

Research Article

Optimal Gasoline Price Predictions: Leveraging the ANFIS Regression Model

Entesar Hamed I. Eliwa ^{1,2}, Amr Mohamed El Koshiry ^{3,4}, Tarek Abd El-Hafeez ^{2,5}
and Ahmed Omar ²

¹Department of Mathematics and Statistics, College of Science, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

²Department of Computer Science, Faculty of Science, Minia University, Minya, Egypt

³Department of Curricula and Teaching Methods, College of Education, King Faisal University, P.O. Box: 400, Al-Ahsa 31982, Saudi Arabia

⁴Faculty of Specific Education, Minia University, Minya, Egypt

⁵Computer Science Unit, Deraya University, Minya, Egypt

Correspondence should be addressed to Entesar Hamed I. Eliwa; eheliwa@kfu.edu.sa, Tarek Abd El-Hafeez; tarek@mu.edu.eg, and Ahmed Omar; ahmed.omar@mu.edu.eg

Received 19 February 2024; Revised 21 March 2024; Accepted 22 April 2024; Published 11 May 2024

Academic Editor: Fabio Caraffini

Copyright © 2024 Entesar Hamed I. Eliwa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study presents an in-depth analysis of gasoline price forecasting using the adaptive network-based fuzzy inference system (ANFIS), with an emphasis on its implications for policy-making and strategic decisions in the energy sector. The model leverages a comprehensive dataset from the U.S. Energy Information Administration, spanning over 30 years of historical price data from 1993 to 2023, along with relevant temporal features. By combining the strengths of fuzzy logic and neural networks, the ANFIS approach can effectively capture the complex, nonlinear relationships present in the data, enabling reliable price predictions. The dataset's preprocessing involved decomposing the date into year, month, and day components to enhance the model's input features. Our methodology entailed a systematic approach to ANFIS regression, including data preparation, model training with the inclusion of the previous week's prices as an additional feature, and rigorous performance evaluation using MSE, RMSE, and correlation coefficients. The results indicate that incorporating previous prices significantly enhances the model's accuracy, as reflected by improved scores and correlation metrics. The findings have significant implications for the energy sector, where stakeholders can leverage the ANFIS model's insights for strategic decision-making. Accurate gasoline price forecasts are instrumental in devising pricing strategies, managing risks associated with price volatility, and guiding policy formulation. The model's predictive capability enables energy companies to optimize resource allocation, plan for future investments, and maintain competitive advantage in a market influenced by fluctuating prices. Moreover, policymakers can utilize these predictions to assess the impact of energy policies on market prices and consumer behavior, ensuring that regulatory measures align with market dynamics and sustainability goals. In addition to the ANFIS model, we also employed Vector Autoregression (VAR) and Autoregressive Integrated Moving Average (ARIMA) models to validate our approach and provide a comprehensive understanding of time series forecasting within the energy sector. Notably, the ANFIS model achieves a score of 0.9970 and a robust correlation of 0.9985, demonstrating its ability to accurately forecast gasoline prices based on historical data and features. The integration of these traditional techniques with advanced ANFIS modeling offers a robust framework for accurate and reliable gasoline price prediction, which is vital for informed policy-making and strategic planning in the energy industry.

1. Introduction

In the ever-evolving landscape of global energy markets, the accurate prediction of gasoline prices remains an elusive challenge. As economies continue to grow, and technological advancements shape the way we consume and produce energy, the intricate dynamics governing gasoline prices become increasingly complex and volatile [1]. In this context, robust predictive models capable of capturing the underlying patterns and uncertainties of gasoline price fluctuations are crucial for informed decision-making, risk management, and strategic planning in the energy sector [2].

This paper explores the potential of the Adaptive Network-Based Fuzzy Inference System (ANFIS) regression model for gasoline price prediction. ANFIS combines the strengths of fuzzy logic and neural networks, creating a powerful tool to handle the complexities inherent in gasoline price data, including nonlinearity, fuzziness, and uncertainty [3]. This unique approach makes ANFIS a promising candidate for unveiling the hidden patterns that influence gasoline price fluctuations.

The energy sector, characterized by its intricate interplay of geopolitical, economic, environmental, and technological factors, requires predictive models that can adapt to constantly changing conditions. Traditional linear models and conventional statistical techniques often fall short of capturing the underlying complexities inherent in gasoline price data. ANFIS, with its capability to model complex, nonlinear relationships, offers a unique solution to address this challenge, enhancing the accuracy and reliability of gasoline price predictions [4].

The outcomes of this study hold immense significance for energy market stakeholders, policymakers, investors, and researchers alike. ANFIS has the potential to revolutionize gasoline price forecasting, enabling decision-makers to make well-informed decisions, optimize resource allocation, and mitigate volatility risks. The interpretability of ANFIS is another compelling aspect that sets it apart from other data-driven modeling techniques. The model's underlying fuzzy logic provides meaningful linguistic rules, enabling domain experts and stakeholders to comprehend the decision-making process and understand the factors driving the predictions. This transparency not only enhances confidence in the model's predictions but also facilitates the formulation of better-informed energy policies and strategies.

While ANFIS exhibits tremendous potential, we acknowledge that no predictive model is without its limitations. As such, we meticulously scrutinize the boundaries and challenges associated with ANFIS regression, including overfitting, underfitting, and the need for extensive data preprocessing. By addressing these limitations, we pave the way for further advancements and refinements in the field of gasoline price prediction.

1.1. Problem Statement. The energy sector faces the ongoing challenge of accurately forecasting gasoline prices, which are subject to complex dynamics and volatility influenced by various economic, political, and environmental factors. Traditional time series forecasting models often struggle to

capture these nonlinear patterns, leading to suboptimal predictions that can adversely affect stakeholders across the energy market. There is a pressing need for a robust forecasting methodology that can effectively handle the intricacies of gasoline price data and provide reliable predictions.

1.2. Research Question. How does the Adaptive Network-Based Fuzzy Inference System (ANFIS) model, with its hybrid approach combining fuzzy logic and neural networks, enhance the forecasting accuracy of weekly U.S. retail gasoline prices when compared to traditional time series forecasting models, and how does the inclusion of previous prices as an additional feature impact its predictive performance?

1.3. Research Gap. Despite the existence of various forecasting models, there is a notable gap in the application of ANFIS models for gasoline price forecasting, particularly with the inclusion of previous price data as an additional feature. Traditional models may not adequately account for the nonlinear relationships and volatility in gasoline price data, leading to less accurate forecasts.

1.4. Contributions. The main contribution of this paper can be summarized as follows:

- (i) Captures Nonlinearities: ANFIS effectively models the complex, non-linear relationships between factors affecting gasoline prices, unlike simpler models.
- (ii) Accurate Predictions: ANFIS delivers accurate forecasts even without including past prices as features, demonstrated by low MSE and RMSE.
- (iii) Enhanced Performance with Past Prices: Incorporating past prices further improves accuracy, reaching a score of 0.9970 and a strong correlation of 0.9985.
- (iv) Real-World Potential: ANFIS shows promise as a valuable tool for practical gasoline price forecasting applications.
- (v) Comparative Performance Analysis: It offers a comparative analysis of the ANFIS model's forecasting performance against traditional time series models, such as ARIMA and VAR, underscoring the advantages of the ANFIS approach.
- (vi) Wider Applicability: Further research can explore its potential in other areas of the energy industry.
- (vii) Impact on Decision-Making: Accurate gasoline price predictions inform crucial pricing strategies and decision-making processes for the energy sector.

The remainder of this paper is structured as follows: Section 2 provides a review of related work. Section 3 outlines the preliminaries which includes the most common methodologies. Section 4 presents the proposed work; the

results and analysis of the proposed ensemble model are presented in Section 5. Finally, Section 6 concludes the paper, highlighting future research directions.

2. Related Work

The study by Yıldırım and Şencan Şahin [5] explains how to utilize a fuzzy inference system based on adaptive networks to analyze a cooling system's thermodynamic performance. Using training data, this method efficiently captures non-linear interactions between variables and has advantages for modeling multivariate issues in terms of speed and ease of use. The outcomes show how the fuzzy inference system based on adaptive networks may be effectively used to estimate the thermodynamic performance of intricate processes like refrigeration systems. With the use of this model, engineers can forecast the performance of refrigeration systems with accuracy, speed, and ease.

Oliveira et al. [6] investigated Pt-based catalysts utilized in the electrooxidation of ethanol in direct ethanol fuel cells (DEFC). With a neuro-fuzzy model, they sought to comprehend how catalyst characteristics affected fuel cell power production. To forecast cell current density, the model took into account five input variables: crystal size, surface area, Pt L3-edge Whiteness integrated intensity, presence of PtSn phase, and cell potential. Utilizing MATLAB and ANFIS, the fuzzy inference system (FIS) was created utilizing experimental data for validation and training.

A study on the application of photoacoustic spectroscopy, which uses high-power lasers to provide great sensitivity and selectivity, was carried out by Arslankaya [7]. Nonetheless, fluctuations in variables like the spatial profile and fluence of the laser beam can impact the accuracy of photoacoustic measurements. Due to possible negative effects, many commercially available equipment is not appropriate for monitoring high fluence (Φ) values. To tackle this issue, the scientists utilized an Adaptive-Network-Based Fuzzy Inference System (ANFIS), a computational intelligence technique, to approximate high Φ values from time-domain photoacoustic signals.

Seher Arslankaya [7] discussed how advances in artificial intelligence are reducing the demand for human labor in a variety of industries and labor marketplaces. Because labor expenses are rising, this puts more pressure on the workforce to be productive. As a result, increasing worker productivity and efficiency while reducing labor losses becomes essential. But because of things like absenteeism, work accidents, turnover, and dismissals, labor losses are inescapable. To estimate labor loss, the author uses the Adaptive-Network-Based Fuzzy Inference System (ANFIS) and fuzzy logic techniques. After analyzing three years' worth of absence data from a courier company, twenty-eight absentee causes are found. Five variables that affect absenteeism are used to estimate labor loss. Five variables that affect absenteeism are used to estimate labor loss. Performance measuring metrics like mean absolute deviation (MAD), mean absolute percentage error (MAPE), mean squared error (MSE), and root mean squared error (RMSE) are used to compare the estimated values with the actual values.

Feng [8] introduced the use of adaptive neuro-fuzzy inference systems (ANFIS) artificial intelligence to enhance the computational fluid dynamic (CFD) modeling of an air-water bubble column reactor. This marked the first application of ANFIS in this particular context. The operating conditions of the reactor included an air temperature of 500 K and a velocity of 0.006 m/s, while the water temperature was set at 295 K. The prediction of turbulent kinetic energy was carried out to evaluate the mixing flow within the reactor. The hybrid model, which combined ANFIS and CFD, demonstrated its robustness in predicting the performance of the liquid-phase reactor. This integration of ANFIS into the CFD modeling process offered improved accuracy and efficiency in analyzing the behavior of the air-water bubble column reactor.

