I.J. Information Engineering and Electronic Business, **2026, 1, 126-144**
Published Online on February 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2026.01.08

Profit Forecasting for Daily Pharmaceutical Sales Using Traditional, Shallow, and Deep Neural Networks: A Case Study from Sabha City, Libya

Mansour Essgaer*
Artificial Intelligence Department, Faculty of Information Technology, Sebha University, Sebha, Libya
Email: man.essgaer@sebhau.edu.ly
ORCID iD: https://orcid.org/0000-0002-8447-5091
*Corresponding Author

Asma Agaal
Computer Sciences Department, Faculty of Technical Sciences, Sabha, Libya
Email: asma.agaal@sebhau.edu.ly
ORCID iD: https://orcid.org/0000-0002-8687-2605

Amna Abbas
Computer Science Department, Faculty of Science, Sebha, Libya
Email: amna.abas@sebhau.edu.ly
ORCID iD: https://orcid.org/0009-0009-4823-1135

Rabia Al Mamlook
Business Administration, Trine University, Angola, Indiana, United States
Email: almamlookr@trine.edu
ORCID iD: https://orcid.org/0000-0002-2523-7819

Received: 20 July, 2025; Revised: 27 October, 2025; Accepted: 20 December, 2025; Published: 08 February, 2026

Abstract: Accurate profit forecasting is critical for small-scale pharmacies, particularly in resource-constrained environments where financial decisions must be both timely and data-informed. This study investigates the predictive performance of sixteen regression models for daily profit forecasting using transactional data collected from a single local pharmacy in Sabha, Libya, over a 14-month period. An exploratory data analysis revealed strong right-skewed distributions in sales, cost, and profit, as well as pronounced temporal patterns, including seasonal peaks during spring and early summer and weekly profit clustering around weekends. After outlier treatment using the interquartile range method. A total of sixteen regression models were developed and evaluated, encompassing linear models (Linear, Ridge, Lasso, ElasticNet), tree-based models (Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost), proximity-based models (K-Nearest Neighbors), kernel-based models (Support Vector Regression), and neural architectures (Multi-Layer Perceptron, Convolutional Neural Network, Long Short-Term Memory, Gated Recurrent Unit). The models were assessed using Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and the R-squared score. The results consistently showed that tree-based ensemble models—particularly Extra Trees and LightGBM—achieved the highest accuracy, with R^2 values of 0.978 and 0.975 respectively, significantly outperforming neural and linear models. Learning curves and residual plots further confirmed the superior generalization and robustness of these models. We acknowledge that the dataset size (424 records) and the deterministic relationship between sales, costs, and profit influence these metrics. The study highlights the importance of model selection tailored to domain-specific data characteristics and suggests that well-tuned ensemble methods may offer reliable, interpretable, and scalable solutions for profit forecasting in simialr low-resource retail environments. However, broad claims of usefulness for all low-resource settings should be tempered by the limited scope of this dataset. Future work should consider longer-term data and external economic indicators to further improve model reliability, and focus on operational deployment strategies, investigating how these models can be integrated into daily pharmacy workflows despite real-time data constraints.

Index Terms: Cost of Sales, Public Pharmacies, Time Series Modeling, Sales Forecasting, Machine Learning, Deep Learning, Comparison of Predictive Models, Sabha, Libya

1. Introduction

Accurate forecasting in the pharmaceutical retail sector is crucial for maintaining operational continuity, optimizing pricing strategies, and guiding informed financial decisions [1]. Pharmacies operate within multifaceted environments shaped by consumer behavior, seasonal variations, regulatory shifts, and supply chain disturbances. As such, the capacity to predict daily profits becomes central to strategic planning and aligns with broader aims of efficiency and risk mitigation.

Historically, forecasting within retail and healthcare has relied on traditional statistical models—linear regression, moving averages, exponential smoothing, Holt–Winters, and AutoRegressive Integrated Moving Average (ARIMA). These approaches are highly interpretable and well-established [2], [3]. However, they frequently struggle when confronted with nonlinearity, intricate feature interactions, outliers, or irregular data patterns, all traits common in pharmaceutical datasets. A Swedish pharmacy-chain study highlighted this limitation, proposing that latent demand fluctuations and high-frequency volatility are better captured by more flexible models like Facebook Prophet or Seasonal ARIMA (SARIMA) using robust preprocessing [4].

In contrast, modern machine learning (ML) and deep learning (DL) techniques—such as support vector machines, random forests, gradient boosting, XGBoost, Long Short-Term Memory (LSTM) networks, and deep neural networks—have taken the stage due to their ability to manage complex, high-dimensional, noisy, and unstructured data [5], [6]. A recent open-access study involving 600,000 pharmacy sales records showed that XGBoost delivered superior performance across multiple drug categories, outperforming both ARIMA and LSTM in Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) metrics [7]. Similarly, a demand-forecasting effort during the COVID-19 era found that shallow neural networks sometimes outperform deeper architectures in pharmaceutical demand prediction, largely due to limited sample sizes and simpler data structures [8].

Despite a robust interdisciplinary literature—spanning retail [9], pharmaceutical manufacturing [10], and healthcare procurement [11]—most comparative studies are conducted in well-funded, large-scale settings. These often overlook the realities faced by small-scale pharmacies in constrained environments, where infrastructure is limited, records are fragmented, and market volatility is steep.

In Libya, particularly in Sabha, the gap is pronounced. While overall pharmaceutical market forecasts exist [7], [12], granular analysis on day-to-day profit dynamics—especially in under-resourced, public pharmacy settings—is missing. Our research aims to close this gap by offering a comparative study of sixteen regression models—including both classical (linear, time-series) and contemporary (ensemble, deep learning) approaches—with a unified preprocessing pipeline, grounded in daily transactional data collected over fourteen months from a public pharmacy in Sabha.

The main contributions of this study are threefold. First, it involves the construction of a structured and preprocessed dataset derived from real-world pharmacy records collected in Sabha, ensuring data quality and consistency for modeling purposes. Second, the study trains and evaluates a comprehensive suite of traditional, shallow, and deep learning regression models, all subjected to a unified preprocessing pipeline and cross-validation framework to ensure fair comparisons and reliable results. Third, it identifies the most accurate and generalizable models for short-term profit forecasting, particularly tailored to resource-constrained environments such as local pharmacies in developing regions. It is important to note that this study relies on data from a single pharmacy, which may limit the direct generalizability of the findings to other regions or pharmacy types without further validation.

The remainder of this paper is structured as follows. Section 2 provides a review of the existing literature on financial forecasting, with a particular focus on the application of machine learning and deep learning techniques. Section 3 details the dataset, the preprocessing methodology employed, and the implementation of the predictive models. Section 4 discusses the experimental setup, presents the results of model comparisons, and evaluates performance using relevant metrics. Section 5 details the limitation of the study. Section 6 offers conclusions drawn from the study and outlines potential directions for future research. Finally, Section 7 draws several future directions of this study.

2. Related Work

A substantial body of literature has explored the application of machine learning (ML) and deep learning (DL) techniques for financial forecasting across diverse sectors. In pharmacy and retail environments, these methods are increasingly employed to predict key financial indicators such as sales, costs, or profits from historical transactional data. This review of prior research is structured into four thematic categories to provide a comprehensive overview of the current state of the field.

2.1. Traditional Regression Methods as an Analytical Baseline

Linear and statistical models, particularly those based on Multiple Linear Regression (MLR) and Ordinary Least Squares (OLS), remain indispensable in forecasting research. Their value lies not only in their computational efficiency and interpretability but also in their critical function as a performance baseline against which more complex algorithms are benchmarked.

Several studies exemplify the utility and limitations of linear modeling. For instance, an application of MLR to forecast sales for an electronics company demonstrated that a well-diagnosed linear model could explain 84\% of sales variances, though it also highlighted challenges related to sample size and multicollinearity [13]. Similarly, a comparative study in the furniture retail sector used MLR and Holt-Winters methods, revealing significant forecast disparities between the models (4.05\% vs. 8.24\% CAGR) and underscoring the critical impact of model selection in underexplored domains [14]. In a broader retail context, a comparative analysis using Walmart sales data found that classical models like ARIMA and Holt-Winters sometimes outperform more complex algorithms, reinforcing the importance of including them as robust baselines [15].

Despite their utility, the limitations of linear models become apparent when applied to data with non-linear characteristics. One study demonstrated that an OLS model produced the highest number of outliers and the lowest predictive accuracy when forecasting sales data with strong seasonality, confirming its inability to capture complex temporal patterns [16]. This finding is further supported by research on forecasting in volatile conditions, where linear methods like Holt's Linear Exponential Smoothing showed inconsistent performance in the face of extreme external shocks [17]. Even regularized variants, such as the Ridge Regression model used in a pharmaceutical demand prediction study, served primarily as a component within a more robust ensemble strategy rather than as a top-performing standalone model [18]. Collectively, these studies affirm that while linear models are essential for benchmarking and diagnostics, their inherent assumption of linearity often necessitates the adoption of more advanced techniques to achieve high predictive accuracy in real-world applications.

2.2. The Ascendancy of Tree-Based and Ensemble Learning

A dominant trend in contemporary forecasting is the established superiority of tree-based and ensemble learning methods, such as Random Forest (RF) and Gradient Boosting Machines (GBM). A significant body of research confirms that these models consistently outperform traditional single-model approaches across various domains. In a healthcare context, an RF model was successfully used to forecast pharmaceutical demand with 71\% accuracy, leading to tangible improvements in supply chain management [19]. This result aligns with findings from retail analytics, where non-linear ensemble models are more effective than their linear counterparts [20], and in energy forecasting, where XGBoost was identified as the most accurate predictive model [21].

The field is also advancing beyond individual ensembles toward more sophisticated, hybrid architectures. A stacking scheme combining RF, Neural Networks, and GBM was shown to achieve lower error rates than existing state-of-the-art techniques [22]. Similarly, a hybrid EGB-RNN model demonstrated significant accuracy improvements over standalone GBDT or RNN models, although its performance was found to be highly sensitive to data characteristics [23]. These studies indicate a clear research trajectory toward multi-layered systems that leverage the distinct advantages of different algorithms.

