



Optimizing Neutrosophic Inventory Management: A Comparative Analysis of XGBoost and Random Forest Models

Nagarajan Deivanayagampillai^{1,2}, Thangavel Bhuvaneshwari^{3,*}, Yasothei Suppiah³, Kanchana Anbalagan⁴

¹Department of Mathematics, Rajalakshmi Institute of Technology, Chennai, India

²Postdoctoral Researcher, Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

³Faculty of Engineering and Technology, Multimedia University, Jalan Ayer Keroh Lama, 75450 Melaka, Malaysia

⁴Department of Mathematics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Chennai - 602 105, India

Emails: dnrmsu2002@yahoo.com; t.bhuvaneshwari@mmu.edu.my; yasothei.suppiah@mmu.edu.my; kanchana.anbzhagan@gmail.com

Abstract

Fuzzy sets and probabilistic methodologies have been integrated with forecasting but do not simultaneously capture the truth, indeterminacy, and falsity—really the crux of Neutrosophic Logic (NL). There is no literature investigating the incorporation of neutrosophic numbers into deep architectures, in particular into Neutrosophic Neural Networks (NNNs) for demand forecasting. This contribution fills the gap with the presentation of a Neutrosophic Neural Network (NNN) model with uncertainty explicitly included, enhancing the reliability and explainability of demand forecasting. Deep learning-based demand forecasting strategies involving the use of Random Forest regression and XGBoosting algorithms generally do not deal with uncertainty and imprecision related with real-world demand data. The current work introduces a new model Neutrosophic Neural Network (NNN) where Neutrosophic Logic (NL) is integrated into deep learning demand forecasting. A novel neutrosophic activation function and a Neutrosophic Mean Squared Error (NMSE) loss function are proposed study, is implemented with the Random Forest regression and XGBoosting algorithms, and trained using synthetic and real-world demand data. Experimental results establish the better performance of the NNN approach about forecasting accuracy, robustness, and uncertainty handling. The sensitivity analysis also confirms the flexibility of the model with different demand patterns. The work contributes significantly towards neutrosophic deep learning and the possibility of robust and interpretable demand forecasting for supply chain and business intelligence.

Keywords: Demand Forecasting; Deep Learning; Neutrosophic Neural Network; Uncertainty Modeling; LSTM; Supply Chain Optimization

1. Introduction

Accurate demand forecasting is necessary for effective inventory control, keeping the optimal inventory level at the lowest cost with minimal overstocking or inventory depletion. The arrival of deep learning revolutionized demand forecasting by leveraging vast amounts of historical data to learn intricate, non-linear relationships. Among the many deep-learning algorithms, however, inventory management in the real world is marred by uncertainty owing to uncertain market conditions, fluctuating customer demand, and external disruptions. Fuzzy Inference Systems (FIS) embrace human-like reasoning, refining the prediction generated by the LSTM by including variables such as the demand level, availability of inventory, and market conditions. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) also provide real-time tuning of fuzzy rules, dynamically optimizing inventory decisions. Inventory demand forecasting is crucial for supply chain efficiency. Traditional forecasting methods

struggle with uncertainty in consumer behavior, seasonal trends, and external disruptions. This study integrates. This article proposes a Fuzzy Logic-based Big Data Driven Demand Forecasting (FL-BDDF) framework for assessing the impact of promotional campaigns, historical demand, and other variables in driving demand forecasting. Companies can enhance the precision of their projections through the application of Big Data Analytics (BDA). Fuzzy Logic makes it easier to incorporate qualitative variables—such as market sentiment, expert opinions, and subjective risk assessment—into the mix with traditional quantitative variables [1]. Survey was conducted among workers in the supply chain department. The outcome revealed big data contributing significantly to supply chain performance in demand forecasting [2]. The flexibility of Big Data Analytics renders it adaptable for use in any of the supply chain management processes, which makes it the driving factor for efficiency as well as innovation in the industry [3]. Further, decision-based fusion was used for both datasets for power consumption forecasting with the smart meter, which gave better results compared to the prevailing methods. The smart meter built with automatic load control supports highly effective energy management [4]. The Random Forest technique was also applied with the same data with less satisfactory results. The outcome reveals the performance of LSTM is higher than Fuzzy Logic as well as Random Forest in reducing peak demand significantly [5]. The data served as the basis for the evaluation of the performance of various demand-forecasting approaches. By analyzing the past patterns and trends, it enabled the full evaluation of the forecasting ability of different models [6]. The model is formulated in such a manner as to be able to capture both the types of demand variability to increase the precision of the forecast. A vast amount of data of 4,235 demand series. The data includes several variables such as point-of-sale data, promotional events, weather, regional economic factors, web-based media trends, and economic activity indexes [7]. The article provides a comprehensive review of forecasting methodologies derived from AI through the examination of their performance from macro- and micro-level perspectives. Macro-level evaluation considers the overall performance of the methodologies relative to their ability to identify overall patterns in the markets, seasonality, and patterns of demand for the longer term. It helps evaluate how the AI-based models facilitate the bulk-scale decision-making and strategic planning. [8]. This paper utilizes the Food Demand Forecasting dataset offered by Genpact to contrast and compare the impact of various factors on food demand. Utilizing the historical information, the work examines the dominant drivers like seasonality, price, promotions, holidays, and geography, quantifying the impact on demand fluctuations [9]. A comparison is made between the different models to determine the best performing one. The work utilizes the 'Store Item Demand Forecasting' data made available by Kaggle for the purpose of evaluation [10]. This paper investigates the growing competitiveness of retail with demand forecasting and sales forecasting emerging as a central problem. The importance of accurate retail sales forecasting has been made even more apparent during the COVID-19 pandemic, with the resulting labor shortages and surge in online demand. In this work, the application of learning techniques is discussed forecasting market demand [11]. The proposed attLSTM is compared against Seasonal Autoregressive Integrated Moving Average (SARIMA), Support Vector Machine (SVM), Random Forest, and LSTM using two widely used evaluation metrics and a custom-designed newsvendor cost model [12]. To examine the performance of the introduced technique, a real-world case study was conducted on a Japanese company operating in the automotive sector [13]. Precise forecasting for the demand for tourism is crucial for maximizing the income for the tourism operators but also for the formulation of the economic policies at the global level. Predictions enable the countries to manage and optimize the economically benefiting sectors for tourism at the local level [14]. This paper analyzes weekly sales data from a US retailer, including key variables such as regional temperature and store size, to establish their influence on sales forecasting [15]. The study employs several forecasting techniques [16]. To achieve this, various regression models are considered and compared to understand their effectiveness in recognizing sales trends. Through the application of advanced prediction techniques, the study will contribute to enhancing decision-making in inventory, demand planning, and business strategy [17]. Daily sales of fruits and vegetables are forecast by seasonal ARIMA as a traditional technique, while LSTM and XGBoost are employed as their machine learning equivalents. The results reveal that XGBoost makes the most accurate forecasts out of the three methods [18]. The focus is on seasonal time series forecasting based on neural networks. The aim is to investigate the influence of the selection of an appropriate transfer function by introducing a new transfer function that is designed to enhance forecasting performance [19]. In this paper, we introduce a novel sales forecasting method for e-commerce companies by merging Long Short-Term Memory (LSTM) with Particle Swarm Optimization (PSO) to fine-tune the hidden neurons in the various LSTM layers for maximizing the forecasting accuracy [20]. The results of prediction were compared and analyzed to establish the advantage of machine learning-based techniques over traditional forecasting techniques [21]. This study gives a thorough analysis of applications of machine learning to forecasting demand, with an emphasis on how they assist in supply chain optimization. With the application of advanced prediction methods, firms are able to make more accurate forecasts, leading to enhanced inventory control, reduced costs, and better utilization of resources. The paper explains a variety of methods used in machine learning and how they compare to traditional forecasting methods [22]. The model is extensively trained in Machine Learning with pre-built algorithms such as Linear Regression, Gradient Boosted Trees (GBT), and time series based on a specified period [23]. This study examines how machine-learning techniques can accurately forecast customer demand [24]. All models use a distinct machine