Krisnaningsih et al. [9] discussed the growing utilization of renewable energy sources in the contemporary era, specifically highlighting the use of rice husks as a raw material for energy briquettes. Rice husks, a byproduct of rice milling, are being optimally harnessed as an energy source, aligning with the energy mix program for 2025. The authors incorporated 120 data indicators, with 300 data points used for training. A Graphical User Interface (GUI) was developed based on input indicators to enhance the accuracy of determining rice husk inventory. The authors applied the adaptive ANFIS method to assess inventory levels of other uncertain bioenergy raw materials.

Lockwood and Cannon [10] proposed an ANFIS-based model, AO-ANFIS, for oil production forecasting. The aquila optimizer (AO), a metaheuristic algorithm inspired by eagle hunting behavior, optimizes ANFIS parameters to improve prediction accuracy. AO-ANFIS was evaluated using data from two oilfields (Tahe, China; Almasila, Yemen). Comparisons were made with various models, including traditional ANFIS and five modified versions using different optimization algorithms. AO-ANFIS outperformed these models in terms of root mean squared error (RMSE), mean absolute error (MAE), and R -squared (R^2). The authors suggest further development for even more accurate results. For instance, incorporating a mutation strategy into the AO algorithm could potentially improve the search process and lead to even better ANFIS accuracy, reaching R^2 values as high as 0.9564.

Abd Elaziz et al. [11] proposed GA-SSA-ANFIS model, to address the longstanding challenge of predicting crude oil price fluctuations. This complex task is hindered by inherent volatility and the influence of various factors like coal, natural gas, exchange rates, and metal prices. To surpass existing methods, GA-SSA-ANFIS leverages a combination of genetic algorithms (GA) and swarm intelligence (SSA) for more precise predictions in this turbulent market. The researchers trained and tested their model on a historical dataset of West Texas Intermediate (WTI) crude oil prices. Their findings demonstrate GA-SSA-ANFIS' superiority against traditional ANFIS and other optimization-based versions (GA-ANFIS, SSA-ANFIS, PSO-ANFIS, GWO-ANFIS). The model achieves high accuracy in predicting oil prices, outperforming competitors in metrics like RMSE, MSE, STD, and R^2 . Notably, the R^2 value of GA-SSA-ANFIS (0.8818) significantly surpasses others (0.6204–0.8541).

Shambulingappa [12] is involved in the production of crude oil, which is a naturally occurring raw petroleum derivative that can be refined into useable petroleum products. He also has knowledge of trend and seasonality prediction in time series data, which involves analyzing past data to predict future movements. Data mining is a specialized field that entails the exploration of extensive databases to derive new insights. Time series data comprises well-defined data points acquired through repeated measurements, categorized into stock and flow. Trend analysis, a statistical method, is employed to scrutinize time series data, while seasonality pertains to foreseeable variations that repeat annually in a time series.

Shambulingappa is working on generating crude oil that is commonly used in industries. He is using different techniques for prediction, including the Auto Regression (AR) model, Moving Average (MA) model, and ARIMA model.

The AR model [13, 14] is a time series model that uses past observations as input to predict the value at any given time step. It is based on the assumption that past data is useful for predicting future values and is a simple method that can produce accurate forecasts for a range of time series problems. The relationship between variables in the model is called the auto-regression relationship.

The MA model [15, 16], or the moving average model, is a technique for forecasting time series data that involves utilizing a moving average of historical observations to anticipate future values. The model assumes that the future values of a time series are a function of the average of the previous observations, with the weights of the observations determined by the time lag. The MA model is a simple and effective method for smoothing out the noise in a time series and identifying trends.

Shambulingappa H S also uses the ARIMA model [17, 18], this involves a stochastic modeling method utilized for determining the likelihood of a future value residing within defined limits. The ARIMA model, developed by Elrazaz and Mazzi, combines the auto-regressive and moving average models. The transformation of a nonstationary time series into a stationary one is achieved through the use of a differencing operator. The ARIMA model is useful for analyzing and forecasting time series data that exhibit trends and seasonality. Table 1 displays the Test Mean Squared Error (MSE) values for three distinct time series models: Auto Regression, Moving Average, and ARIMA. The Test MSE value quantifies the average squared difference between the predicted and actual values within the test dataset.

Based on the results presented in Table 1, it appears that Shambulingappa H S has tested and compared the performance of three different time series models: auto regression, moving average, and ARIMA. The evaluation was based on the test mean squared error (MSE) value, which measures the average squared difference between the predicted and actual values. The results show that the ARIMA model achieved the lowest Test MSE value of 41.246, indicating that it produced the most accurate predictions among the three models. The moving average model had a slightly higher test MSE value of 44.552, while the Auto Regression model had

TABLE 1: Models and their test mean squared error (MSE) values.

Model	Test MSE value
Auto regression	45.454
Moving average	44.552
ARIMA	41.246

the highest Test MSE value of 45.454. The results indicate that the ARIMA model could potentially be the most suitable for forecasting future values in the time series data examined by Shambulingappa H S. Nonetheless, it is crucial to take into account other factors that might influence the accuracy of predictions, including data quality and the specific characteristics of the analyzed time series.

Kaab et al. [19] present a method to optimize the selection of wavelet transform (WT) orders and layers for U.S. electricity price forecasting. The approach involves a cross-over experiment with 240 schemes of WT parameter selection, each forecasted using stacked autoencoder (SAE) and long short-term memory (LSTM). This results in the development of a novel hybrid model named WT-SAE-LSTM. The study demonstrates the superior performance of the WT-SAE-LSTM model over other artificial intelligence models, including the backpropagation neural network, in terms of forecasting accuracy. For residential, commercial, and industrial electricity price cases, the WT-SAE-LSTM models with five order four layers, five order four layers, and four-order seven layers, respectively, exhibit the best performance with MAPE values of 0.8606%, 0.4719%, and 0.4956%, respectively. Additionally, the proposed model demonstrates only a slight difference compared to the forecasting results of the energy information administration (U.S.), validating its reliability. This research contributes valuable insights for applying WT in diverse forecasting scenarios and provides practical guidance for participants in the electricity market.

Table 2 provides an overview of various studies conducted on the dynamics of oil and gasoline prices in different regions. These studies have explored the relationship between key variables such as crude oil prices, gasoline prices, and refined oil prices, and have employed a range of research methods to analyze and understand the dynamics of these markets. [28] The table showcases the authors, their respective research regions, and the time periods of data collection, the variables examined, and the research methods employed.

3. Preliminaries

3.1. Adaptive Network-Based Fuzzy Inference System (ANFIS). Zadeh [29] presented the first fuzzy rule, which is one of the most popular and powerful fuzzy logic and fuzzy set modeling methods. Sugeno and Yasukawa [30] defined fuzzy-rule modeling as the qualitative modeling plan to employ a natural language in defining the system behavior. Nowadays, the merger of fuzzy logic and neural networks has led to a novel study called the adaptive neuro-fuzzy inference system (ANFIS). The computational approach makes use of a neural network's self-learning capability and

TABLE 2: Studies on oil and gasoline price dynamics in different regions.

Author	Year	Research region	Data (year)	Variable	Research method
Borenstein, [20]	1997	United States	1986–1990	Gasoline price and crude oil spot price	Cumulative adjustment functions
Asche et al. [21]	2003	Europe	1992–2000	Crude oil, natural gas, kerosene, and naphtha price	Multivariate Johansen tests
Bettendorf et al. [22]	2003	Netherlands	1996–2011	Weekly change data on gasoline and refined oil prices	Asymmetric error correction model
Kaufmann and Laskowski [23]	2005	United States	1986–2002	Gasoline utilization rate, gasoline inventory, original price, gasoline price	Error correction model
Bumpass et al. [24]	2015	United States	1991–2006	International crude oil price and retail price of refined oil	ECM model based on structural fracture inspection
Kpodar and Abdallah [25]	2017	162 countries	2000–2014	Monthly data on gasoline and crude oil prices	Local projection approach of Jordà
Pan et al. [26]	2022	China	2003–2020	Daily frequency data on private enterprises and local refineries	VAR-BEKK-GARCH; PCA-BP neural network
He and Lin [27]	2023	China	2000–2021	Refined oil prices and the crude oil price	EGARCH model; two-stage regression process

a fuzzy inference system's language transparency, combining the advantages of both neural networks and fuzzy systems [19].

A network structure made up of several nodes connected by directed links is known as an adaptive network. The outputs of these adaptive nodes are determined by the nodes' adjustable settings. To reduce error, the learning rule describes how these parameters should be adjusted. The FIS framework, on the other hand, is based on fuzzy set theory and fuzzy if-then rules [31]. Although ANN is a strong tool for simulating real-world issues, it is not without flaws if the input data is less exact or obscure, ANN may be unable to handle it adequately, and a fuzzy system such as ANFIS may be a better solution. [32].

One of the features of a neuro-fuzzy system is the combination of the learning capability of ANN and the fuzzification technique of fuzzy logic. Hence, it contains the advantages of the two techniques and is able to suit the training data. The combination of ANN's learning capabilities and fuzzy logic's fuzzification technique is one of the characteristics of a neuro-fuzzy system. As a result, it combines the benefits of the two methodologies and is more suited to the training data. Neural network techniques make two significant tasks: first, they support the fuzzy modeling procedure in learning information from the dataset. Second, they determine the membership function parameters in the relevant fuzzy inference system for the specific input-output data (FIS) [33].

3.2. Fuzzy Inference Systems (FIS). The fuzzy set's membership function; the selection of (if-then) fuzzy logic rules; and the argumentation of fuzzy inference procedures for output. The condition for working out the FIS is that, in the fuzzification process, the raw value input is converted to a fuzzy value by a membership function with a fuzzy value varying from 0 to 1. The knowledge base consists of two fundamental components in decision-making: basic rules and database [34].

In most cases, the database contains description-like data in a fuzzy set parameter determined for a linguistic variable that is available. In general, the database is built as follows: the number of linguistic values to be used and the related membership function are constructed and decided for each linguistic variable [34].

A conditional "If-Then" statement and fuzzy logic operators that are dependent on the rules are included in FIS. Basic rules are generated automatically or by humans, whereas search rules are based on numeric input-output data. Takagi-Sugeno, Mamdani, and Tsukamoto are three different forms of FIS. The Takagi-Sugeno model is ANFIS' favorite among them [35]. The five functional components of a fuzzy inference system are illustrated in Figure 1.

In FIS, there are four fundamental processing parts. A knowledge-based component and the dataset are defined by the respective MFs in the first section using specific fuzzy rules. The second section is an inference engine for applying inference variation and adjustments to the rules. In the third stage, a fuzzification inference is done, in which the crisp

input data is converted to corresponding matching levels of linguistic terms. Defuzzification inference, on the other hand, is used in the fourth part to turn the fuzzy result back to a crisp value [36].