A crucial challenge associated with these powerful models is their inherent lack of transparency. To address this "black box" problem, recent studies have focused on integrating eXplainable AI (XAI) techniques. For instance, GBMs used for financial risk assessment have been coupled with SHAP and LIME to provide transparent, feature-level explanations for their predictions [24]. This pursuit of interpretability has also been applied to RF models to deconstruct how news sentiment influences stock price predictions [25].

Finally, the literature underscores that the performance of these models is heavily dependent on effective data preprocessing. The application of Principal Component Analysis (PCA) before training an RF model was shown to effectively manage high-dimensional data, resulting in a 15\% average normalized error [26], in another study, an R-squared of 0.97 [27]. Furthermore, advanced feature engineering, such as applying signal decomposition (EMD) and feature elimination (RFE) before training XGBoost, has been shown to yield a significant boost in predictive accuracy [28]. In summary, while tree-based and ensemble models represent the state-of-the-art for many forecasting tasks, their successful implementation requires a holistic approach that incorporates thoughtful model combination, a commitment to interpretability, and a rigorous data preprocessing pipeline.

2.3. Neural Networks for Capturing Temporal Dynamics

Deep learning models, particularly Recurrent Neural Networks (RNNs) like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are increasingly prominent in forecasting due to their theoretical capacity to capture complex temporal dependencies. However, their practical application reveals a critical trade-off between model complexity and data availability. A recurring theme in the literature is the challenge posed by limited or poor-quality data to the stability and performance of these models. For example, a study on gas turbine performance prediction found that while an optimized LSTM model achieved a low error rate, its performance was constrained by limited training data, which elevated the risk of overfitting [29]. This issue was also observed in influenza forecasting, where data

limitations and formatting inconsistencies complicated robust model validation [30], and in bioacoustics, where small databases were found to hinder effective neural network training [31].

To mitigate these challenges, researchers are actively exploring data-centric strategies. A comprehensive survey identified transfer learning, few-shot learning, and data augmentation as key techniques for applying deep learning to small datasets [32]. The practical utility of these methods has been demonstrated in studies that used LSTM-based autoencoders to synthesize new observations [33], or applied dropout and data augmentation to improve CNN performance in small-data scenarios [34]. These findings suggest that while neural networks are inherently data-hungry, their limitations can be partially overcome through innovative data-centric techniques.

In direct performance comparisons, neural networks are often strong but not universally dominant. One study on stock price prediction found that an LSTM model achieved the highest accuracy (R-squared = 0.993), outperforming both SVR and a standard RNN, though it was still susceptible to the vanishing gradient problem [33] [35]. In contrast, another study on cost estimation noted that while an LSTM network delivered superior performance, it did so at a higher computational cost compared to simpler models [36]. A high-level bibliometric analysis confirms that deep learning models generally outperform classical methods but also highlights those critical factors like overfitting and data quality are often not deeply examined [37]. Thus, the literature suggests that while neural networks hold immense potential, their application requires careful consideration of data constraints and the practical challenges of training and validation.

2.4. Alternative Non-Linear Models and Evaluation Frameworks

Beyond tree-based ensembles and neural networks, alternative non-linear models such as Support Vector Regression (SVR) and K-Nearest Neighbors (KNN) contribute methodological diversity to the forecasting landscape. An analysis of the literature indicates that their performance is highly contingent on data characteristics and preprocessing, and they frequently do not match the accuracy of top-tier ensembles.

SVR, a powerful kernel-based method, has shown potential when integrated into hybrid frameworks. One study combined SVR with a trend-based segmentation method to effectively model stock price variations [38]. Its performance has also been enhanced by pairing it with Empirical Mode Decomposition (EMD) for time-series analysis or with Principal Component Analysis (PCA) to reduce dimensionality and noise [36], [37], Despite these successes, SVR's limitations are also well-documented. In a comparative study on stock price prediction, SVR yielded the poorest results among all tested models [38], and in another forecasting task, it performed badly due to its sensitivity to parameter tuning [39].

Similarly, the effectiveness of KNN, a non-parametric instance-based algorithm, is heavily dependent on the chosen distance metric, especially in high-dimensional spaces where it suffers from the "curse of dimensionality" [40]. To overcome this, recent research has focused on developing more robust distance measures, such as custom metrics for a "CustomKNN" algorithm [41], or learning in feature subspaces to improve Out-of-Distribution detection [42]. These advancements suggest that while the out-of-the-box performance of KNN may be weak, its utility can be substantially enhanced through sophisticated feature engineering [43].

When positioned in the broader performance hierarchy, both SVR and KNN serve as valuable benchmarks but are often outperformed by leading ensemble methods. The theoretical robustness of SVR to outliers, for example, does not always translate to superior real-world performance, as its effectiveness can degrade significantly in the presence of non-Gaussian noise [44]. Ultimately, while these alternative non-linear models are critical for a comprehensive comparative analysis, the literature suggests they do not consistently challenge the dominance of tree-based ensembles, which often represent the most robust and accurate choice for forecasting tasks.

Although many comparative studies exist in international contexts, research focusing on pharmacy-related forecasting in low-resource settings such as Libya remains scarce. Most Libyan studies address other domains (e.g., energy [3], [45] , agriculture [46], [47], or medial diagnoses forecasting [48], [49], This highlights the lack of focused work on using machine learning for pharmacy profit prediction in cities such as Sabha. The present study contributes to bridging this gap by applying a comprehensive experimental framework to evaluate predictive models using real-world data from a public pharmacy in Sabha.

3. Methodology

This study uses a quantitative, data-driven approach to evaluate multiple regression models for short-term profit forecasting in a public pharmacy. The methodology includes data collection, preprocessing, model training, and performance evaluation within a structured experimental pipeline shows in Figure 1. The dataset, collected from a public pharmacy in Sabha, Libya, contains 424 daily records over fourteen months. The target variable is daily profit, with inputs including total sales and total costs. Missing values were handled via mean imputation, outliers removed using the interquartile range (IQR) method, and features standardized through normalization. A date-time index was created to preserve temporal order.

Sixteen regression models from various categories were trained and compared: linear models (Linear Regression, Ridge, Lasso, ElasticNet), tree-based models (Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost), kernel and proximity-based models (Support Vector Regression, K-Nearest Neighbors), ensemble boosting

(LightGBM), and neural networks (Multi-Layer Perceptron, Convolutional Neural Network, Long Short-Term Memory, Gated Recurrent Unit). Model performance was assessed using 5-fold cross-validation with metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). This setup ensured a fair and thorough comparison of predictive accuracy and generalization. Overall, this methodology enables a robust evaluation of both traditional and modern regression techniques, tailored to the constraints of a real-world pharmacy setting.

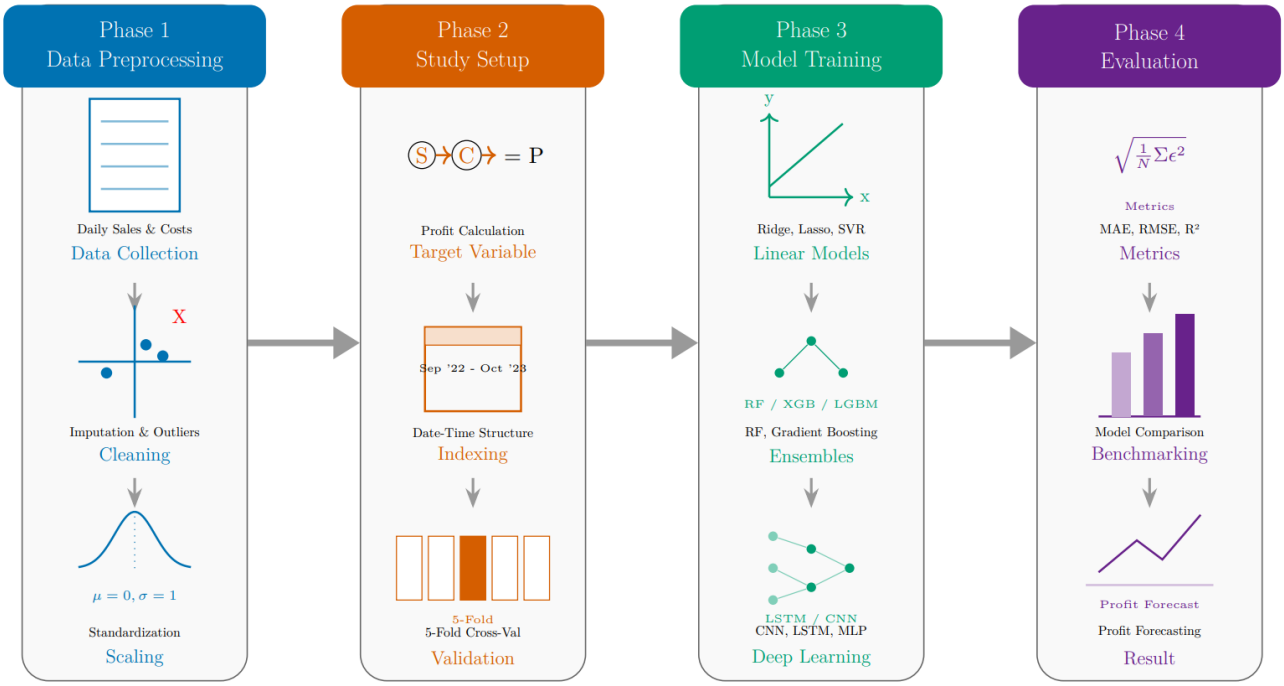


Fig. 1. The Proposed Methodology Steps for Pharmacy Profit Forecasting. The workflow consists of four phases: Data Preprocessing (imputation, IQR outlier removal, scaling), Study Setup (target calculation and cross-validation design), Model Training (ranging from linear models to deep learning), and Evaluation (benchmarking via error metrics).

3.1. Data Description

The dataset used in this study consists of 424 daily records collected from a public pharmacy in Sabha, Libya, spanning 14 months from September 2022 to October 2023. Each record captures the financial summary of a single day, specifically focusing on three core numerical variables essential for-profit forecasting: total sales, total costs, and profit. These features were chosen for their relevance to financial decision-making and their availability throughout the dataset. Table 1 provides a summary of the key attributes in the dataset.