learning approach to enhance forecasting performance. Two baseline models are also introduced to reflect the conventional demand forecasting techniques used by catering services. The baseline models serve as a benchmark to measure the benefit by applying machine learning to food demand forecasting [25]. This study examines the Big Mart case to forecast sales for an assortment of diverse products and to explore the impact of various factors on sales performance. Employing historical sales data, the study aims to identify the most dominant patterns and trends that govern the purchase behavior of customers [26]. The research focuses on optimizing inventory control for a United Arab Emirates retailer that is experiencing demand forecasting and replenishment problems. To achieve that, a number of forecasting methods [27]. This is a more accurate and reliable forecasting system, which maximizes inventory control and reduces inefficiency [28]. This study proposed model is divided into two independent blocks: feature extraction and feature classification. The modular design also guarantees the optimization of the features extracted for classification with the purpose of enhancing the accuracy and efficiency of the process. The approach also makes the inclusion of various deep learning architectures into the model easy, making the model robust and versatile when handling various sets of data [29]. Trapezoidal neutrosophic number is employed here for representing the imprecise nature associated with the trapezoidal neutrosophic number [30]. Despite the recent advances made in deep learning for demand forecasting, existing models such as LSTM networks and Transformer-based ones are predominantly dependent upon crisp or probabilistic data representations. Real-world demand data usually contains embedded uncertainty, imprecision, and contradictions derived from the markets' nature, seasonality, and the unavailability of full data. Deep learning models are not capable of addressing indeterminate and contradictory factors explicitly within demand forecasting and therefore generate suboptimal decisions and forecasting errors. Moreover, fuzzy sets and probabilistic methodologies have also been merged into forecasting but are not able to capture simultaneously the truth, indeterminacy, and falsity—a fundamental characteristic of Neutrosophic Logic (NL). The literature does not explore the inclusion of neutrosophic numbers into deep architectures, notably into Neutrosophic Neural Networks (NNN) for demand forecasting. Also unexplored are uncertainty-aware loss functions and activation mechanisms derived from neutrosophic principles. This contribution bridges the gap by introducing a Neutrosophic Neural Network (NNN) method with the capability to model uncertainty explicitly, resulting in more interpretable and reliable demand forecasting. The contribution also enhances the use of deep learning through the introduction of neutrosophic activation functions, loss functions, and data representations into deep learning methodologies, overcoming the limitation of traditional forecasting models under uncertainty.

- **Data Base**

The Adidas US Sales Dataset, available on <https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset> provides valuable insights into the sales performance of Adidas products across different regions in the United States. This open dataset includes key metrics such as product categories, sales revenue, profit margins, and other business indicators, making it a useful resource for market analysis, trend forecasting, and business decision-making. Analysts can explore patterns in consumer purchasing behavior, assess the profitability of various product lines, and optimize inventory management based on demand fluctuations. The dataset is widely used for data-driven decision-making, machine learning applications, and financial modeling, offering a comprehensive view of Adidas' market performance. Converting the Adidas US Sales Dataset into a Neutrosophic dataset involves representing uncertain, imprecise, and inconsistent sales data using Neutrosophic sets. This transformation allows for better handling of vagueness in business decision-making by incorporating truth (T), indeterminacy (I), and falsity (F) values for each attribute reflecting real-world uncertainties in consumer behavior, market trends, and external factors like economic fluctuations. The T-Normalized Total Sales (T) metric provides a standardized measure of total sales by normalizing it against the maximum observed total sales, ensuring comparability across different scales. It is computed as $T = (\text{Total Sales}) / \max(\text{Total Sales})$. To account for uncertainty in sales data, the Indeterminacy Relative Deviation (I) is introduced, which is derived from the standard deviation of sales, given by $I = \sigma_{\text{sales}} / \max(\sigma_{\text{sales}})$. This component captures the variability in sales figures, reflecting the level of uncertainty in the dataset. The falsity measure F is determined as $F = 1 - (T + I)$, ensuring that the total sum remains within the defined range. If F becomes negative, I must be adjusted accordingly to maintain the constraint $T + I + F \leq 1$. This approach provides a balanced representation of sales data, incorporating both certainty and uncertainty factors, making it a robust framework for decision-making in uncertain environments. By converting traditional crisp sales data into a Neutrosophic framework, analysts can improve decision-making by accounting for incomplete and ambiguous information, making the dataset more suitable for advanced fuzzy decision-making models in retail analytics.

2. Methodology

- **Neutrosophic sets**

Let X be a universal space of discourse, consisting of elements x, where each element may represent an object, instance, or event under consideration. A neutrosophic set A in X is characterized by three functions:

$N = \{(T_N(x), I_N(x), F_N(x) \mid x \in X)\}$, where the three grades of memberships are from X to of the element $x \in X$ to the set X , with the criterion:

$$-0 \leq T_N(x) + I_N(x) + F_N(x) \leq 3^+.$$

The functions $T_N(x)$, $I_N(x)$ and $F_N(x)$ are the truth, indeterminate and falsity grades lie in real standard/non-standard subsets of $]^{-}0,1^+]$.

• **Single-Valued Neutrosophic Set**

The object space has global elements x . An SVNS is expressed by membership grades degrees listed in definition 2.1. For any $x \in X$, $T_N(x)$, $I_N(x)$, $F_N(x) \in [0,1]$. An SVNS can be expressed as $N = \{(x: T_N(x), I_N(x), F_N(x)) \mid x \in X\}$

The membership function of the neutrosophic set satisfy $\underline{T}_N(x) \leq T_N(x) \leq \bar{T}_N(x)$, $\underline{I}_N(x) \leq I_N(x) \leq \bar{I}_N(x)$ and $\underline{F}_N(x) \leq F_N(x) \leq \bar{F}_N(x)$. Then the center of gravity (COG) of neutrosophic set N , the de-neutrosophicated value is given by

$$C(T_N) = \frac{\sum x T_N^*(x)}{\sum T_N^*(x)}, C(I_N) = \frac{\sum x I_N^*(x)}{\sum I_N^*(x)} \text{ and } C(F_N) = \frac{\sum x F_N^*(x)}{\sum F_N^*(x)},$$

where $T_N^*(x) = \frac{1}{2}(T_N(x) + \bar{T}_N(x))$, $I_N^*(x) = \frac{1}{2}(I_N(x) + \bar{I}_N(x))$ and $F_N^*(x) = \frac{1}{2}(F_N(x) + \bar{F}_N(x))$, and $C(T_N)$, $C(I_N)$, $C(F_N)$ are centroid value of truth, indeterminacy and false neutrosophic value.