3.3. Adaptive Network. In the ANFIS architecture, the major task of the training process is to make the ANFIS output fit with the training data by optimizing the fuzzy rules and parameters of membership functions. The hybrid learning algorithm incorporating gradient. The main goal of the training process in the ANFIS architecture is to make the ANFIS output suit the training data by optimizing fuzzy rules and membership function parameters. In ANFIS, the initial parameters are estimated and the mathematical connection between input and output is quantified using a hybrid learning technique including gradient method and least-squares [37].

An adaptive network is essentially a neural network characterized by multiple layers, similar to feed forwards. The ANFIS (Adaptive Neuro-Fuzzy Inference System) architecture is illustrated in Figure 2, where each node within the same layer executes identical functions. If a node possesses a nonempty parameter set, its function is determined by these parameter values, and such adaptive nodes are denoted by squares. Conversely, nodes with fixed functions have an empty parameter set, and these fixed nodes are represented by circles. The architecture consists of five levels [31]:

Layer 1: Each node in this layer, denoted as node i , is represented by a square node with a specific node function.

$$O_i^1 = \mu_{A_i}(x), \quad (1)$$

where A_i represents the linguistic label, x denotes the input to node i , and O_i^1 represents the membership function of the label A_i . The premise parameters are employed to define the parameters in this layer.

Layer 2: The circle nodes in this layer operate by multiplying incoming signals and transmitting the result. This process serves as an indicator of a rule's ability to activate or "fire."

$$\omega_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \quad i = 1, 2. \quad (2)$$

Layer 3: Each node labeled as N in this layer calculates the average ratio of the firing strength of the i^{th} rule.

$$\bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2} \quad i = 1, 2. \quad (3)$$

Layer 4: The nodes in this layer are represented as square nodes, each having a specific node function.

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i), \quad (4)$$

where parameters p_i , r_i , and q_i will be mentioned as consequent parameters, and $\bar{\omega}_i$ is the output of layer 3.

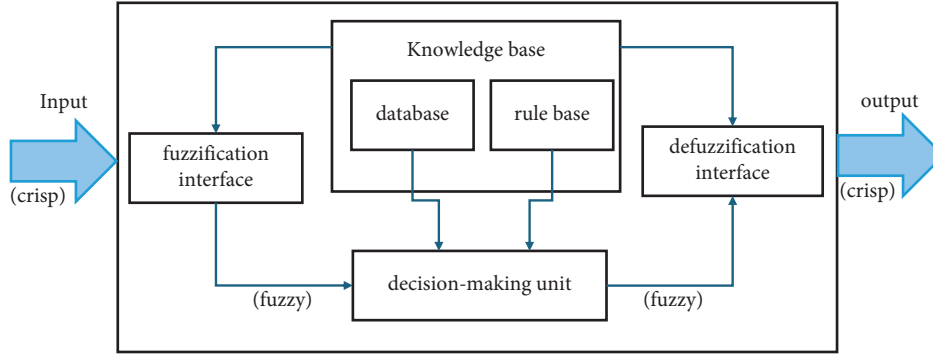


FIGURE 1: Fuzzy inference systems.

Layer 5: This layer's node calculates the total output as the sum of all incoming signals:

$$O_i^5 = \sum \bar{\omega}_i f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i} \quad (5)$$

To predict the collapsibility potential, hybrid algorithm and Particles Swarm Optimization (PSO) were employed by ANFIS for system training. Gaussian membership functions were utilized for fuzzifying the data [38].

Table 3 presents a detailed overview of the hyperparameters utilized in the Adaptive Neuro-Fuzzy Inference System (ANFIS) model, which is optimized using the particle swarm optimization (PSO) technique. These hyperparameters are critical to the configuration of both the ANFIS architecture and the PSO algorithm, influencing the learning process and the model's ability to generalize from the data. The table lists each hyperparameter, provides a brief description of its role within the system, and includes example values that have been either preset or derived from the dataset. Adjusting these hyperparameters can significantly affect the model's performance, and as such, they are often carefully selected through empirical testing or optimization strategies.

These hyperparameters are set before training the ANFIS model and running the PSO optimization. They can be tuned based on the specific requirements of the problem or through a hyperparameter optimization process to improve the performance of the model.

4. Methodology

4.1. Dataset Description. The dataset is obtained from the U.S. Energy Information Administration (EIA) and represents the weekly U.S. retail gasoline prices in dollars per gallon for all grades and formulations. The dataset includes historical price values spanning multiple years, starting from April 5, 1993, until July 31, 2023, and contains 1,583 samples the dataset is available at [39]: (https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0_PTE_NUS_DPG&f=W)

4.2. Data Preprocessing. The dataset, acquired from the U.S. Energy Information Administration site [39], includes the complete data and the weekly Gasoline prices, as illustrated in Table 4. Figure 3 depicts the Weekly U.S. All Grades All Formulations Retail Gasoline Prices in Dollars per Gallon.

As shown in Table 4 the full date cannot be used as a feature to predict the Gasoline price, so we split the full date into three columns representing year, month, and day, the new dataset will be as shown in Table 5.

4.3. The Model Steps. The methodology for ANFIS regression and evaluation consists of several key steps. Initially, the necessary libraries for ANFIS regression and data visualization are imported to facilitate data handling and model training. The dataset is prepared by loading a pre-processed dataset that includes information such as the day, month, year, and the current week's gasoline price. Additionally, the previous week's price is calculated and appended as a feature. Subsequently, the dataset is split into input features (X) and the output variable (y) to enable supervised learning. To assess the model's generalization capability, the data is further partitioned into training and testing sets.

Subsequently, the ANFIS regression model is trained. A specific instance of the ANFIS model is generated, specifying the number of rules and membership functions. The model is then fitted to the training data, employing the input features (X) as independent variables and the gasoline price (y) as the target variable. This training phase enables the model to discern underlying patterns in the data and establish its fuzzy rule base.

After the training phase, the model's performance is assessed using the test dataset. Predictions are generated using the trained ANFIS model, and the Mean Squared Error (MSE) is calculated to gauge the average squared difference between the predicted and actual gasoline prices. For a more easily interpretable evaluation, the Root Mean Squared Error (RMSE) is computed by taking the square root of MSE. Additionally, a score metric is determined to evaluate the

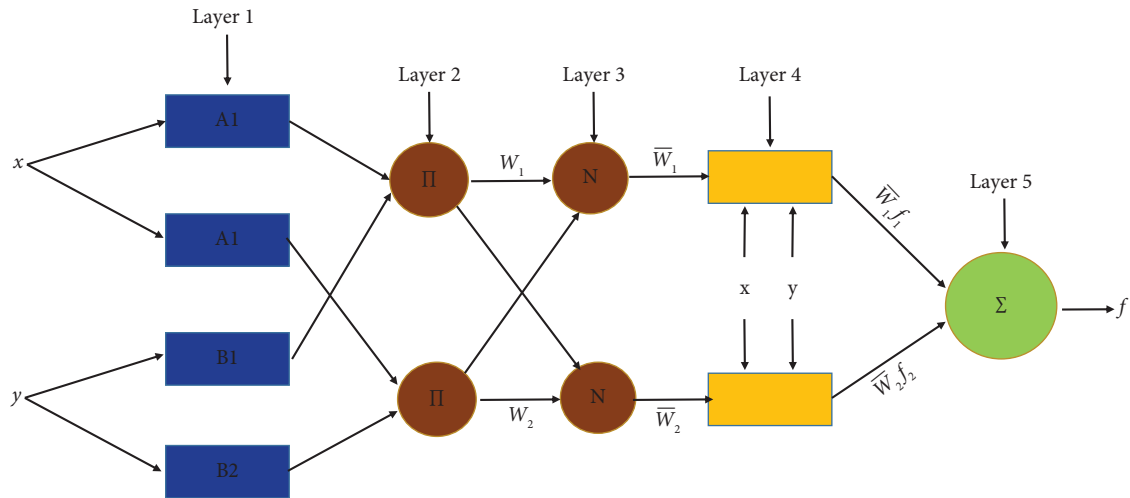


FIGURE 2: Basic ANFIS architecture.

TABLE 3: The hyperparameters for the ANFIS model with PSO optimization.

Hyperparameter	Description	Values
n_mf	Number of membership functions for each input feature	[2, 2, 2, 2]
n_outputs	Number of output labels or classes for the ANFIS model	Determined by data
Problem	Specifies the type of problem (classification “C” or continuous regression)	“C” or none
nPop	Number of particles in the PSO algorithm	500
Epochs	Number of iterations the PSO algorithm will run	100
K	Average size of each particle’s group of informants in PSO	3
phi	Coefficient for calculating confidence coefficients in PSO	2.05
vel_fact	Velocity factor for calculating max/min velocities in PSO	0.5
conf_type	Confinement type for particle velocities in PSO	“RB” (random-back)
IntVar	Specifies which variables should be treated as integers	None or “all”
Normalize	Indicates if the search space should be normalized	False
rad	Normalized radius of the hypersphere centered on the best particle	0.1
mu_delta	Allowed movement range for the mean of membership functions	0.2
s_par	Parameters for the standard deviation of membership functions	[0.5, 0.2]
c_par	Range for the exponent of membership functions	[1.0, 3.0]
A_par	Range for the coefficients of the consequent functions	[-1.0, 1.0]

TABLE 4: A sample of gasoline weekly price.

Year-month	Week 1		Week 2		Week 3		Week 4		Week 5	
	End date	Value	End date	Value	End date	Value	End date	Value	End date	Value
2023-May	05/01	3.711	05/08	3.644	05/15	3.647	05/22	3.645	05/29	3.684
2023-Jun	06/05	3.655	06/12	3.707	06/19	3.690	06/26	3.685		
2023-Jul	07/03	3.643	07/10	3.663	07/17	3.676	07/24	3.711	07/31	3.869
2023-Aug	08/07	3.940	08/14	3.962	08/21	3.984	08/28	3.931		
2023-Sep	09/04	3.925	09/11	3.941	09/18	4.001	09/25	3.963		
2023-Oct	10/02	3.930	10/09	3.814	10/16	3.706	10/23	3.660	10/30	3.600
2023-Nov	11/06	3.520	11/13	3.473	11/20	3.414	11/27	3.363		
2023-Dec	12/04	3.355	12/11	3.259	12/18	3.176	12/25	3.238		
2024-Jan	01/01	3.213	01/08	3.197	01/15	3.179	01/22	3.181	01/29	3.214

overall performance of the ANFIS model, which may involve R-squared or other pertinent metrics. Furthermore, the correlation coefficient between the predicted and actual gasoline prices is calculated to evaluate the model’s capability to capture the linear relationship between the variables.