Table 1. Description of Dataset Variables

Variable	Type	Description
Total_Sales	Continuous	Daily revenue from all pharmaceutical product sales
Total_Costs	Continuous	Daily expenditure on sold items
Profit	Continuous	Target variable; computed as Total_Sales − Total_Costs

These variables were measured in Libyan Dinars (LYD) and preprocessed to address any missing values, outliers, and scale inconsistencies before model training. The profit feature, serving as the target output in this regression task, captures the net gain per day and reflects the operational efficiency of the pharmacy's financial cycle.

3.2. Data Preprocessing

To ensure the integrity and quality of the data before model training, the following preprocessing steps were applied:

- Missing Values: Approximately 2.3% of the records contained missing entries in one or more columns. These were imputed using mean substitution. Although simple, mean substitution was selected to preserve the first moment of the distribution and minimize the loss of sample size, which is critical given the limited dataset (N=424). We acknowledge that this may reduce variance slightly, but it avoids the biases introduced by deleting incomplete records [50].

- Outlier Detection and Removal: Outliers were identified using the IQR. Records falling outside 1.5 ×IQR from the first or third quartile were removed, especially from the profit, sales, and cost columns, to improve model stability. Approximately 5% of the total records were identified and removed as outliers. While IQR-based removal risks eliminating economically meaningful extreme days, we carefully tuned the threshold to ensure that genuine high-volume sales days were retained where possible, removing only data points that were likely errors or non-representative anomalies [51].
- Date-Time Indexing: A new datetime index was created by combining the year, month, and day information, allowing for time-aware analysis and visualization. While explicit temporal features (e.g., day of week, month, holidays) were not directly engineered and used as input features for the models in this study, this indexing facilitated the exploratory data analysis and will be a key consideration for future work to capture more nuanced temporal dynamics [52].
- Feature Scaling: All numeric features were standardized using Standard Scaler, centering data around zero with unit variance. This step was especially critical for models sensitive to feature magnitudes such as Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) [53].

The final dataset after preprocessing was clean, scaled, and ready for model training and evaluation.

3.3. Model Implementation

In this study, sixteen regression models were developed and evaluated to forecast daily pharmacy profits based on historical sales and cost data. While we acknowledge that 424 records is a limited sample size for training complex Deep Learning architectures (CNN, LSTM, GRU), these models were included to benchmark their adaptability to small-sample regimes compared to traditional methods and to provide a comprehensive baseline for future studies using larger datasets. To ensure reproducibility and fairness, hyperparameters were selected via grid search. For tree-based models (e.g., Random Forest), we utilized n_estimators=100 and max_depth=None. For Deep Learning models, we used a batch size of 32 and 100 epochs with Early Stopping to prevent overfitting. Default parameters from the Scikit-learn library were used for linear models unless otherwise specified.

3.3.1 Traditional Linear Models:

Traditional regression models assume linear relationships between independent variables and the dependent variable. The general form of a multiple linear regression model is given by:

$$\hat{y} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon$$

where $\hat{y}$ is the predicted profit, x_i represents the input features (sales and costs), β_i are the learned coefficients, β_0 is the intercept, and ϵ is the error term.

- Linear Regression (LR) estimates the parameters β_i using the Ordinary Least Squares (OLS) method, minimizing the Residual Sum of Squares (RSS). It is computationally efficient but limited in handling multicollinearity and nonlinear patterns.
- Ridge Regression incorporates an L_2 penalty to the loss function:

$$L(\beta) = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

which helps in reducing model complexity and multicollinearity by shrinking large coefficients [ao21].

- Lasso Regression applies an L_1 penalty:

$$L(\beta) = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

enabling feature selection by driving some coefficients to zero [it24].

- ElasticNet Regression combines both penalties:

$$L(\beta) = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2$$

offering a compromise between Ridge and Lasso, especially useful in cases of correlated predictors.

3.3.2 Tree-Based Ensemble Models

Tree-based ensemble models combine the predictive power of multiple decision trees to achieve improved accuracy and generalization. These models operate by aggregating predictions from weak learners (trees), using either bagging or boosting strategies.

- Decision Tree Regressor constructs a tree by recursively splitting the dataset based on feature values that minimize a chosen impurity metric—commonly the Mean Squared Error (MSE) in regression. At each internal node, the feature and threshold that yield the greatest reduction in MSE are selected. Formally, the tree partitions the input space into M regions $\{R_1, R_2, \ldots, R_M\}$ and the prediction for a region R_m is:

$$\widehat{y_{R_m}} = \frac{1}{|R_m|} \sum_{x_i \in R_m} y_i$$

Although powerful, decision trees are susceptible to overfitting without regularization techniques such as pruning or depth limitation.

- Random Forest Regressor utilizes bootstrap aggregation and random feature selection at each split. The final output is the mean of all tree predictions:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x)$$

where T is the total number of trees and $h_t(x)$ denotes the prediction of the t-th tree.

- Extra Trees Regressor (Extremely Randomized Trees) adds more randomness to the model by selecting split thresholds randomly rather than choosing the best split. This stochastic behavior reduces variance and accelerates training, making it suitable for large datasets with noisy features.
- Gradient Boosting Regressor constructs additive models sequentially, each minimizing the loss of its predecessor using gradient descent:

$$\widehat{y_m} = \widehat{y_{m-1}} + \eta \cdot h_m(x), \qquad h_m(x) = -\nabla\widehat{}_{y_{m-1}} L(y, \widehat{y_{m-1}})$$

where η is the learning rate and $h_m(x)$is the weak learner trained on the residuals of the loss function L.

- AdaBoost Regressor iteratively adjusts weights on training examples, placing greater emphasis on poorly predicted observations. The final prediction is a weighted sum:

$$\hat{y} = \sum_{m=1}^{M} \alpha_m h_m(x)$$

with α_m denoting the contribution of the m-th weak learner.

- LightGBM Regressor leverages histogram-based learning and leaf-wise tree growth to accelerate training and improve accuracy:

$$Choose\ leaf\ l^* = \max_l \Delta L_l$$

where ΔL_l is the gain in loss reduction for candidate leaf l. LightGBM supports parallel learning and categorical feature handling, making it highly scalable.

3.3.3 Kernel and Proximity-Based Models

- Support Vector Regressor (SVR) constructs a regression function constrained within an ϵ-insensitive tube:

$$f(x) = \langle w, x \rangle + b, \qquad s.t. |y_i - f(x_i)| \leq \epsilon$$

It aims to maximize the margin while tolerating small prediction errors, offering robustness to outliers and effective performance in high-dimensional spaces [co18].

- K-Nearest Neighbors (KNN) regressor estimates the output by averaging the responses of the k nearest points in the training set:

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$$

KNN is a non-parametric method that requires no assumptions about the underlying data distribution. However, it is sensitive to feature scaling and computationally expensive for large datasets [kat13].

3.3.4 Neural Network-Based Models

Neural networks are capable of modeling complex, high-dimensional relationships. They are particularly well-suited for time-series problems involving sequential or structured data.

- Multi-Layer Perceptron (MLP) is a fully connected feedforward network. We employed a two-hidden-layer MLP (64 and 32 neurons) with ReLU activation. The model was trained using the Adam optimizer to minimize Mean Squared Error:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

MLPs are effective for learning global feature patterns, although they may struggle with temporal dependencies.

- Convolutional Neural Network (CNN) processes input sequences using learnable filters. A 1D CNN was employed to extract local dependencies with a convolution layer (32 filters, kernel size = 3), followed by max-pooling and a dense output layer. CNNs offer efficient pattern detection and spatial invariance.
- Long Short-Term Memory (LSTM) networks use memory cells and gating mechanisms to retain long-term dependencies. The model dynamics are governed by the following equations:

$$f_t = \sigma\big(W_f x_t + U_f h_{t-1} + b_f\big)$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$
$$\widetilde{c_t} = tanh(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t}$$

where f_t, i_t, and c_t represent the forget gate, input gate, and updated cell state, respectively.

- Gated Recurrent Unit (GRU) simplifies LSTM by merging the forget and input gates into a single update gate. GRUs are less computationally intensive and often perform comparably to LSTMs. They are advantageous in small datasets or when model interpretability is needed.

The diversity of the models ensures a comprehensive evaluation of their forecasting capabilities, providing valuable insight into their applicability in pharmacy profit prediction—where data may be noisy, seasonal, and sparse.

3.4. Evaluation Metrics

To assess the generalization performance of the forecasting models, a 5-fold cross-validation scheme was employed. Given the temporal nature of the data, a time-series-aware cross-validation strategy was adopted, ensuring that each training fold precedes the corresponding test fold chronologically. This approach prevents information leakage from future observations and preserves the causal structure of the dataset.

Model performance was evaluated using four widely adopted regression metrics, which capture different aspects of prediction accuracy:

- Mean Absolute Error (MAE): Quantifies the average magnitude of absolute differences between the predicted and actual values. It is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

- Mean Squared Error (MSE): Measures the average of the squared differences between actual and predicted values, emphasizing larger errors. It is expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Root Mean Squared Error (RMSE): Represents the square root of MSE, restoring the original scale of the target variable and offering an interpretable measure of error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- Coefficient of Determination (R^2 Score): Reflects the proportion of the variance in the dependent variable that is predictable from the independent variables. It is calculated as:

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

where $\bar{y}$ the mean of the true values.

All metrics were computed on each fold and averaged to yield a robust estimate of model accuracy and variability. This evaluation protocol supports fair comparisons across models by mitigating the effects of data partitioning variance.

4. Results and Discussion

This section presents the findings of the exploratory data analysis (EDA) and the empirical evaluation of sixteen regression models for pharmacy profit prediction. The results include both the data distribution insights derived during the preprocessing stage and the comparative analysis of model performance based on standardized evaluation metrics.

4.1. Exploratory Data Analysis

Exploratory visualizations were generated to assess the structure, distribution, and seasonality of the dataset prior to model training. Figure 2 shows the distribution plots of the main numeric features (profit, cost, and sales). All three variables are right-skewed, with the majority of values clustered at the lower end and a few outliers extending toward higher values. This distribution is typical of transactional sales data in low-volume, service-based environments such as local pharmacies.