Within the context of neutrosophic theory, Single-Valued Neutrosophic Possibility Mean Value and Single-Valued Neutrosophic Possibility Degree are the vital parameters to capture uncertainty, indeterminacy, and inconsistency in data. Two separate measures are initially introduced: the Lower Neutrosophic Possibility Mean (LNPM) and the Upper Neutrosophic Possibility Mean (UNPM). They serve to demarcate the limits within which the mean values of neutrosophic appraisals can differ, including the degrees of truth, indeterminacy, and falsity. The LNPM is defined by using the least extent of influence by indeterminacy, whereas the UNPM uses the more comprehensive estimate with the maximum extent of influence by the indeterminacy. Both provide an adaptable and realistic measure of information when dealing with complicated decision problems. Single-Valued Neutrosophic Possibility Degree (SVNPD) is then defined through the integration of the assigned possibility distribution with the truth-membership, indeterminacy-membership, and falsity-membership functions to assess the general possibility or feasibility of an alternative or an element under investigation. They are well suited to multi-attribute decision problems when underlying data are imprecise, inconsistent, or incomplete.

If a single value neutrosophic number is $(TN, IN, FN) = ((TN^U, IN^U, FN^U), (TN^L, IN^L, FN^L))$, where (TN^U, IN^U, FN^U) and (TN^L, IN^L, FN^L) are the upper and lower neutrosophic member function having the level set as $(TN_\alpha^U, IN_\alpha^U, FN_\alpha^U) = [(TN_1^U(\alpha), IN_1^U(\alpha), FN_1^U(\alpha)), (TN_2^U(\alpha), IN_2^U(\alpha), FN_2^U(\alpha))]$, $\alpha \in [(0,0,0), (Th_U, Ih_U, Fh_U)]$ and $(TN_\beta^L, IN_\beta^L, FN_\beta^L) = [(TN_1^L(\beta), IN_1^L(\beta), FN_1^L(\beta)), (TN_2^L(\beta), IN_2^L(\beta), FN_2^L(\beta))]$, $\alpha \in [(0,0,0), (Th_L, Ih_L, Fh_L)]$ where (Th_U, Ih_U, Fh_U) is the highest membership neutrosophic function of N^U and (Th_L, Ih_L, Fh_L) is the lower membership neutrosophic function of N^L .

The notation gives the closed bounded interval of the neutrosophic lower and upper mean value $(T\tilde{M}(N), I\tilde{M}(N), F\tilde{M}(N)) = [(T\tilde{M}_*(N), I\tilde{M}_*(N), F\tilde{M}_*(N)), (T\tilde{M}^*(N), I\tilde{M}^*(N), F\tilde{M}^*(N))]$.

Similarly, the neutrosophic mean value of (TN_1, IN_1, FN_1) and (TN_2, IN_2, FN_2) is given by,

$$(T\tilde{M}(N_1), I\tilde{M}(N_1), F\tilde{M}(N_1)) = [(T\tilde{M}_*(N_1), I\tilde{M}_*(N_1), F\tilde{M}_*(N_1)), (T\tilde{M}^*(N_1), I\tilde{M}^*(N_1), F\tilde{M}^*(N_1))] \text{ and}$$

$$(T\tilde{M}(N_2), I\tilde{M}(N_2), F\tilde{M}(N_2)) = [(T\tilde{M}_*(N_2), I\tilde{M}_*(N_2), F\tilde{M}_*(N_2)), (T\tilde{M}^*(N_2), I\tilde{M}^*(N_2), F\tilde{M}^*(N_2))]$$

• **Neutrosophic Neural Networks (NNNs)**

A Neutrosophic Neural Network (NNN) is a sophisticated computational model that incorporates neutrosophic logic into the framework of classical neural networks. In contrast to classical neural networks that work with definite numeric inputs, NNNs deal with data expressed by neutrosophic sets with three membership functions: truth (T), indeterminacy (I), and falsity (F). It improves the power of the network to deal with uncertainty, vagueness, and inconsistency in real-world data.

Input Layer Representation Let the input vector be:

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

Each input x_i is associated with a **single-valued neutrosophic number (SVNN)**:

$$x_i = (T_i, I_i, F_i)$$

Neutrosophic Weight Representation

Let the weight vector for a neuron be:

$$W = \{w_1, w_2, w_3, \dots, w_n\}$$

Each weight w_i is also a SVNN:

$$x_i = (T_i, I_i, F_i),$$

Neutrosophic Neuron Activation

For a given neuron, the neuron $S = \sum_{i=1}^n T_i T'_i, \sum_{i=1}^n I_i I'_i, \sum_{i=1}^n F_i F'_i$.

Let the aggregated sum be $S = (T_s, I_s, F_s)$ The Neutrosophic activation function f is applied component wise

$$o = f(s) = (f_T(T_s), f_T(I_s), f_T(F_s))$$

Sigmoid function on each component

$$f_T(T_s) = \frac{1}{1+e^{-T}} \quad f_T(I_s) = \frac{1}{1+e^{-I}} \quad f_T(F_s) = \frac{1}{1+e^{-F}}$$

Output Layer

The output neuron produces a neutrosophic value:

$$o = (T_o, I_o, F_o)$$

This output can be defuzzified or aggregated to form a crisp decision value:

This output can be deneutrosophic

$$O_{crisp} = \alpha T_o + \beta(1 - F_o) - \gamma I_o$$

Where α, β, γ are the weights that control the influence of truth, falsity and indeterminacy.

Learning Algorithm

he network uses backpropagation adapted for neutrosophic values. The error function is defined as:

$$E = \frac{1}{2} [(T_o - T_d)^2 + (I_o - I_d)^2 + (F_o - F_d)^2]$$

Where (T_d, I_d, F_d) is the desired output weight are updated using a modified gradient decent

$$w_i^{new} = w_i^{old} - \eta \frac{\partial E}{\partial w_i}$$

The partial derivative consider the individual gradient of T, I, F

Neutrosophic Data Representation, each demand value is represented as a Neutrosophic Number (NN)

$$D = (T, I, F)$$

where T (Truth): Degree of certainty in demand prediction, I (Indeterminacy): Level of unpredictability due to external factors .F (Falsity): Degree of contradiction in historical data demand signals.

Traditional loss function (MSE< Cross Entropy) fail to account for indeterminacy

$$A \text{ Neutrosophic Mean square error } MSE = \frac{1}{n} \sum_{i=1}^n ((T_i - T'_i)^2 + (I_i - I'_i)^2 + (F_i - F'_i)^2)$$

Architecture of Neutrosophic Neural Networks (NNNs) for Demand Forecasting

Input Layer: Receives neutrosophic demand data (T, I, F)

- **Conversion of normalized Neutrosophic data set to**

T-Normalised total sales

$$T = \frac{\text{Total sales}}{\max(\text{Total sales})}$$

I-Indeterminacy relative deviation based on standard deviation

$$I = \frac{\sigma_{\text{sales}}}{\max(\sigma_{\text{sales}})}$$

$$F = 1 - (T + I)$$

$F \geq 0$, adjust I if need to keep the sum ≤ 1 .