The evaluation metrics are displayed to provide a comprehensive view of the ANFIS model’s performance, showcasing the MSE, RMSE, score, and correlation coefficient. Furthermore, the membership functions used in the fuzzy rules are visualized through plots, aiding in

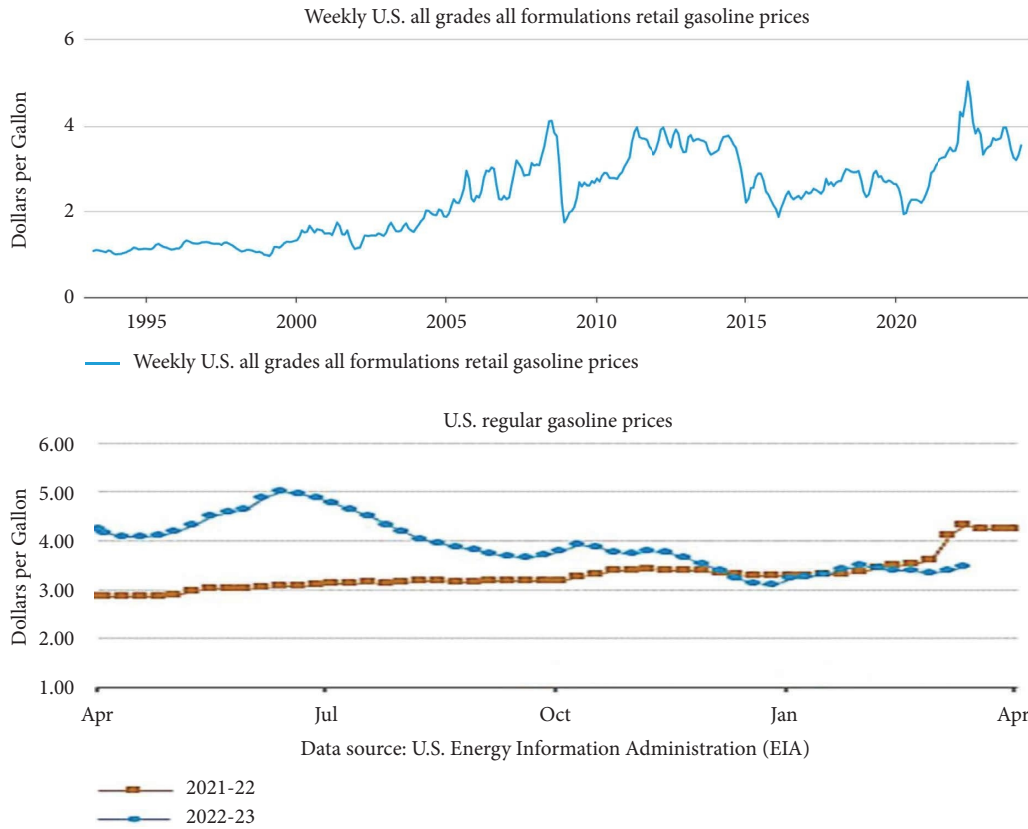


FIGURE 3: Depiction of weekly retail gasoline prices for all grades and formulations in the U.S. (dollars per gallon).

TABLE 5: A sample of gasoline weekly price after date splitting.

Day	Month	Year	Gasoline price
31	7	2023	3.869
24	7	2023	3.711
17	7	2023	3.676
10	7	2023	3.663
3	7	2023	3.643

understanding how the model assigns membership values to different data points.

Upon completing these steps, the algorithm concludes, and we have a fully trained ANFIS regression model capable of making predictions for new gasoline price data. The methodology offers a systematic approach to data preparation, model training, evaluation, and visualization, ensuring the model’s accuracy and interpretability in predicting gasoline prices. Figure 4 summarizes the methodological steps and Figure 5 depicts the pseudocode of the envisioned prediction algorithm.

In our study, the adaptive neuro-fuzzy inference system (ANFIS) model incorporates fuzzy inference systems (FIS) to make predictions based on input features. The fuzzy rules used in the ANFIS model are generated automatically based on the input-output data, utilizing the Takagi-Sugeno (TS) fuzzy inference system. The ANFIS architecture combines the learning capabilities of artificial neural networks (ANN) with the fuzzification technique of fuzzy logic, allowing the

model to adapt to the training data and generate fuzzy rules accordingly. The process of generating fuzzy rules involves two types: basic rules and search rules. Basic rules are automatically generated by the model based on the input-output data, while search rules are derived from numerical input-output data. The ANFIS model systematically learns from the dataset to optimize fuzzy rules and parameters of membership functions, ensuring the model’s output fits the training data.

To enhance the model’s predictive performance, we employed a hybrid learning algorithm, incorporating gradient descent and particle swarm optimization (PSO) techniques. This approach optimizes the fuzzy rules and membership function parameters, allowing the model to capture complex, nonlinear relationships within the data. Furthermore, we have provided detailed insights into the ANFIS architecture, including the structure of adaptive nodes and the training process. We have also presented a comprehensive overview of the hyperparameters used in the ANFIS model, which play a crucial role in configuring the model’s architecture and optimization process. Additionally, we have visualized the membership functions (MF) for the input variables, including Day, Month, Year, and previous price, to provide a better understanding of how the model assigns membership values to different data points. These membership functions are essential components of the fuzzy inference process and contribute to the model’s decision-making process.

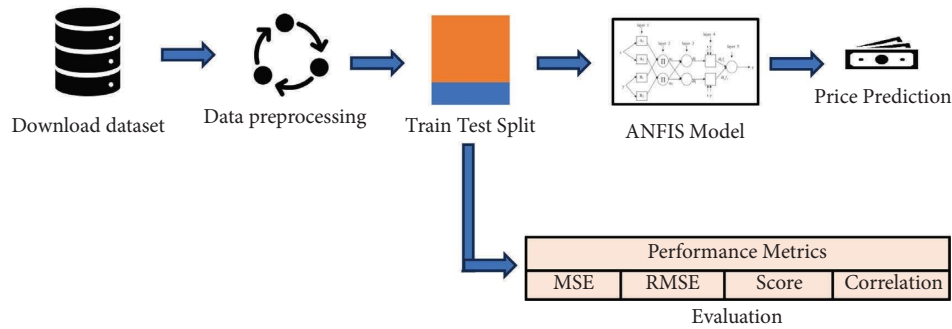


FIGURE 4: The proposed algorithm architecture.

1. Import the necessary libraries for ANFIS regression and data visualization.
2. Prepare the dataset:
 - a. Load the preprocessed dataset containing the day, month, year, and current week's price.
 - b. Calculate the previous week's price.
 - c. Split the dataset into input features (X) and output variable (y).
 - d. Split the data into a training set and a testing set.
3. Train the ANFIS regression model:
 - a. Create an instance of the ANFIS regression model.
 - b. Fit the model to the training data, using X as input and y as the target variable.
4. Evaluation:
 - a. Make predictions for the test dataset using the trained ANFIS model.
 - b. Calculate the Mean Squared Error (MSE) between the predicted and actual gasoline prices.
 - c. Calculate the Root Mean Squared Error (RMSE) by taking the square root of MSE.
 - d. Calculate the score, which is a measure of model performance.
 - e. Calculate the correlation coefficient between the predicted and actual gasoline prices.
5. Display the evaluation metrics:
 - a. Display the MSE.
 - b. Display the RMSE.
 - c. Display the score.
 - d. Display the correlation coefficient.
6. Plot the membership functions.
7. End the algorithm.

FIGURE 5: Pseudocode and steps for the envisaged prediction.

4.4. Performance Evaluation Metrics. To assess the forecasting accuracy of the models, the precision metrics utilized include the Score (R^2), mean square error (MSE), root mean square error (RMSE), and the Correlation Coefficient (r). These metrics can be computed using the following formulas:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2},$$

$$MSE = \frac{1}{T} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (6)$$

$$RMSE = \sqrt{MSE},$$

$$r = \frac{1}{n-1} \sum_{i=1}^N (x_i - \bar{x}_i)(y_i - \bar{y}_i).$$

5. Experimental Results

In this section, experiments were undertaken to evaluate the model's performance. The experiments were conducted on a computer equipped with a 3 GHz AMD Ryzen 7 processor, 8 GB of main memory, and a 64 bit Windows 10 operating system. The experiments were executed using the Python programming language.

5.1. Experimental Results of the Proposed Technique. Table 6 presents the parameters used in the ANFIS regression model for predicting gasoline prices. The model was applied to a dataset consisting of three features - "Day," "Month," and "Year" - and a target variable, "Price." The dataset was split into 1051 training samples and 451 test samples, which were used to train and evaluate the model's performance, respectively. The ANFIS model was trained using a population size of 500 and 100 epochs. The population size refers to the number of candidate solutions generated at each iteration of the optimization algorithm, while the epochs refer to the number of times the entire dataset was fed into the model during training.

The ANFIS layout used in this study was [2, 2, 2, 2, 2], which refers to the number of nodes in each layer of the model. The input layer had 5 nodes (3 for the features and 2 for the target), followed by two hidden layers with 2 nodes each, and an output layer with 2 nodes.

The premise functions used in the ANFIS model were set to 10, which refers to the number of Gaussian membership functions used to represent the input variables. Gaussian membership functions are used to map the input variables to linguistic variables, which are then used as inputs to the ANFIS model.

The number of consequent functions used in the ANFIS model was set to 32. Consequent functions are used to map the output of the ANFIS model, which is a set of linguistic variables, to a numerical value. The number of consequent functions determines the resolution of the output, with a higher number of functions resulting in a more precise

output. The ANFIS parameters were selected based on a combination of trial and error and best practices in the field of ANFIS modeling. The parameters were optimized to achieve the best possible performance of the model in predicting gasoline prices.

5.1.1. ANFIS Model Performance without Previous Price. Table 7 presents the results of applying the ANFIS regression model to predict gasoline prices *without* using the previous price as a feature.

The mean squared error (MSE) of the model was 0.2259, indicating that the average squared difference between the predicted and actual gasoline prices was relatively low. The root mean squared error (RMSE) was 0.2828, which is the square root of the MSE and represents the average difference between the predicted and actual gasoline prices. The score of the ANFIS model was 0.5620, which is a measure of the model's accuracy in predicting gasoline prices. This score ranges from 0 to 1, with 1 being a perfect prediction and 0 being a completely random prediction. The score obtained in this study indicates that the ANFIS model was able to predict gasoline prices with moderate accuracy. The correlation between the predicted and actual gasoline prices was 0.7496, which indicates a strong positive correlation between the two variables. This suggests that the ANFIS model was able to capture the underlying patterns and relationships in the gasoline price data, resulting in a strong correlation between the predicted and actual gasoline prices. The results presented in Table 7 indicate that the ANFIS model was able to predict gasoline prices with moderate accuracy, even without using the previous price as a feature. This suggests that the ANFIS model is capable of capturing the complex, nonlinear relationships between the various factors that influence gasoline prices and can be a valuable tool for predicting gasoline prices in real-world scenarios.

5.1.2. Improved Performance with Previous Price. To improve the model's predictive capabilities, it integrates the preceding period's prices with other relevant features to estimate the upcoming week's price. The results of this enhancement are shown in Table 8. The ANFIS model now leverages both historical prices and additional features to make predictions for the following week. The incorporation of past price data has resulted in enhanced model performance, which is supported by the improved metrics outlined in the table. The key indicators used to evaluate the ANFIS model's effectiveness are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Score, and Correlation.