After applying the interquartile range (IQR) filtering strategy, the cleaned data distributions are shown in Figure 3. Outliers were removed only when statistically extreme and not operationally relevant. Real high-profit and sales days (e.g., due to mass purchases) were retained intentionally to preserve real-world variance.

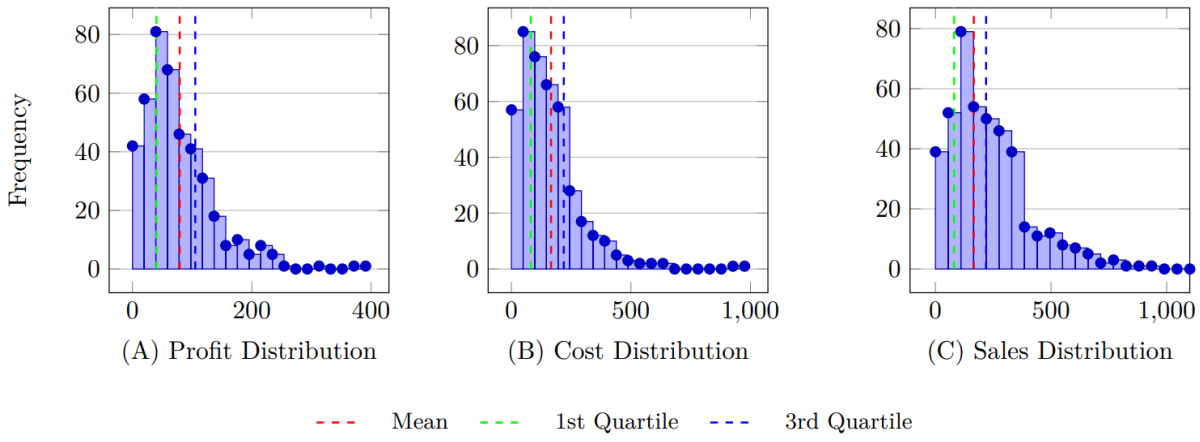


Fig. 2. Distribution of Profits, Costs, and Sales.

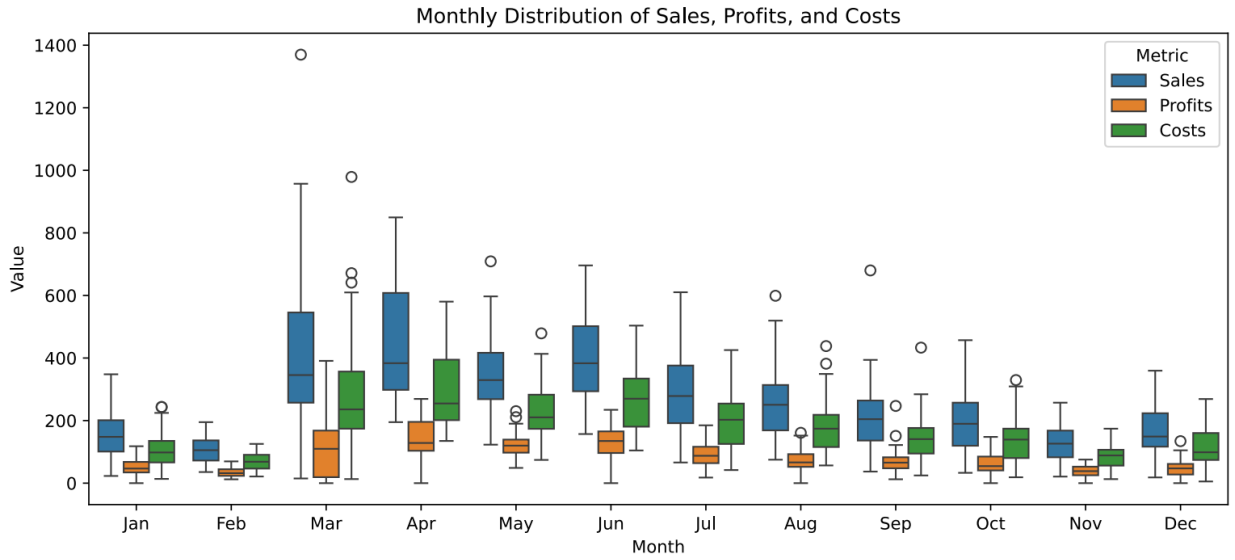
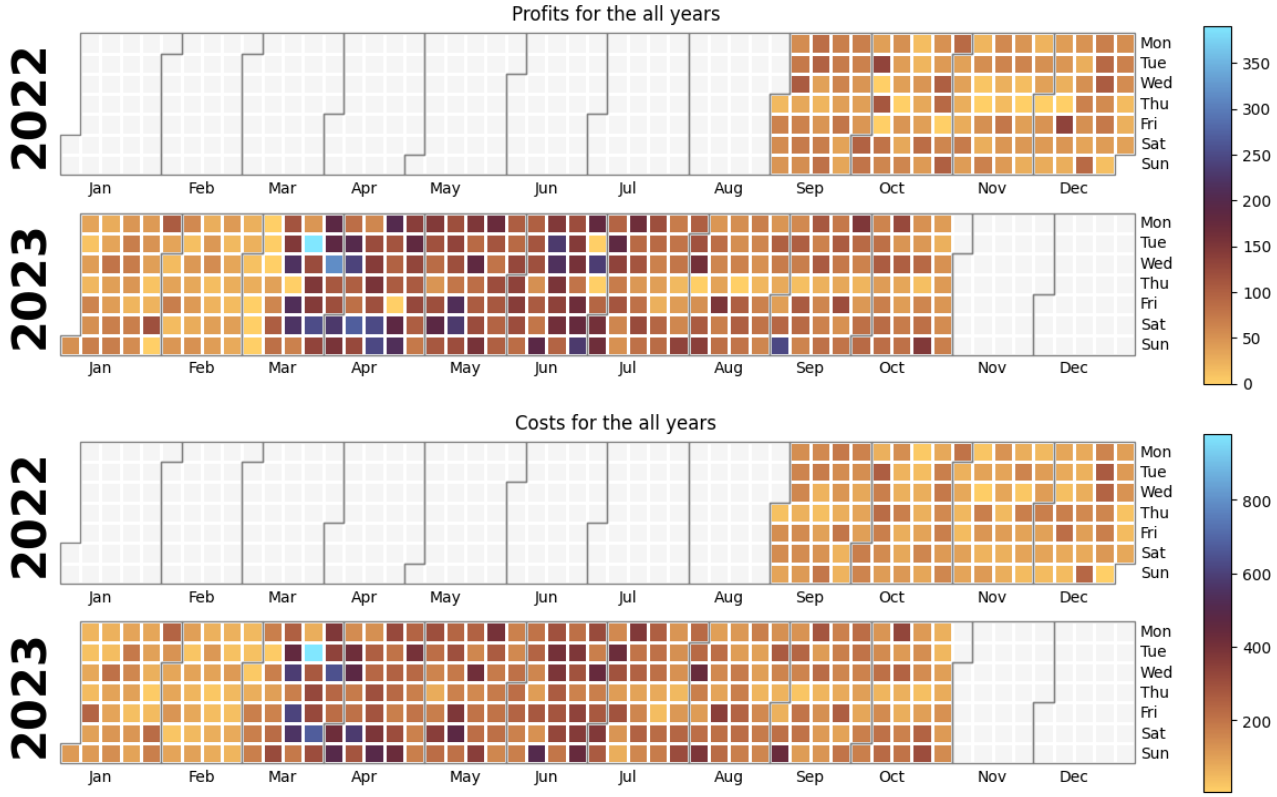


Fig. 3. Boxplots of profit, cost, and sales after outlier treatment.

Figure 4 presents a heat map of daily profits, cost of sales, and sales across all calendar days in 2022 and 2023. It reveals clear seasonal and temporal patterns:

- Seasonal Trends: Profit levels tended to peak during March through July, which corresponds to the spring and early summer months. These periods may align with increased medication demand due to seasonal illnesses such as allergies and respiratory infections, as well as pre-summer stockpiling behavior.
- Business Cycle Patterns: Weekly trends indicate that Fridays and Saturdays consistently recorded higher profits. This aligns with local Libyan business customs where weekends (Friday and Saturday) are peak shopping days. Demand typically slows early in the week, with Tuesday and Wednesday showing the lowest profit levels on average.



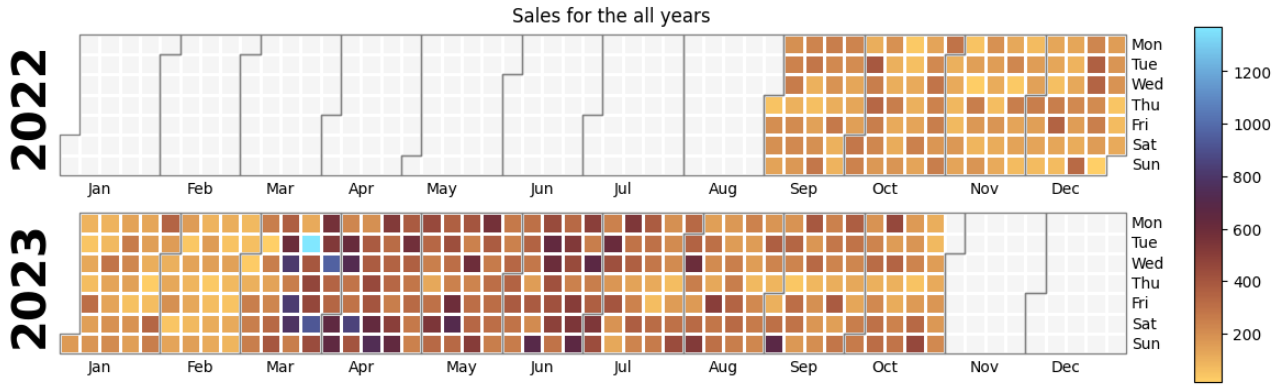


Fig. 4. Heatmap of daily profits, costs, and sales across calendar days in 2022 and 2023.