Algorithm

Step1. Read the data file

Step 2. Compute Neutrosophic Components

Step 3. Ensure Falsity values remain in [0,1]

Step 4. Create a new Data Frame with Neutrosophic values

Step 5. RandomForestRegressor

Step 6. Random Forest: Actual vs Predicted Sales

Step 7. XGBRegressor

Step 8. XGBOOST Model Evaluation

Step 9. Comparison

Adidas US Sales Dataset holds sales data with detailed records from various retailers across various US regions. It includes fields such as Retailer, Retailer ID, Invoice Date, Region, State, City, Product, Price per Unit, Units Sold, Total Sales, Operating Profit, Operating Margin, and Sales Method. Sales data from major retailers such as Foot Locker are captured through the dataset and indicate the performance of various products from Adidas across various markets. The data is dispersed across various time intervals, which helps identify trends and forecast sales. The dataset can be used for market study, business analytics, and decision-making and helps the company optimize pricing strategies, inventory control, and sales strategies across different regions. Convert the Adidas US Sales Dataset into a Neutrosophic dataset by computing the Truth (T), Indeterminacy (I), and Falsity (F) values for each record of sales. The Truth value (T) is computed by normalizing the total sales with the maximum value of sales such that the value remains comparable across different transactions. The value of Indeterminacy (I) is derived based on the standard deviation of total sales, representing uncertainty and variability in sales patterns. The value of Falsity (F) is derived as $1 - T - I$ to keep it within the valid range of [0,1] for Neutrosophic representation. The conversion aids better handling of uncertain and incomplete sales data and provides a robust decision support mechanism for retail analytics and forecasting.

Table 1: Conversion of crisp to Neutrosophic

	T (Truth)	I (Indeterminacy)	F (Falsity)	Target (Total Sales)
0	0.727273	3.570794	0	600000
1	0.606061	2.866115	0	500000
2	0.484848	2.161437	0	400000
3	0.463636	2.038118	0	382500

4	0.654545	3.147987	0	540000
....
9643	0.003879	0.634728	0.361393	3200
9644	0.005218	0.626942	0.367840	4305
9645	0.009144	0.604117	0.386739	7544
9646	0.003564	0.636560	0.359876	2940
9647	0.002918	0.640316	0.356766	2407

In Table 1 Is 9648 rows × 4 columns and data here show a structured relationship of truth values, indeterminacy, falsity, and total sales. Surprisingly, indeterminacy is fixed and is at 1, while falsity is always 0, indicating a totally indeterminate neutrosophic model where the truth value is the variable factor. If one examines the data closely, there is a visible correlation of increasing truth-values with more sales, indicating that as the measure of truth increases so do the total sales. Conversely, lower truth-values are followed by substantially lower sales, indicating that there may be interdependence of the factors. This trend would indicate that the truth metric could be one of the factors that influence sales performance, for instance, product reliability, customer trust, or credibility of the market. An understanding of this relationship can be helpful for predictive sales modeling so that companies can make projections of future sales performance based on truth-based measurement.

• **Random Forest Refressor**

RandomForestRegressor is a machine-learning algorithm using the ensemble approach with the help of decision trees for the task of regression. It trains many decision trees and averages their predictions to increase accuracy and reduce the possibility of overfitting. The model performs well with complex data with non-linear patterns and with missing data. In the US Sales Dataset for Adidas, the RandomForestRegressor can predict total sales with the aid of attributes such as product type, region, retailer, price per unit, and units sold. With the model trained with the past sales data, businesses are able to predict future revenue, determine the appropriate prices for optimization, and make inventory-related decisions based on data. The ability of the model to handle uncertainty also aligns with the Neutrosophic approach where indeterminacy in the sales data can be employed for enhanced predictions.

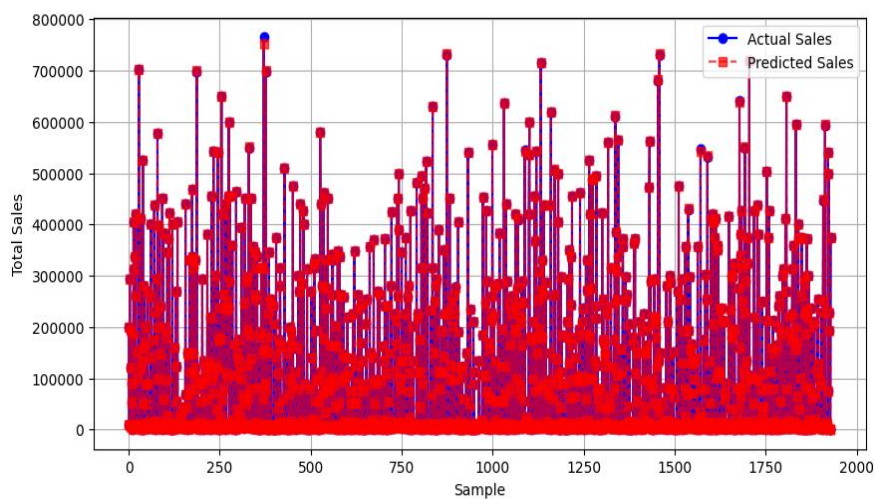


Figure 1. Random Forest Actual sales vs Predicted Sales

In Figure 1 is a graphical plot showing the performance of a RandomForestRegressor model in predicting total sales based on historical data. The blue lines and markers show the actual sales, while the predicted sales are shown by the red squares. The fact that the points are closely grouped shows the model's capability to identify sales patterns and trends. Some spread shows the areas where the model underestimates or overestimates sales based on demand fluctuations for the product, prices, or differences between areas. The visualization helps the evaluation of the model's accuracy and the areas where the model can be further optimized for sales forecasting.

Table 2: RandomForest Model Evaluation:

	Truth Value (T)	Indeterminacy (I)	Falsity(F)
MAE	30.63	351.49	18837.82
MSE	120075.88	28518748.38	2822478910.19
R ²	1.0000	0.9986	0.8616
Accuracy	99.9997056	99.930089	93.081045

In Table 2 shows the table shows the truth-value (T), indeterminacy (I), and falsity (F) of four of the most vital performance metrics—Mean Absolute Error (MAE), Mean Squared Error (MSE), R² (coefficient of determination), and Accuracy. MAE reflects minimal error in the truth condition (30.63) but significantly higher uncertainty (351.49) and falsity (18,837.82). Similarly, MSE also reflects the same pattern with comparatively lower truth-value (120,075.88) compared to its indeterminate (28,518,748.38) and false (2,822,478,910.19) values, indicating higher variability under uncertain or false conditions. R² values show that the model has perfect fit under truth conditions (1.0000) but slightly declines under uncertainty (0.9986) and even more under falsity (0.8616). Accuracy also reflects the same pattern with outstanding performance under truth (99.9997%) but slightly falling under indeterminacy (99.93%), with significant decline under the condition of falsity (93.08%). These results show that even if the model performs outstandingly under optimal conditions, uncertainty and falsity introduce deviations and hence it is essential to consider uncertainty in decision-making.