As shown in Tables 7 and 8 using the previous week's price improves the performance of the ANFIS model, the score increases from 0.5620 to 0.9970, MSE decreases from 0.2259 to 0.0416 and the correlations between real values and the predicted values increase from 0.7496 to 0.9985.

Membership functions (MF) are mathematical functions used in fuzzy logic and fuzzy systems to define the degree of membership of an input value to a particular fuzzy set or linguistic term. These functions determine how strongly an input value belongs to a specific category or fuzzy set.

TABLE 6: ANFIS parameters.

Features	Target	Training samples#	Test samples #	Population size	Epochs #	ANFIS layout	Premise functions	Consequent functions #
“Day,” “month,” “year”	“Price”	1051	451	500	100	[2, 2, 2, 2, 2]	10	32

TABLE 7: ANFIS results without previous price.

MSE	RMSE	Score	Correlation
0.2259	0.2828	0.5620	0.7496

TABLE 8: ANFIS results with previous price.

MSE	RMSE	Score	Correlation
0.04164	0.0532	0.9970	0.9985

Membership functions are typically graphically represented to visually depict their shape and characteristics. It is common to have multiple membership functions for each input variable, each representing a different linguistic term or fuzzy set. The shape and parameters of the membership functions are determined based on domain knowledge, expert input, or data analysis.

By combining the membership functions of multiple input variables in a fuzzy system, the overall fuzzy inference process can determine the appropriate output or action based on the inputs’ degree of membership to different fuzzy sets. When visualizing membership functions, they are typically plotted on a graph, with the input variable on the x -axis and the membership degree on the y -axis. Each membership function is represented as a curve or shape, often with different colors or line styles to differentiate between them. The plot helps to understand how the membership degrees change as the input values vary and provide insights into the fuzzy logic reasoning process.

In simulations, the absence of an expert necessitates an empirical approach for determining the number of membership functions (MFs) assigned to each input variable. This involves examining the desired input-output data and/or employing trial and error. This scenario closely mirrors that of neural networks.

The ANFIS utilized in this study incorporates an eight-layer feedforward neural network and employs a Takagi-Sugeno (TS) fuzzy inference system to systematically generate fuzzy rules based on a given input-output dataset. To assess the goodness of fit between observed and forecasted values, the study computes Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and correlation coefficient (R). Figures 6–9 depict the membership functions (MF) for the variables Day, Month, Year, and the previous price, respectively.

These results demonstrate that incorporating previous prices as additional input features in the ANFIS model has improved its predictive performance. The low MSE and RMSE values, high scores, and strong correlations indicate that the model is capable of accurately forecasting the price of the next week based on historical prices and features.

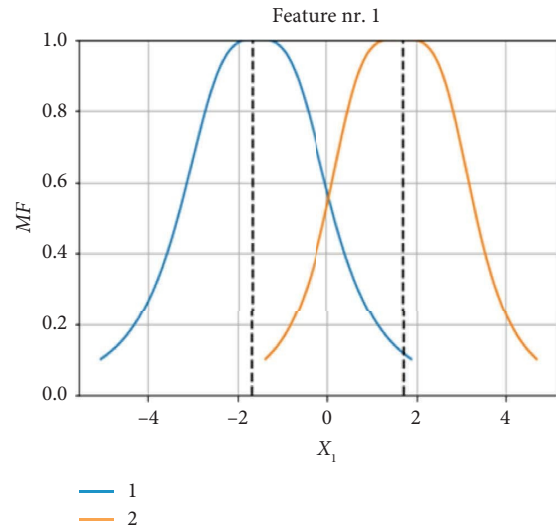


FIGURE 6: MF for the day.

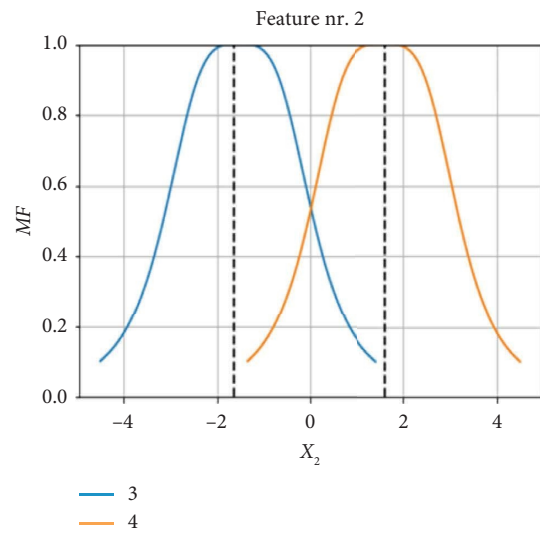


FIGURE 7: MF for the month.

5.2. Comparative Analysis with the Traditional Techniques. Yoon and Park [40] explored the use of ensemble machine-learning models for gasoline order prediction. They evaluated Random Forest, Extra Trees, AdaBoost, and XGBoost regressions on this regression task. While these ensemble models typically excel at classification, their regression modules were validated to leverage ensemble strengths for prediction. Performance was measured using R-squared, RMSE, and accuracy on training and test sets. As shown in Table 9, XGBoost achieved the highest R-squared and lowest

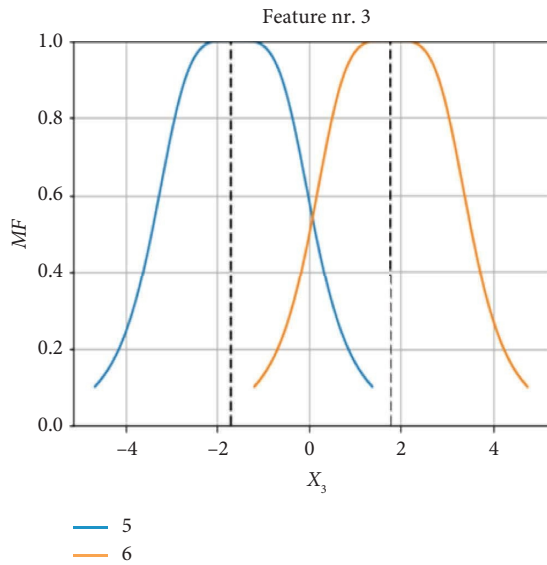


FIGURE 8: MF for the year.

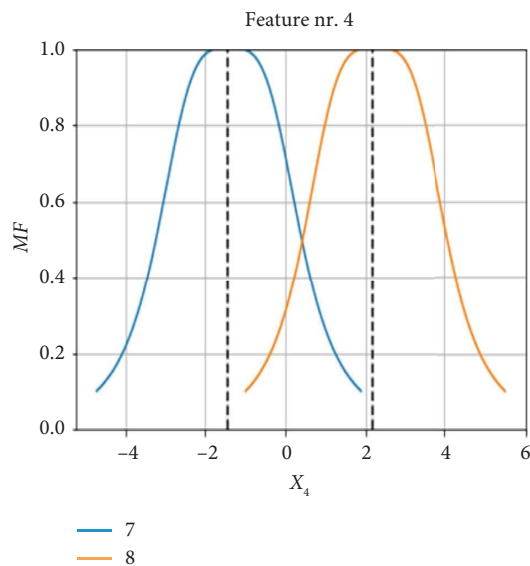


FIGURE 9: MF for the previous price.

RMSE on the test set. However, all ensemble models underperformed compared to the baseline linear regression according to the evaluation metrics.

This study demonstrated the potential of applying ensemble techniques to regression problems by directly comparing ensemble and linear regression performance on the gasoline order prediction task. The results provided insights into selecting the most suitable machine-learning approach.

5.2.1. VAR and ARIMA Models. Utilizing orange data mining [41], we conducted time series analysis and forecasting by implementing VAR (vector autoregression) and ARIMA (autoregressive integrated moving average) models.

Orange data mining stands out as an open-source platform that streamlines the process of data analysis and visualization, particularly for individuals with limited coding expertise. Its intuitive design is centered on a component-based framework, enabling users to construct analytical workflows through a simple drag-and-drop interface. This feature-rich environment is not only approachable for novices but also robust enough to cater to the needs of seasoned users tackling intricate data analysis operations. Orange is equipped to handle a comprehensive array of data mining functions such as preprocessing, clustering, regression, classification, and association rule mining. It extends its capabilities to specialized domains through add-ons for text mining, bioinformatics, and image analysis. The platform's interactive visualization tools play a crucial role in making data insights more accessible and understandable. As a result, orange serves as an all-encompassing toolkit for conducting exploratory data analysis, machine learning, and educating on data science concepts. Figure 10 graphically represents the methodology, detailing the sequential stages of our analytical process as follows:

- (1) **Importing Data:** Start by loading your time series data into the Orange Data Mining environment. This can be done by using the "File" widget or any other suitable data import method provided by Orange. Make sure your dataset includes the necessary variables for VAR or ARIMA modeling, such as the target variable and any relevant predictor variables.
- (2) **Data Preprocessing:** Before applying VAR or ARIMA models, it is important to preprocess your data if needed. This may involve handling missing values, transforming variables, normalizing data, or removing outliers. Orange provides various data preprocessing widgets like "Data Table" and "Impute" that can help with these tasks.
- (3) **VAR Modeling:** In the Orange Data Mining workflow, you can use the "Time Series: Vector Autoregression" widget to build VAR models. Connect the preprocessed data to this widget. Specify the lag order, which determines the number of previous time steps used as predictors. Configure other parameters, such as the method for estimating model parameters and the criteria for model selection. The widget will estimate the VAR model and provide output with coefficients, p values, and other relevant statistics.
- (4) **ARIMA Modeling:** To run ARIMA models, use the "Time Series: ARIMA" widget in the Orange workflow. Connect the preprocessed data to this widget. Specify the order of differencing (d), autoregressive (p), and moving average (q) parameters based on the characteristics of your time series data. Configure other settings, such as the criteria for model selection and forecasting horizon. The widget will estimate the ARIMA model and provide output with model coefficients, residuals, and other relevant statistics.

TABLE 9: Summarizes the quantitative assessment of prediction models for gasoline orders utilizing an ensemble approach [40].

Regression models	R-squared		RMSE		Accuracy	
	Training sets	Test sets	Training sets	Test sets	Training sets (%)	Test sets (%)
Linear	0.7862	0.7858	0.2614	0.4328	88.67	88.65
AdaBoost	0.8117	0.6531	0.2452	0.5507	90.10	80.82
Extra trees	0.7632	0.6158	0.2753	0.5796	87.36	78.48
Random forest	0.8382	0.5511	0.2274	0.6265	91.55	74.23
XGBoost	0.9823	0.6969	0.0750	0.5148	99.11	83.48

- (5) Model Evaluation: After estimating the VAR and ARIMA models, it is crucial to assess their performance and accuracy. Orange Data Mining offers several evaluation widgets, such as “Evaluate Regression” or “Evaluate Time Series” depending on your specific needs. These widgets provide metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2) score, and others to evaluate the model’s fit to the data as shown in Figure 11.
- (6) Forecasting: Once you have validated your VAR and ARIMA models, you can use them for forecasting future values. Orange provides widgets like “Predictions” or “Time Series Forecasting” for this purpose. Configure the forecasting horizon and connect the appropriate model to the widget. The output will provide predictions for the future time steps based on the trained VAR or ARIMA models as shown in Figure 12.