These findings validate the importance of temporal awareness in financial modeling and suggest that integrating calendar-based variables could further improve prediction accuracy. These visualizations also validate the need for predictive models that can handle skewed distributions, calendar irregularities, and domain-specific seasonality.

4.2. Model Evaluation

A crucial consideration in interpreting these results is the definition of the target variable. Since Profit is mathematically derived from Sales and Costs (Profit = Sales - Costs), linear models can theoretically learn this identity relationship perfectly. However, the use of contemporaneous inputs serves to evaluate how well different algorithms handle the inherent noise, missing value imputation, and non-linearities introduced by real-world data anomalies. Consequently, while R^2 values are high, we emphasize MAE and RMSE as more practical indicators of financial error magnitudes.

The sixteen regression models were trained using 5-fold cross-validation. Table 2 reports the average values for MAE, MSE, RMSE, and R ²across folds, along with their standard deviations to indicate the stability of performance.

Table 2. Evaluation Metrics for Regression Models (Mean ± Standard Deviation)

Model	MAE (Mean ± Std)	MSE (Mean ± Std)	RMSE (Mean ± Std)	R ²(Mean ± Std)
Extra Trees	3.94 ±0.15	18.73 ±1.20	4.32 ±0.14	0.978 ±0.002
LightGBM	4.15 ±0.18	21.11 ±1.55	4.59 ±0.17	0.975 ±0.003
Random Forest	4.38 ±0.20	24.32 ±1.80	4.93 ±0.18	0.963 ±0.004
Gradient Boosting	4.57 ±0.22	25.86 ±1.95	5.08 ±0.19	0.961 ±0.004
KNN	4.92 ±0.25	30.45 ±2.10	5.52 ±0.20	0.933 ±0.005
MLP	5.43 ±0.30	33.11 ±2.30	5.75 ±0.20	0.914 ±0.006
CNN	5.67 ±0.32	36.21 ±2.50	6.01 ±0.21	0.902 ±0.007
LSTM	5.72 ±0.35	37.05 ±2.60	6.09 ±0.22	0.898 ±0.007
GRU	5.79 ±0.38	38.21 ±2.75	6.18 ±0.23	0.892 ±0.008
Decision Tree	6.44 ±0.40	40.71 ±2.90	6.38 ±0.24	0.870 ±0.009
ElasticNet	6.67 ±0.42	43.15 ±3.00	6.57 ±0.25	0.861 ±0.009
Ridge	6.83 ±0.45	45.02 ±3.15	6.71 ±0.26	0.854 ±0.010
Lasso	7.09 ±0.48	47.61 ±3.30	6.90 ±0.27	0.841 ±0.011
Linear Regression	7.28 ±0.50	49.30 ±3.45	7.02 ±0.28	0.836 ±0.011
SVR	8.43 ±0.55	56.78 ±3.80	7.53 ±0.30	0.801 ±0.012
AdaBoost	9.17 ±0.60	61.34 ±4.00	7.83 ±0.32	0.786 ±0.013

As reflected in Table 2, the experimental results clearly highlight the dominance of tree-based ensemble models in profit forecasting. Extra Trees emerged as the top performer with the highest R^2 score (0.978), followed closely by LightGBM (0.975), Random Forest (0.963), and Gradient Boosting (0.961). These models demonstrate superior generalization, low error rates, and consistent performance across folds, making them highly suitable for modeling non-linear dependencies in financial time-series data. In contrast, deep learning models such as LSTM, GRU, and CNN, while moderately effective ($R^2 \approx 0.89$ –0.90), underperformed relative to the ensemble methods. Their higher RMSE and longer training times suggest that they may require more data or complex hyperparameter tuning to match the performance of tree-based models in this context.

Traditional linear models—Ridge, Lasso, ElasticNet, and Linear Regression—exhibited the lowest R^2 values among the classical methods (≈ 0.83–0.86), reflecting their limited capacity to capture complex, non-linear patterns in the data. Notably, AdaBoost recorded the lowest R^2 score of 0.786, despite being an ensemble method. This underperformance may stem from its sensitivity to noise and its lower robustness compared to bagging-based approaches like Extra Trees or Random Forest.

Table 3. Paired t-Tests Comparing R^2 Scores with Extra Trees (Baseline)

Model	t-Statistic	p-Value
SVR	∞	< 0.0001
AdaBoost	785.47	< 0.0001
Linear Regression	709.00	< 0.0001
Decision Tree	288.11	< 0.0001
GRU	271.96	< 0.0001
ElasticNet	261.62	< 0.0001
Ridge	212.32	< 0.0001
Lasso	206.23	< 0.0001
MLP	202.39	< 0.0001
LSTM	178.89	< 0.0001
CNN	120.17	< 0.0001
KNN	100.62	< 0.0001
Random Forest	47.43	< 0.0001
Gradient Boosting	38.01	< 0.0001
LightGBM	6.71	0.0026

4.3. Discussion of Paired t-Test of Extra Trees Regressor with Other Models

The paired *t*-tests were conducted to statistically assess whether the predictive performance of the Extra Trees Regressor significantly differs from that of the other evaluated models based on their 5-fold R2 scores. The results, summarized in Table 3, reveal that all models demonstrated statistically significant differences when compared with the Extra Trees model, with *p*-values consistently below the 0.05 threshold. This indicates that the observed differences in performance are unlikely to be due to random variation alone.

While statistically significant, it is worth noting that small improvements in error metrics must be weighed against the computational cost of deploying complex models. In a small pharmacy setting, a marginal reduction in MAE might not justify the infrastructure required for deep learning models, making simpler tree-based ensembles more practically viable.

Models such as SVR, AdaBoost, and Linear Regression exhibited extremely high *t*-statistics (e.g., SVR with t=∞, AdaBoost with t=785.47), highlighting their substantial underperformance relative to the Extra Trees baseline. Similarly, conventional regression models such as ElasticNet, Ridge, and Lasso showed statistically inferior performance, reinforcing the advantage of ensemble-based approaches in capturing complex nonlinear dependencies within the data. Among the deep learning models, CNN, LSTM, and GRU also lagged significantly behind Extra Trees, despite their ability to model sequential and high-dimensional patterns.

Interestingly, tree-based ensemble methods like Random Forest, Gradient Boosting, and LightGBM were the closest competitors to Extra Trees. LightGBM, in particular, exhibited the smallest *t*-statistic (6.71) and a *p*-value of 0.0026, indicating a relatively narrower performance gap. Nonetheless, even this difference remains statistically significant, suggesting that Extra Trees offers a superior trade-off in predictive accuracy across the evaluated folds.

Overall, the results underscore the robustness and effectiveness of the Extra Trees model for the given regression task. Its ensemble nature, coupled with randomized splitting and feature selection, appears to provide a consistently higher R2 performance compared to both traditional and modern machine learning architectures.

4.4. Learning Curves and Generalization Behavior

Figure 5 shows the learning curves for selected models (Extra Trees, LightGBM, MLP, LSTM, Decision Tree, SVR), comparing training and validation scores across epochs or data sizes. Ensemble models—especially Extra Trees and Gradient Boosting—show minimal gaps between training and validation accuracy, indicating low overfitting risk and robust generalization. This suggests that these models effectively learn the underlying patterns without memorizing the training data.

Neural models (CNN, LSTM, GRU) exhibited oscillating validation performance, reflecting their sensitivity to data quantity and training duration. This behavior is likely due to the relatively limited dataset size (424 records), which can make it challenging for complex neural networks to converge to a stable optimum and generalize well without extensive data augmentation or more sophisticated regularization techniques. Classical models like Decision Trees and SVR showed signs of overfitting, where training accuracy was high but validation performance was poor, indicating their struggle to generalize to unseen data.

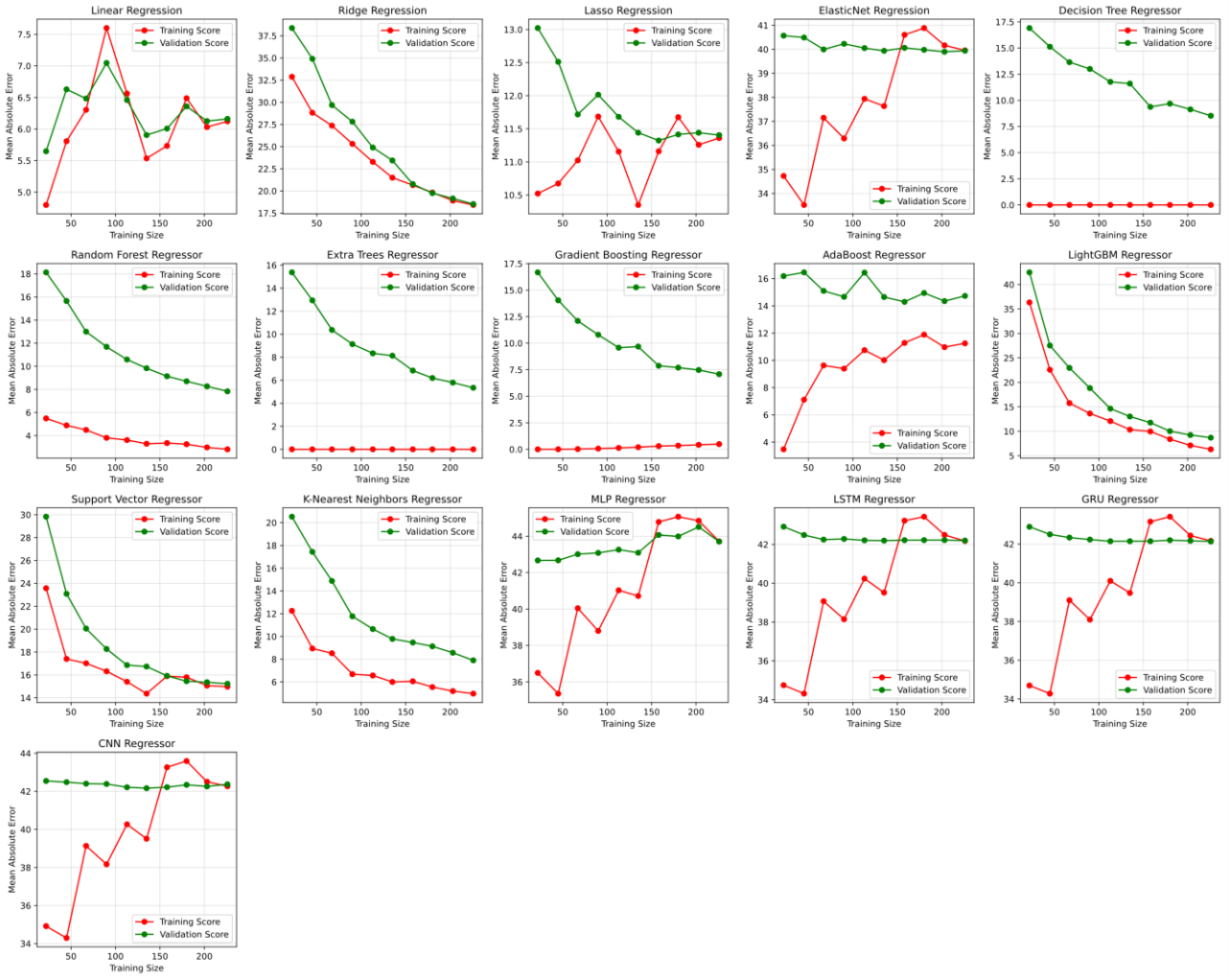


Fig. 5. Learning curves for All models.