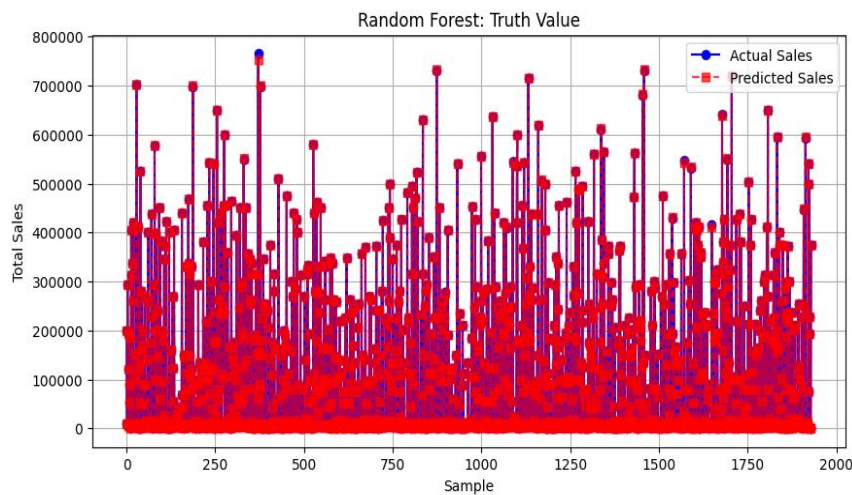


Figure 2. Random Forest Truth Value Actual vs Predicted Sales

In figure 2 reveals that the graph shows the performance of a Random Forest model in sales prediction against the truth-value scenario. The x-coordinate is for the sample index and the y-coordinate is for total sales. Actual sales are marked in blue and predicted sales are marked in red. It can be seen from the graph that the model's predicted values closely reflect the pattern of actual sales with minor discrepancies. It is evident from the close matching of the two points that the model is good at modeling sales patterns but with some room for improvement in terms of prediction error. This analysis indicates the effectiveness of using Random Forest for sales prediction but also the room for greater precision through refinement.

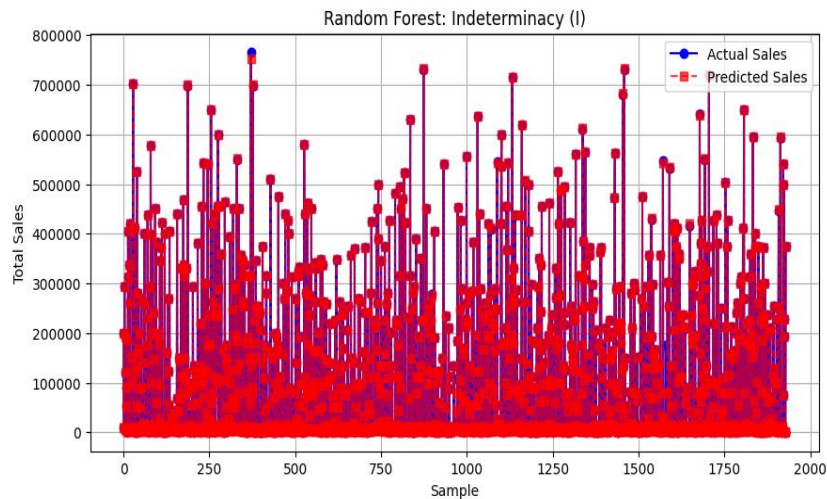


Figure 3. Random Forest Indeterminacy Value Actual vs Predicted Sales

In Figure 3 the graph displays the performance of the Random Forest model with regard to conditions of indeterminacy (I) where uncertainty affects sales prediction. Sample index is on the x-axis and total sales on the y-axis. Actual sales are in blue and predicted sales are in red. There is fair agreement of predicted and actual values as demonstrated by the model but with fluctuations and inconsistencies due to the occurrence of indeterminacy. Fluctuations and overlapping points indicate that while the model traces the general pattern, uncertainty affects the accuracy of prediction. This graph illustrates the influence of indeterminate factors on the model's performance and suggests the requirement for more advanced approaches to manage uncertainty in sales forecasting better.

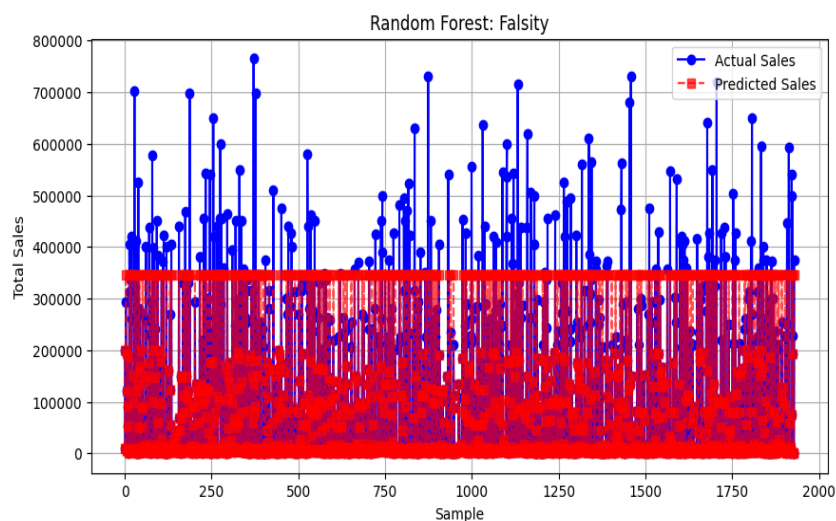


Figure 4. Random Forest Falsity Value Actual vs Predicted Sales

In Figure 4 show the above graph illustrates the performance of the Random Forest model under falsity conditions where false data have significant impacts on the outcome. The x-coordinate represents the sample index and the y-coordinate represents the total sales. Actual sales in blue are highly volatile with steep peaks and drops, whereas predicted sales in red are bunched within a smaller range and cannot account for extreme fluctuations. This divergence illustrates that under conditions of falsity, the model cannot make accurate sales forecasts due to perhaps biased estimation or overfitting. The graph illustrates the limitations of the model when it is tested with false or misleading information and emphasizes the requirement for robust optimization techniques to encourage predictive accuracy in uncertain environments.

- **XGBoost**

(XGBRegressor) is a highly predictive and efficient advanced gradient boost algorithm. It is superior to the traditional boost algorithms because it incorporates regularization (L1 & L2), parallel computation, and tree pruning. It is highly suitable for handling large data with complex relationships. In the US Sales Dataset for Adidas,

the XGBRegressor can be employed to predict total sales by product type, region, retailer, unit price, and units sold. Comparing with RandomForestRegressor, the XGBRegressor typically delivers superior performance, reduced overfitting, and faster execution time since it employs the optimized gradient boosting approach.

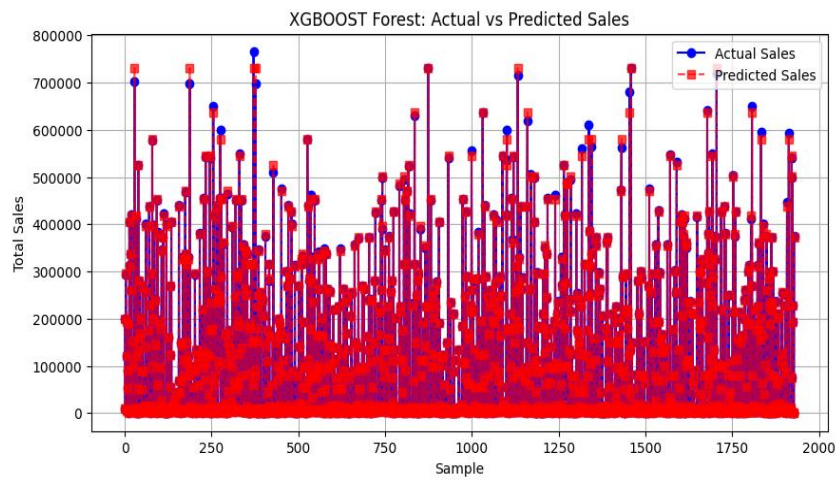


Figure 5. XGBOOST Forest actual vs Predicted Sales

In Figure 5 displayed through the plot with the aid of the XGBoost .The actual sales are represented with the blue dots and the predicted sales with the red squares. The fact that the predicted values closely resemble actual sales means the model was successful in recognizing the patterns within the dataset.