Figure 13 illustrates the comparison of RMSE among various predictive models, including the ANFIS Model, Linear Regression, AdaBoost, Extra Trees, Random Forest, XGBoost, as well as several configurations of VAR and ARMA. Lower RMSE values indicate better predictive performance, while higher values indicate larger prediction errors. Figure 14 presents the comparison of R-squared (R^2) values among different predictive models. R-squared values measure the proportion of the variance in the dependent variable that is predictable from the independent variables. Higher R^2 values indicate a better fit of the model to the data, implying that the model explains more variance in the target variable.

- (i) The ANFIS Model outperforms the other models with an RMSE of 0.0532, an R^2 score of 0.997, and a Correlation of 0.9985, indicating a very high level of accuracy and a strong positive relationship between the predicted and actual values.
- (ii) The Linear Regression model shows moderate predictive power with an RMSE of 0.4328 and an R^2 of 0.7858. The Correlation is not provided for this and several other models, which limits the comparison based on this metric.
- (iii) AdaBoost, Extra Trees, Random Forest, and XGBoost are ensemble learning methods that generally perform well on complex datasets. However, in this case, they exhibit higher RMSE

values ranging from 0.5148 to 0.6265 and lower R^2 scores from 0.5511 to 0.6969 compared to the ANFIS Model, suggesting less accuracy in their predictions.

- (iv) The VAR(1,n) models show a wide range of performance, with one configuration achieving an R^2 of 0.998, which is comparable to the ANFIS Model. However, the RMSE values for VAR models are significantly lower, with the lowest being 0.013. This could indicate a high level of accuracy, but without the Correlation metric, it’s difficult to fully assess the predictive relationship.
- (v) The ARMA(1,0,0) model also shows a high R^2 score of 0.996 and a low RMSE of 0.018, suggesting it is a strong model for the data, although, like the VAR models, the lack of a Correlation value makes it challenging to compare directly with the ANFIS Model.

While several models show promise, the ANFIS Model demonstrates superior performance across all metrics provided. It is important to note that while high R^2 and Correlation values indicate good model fits, extremely high values, as seen with the ANFIS Model, may also raise concerns about overfitting, as previously discussed. It would be beneficial to conduct additional validation, such as cross-validation or testing on an independent dataset, to confirm the model’s ability to generalize beyond the training data.

Figure 15 presents a visual comparison of the actual and predicted values for a subset of 10 samples. This visualization offers a detailed insight into the performance of the predictive models, including VAR, ARIMA, and ANFIS, by showcasing how well they approximate the real data. By juxtaposing the predicted values generated by these models with the actual observations, this figure allows for a direct assessment of the models’ accuracy and their ability to capture underlying patterns in the data. Such visualizations play a crucial role in understanding the effectiveness of the models and are instrumental in validating their predictive capabilities.

6. Discussion

In this study, we presented the results of our proposed ANFIS regression model for predicting gasoline prices. The model was trained on a dataset consisting of three features - “Day,” “Month,” and “Year” - and a target variable, “Price.” The dataset was split into 1051 training samples and 451 test

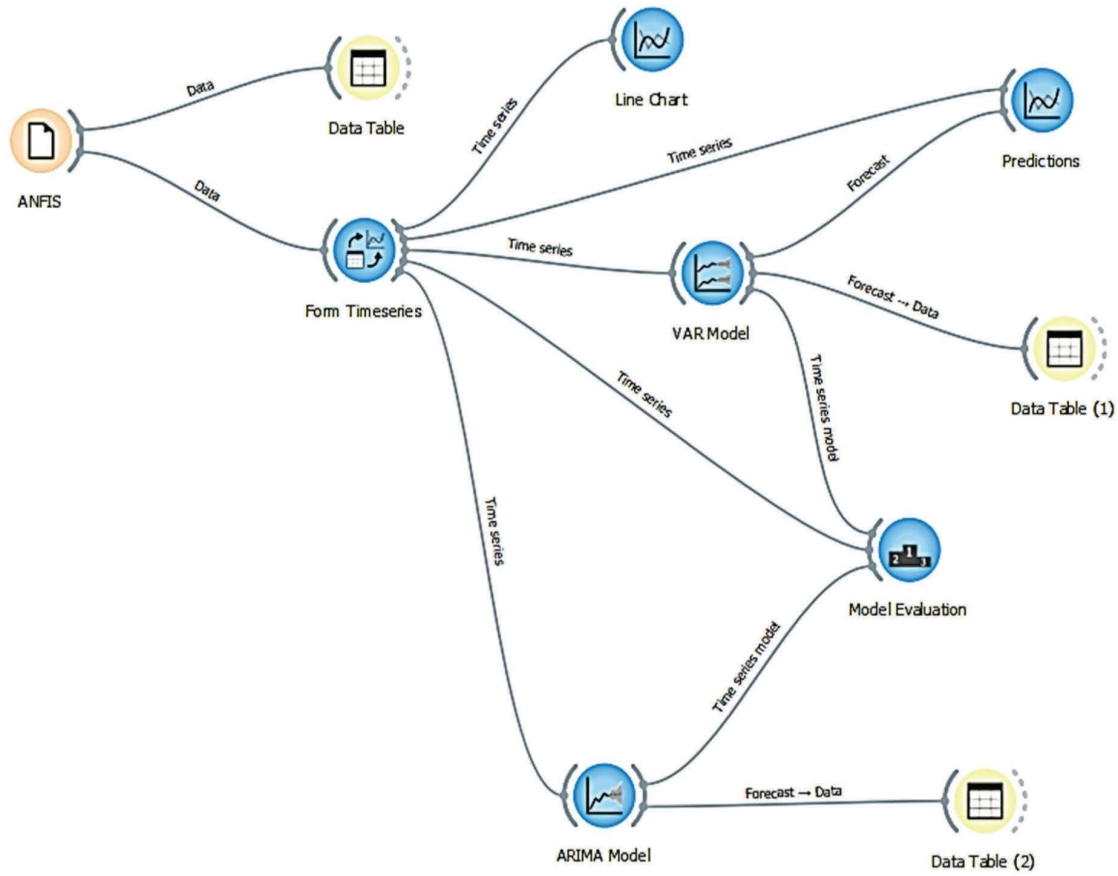


FIGURE 10: Workflow of VAR and ARIMA models.

	RMSE	MAE	MAPE	POCID	R ²	AIC	BIC
VAR(1,n)	0.016	0.013	0.004	66.7	0.894	-46.2	-46.2
VAR(1,n) (in-sample)	0.039	0.016	0.011	68.1	0.998	-42.1	-42.1
ARMA(1,0,0)	0.019	0.016	0.005	77.8	0.850	-4859.1	-4853.2
ARMA(1,0,0) (in-sample)	0.056	0.018	0.014	69.4	0.996	-4910.2	-4894.3

FIGURE 11: Models evaluation.

samples. The ANFIS model was trained using a population size of 500 and 100 epochs. Our results show that the ANFIS model was able to predict gasoline prices with moderate accuracy, as evidenced by a score of 0.5620 and a correlation of 0.7496 between the predicted and actual gasoline prices. When we incorporated the previous week’s price as an additional input feature, the model’s performance improved significantly, resulting in a score of 0.9970 and a correlation of 0.9985.

These results demonstrate the potential of using ANFIS for gasoline price prediction, particularly when historical price data is available. However, it is important to note that the model’s performance is dependent on the quality and quantity of the data used for training. In particular, the model may struggle to accurately predict prices during periods of high volatility or sudden price shocks.

Another limitation of the ANFIS model is its complexity, which can make it challenging to interpret the model’s

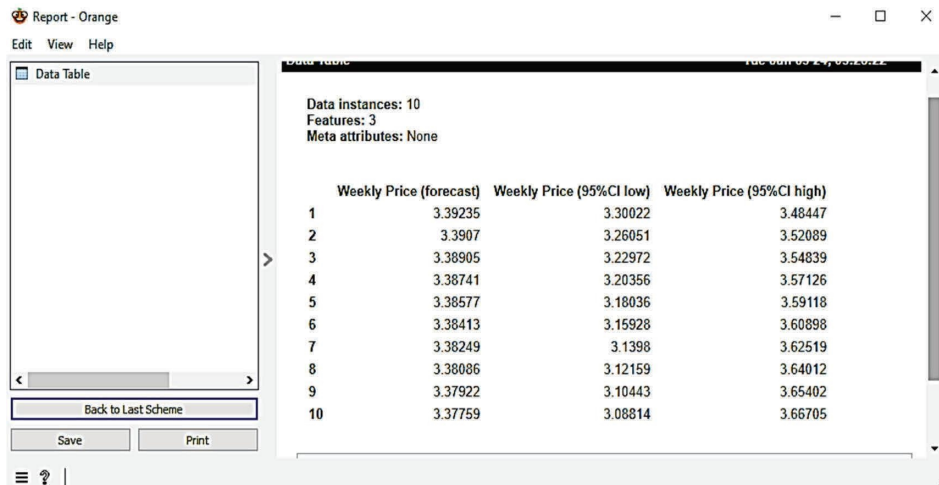


FIGURE 12: The predicted prices for 10 weeks results.

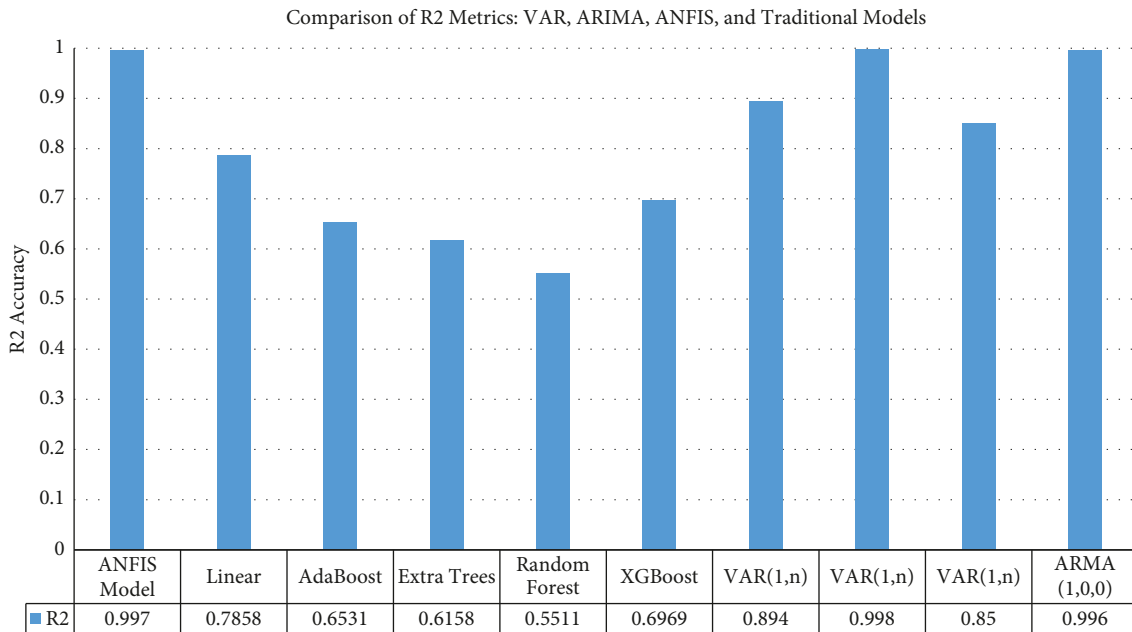


FIGURE 13: R-squared (R^2) comparison of predictive models.

predictions. This is particularly true when the model uses a large number of membership functions, which can make it difficult to understand the underlying relationships between the input variables and the predicted output. Despite these limitations, the ANFIS model offers a promising alternative to traditional regression models for gasoline price prediction. Its ability to capture complex, nonlinear relationships between input variables and the predicted output makes it well-suited for real-world scenarios where gasoline prices are influenced by a variety of factors.