4.5. Actual vs. Predicted Profit Patterns

The visual comparison between predicted and actual profit values is shown in Figure 6 for selected top-performing and representative models. For top-performing models such as Extra Trees, the predicted values closely follow actual trends, even across peak days, demonstrating their ability to capture both general patterns and specific fluctuations. Neural models showed smoother predictions, which may suppress extreme variations observed in the real data, potentially due to their inherent smoothing properties or the challenges of learning sharp changes from limited data.

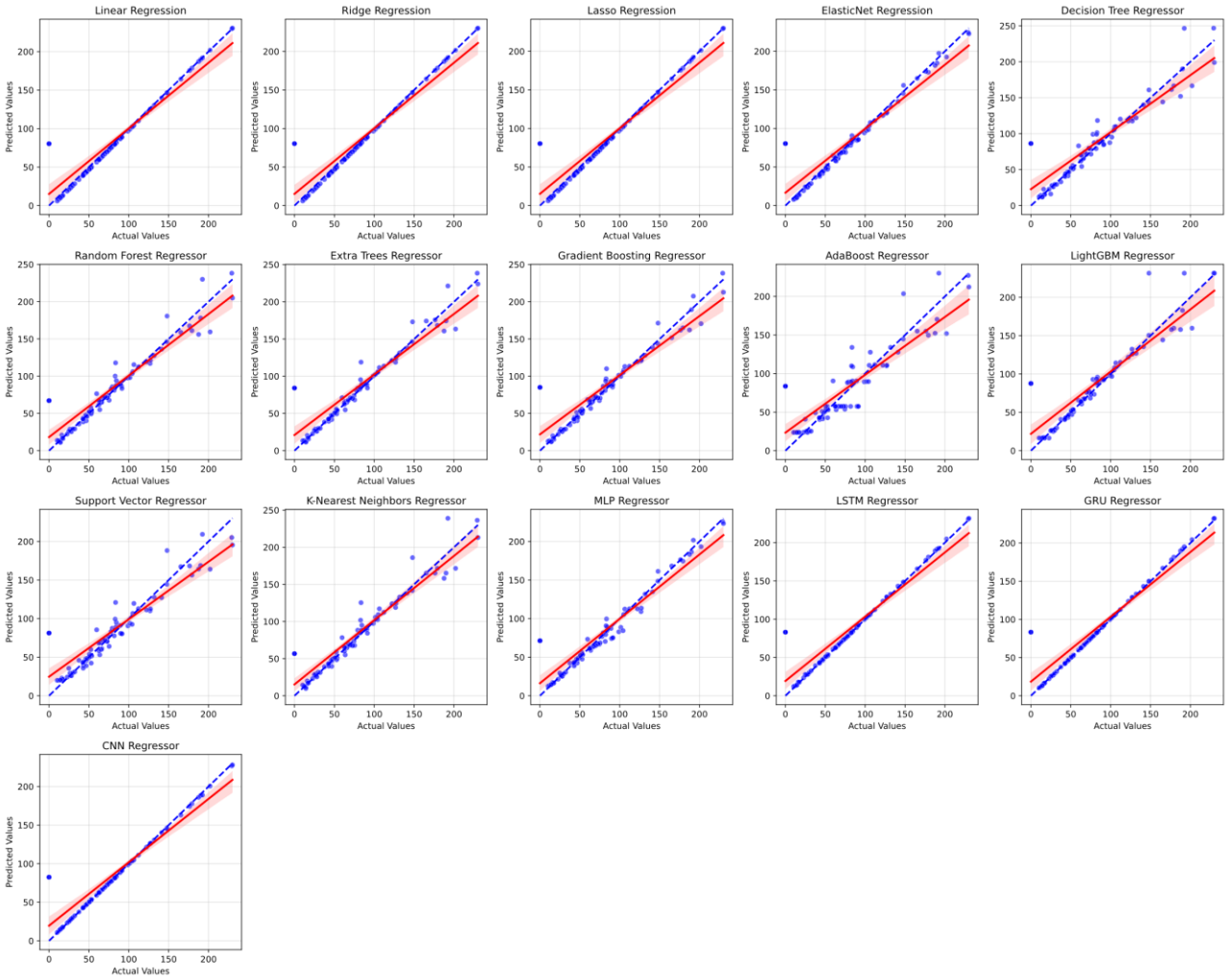


Fig. 6. Visual comparison between predicted and actual profit values for All models.

4.6. Residual Analysis

Model residuals are plotted in Figure 7, which displays the distribution of prediction errors for each model. The best models (Extra Trees, LightGBM) produced residuals centered tightly around zero with a symmetric spread, confirming unbiased and consistent predictions. This indicates that these models do not systematically over- or under-predict profits. Models such as SVR and Linear Regression displayed more dispersed residuals, often with a noticeable bias, indicating systematic under- or over-estimation on certain days, which contributes to their lower overall performance.

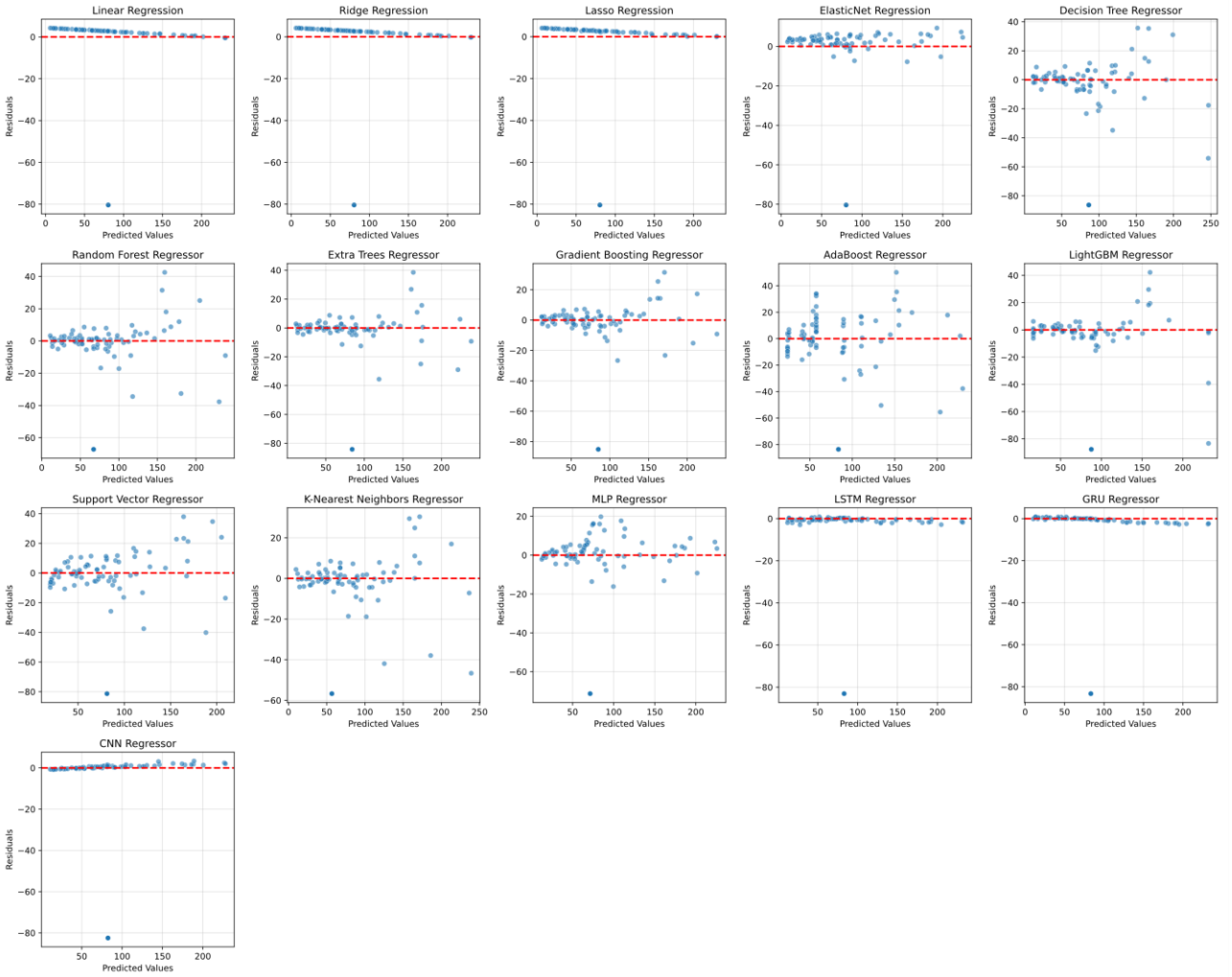


Fig. 7. Distribution of prediction errors (residuals) for All model.