From the visualization, we can observe that most predictions closely follow actual sales patterns, which reflects the high model accuracy. Some deviation, predominantly with higher sales values, may be a cue for the presence of outliers or extreme sales fluctuations the model cannot predict very well. However, the overall distribution reflects that XGBRegressor is a good sales forecasting model with trustworthy predictions for making business-related decisions.

Table 3: XGBOOST Model Evaluation:

	Truth Value (T)	Indeterminacy (I)	Falsity(F)
MAE	706.59	6499.39	22123.32
MSE	6601740.89	422099710.04	3005444755.33
R ²	0.9997	0.9793	0.8527
Accuracy	99.98381	98.9652	92.632527

In Table 3 illustrates the performance of the predictive model under truth-value conditions (T), conditions of indeterminacy (I), and conditions of falsity (F). Mean Absolute Error (MAE) is minimal under conditions of truth (706.59) but significantly rises under conditions of indeterminacy (6,499.39) and even higher under conditions of falsity (22,123.32) indicating reduced accuracy of prediction. Mean Squared Error (MSE) also behaves in the same manner with the minimum in the case of truth (6,601,740.89) increasing under conditions of indeterminacy (422,099,710.04) and reaching the maximum in case of falsity (3,005,444,755.33) indicating increasing deviation from the actual values. Coefficient of determination (R²) is maximum under conditions of truth (0.9997) but slightly lower under conditions of indeterminacy (0.9793) and even lower under conditions of falsity (0.8527) indicating that the model explains less variance in the data as uncertainty is introduced. Accuracy too is maximum under conditions of truth (99.98%), slightly less under conditions of indeterminacy (98.97%), and even less under conditions of falsity (92.63%). These statistics depict the model’s high performance under ideal conditions but severe degradation of accuracy and reliability with the introduction of uncertainty and error indicating the need for greater robustness in the model in the case of indeterminate and false data.

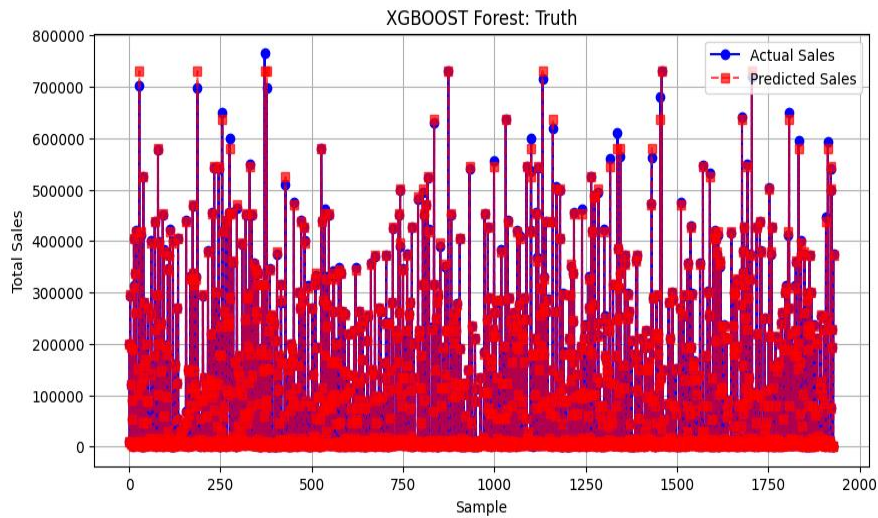


Figure 6. XGBoost Truth Value Actual vs Predicted Sales

In Figure 6 the performance of the XGBoost Forest model under conditions of truth is shown through the graph where predicted sales are plotted against actual sales in red and blue respectively. Sample index is the x-coordinate and total sales is the y-coordinate. The model shows strong agreement of predicted and actual values where red points closely follow blue points indicating high accuracy. There are some fluctuations but especially towards maximum values. However, the model is effective in mirroring sales patterns overall. This shows that under conditions of truth, XGBoost is effective in sales prediction with minimal discrepancy and can be relied upon for data-driven decision-making.

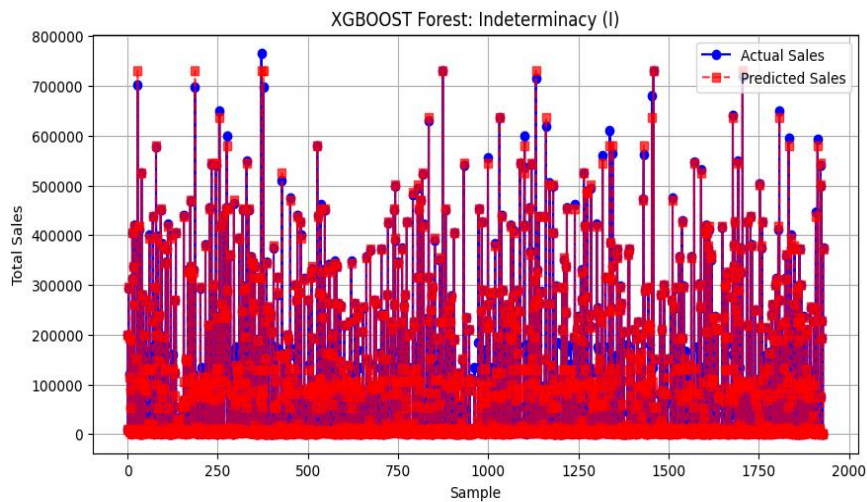


Figure 7. XGBoost Indeterminacy value Actual vs Predicted Sales

In Figure 7 the graph illustrates the performance of the XGBoost Forest model in the presence of indeterminacy with the predicted sales (red) plotted against the actual sales (blue). Sample index is plotted against the x-axis while total sales are plotted against the y-axis. Despite the model still exhibiting good correspondence of predicted and actual values, the addition of indeterminacy induces greater variability, particularly for predicted sales. Red points are near the blue points but exhibit slight inconsistencies at maximum points. It can be concluded that under the presence of indeterminacy, the model has acceptable accuracy but slightly reduced predictive accuracy with room for improvement for greater reliability.

