://www.migrateto.net/5-reasons-vb-net-is-better-than-c/

[18] Simplilearn. (2021). Understanding time series analysis in Python. *Simplilearn Tutorials*. Retrieved from https://www.simplilearn.com/tutorials/python-tutorial/time-series-analysis-in-python

[19] Su, M., Zhang, Z., Zhu, Y., Zha, D., & Wen, W. (2019). Data-driven natural gas spot price prediction models using machine learning methods. *Energies, 12*(9), Article 1680 (or pp. 1-17). https://doi.org/10.3390/en12091680

[20] Sun, L., et al. (2018). Demand forecasting for petrol products in gas stations using clustering and decision tree. *Advance Computational Intelligence and Intelligent Informatics, 22*(3), 387–393. https://doi.org/10.20965/jaciii.2018.p0387

[21] Tyler, C. (2022, April 16). *What is VB.NET? Introduction & features*. GURU99. https://www.guru99.com/vb-net-introduction-features.html#

[22] U.S. Energy Information Administration. (2021, January). *U.S. gasoline and diesel retail prices 1995-2021* [Data set]. Kaggle. https://www.kaggle.com/datasets/mruanova/us-gasoline-and-diesel-retail-prices-19952021

[23] Uddin, S. M., Rahman, A., & Ansari, E. U. (2018). Comparison of some statistical forecasting techniques with GMDH predictor: A case study. *Journal of Mechanical Engineering*, 47(1), 16–21. https://doi.org/10.3329/jme.v47i1.35354

[24] U.S. Energy Information Administration (January 2021). *U.S. gasoline and diesel retail prices 1995-2021*. https://www.kaggle.com/datasets/mruanova/us-gasoline-and-diesel-retail-prices-19952021

[25] Vandeput, N. (2019). Forecast KPIs: RMSE, MAE, MAPE and bias. *Data Science for Supply Chain Forecasting*. https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

[26] Varshney, P. (2020). Measure performance for a time series model (ETS or ARIMA). *Towards Data Science.*

[27] Vijay, A. M., Devi, T., Poornima, N., & Gnanavel, R. (2021). Accuracy acquirement for petrol prices prediction using machine learning enhanced random forest algorithm. *Turkish Online Journal of Qualitative Inquiry*, 12(3), 2468-2474.

[28] Vijiyan, S. (2022). Time series analysis of average fuel prices (2000-2021). https://rpubs.com/SV93/forecasting_ts_final

[29] Zhao, Z., Fu, C., & Wang, C. (2018). Improvement to the prediction of fuel cost distributions using ARIMA model. *Department of Electrical and Computer Engineering*. [Report/Working paper]. https://www.researchgate.net/publication/322306188

Acknowledgment

The researchers would like to express their gratitude to the U.S. Energy Information Administration (January 2021) for making the gasoline and diesel retail prices dataset available for this study.