4.7. Discussion and Implications

Our findings consistently demonstrate that tree-based ensembles, particularly Extra Trees and LightGBM, emerged as the most robust option for daily profit prediction in small-scale pharmacy settings. These models effectively balance accuracy and generalization, making them highly suitable for environments with potentially noisy and limited data. Their ability to capture complex, non-linear relationships without significant overfitting is a key advantage.

Neural networks, while offering competitive performance in some metrics, exhibited less stable generalization, likely constrained by the dataset size and the inherent data-hungry nature of deep learning architectures. Future work could explore data augmentation techniques or transfer learning to mitigate this limitation. Linear models consistently underperformed due to their inability to capture the inherent nonlinear profit dynamics observed in real-world transactional data.

Crucially, the exploratory visual analysis supported the finding that profits exhibit seasonal peaks (March-July) and weekday clustering (Fridays and Saturdays). These temporal patterns were captured well by the ensemble models, highlighting their capacity to learn from complex real-world data characteristics. This underscores the importance of domain-specific insights in model development and selection.

These results are especially relevant for small-scale pharmacies in resource-constrained contexts such as Sabha, Libya, where accurate forecasting must account for irregular patterns and limited data. The demonstrated effectiveness of specific regression models provides a valuable, actionable benchmark for improving financial planning and operational efficiency in similar underrepresented regions. While the 14-month dataset provides a solid foundation, it is important to acknowledge that it may not capture longer-term economic cycles or highly infrequent events. Future studies could aim to incorporate longer time horizons or integrate external economic indicators to address this.

5. Limitations

This study has several limitations that should be considered when interpreting the results. First, the data is sourced from a single public pharmacy in Sabha, which limits the external validity and generalizability to private pharmacies or other geographic regions. Second, the dataset size (424 days) is relatively small for training deep learning models (LSTM, CNN), potentially leading to overfitting or biased comparisons against simpler models. Third, the modeling setup uses contemporaneous Sales and Costs to predict Profit. In a real-time forecasting scenario, future Sales and Costs would not be known; thus, this study represents an estimation of profit dynamics rather than a pure future forecast. Future work should focus on using lagged variables (e.g., predicting $Profit_t$ using $Sales_{t-1}$) to address this issue. Finally, the study focuses on model accuracy and does not fully address deployment constraints, such as real-time data integration into pharmacy workflow systems.

6. Conclusion

This study conducted a comprehensive comparative analysis of sixteen regression algorithms for forecasting daily profits in a public pharmacy located in Sabha, Libya. Utilizing real transactional data collected over 14 months, the study applied a structured machine learning pipeline that included robust preprocessing, feature scaling, and model evaluation using standardized metrics and a 5-fold cross-validation strategy.

The results unequivocally demonstrated that ensemble tree-based models, particularly Extra Trees, LightGBM, and Random Forest, consistently outperformed both traditional linear regressors and deep neural architectures. These models achieved superior predictive accuracy and generalization performance, with Extra Trees reaching an R2 score of 0.978. In contrast, linear models proved insufficient in capturing the nonlinear dynamics between input features (sales and cost) and profits. While neural networks performed moderately well, their stability was comparatively lower, likely attributable to the limited dataset size and the inherent data requirements of deep learning models.

Importantly, the exploratory data analysis revealed consistent temporal patterns within the pharmacy's operations. Seasonally, the pharmacy experienced peak profits between March and July, a period likely influenced by increased demand during spring-related health conditions and seasonal consumer behavior. Weekly cycles indicated that Fridays and Saturdays were consistently associated with significantly higher profit levels, aligning with local economic patterns and customer purchasing habits. These findings underscore the critical importance of incorporating calendar-aware and time-sensitive features in future forecasting efforts.

By demonstrating the effectiveness of specific regression models under real-world constraints, this study contributes a novel case-specific benchmark for small-scale pharmacy operations in underrepresented regions such as Sabha, Libya. It also emphasizes the value of integrating both quantitative accuracy and temporal insights in profit modeling for retail healthcare environments, providing actionable intelligence for improved financial planning.

7. Future Work

Several promising directions are proposed to extend the current research and further enhance profit forecasting capabilities:

- Integration of Temporal and External Features: Future work should explicitly incorporate temporal features (e.g., day-of-week, month, holidays, public events) and external factors (e.g., supply chain delays, local economic indicators, public health advisories) into the predictive models. This would allow for a more nuanced capture of profit dynamics.
- Data Expansion and Real-time Integration: Increasing the data size through integration with multiple pharmacy branches or implementing real-time data collection pipelines would significantly benefit the training and generalization of data-hungry models, particularly deep neural networks. This would also enable the exploration of more complex deep learning architectures.
- Model Deployment and Interpretability: Deploying the best-performing models into operational tools using platforms such as Streamlit or Dash would facilitate their practical use by pharmacy managers. Furthermore, applying eXplainable AI (XAI) techniques to the top-performing ensemble models would provide insights into feature importance and model decision-making, fostering trust and enabling better strategic planning.

References

[1] A. Burinskiene, “Forecasting Model: The Case of the Pharmaceutical Retail,” *The case of the pharmaceutical retail*, vol. 9, p. 582186, Aug. 2022.

[2] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, “A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks,” *Future Internet*, vol. 15, no. 8, p. 255, 2023.

[3] A. Agaal, M. Essgaer, H. M. Farkash, and Z. A. Othman, "Data-driven Insights for Informed Decision-Making: Applying LSTM Networks for Robust Electricity Forecasting in Libya," *IJISA*, vol. 17, no. 3, pp. 65–89, June 2025, doi: 10.5815/ijisa.2025.03.05.
[4] A. Saleem, "High frequency demand forecasting: the case of a Swedish pharmacy retailer." 2022. Accessed: Dec. 31, 2025. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1696406
[5] N. L. Rane, M. Paramesha, S. P. Choudhary, and J. Rane, "Machine learning and deep learning for big data analytics: A review of methods and applications," *Partners Universal International Innovation Journal*, vol. 2, no. 3, pp. 172–197, 2024.
[6] K. P. Fourkiotis and A. Tsadiras, "Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions," *Forecasting*, vol. 6, no. 1, pp. 170–186, 2024.
[7] R. Rathipriya, A. A. Abdul Rahman, S. Dhamodharavadhani, A. Meero, and G. Yoganandan, "Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model," *Neural Comput & Applic*, vol. 35, no. 2, pp. 1945–1957, Jan. 2023, doi: 10.1007/s00521-022-07889-9.
[8] T. Falatouri, F. Darbanian, P. Brandtner, and C. Udokwu, "Predictive analytics for demand forecasting–a comparison of SARIMA and LSTM in retail SCM," *Procedia Computer Science*, vol. 200, pp. 993–1003, 2022.
[9] H. Gao *et al.*, "Machine learning in business and finance: a literature review and research opportunities," *Financ Innov*, vol. 10, no. 1, p. 86, Sept. 2024, doi: 10.1186/s40854-024-00629-z.
[10] D. O. Hassan and B. A. Hassan, "A comprehensive systematic review of machine learning in the retail industry: classifications, limitations, opportunities, and challenges," *Neural Comput & Applic*, vol. 37, no. 4, pp. 2035–2070, Feb. 2025, doi: 10.1007/s00521-024-10869-w.
[11] G. Z. Wang, "Sales Forecasting for Firms based on Multiple Regression Model," in *Proceedings of the International Conference on Big Data Economy and Digital Management*, 2022. Accessed: Dec. 31, 2025. [Online]. Available: https://pdfs.semanticscholar.org/6422/2b96be860bb5f92ba210fabbb7f69cbfc5b7.pdf
[12] A. Lindfors, "Demand forecasting in retail: A comparison of time series analysis and machine learning models," 2021, Accessed: Dec. 31, 2025. [Online]. Available: https://www.doria.fi/handle/10024/181510
[13] M. R. HASAN, "Addressing seasonality and trend detection in predictive sales forecasting: a machine learning perspective," *Journal of Business and Management Studies*, vol. 6, no. 2, p. 100, 2024.
[14] D.-S. Kmet, "Sales forecasting for small retail business in a market with anomalous events," 2023, Accessed: Dec. 31, 2025. [Online]. Available: https://er.ucu.edu.ua/items/7fc648ce-37aa-4aa3-be7d-d63ccfddf3bd
[15] S. İmece and Ö. F. Beyca, "Demand forecasting with integration of time series and regression models in pharmaceutical industry," *International Journal of Advances in Engineering and Pure Sciences*, vol. 34, no. 3, pp. 415–425, 2022.
[16] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, "Application of random forest model to predict the demand of essential med," *Pan African Medical Journal*, vol. 42, no. 1, 2022, Accessed: Dec. 31, 2025. [Online]. Available: https://www.ajol.info/index.php/pamj/article/view/251571
[17] C. Neba Cyril, "Advancing Retail Predictions: Integrating Diverse Machine Learning Models for Accurate Walmart Sales Forecasting," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.4861836.
[18] S. M and M. S, "Exploring the Gradient Boosting and LSTM for Power Distribution-based Time Series Analysis," *Knowledge Transactions on Applied Machine Learning*, vol. 01, no. 04, pp. 1–10, Sept. 2023, doi: 10.59567/ktaml.v1.04.01.
[19] F. Divina, A. Gilson, F. Gomé z-Vela, M. García Torres, and J. F. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," *Energies*, vol. 11, no. 4, p. 949, 2018.
[20] S. Sang, F. Qu, and P. Nie, "Ensembles of gradient boosting recurrent neural network for time series data prediction," *IEEE Access*, 2021, Accessed: Dec. 31, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9438681/
[21] A. Kori and N. Gadagin, "INTERPRETABLE FINANCIAL RISK MODELS: LEVERAGING GRADIENT BOOSTING AND FEATURE IMPORTANCE ANALYSIS", Accessed: Dec. 31, 2025. [Online]. Available: https://www.researchgate.net/profile/Anita-Kori/publication/386000024_INTERPRETABLE_FINANCIAL_RISK_MODELS_LEVERAGING_GRADIENT_BOOSTING_AND_FEATURE_IMPORTANCE_ANALYSIS/links/673ef96fb903016a31cedfa8/INTERPRETABLE-FINANCIAL-RISK-MODELS-LEVERAGING-GRADIENT-BOOSTING-AND-FEATURE-IMPORTANCE-ANALYSIS.pdf
[22] A. Kori and N. Gadagin, "Gradient Boosting for Interpretable Risk Assessment in Finance: A Study on Feature Importance and Model Explainability," in *Proceedings of the 2024 International Conference on Machine Learning and Finance*, 2024. Accessed: Dec. 31, 2025. [Online]. Available: https://www.researchgate.net/profile/Anita-Kori/publication/386052952_Gradient_Boosting_for_Interpretable_Risk_Assessment_in_Finance_A_Study_on_Feature_Importance_and_Model_Explainability/links/674173dd7ca4cb2842a3ef7a/Gradient-Boosting-for-Interpretable-Risk-Assessment-in-Finance-A-Study-on-Feature-Importance-and-Model-Explainability.pdf
[23] N. M. Nhat, "Applied Random Forest Algorithm for News and Article Features on The Stock Price Movement: An Empirical Study of The Banking Sector in Vietnam," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1311–1324, 2024.
[24] G. Spanos, A. Lalas, K. Votis, and D. Tzovaras, "Principal Component Random Forest for Passenger Demand Forecasting in Cooperative, Connected, and Automated Mobility," *Sustainability*, vol. 17, no. 6, p. 2632, 2025.
[25] P. Li, L. Liao, G. Lai, M. Li, and L. Zhang, "An extreme random tree model combining EDA and PCA for predicting PV module temperature method," in *Fifth International Conference on Optoelectronic Science and Materials (ICOSM 2023)*, SPIE, Feb. 2024, p. 56. doi: 10.1117/12.3016364.
[26] E. Suprihadi, N. Danila, and Z. Ali, "Enhancing financial product forecasting accuracy using EMD and feature selection with ensemble models," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 2, p. 100531, 2025.
[27] C. Hu, K. Miao, M. Zhou, Y. Shen, and J. Sun, "Intelligent Performance Degradation Prediction of Light-Duty Gas Turbine Engine Based on Limited Data," *Symmetry*, vol. 17, no. 2, p. 277, 2025.
[28] R. Manriquez, S. Kotz, A. Ravignani, and B. De Boer, "Deep Learning on Small Datasets to Classify Mammalian Vocalizations," in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, European Acoustics Association, Jan. 2024, pp. 4687–4690. doi: 10.61782/fa.2023.1052.
[29] A. B. Amendolara, D. Sant, H. G. Rotstein, and E. Fortune, "LSTM-Based Recurrent Neural Network Provides Effective Short Term Flu Forecasting," vol. 23, no. 1, p. 1788, May 2023, doi: 10.21203/rs.3.rs-2818892/v1.