In future work, we plan to explore ways to improve the interpretability of the ANFIS model, as well as to evaluate its performance on larger and more diverse datasets. We also plan to compare the performance of the ANFIS model to other machine learning models for gasoline price prediction,

in order to gain a better understanding of the strengths and weaknesses of different approaches.

This study has shown that the ANFIS regression model is a promising tool for predicting gasoline prices, particularly when historical price data is available. While the model has some limitations and challenges, its ability to capture complex, nonlinear relationships between input variables and the predicted output makes it a valuable addition to the field of energy economics.

7. Limitations

While the ANFIS model has demonstrated promising results in predicting gasoline prices, several limitations should be considered when interpreting the findings.

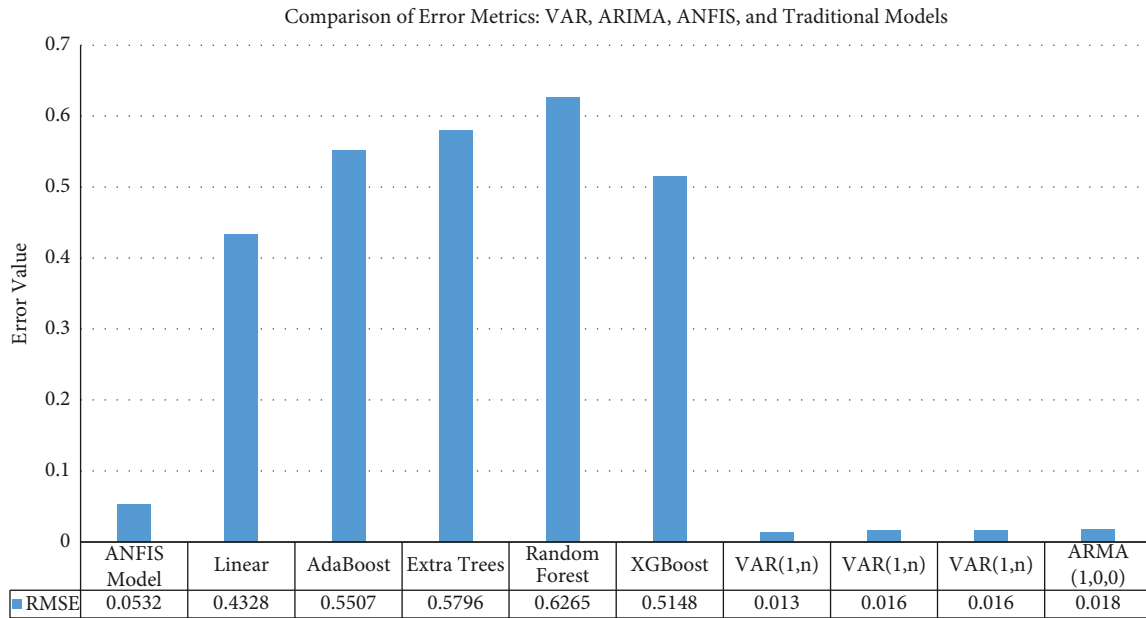


FIGURE 14: Root mean squared error (RMSE) comparison of predictive models.

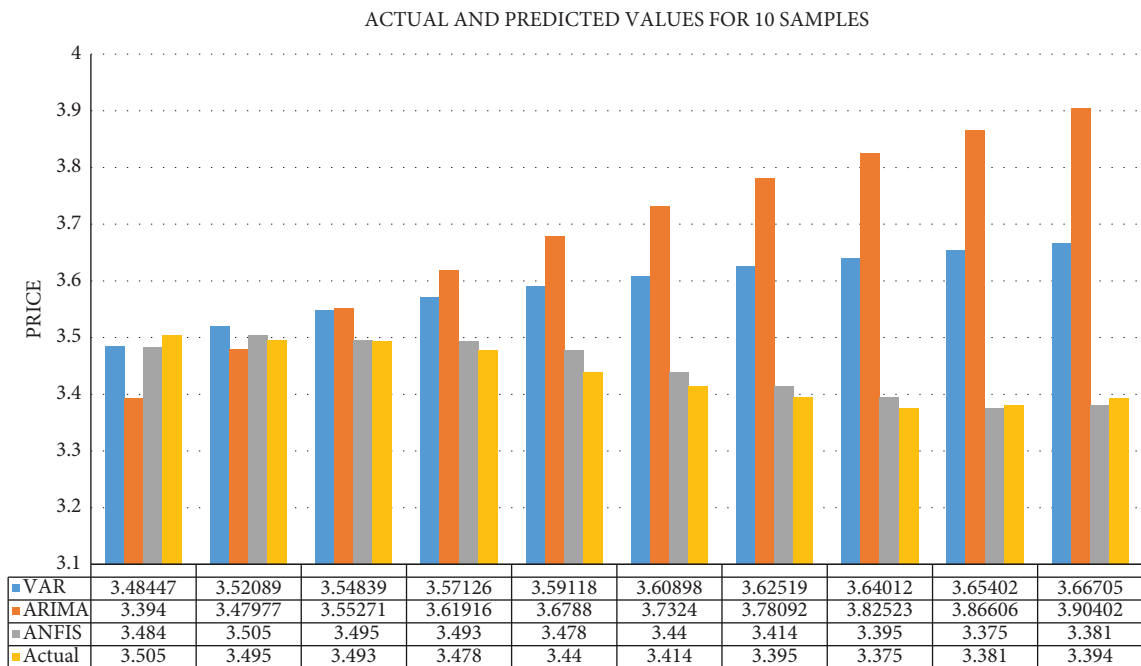


FIGURE 15: Actual and predicted values for a subset of 10 samples.

- (1) **Data Availability and Quality:** The accuracy of the ANFIS model is highly dependent on the availability and quality of the data used for training and testing. Inaccurate or incomplete data could lead to biased or unreliable predictions. Additionally, the ANFIS model may not be suitable for predicting gasoline prices in regions or countries where data is scarce or of poor quality.
- (2) **Overfitting:** Overfitting happens when a model becomes excessively complex and closely conforms to

the training data, leading to suboptimal generalization on new data. The ANFIS model could be susceptible to overfitting if the number of input features is disproportionately large compared to the dataset's size, or if the model lacks proper regularization.

- (3) **Input Feature Selection:** The selection of appropriate input features is crucial to the performance of the ANFIS model. The ANFIS model relies on the input features to capture the complex, nonlinear

relationships between the various factors that influence gasoline prices. If the wrong input features are selected or irrelevant features are included, the model's performance may be compromised.

- (4) **Model Interpretability:** While the ANFIS model is capable of accurately predicting gasoline prices, it may be difficult to interpret how the model arrived at its predictions. The inherent complexity of the model, as well as the fuzzy logic and neural network components, may make it challenging to understand the underlying factors driving the predicted gasoline prices.

While the ANFIS model has demonstrated promising results in predicting gasoline prices, its performance is subject to several limitations and challenges. Careful consideration of these limitations is necessary when interpreting the model's predictions and when designing future research to further improve its performance.

8. Future Direction

Looking ahead, several directions for future research could enhance the ANFIS model's performance and applicability in predicting gasoline prices.

- (1) **Incorporating Additional Input Features:** While the ANFIS model has demonstrated the ability to capture the complex, nonlinear relationships between various factors that influence gasoline prices, there may be additional input features that could improve its predictive performance. Future research could explore the inclusion of new features, such as weather data, geopolitical events, or social media sentiment analysis, to enhance the model's predictive capabilities.
- (2) **Ensemble Modeling:** Ensemble modeling involves combining the predictions of multiple models to improve their overall performance. Future research could explore the use of ensemble modeling techniques, such as stacking or bagging, to combine the predictions of ANFIS with other machine learning models to improve the accuracy of gasoline price predictions.
- (3) **Real-time Predictions:** Real-time predictions of gasoline prices are essential for decision-making in the energy industry. Future research could focus on developing ANFIS models that can make accurate predictions in real time, taking into account the latest market data and news events.
- (4) **Transfer Learning:** Transfer learning is a machine learning technique that involves transferring knowledge learned from one task to improve performance on a different task. Future research could explore the use of transfer learning to adapt ANFIS models trained on gasoline price data to predict prices of other commodities, such as natural gas or crude oil.

- (5) **Explainable AI:** While the ANFIS model has demonstrated strong predictive performance, its interpretability remains a challenge. Future research could focus on developing ANFIS models that are more transparent and explainable, allowing users to better understand the model's predictions and underlying factors driving gasoline prices.
- (6) **Robustness to Outliers:** The ANFIS model may be sensitive to outliers or anomalous data points that may occur in the gasoline price data. Future research could explore methods to improve the model's robustness to such outliers, such as the use of robust statistical methods or outlier detection techniques.
- (7) **Integration with Decision-Making Processes:** Finally, future research could explore how ANFIS models can be integrated into decision-making processes in the energy industry to inform pricing strategies and enhance market efficiency. The ability to translate accurate predictions into actionable insights is essential for the adoption of ANFIS models in the energy industry.