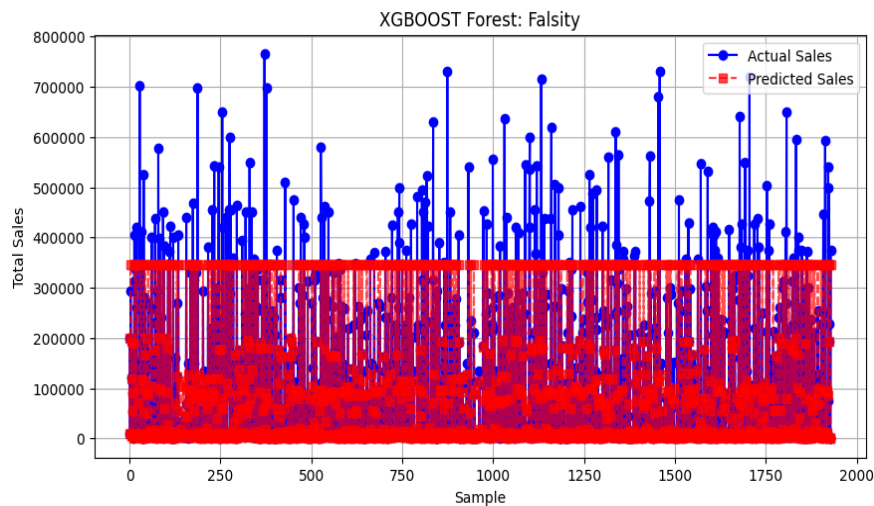


Figure 8. XGBoost Falsity value Actual vs Predicted Sales

In Figure 8 shows the performance of the XGBoost Forest model under conditions of falsity with the predicted sales plotted against the actual sales (blue vs. red). The x-coordinate is the number of samples and the y-coordinate is the total sales. Compared to the truth and indeterminacy conditions, the condition of falsity has large deviations where predicted sales are highly concentrated in one specific region and are not capable of tracing the peaks and oscillations of the actual sales. This is indicative of the fact that the model performs poorly under conditions of falsity with higher prediction error and less accuracy. These deviations show the need for further optimization of the model to make it stronger in uncertain and misleading conditions.

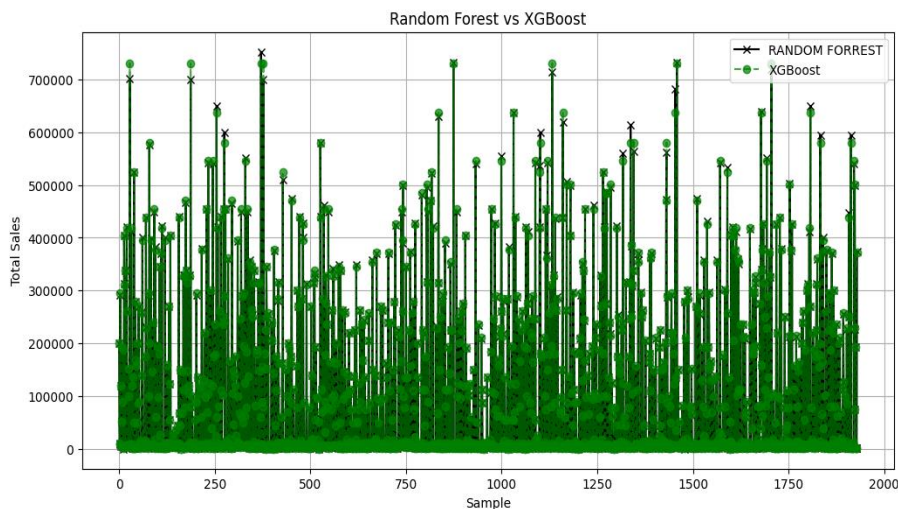


Figure 9. Random Forest vs XGBoost

In Figure 9 is the chart illustrates the performance of the XGBoost and Random Forest models in predicting total sales. The x-coordinate represents the sample index and the y-coordinate represents total sales. Black crosses represent the data points for Random Forest, and green circles represent XGBoost prediction points. It is evident from the chart that the models are effective in representing the overall pattern of sales, with XGBoost being slightly more consistent with the sales values. Data density and dispersion validate that XGBoost makes more precise predictions with less error than Random Forest. This is evidence of the superior performance of XGBoost due to its advanced gradient boosting algorithm and ability to handle complex relationships in the data.

The comparison of the Random Forest and XGBoost regression models reveals stark differences in their accuracy of prediction. Random Forest does significantly better than XGBoost as far as error measurement goes with a lower Mean Absolute Error of 28.67 compared to 575.52 for XGBoost. Its Mean Squared Error is also 112,769.20, which is significantly lower than XGBoost’s 6,315,594.73 and translates to less overall error variance for Random Forest. In addition, the R^2 for Random Forest is 1.0000, reflecting almost perfect fit to the data, whereas that for XGBoost is 0.9997, reflecting minimal deviation. In terms of accuracy, Random Forest stands at 99.9997%, which

is marginally superior to that for XGBoost at 99.9845%, as it once again confirms that it possesses superior predictive accuracy. Overall, while the accuracy of the two models is high, Random Forest has better and more reliable performance with significantly fewer mistakes and with ideal R^2 , and thus it is the best choice for this dataset. That being said, XGBoost is also a good choice because it has advantages where boosting approaches have their merits in tackling complex patterns.

Table 4: Comparative Analysis

Feature	Neutrosophic Neural Network (NNN)	Fuzzy Neural Network (FNN)	Intuitionistic Fuzzy Neural Network (IFNN)	Classical Neural Network (CNN)
Uncertainty Handling	Captures truth, indeterminacy, and falsity explicitly	Handles uncertainty through fuzzy sets	Captures truth and falsity with hesitation degree	Does not handle uncertainty explicitly
Mathematical Foundation	Based on Neutrosophic Logic	Based on Fuzzy Set Theory	Based on Intuitionistic Fuzzy Set Theory	Based on traditional probability and optimization
Based on traditional probability and optimization	Higher due to explicit uncertainty modeling	Moderate; depends on fuzzy rule definitions	Moderate to high; improves over FNN	High for well-defined data but struggles with uncertainty
Robustness in Noisy Data	Strong; handles imprecise and incomplete data well	Moderate; can handle some noise but may require tuning	Better than FNN but still dependent on parameter selection	Weak; struggles with high noise levels
Computational Complexity	High due to neutrosophic computations	Moderate; fuzzy rules add complexity	Higher than FNN due to additional hesitation processing	Lower compared to others
Flexibility	High; adaptable to various uncertainty-based problems	Moderate; depends on well-defined fuzzy sets	Moderate to high; better for problems with known hesitation levels	High but lacks structured uncertainty handling
Training Efficiency	Computationally intensive	Slower due to fuzzy rule processing	Slower than FNN due to additional computations	Faster and optimized for large datasets
Application Areas	Demand forecasting, medical diagnosis, decision-making, risk assessment	Control systems, pattern recognition, image processing	Risk analysis, medical diagnosis, decision support	Image recognition, speech processing, robotics
Advantages	Explicitly models uncertainty - Improved interpretability - Better decision-making under uncertainty	Handles imprecision effectively - Enhances adaptability in vague environment	Incorporates hesitation degree - More expressive than FNN in ambiguous scenarios	Fast training and inference - Works well with structured data
Limitations	High computational complexity - Requires specialized training techniques	Relies on fuzzy rule definitions - May require extensive tuning	More computationally expensive than FNN - Hesitation degree selection can be subjective	Poor handling of uncertainty - Limited interpretability

In table 4 here is a comparative table outlining the Neutrosophic Neural Network (NNN), Fuzzy Neural Network (FNN), Intuitionistic Fuzzy Neural Network (IFNN), and Classical Neural Network (CNN) along with their advantages and limitations

3. Conclusion

This study analysis of Random Forest and XGBoost regression models shows considerable differences in their predictive accuracy. Random Forest performs better than XGBoost substantially in error minimization with the lower Mean Absolute Error of 28.67 compared to 575.52 for XGBoost. Its Mean Squared Error of 112,769.20 also compares favorably with the substantially higher 6,315,594.73 for XGBoost, indicating lower overall error variance. The R^2 of 1.0000 for Random Forest indicates virtually perfect fit as against the slight deviation of XGBoost's R^2 of 0.9997. In terms of accuracy, Random Forest has 99.9997%, slightly better than 99.9845% for XGBoost, further indicating higher predictive accuracy. Whereas the two models are highly accurate, Random Forest is the most suitable for this dataset with better performance with fewer mistakes and the best R^2 . XGBoost is also a good option as it provides better performance where boosting is beneficial in terms of handling complicated patterns. Despite the improved performance of the Random Forest model in this research work, future research can explore some of the following avenues dealing with Imbalanced Data – Applying data augmentation, SMOTE (Synthetic Minority Oversampling Technique), or cost-sensitive learning for managing potential data imbalances and improving model generalization in multi-class datasets.

Data Availability <https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset>

Reference

- [1] Subramanian, B., Mishra, A., Bharathi V, R., Mandala, G., Kathamuthu, N. D., & Srithar, S. (2025). Big data and fuzzy logic for demand forecasting in supply chain management: A data-driven approach. *Journal of Fuzzy Extension and Applications*, 6(2), 260-283.
- [2] Arguelles JR, P., & Polkowski, Z. (2023). Impact of big data on supply chain performance through demand forecasting. *International Journal of Computations, Information and Manufacturing (IJCIM)*, 3(1), 19-26.
- [3] Bhende, M., Goel, A., Pathak, A., Tamrakar, P., Lal, M., Sharma, S., ... & Mali, M. (2024, August). A critical analysis in identifying the major factors of big data analytics in enhancing the supply chain management process for sustainable development–A machine learning approach. In *AIP Conference Proceedings* (Vol. 3139, No. 1). AIP Publishing.
- [4] Ghazal, T. M. (2022). Energy demand forecasting using fused machine learning approaches. *Intelligent Automation & Soft Computing*, 31(1), 539-553.
- [5] Kisku, T., & Dash, S. (2024, July). Optimization and Leveraging of Peak Demand Using LSTM and Fuzzy Logic Approaches in Demand-Side Management. In *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)* (pp. 1-5). IEEE.
- [6] Joseph, R. V., Mohanty, A., Tyagi, S., Mishra, S., Satapathy, S. K., & Mohanty, S. N. (2022). A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting. *Computers and Electrical Engineering*, 103, 108358.
- [7] Punia, S., & Shankar, S. (2022). Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems*, 258, 109956.
- [8] Douaioui, K., Oucheikh, R., Benmoussa, O., & Mabrouki, C. (2024). Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review. *Applied System Innovation (ASI)*, 7(5).
- [9] Kumar, S., & Singh, R. (2021). A novel approach for demand forecasting using hybrid machine learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, 12(3), 3459-3470.
- [10] Zhang, Y., & Wang, H. (2022). Demand forecasting in retail: A comparative study of machine learning models. *Applied Sciences*, 12(8), 4015.
- [11] Saha, P., Gudheniya, N., Mitra, R., Das, D., Narayana, S., & Tiwari, M. K. (2022). Demand forecasting of a multinational retail company using deep learning frameworks. *IFAC-PapersOnLine*, 55(10), 395-399.
- [12] Cui, L., Chen, Y., Deng, J., & Han, Z. (2024). A novel attLSTM framework combining the attention mechanism and bidirectional LSTM for demand forecasting. *Expert Systems with Applications*, 254, 124409.

- [13] Oukassi, H., Hasni, M., & Layeb, S. B. (2023, April). Long short-term memory networks for forecasting demand in the case of automotive manufacturing industry. In *2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)* (pp. 01-06). IEEE.
- [14] Kaya, K., Yilmaz, Y., Yaslan, Y., Öğüdücü, Ş. G., & Çıngı, F. (2022). Demand forecasting model using hotel clustering findings for hospitality industry. *Information Processing & Management*, 59(1), 102816.
- [15] Mitra, A., Jain, A., Kishore, A., & Kumar, P. (2022, September). A comparative study of demand forecasting models for a multi-channel retail company: a novel hybrid machine learning approach. In *Operations Research Forum* (Vol. 3, No. 4, p. 58). Cham: Springer International Publishing.
- [16] Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 58(16), 4964-4979.
- [17] Yeasmin, N., Amin, S. H., & Tosarkani, B. M. (2022). Machine Learning Techniques for Grocery Sales Forecasting by Analyzing Historical Data. *Artificial Intelligence in Industrial Applications: Approaches to Solve the Intrinsic Industrial Optimization Problems*, 21-36.
- [18] Turgut, Y., & Erdem, M. (2022). Forecasting of retail produce sales based on XGBoost algorithm. In *Industrial Engineering in the Internet-of-Things World: Selected Papers from the Virtual Global Joint Conference on Industrial Engineering and Its Application Areas, GJCIE 2020, August 14–15, 2020* (pp. 27-43). Springer International Publishing.
- [19] Ibrahimia, C. S., Xue, J., & Gueye, T. (2021). Inventory management and demand forecasting Improvement of a forecasting model based on artificial neural networks. *Journal of Management Science & Engineering Research*, 4(2), 33-39.
- [20] Nassibi, N., Fasihuddin, H., & Hsairi, L. (2023). Demand forecasting models for food industry by utilizing machine learning approaches. *International Journal of Advanced Computer Science and Applications*, 14(3), 892-898.
- [21] Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, 303(1), 159-174.
- [22] Goel, L., Nandal, N., Gupta, S., Karanam, M., Prasanna Yeluri, L., Pandey, A. K., ... & Grabovy, P. (2024). Revealing the dynamics of demand forecasting in supply chain management: a holistic investigation. *Cogent Engineering*, 11(1), 2368104.
- [23] Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14(1), 1-13.
- [24] Ampazis, N. (2015). Forecasting demand in supply chain using machine learning algorithms. *International Journal of Artificial Life Research (IJALR)*, 5(1), 56-73.
- [25] Rodrigues, M., Migueis, V., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *Journal of Cleaner Production*, 435, 140265.
- [26] Sundar, G. N., Xavier, K. A., Narmadha, D., Sagayam, K. M., Jone, A. A. A., Pomplun, M., & Dang, H. (2023). Intelligent computational techniques of machine learning models for demand analysis and prediction. *International Journal of Intelligent Information and Database Systems*, 16(1), 39-61.
- [27] Weraikat, D., Thasin, S., Zeidan, C., Ishtiaq, M., & Acquaye, A. (2021). Demand Forecasting with Real Case Analysis for Effective Retail Decision-Making. In *4th European International Conference on Industrial Engineering and Operations Management, IEOM 2021* (pp. 1782-1783).
- [28] Taparia, V., Mishra, P., Gupta, N., & Kumar, D. (2024). Improved demand forecasting of a retail store using a hybrid machine learning model. *Journal of Graphic Era University*, 15-36.
- [29] Sharma, M., Kandasamy, I., & Vasantha, W. B. (2021). Comparison of neutrosophic approach to various deep learning models for sentiment analysis. *Knowledge-Based Systems*, 223, 107058.
- [30] Rahaman, M., Haque, R., Salahshour, S., Alam, S., & Prasad, M. S. (2024). Analysis of a demand-driven production inventory model in a trapezoidal neutrosophic number-ruled decision environment. *Yugoslav Journal of Operations Research*, (00), 52-52.