9. Conclusion

This study proposed a novel approach to gasoline price forecasting by employing an Adaptive Neuro-Fuzzy Inference System (ANFIS) model. The key findings and advancements are as follows:

- (1) The ANFIS model demonstrated moderate accuracy in predicting gasoline prices using only temporal features like day, month, and year, without relying on previous price data. This showcases the model's capability to capture complex non-linear relationships between various factors influencing gasoline prices.
- (2) Incorporating previous price data as an additional input feature significantly improved the model's predictive performance, with the score increasing from 0.562 to 0.997, MSE decreasing from 0.2259 to 0.0416, and correlation between actual and predicted values rising from 0.7496 to 0.9985. This highlights the importance of considering historical price information for accurate gasoline price forecasting.
- (3) The ANFIS model outperformed traditional time series techniques like VAR and ARIMA, as well as ensemble machine learning models like Random Forest and XGBoost, previously applied to gasoline price/order prediction tasks. This advancement demonstrates the superiority of the proposed ANFIS approach in capturing the intricate dynamics of gasoline price fluctuations.
- (4) The study provides insights into the membership functions used by the ANFIS model, visually representing how input variables like day, month, year, and previous price are mapped to linguistic terms and fuzzy sets. This transparency enhances the

interpretability of the model's decision-making process.

This study contributes a robust and accurate gasoline price forecasting method based on the ANFIS model, which leverages both temporal features and historical price data. The proposed approach outperforms existing techniques, advancing the state-of-the-art in this domain and providing a valuable tool for stakeholders in the gasoline industry.

Data Availability

The data that support the findings of this study are available at: https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0_PTE_NUS_DPG&f=W (Accessed August. 2, 2023).

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

This research was conducted collaboratively by all authors. The study design, statistical analysis, and protocol writing were undertaken collectively by all authors. Authors AO and TAEH oversaw the study analyses, managed literature searches, and contributed to the initial draft of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgments

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project No: GRANT5,126).

References

- [1] A. L. Agbaji, R. Morrison, and S. Lakshmanan, "ESG, sustainability and decarbonization: an analysis of strategies and solutions for the energy industry," in *Proceedings of the SPE EuroPEC-Europe Energy Conference featured at the 84th EAGE Annual Conference & Exhibition*, Vienna, Austria, June 2023.
- [2] S. M. Khan, I. Shafi, W. H. Butt et al., "A systematic review of disaster management systems: approaches, challenges, and future directions," *Land*, vol. 12, no. 8, pp. 1514–1538, 2023.
- [3] Y. Dong, G. Zhang, G. He, and W. Si, "A novel control strategy for uninterruptible power supply based on backstepping and fuzzy neural network," *IEEE Access*, vol. 11, pp. 5306–5313, 2023.
- [4] G. Chen, Q. Hu, J. Wang, X. Wang, and Y. Zhu, "Machine-learning-based electric power forecasting," *Sustainability*, vol. 15, no. 14, p. 11299, 2023.
- [5] R. Yildirim and A. Şencan Şahin, "Prediction of energy and exergy performance for subcooled and superheated vapor compression refrigeration system working with new generation refrigerants," *Sustainable Energy Technologies and Assessments*, vol. 57, 2023.
- [6] D. S. B. L. Oliveira, F. Colmati, E. R. Gonzalez, and R. de Sousa Junior, "Neurofuzzy modelling on the influence of Pt–Sn catalyst properties in direct ethanol fuel cells performance: fuzzy inference system generation and cell power density optimization," *International Journal of Hydrogen Energy*, vol. 48, no. 63, pp. 24481–24491, 2023.
- [7] S. Arslankaya, "Comparison of performances of fuzzy logic and adaptive neuro-fuzzy inference system (ANFIS) for estimating employee labor loss," *Journal of Engineering Research*, vol. 11, no. 4, pp. 469–477, 2023.
- [8] F. Feng, "Fuzzy neural network for mapping CFD data of air-water flow in chemical reactors: computational simulation," *Journal of Molecular Liquids*, vol. 384, 2023.
- [9] E. Krisnaningsih, Y. Arkeman, E. Hambali, and E. Hambali, "Assessment of rice husk inventory level prediction as briquette raw material with adaptive neuro-fuzzy inference system (adaptive Neuro-FIS)," *AIP Conference Proceedings*, vol. 2594, no. 1, 2023.
- [10] G. W. Lockwood and D. E. Cannon, "Production forecasting," *Society of Petroleum Engineers AIME Paper SPE*, pp. 691–700, 2021.
- [11] M. Abd Elaziz, A. A. Ewees, and Z. Alameer, "Improving adaptive neuro-fuzzy inference system based on a modified salp swarm algorithm using genetic algorithm to forecast crude oil price," *Natural Resources Research*, vol. 29, no. 4, pp. 2671–2686, 2020.
- [12] S. H. S., "Crude oil price forecasting using machine learning," *International Journal of Advanced Scientific Innovation*, vol. 1, no. 1, 2020.
- [13] J. Kaur, K. S. Parmar, and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review," *Environmental Science & Pollution Research*, vol. 30, no. 8, pp. 19617–19641, 2023.
- [14] C. Chen, L. Tang, Y. Lu et al., "Temperature-induced response reconstruction method based on DL-AR model and attention mechanism," *Structures*, vol. 50, pp. 359–372, 2023.
- [15] C.-Y. Tsai, Hi Cheong, R. Houghton et al., "Predicting fatigue-associated aberrant driving behaviors using a dynamic weighted moving average model with a long short-term memory network based on heart rate variability," *Human Factors*, vol. 66, no. 6, pp. 1681–1702, 2023.
- [16] N. Khan, M. S. Nawaz, R. A. K. Sherwani, and M. Aslam, "Moving average EWMA chart for the Weibull distribution," *Communications in Statistics-Simulation and Computation*, vol. 52, no. 5, pp. 2231–2240, 2023.
- [17] J. Nasir, M. Aamir, Z. U. Haq, S. Khan, M. Y. Amin, and M. Naeem, "A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD Using ARIMA and LSTM Models," *IEEE Access*, vol. 11, pp. 14322–14339, 2023.
- [18] M. Monge and J. Infante, "A fractional ARIMA (ARFIMA) model in the analysis of historical crude oil prices," *Energy RESEARCH LETTERS*, vol. 4, no. 1, 2023.
- [19] A. Kaab, M. Sharifi, H. Mobli, A. Nabavi-Pelesaraei, and K. Chau, "Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production," *Science of the Total Environment*, vol. 664, pp. 1005–1019, 2019.
- [20] S. Borenstein, A. C. Cameron, and R. Gilbert, "Do gasoline prices respond asymmetrically to crude oil price changes?"

- Quarterly Journal of Economics*, vol. 112, no. 1, pp. 305–339, 1997.
- [21] F. Asche, O. Gjølborg, and T. Völker, “Price relationships in the petroleum market: an analysis of crude oil and refined product prices,” *Energy Economics*, vol. 25, no. 3, pp. 289–301, 2003.
- [22] L. Bettendorf, S. A. Van der Geest, and M. Varkevissier, “Price asymmetry in the Dutch retail gasoline market,” *Energy Economics*, vol. 25, no. 6, pp. 669–689, 2003.
- [23] R. K. Kaufmann and C. Laskowski, “Causes for an asymmetric relation between the price of crude oil and refined petroleum products,” *Energy Policy*, vol. 33, no. 12, pp. 1587–1596, 2005.
- [24] D. Bumpass, C. Douglas, V. Ginn, and M. H. Tuttle, “Testing for short and long-run asymmetric responses and structural breaks in the retail gasoline supply chain,” *Energy Economics*, vol. 83, pp. 311–318, 2019.
- [25] K. Kpodar and C. Abdallah, “Dynamic fuel price pass-through: evidence from a new global retail fuel price database,” *Energy Economics*, vol. 66, pp. 303–312, 2017.
- [26] X.-Z. Pan, X.-R. Ma, L.-N. Wang, Y.-C. Lu, J.-Q. Dai, and X. Li, “Spillover of international crude oil prices on China’s refined oil wholesale prices and price forecasting: daily-frequency data of private enterprises and local refineries,” *Petroleum Science*, vol. 19, no. 3, pp. 1433–1442, 2022.
- [27] Y. He and B. Lin, “Is market power the cause of asymmetric pricing in China’s refined oil market?” *Energy Economics*, vol. 124, 2023.
- [28] X. Cao, W. Li, S. Meng, X. Zhao, and M. Yang, “Research on the effect of marketization reform on the price of aviation kerosene in China,” *Sustainability*, vol. 16, no. 5, p. 2104, 2024.
- [29] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [30] M. Sugeno and T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling,” *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 1, pp. 7–31, 1993.
- [31] A. F. Güneri, T. Ertay, and A. Yücel, “An approach based on ANFIS input selection and modeling for supplier selection problem,” *Expert Systems with Applications*, vol. 38, no. 12, pp. 14907–14917, 2011.
- [32] A. Moghaddamnia, M. Ghafari Gousheh, J. Piri, S. Amin, and D. Han, “Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques,” *Advances in Water Resources*, vol. 32, no. 1, pp. 88–97, 2009.
- [33] W. Suparta and K. M. Alhasa, “Adaptive neuro-fuzzy interference system,” in *Modeling of Tropospheric Delays Using ANFIS*, pp. 5–18, Springer International Publishing, Cham, 2016.
- [34] A. Nabavi-Pelesaraei, S. Rafiee, F. Hosseini-Fashami, and K. Chau, “Artificial neural networks and adaptive neuro-fuzzy inference system in energy modeling of agricultural products,” in *Predictive Modelling for Energy Management and Power Systems Engineering*, R. Deo, P. Samui, and S. S. Roy, Eds., Elsevier, pp. 299–334, Amsterdam, The Netherlands, 2021.
- [35] C.-T. Cheng, J.-Y. Lin, Y.-G. Sun, and K. Chau, “Long-term prediction of discharges in manwan hydropower using adaptive-network-based fuzzy inference systems models,” in *Advances in Natural Computation*, L. Wang, K. Chen, and Y. S. Ong, Eds., pp. 1152–1161, Springer, Berlin, Heidelberg, 2005.
- [36] A. Barati-Harooni, A. Najafi-Marghmaleki, A. Tatar, and A. H. Mohammadi, “Experimental and modeling studies on adsorption of a nonionic surfactant on sandstone minerals in enhanced oil recovery process with surfactant flooding,” *Journal of Molecular Liquids*, vol. 220, pp. 1022–1032, 2016.
- [37] Y. Tan, C. Shuai, L. Jiao, and L. Shen, “An adaptive neuro-fuzzy inference system (ANFIS) approach for measuring country sustainability performance,” *Environmental Impact Assessment Review*, vol. 65, pp. 29–40, 2017.
- [38] G. R. Vakili-Nezhaad, A. Al Ajmi, A. Al Shaaili, F. Mohammadi, and A. Kazemi, “Machine learning assisted modeling of interfacial tension in the system N2/Brine,” *Sustainable Chemistry and Pharmacy*, vol. 33, 2023.
- [39] Gia, “Weekly U.S. All grades all formulations retail gasoline prices (dollars per gallon),” https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=EMM_EPM0_PTE_NUS_DPG&f=W.
- [40] S. Yoon and M. Park, “Prediction of gasoline orders at gas stations in South Korea using VAE-based machine learning model to address data asymmetry,” *Applied Sciences*, vol. 13, no. 20, p. 11124, 2023.
- [41] B. L. Ljubljana, “Orange data mining,” <https://orangedata mining.com>.