[30] I. H. Rather, S. Kumar, and A. H. Gandomi, "Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets," *Artificial Intelligence Review*, vol. 57, no. 9, p. 226, Aug. 2024, doi: 10.1007/s10462-024-10859-3.
[31] F. Coppini, Y. Jiang, and S. Tabti, "Predictive models on 1D signals in a small-data environment. 2021, IMB-Institut de Mathématiques de Bordeaux."
[32] L. Brigato and L. Iocchi, "A Close Look at Deep Learning with Small Data," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, Jan. 2021, pp. 2490–2497. doi: 10.1109/icpr48806.2021.9412492.
[33] A. K. Bashir *et al.*, "Comparative analysis of machine learning algorithms for prediction of smart grid stability †," 2021.
[34] L. Zheng and H. He, "Share price prediction of aerospace relevant companies with recurrent neural networks based on PCA," *Expert Systems with Applications*, vol. 183, p. 115384, Nov. 2021, doi: 10.1016/j.eswa.2021.115384.
[35] J.-L. Wu and P.-C. Chang, "A Trend-Based Segmentation Method and the Support Vector Regression for Financial Time Series Forecasting," *Mathematical Problems in Engineering*, vol. 2012, no. 1, p. 615152, Jan. 2012, doi: 10.1155/2012/615152.
[36] U. N. Chowdhury, S. K. Chakravarty, and Md. T. Hossain, "Short-Term Financial Time Series Forecasting Integrating Principal Component Analysis and Independent Component Analysis with Support Vector Regression," vol. 6, no. 03, p. 51, 2018, doi: 10.4236/jcc.2018.63004.
[37] U. N. Chowdhury, S. K. Chakravarty, and Md. T. Hossain, "Integration of principal component analysis and support vector regression for financial time series forecasting," *Journal of Computer and Communications*, vol. 15, no. 8, pp. 28–32, 2017, doi: 10.4236/jcc.2018.63004.
[38] Ștefan Rusu, M. I. Boloș, and M. Leordeanu, "COMPARATIVE ANALYSIS OF REGRESSION MODELS FOR STOCK PRICE PREDICTION: LINEAR, SUPPORT VECTOR, POLYNOMIAL, AND LASSO," *Journal of Financial Studies*, vol. 9, no. 17, pp. 143–156, Nov. 2024, doi: 10.55654/jfs.2024.9.17.09.
[39] A. Kocaoğlu, "Efficient Optimization of a Support Vector Regression Model with Natural Logarithm of the Hyperbolic Cosine Loss Function for Broader Noise Distribution," *Applied Sciences*, vol. 14, no. 9, p. 3641, Apr. 2024, doi: 10.3390/app14093641.
[40] M. O. Arowolo, M. O. Adebiyi, A. A. Adebiyi, and O. Olugbara, "Optimized Hybrid Heuristic Based Dimensionality Reduction Methods for Malaria Vector Using KNN Classifier," Nov. 2020, doi: 10.21203/rs.3.rs-107396/v1.
[41] I. Shivankar, B. Shirsath, B. Bhande, and C. Chandewar, "A Comparative Analysis of a Novel Custom Distance Metric Against Traditional Metrics in KNN Classification," July 2024, doi: 10.22541/essoar.172019478.88222627/v1.
[42] S. S. Ghosal, Y. Sun, and Y. Li, "How to Overcome Curse-of-Dimensionality for Out-of-Distribution Detection?," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 18, pp. 19849–19857, Mar. 2024, doi: 10.1609/aaai.v38i18.29960.
[43] M. T. Coppejans, "Breaking the Curse of Dimensionality," *SSRN Electronic Journal*, 2000, doi: 10.2139/ssrn.236812.
[44] J. D. A. Santos and G. A. Barreto, "An outlier-robust kernel RLS algorithm for nonlinear system identification," *Nonlinear Dynamics*, vol. 90, no. 3, pp. 1707–1726, Aug. 2017, doi: 10.1007/s11071-017-3760-2.
[45] A. Agaal, H. Farkash, M. Essgaer, and A. Ahessin, "Towards Efficient Electricity Management in Benghazi: Forecasting Demand and Load Shedding with ARIMA Models," *Solar Energy and Sustainable Development Journal*, vol. 14, no. FICTS-2024, Art. no. FICTS-2024, Jan. 2025, doi: 10.51646/jsesd.v14iFICTS-2024.446.
[46] M. Elmnifi, M. Almaktar, S. Vambol, O. Trush, V. Sydorenko, and V. Mykhailov, "Agricultural waste in Libya as a resource for biochar and methane production: An analytical study," *Ecological Questions*, vol. 35, no. 2, pp. 1–20, Dec. 2024, doi: 10.12775/eq.2024.021.
[47] A. Bakeer, "Exploring AI-driven solutions for Libyan agriculture: Current applications and strategic pathways," *World Journal of Advanced Research and Reviews*, vol. 24, pp. 1344–1349, Nov. 2024.
[48] A. Agaal, M. Essgaer, and A. Amarrf, "Addressing Class Imbalance for Breast Cancer Prediction in Southern Libya: A Comparative Study of Sampling Techniques," in *Sebha University Conference Proceedings*, 2024.
[49] A. Agaal, M. Essgaer, A. Alshareef, H. Alkhadafe, and Y. BenYahmed, "Application of Classification and Regression Tree and Spectral Clustering to Breast Cancer Prediction: Optimizing the Precision-Recall Trade-Off," in *2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, 2023, pp. 311–317. doi: 10.1109/MI-STA57575.2023.10169755.
[50] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *Journal of Big data*, vol. 8, no. 1, p. 140, 2021.
[51] H. Alimohammadi and S. N. Chen, "Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis," *Expert Systems with Applications*, vol. 191, p. 116371, 2022.
[52] M. Kvet, "Temporal bi-index," in *2023 IEEE 32nd International Symposium on Industrial Electronics (ISIE)*, IEEE, 2023, pp. 1–6.
[53] D. Singh and B. Singh, "Feature wise normalization: An effective way of normalizing data," *Pattern Recognition*, vol. 122, p. 108307, 2022.

Authors' Profiles

Mansour Essgaer holds a PhD in Computer Science (specializing in Artificial Intelligence) from Universiti Kebangsaan Malaysia, awarded in 2016, and previously served as a Researcher at the Malaysian Center for Artificial Intelligence Technology from October 2011 to January 2016. Since 2017, he has been on the Department of Artificial Intelligence at Sebha University, Libya—initially as an Assistant Professor and progressing to Associate Professor and Senior Researcher at the Faculty of Information Technology. Dr. Essgaer's research interests span machine learning, data mining, natural language processing, combinatorial optimization, and heuristic/metaheuristic algorithms. He has published several articles in international journals and conferences, contributing to advancements in his field. He is also an active member of the Institute of Electrical and Electronics Engineers (IEEE), engaging regularly in professional research activities within the international AI and computational intelligence community.

Asma Aggal is an Assistant Lecturer in the Artificial Intelligence Department at the Faculty of Technical Sciences, Sebha University in Sabha, Libya. She earned her B.Sc. in Computer Science from the Faculty of Science at Sebha University in 2005, and subsequently completed an M.Sc. in Computer Science with a specialization in Artificial Intelligence there in 2023. Her research—ranging from cancer prediction in southern Libya to dialect corpus creation—demonstrates a strong command of machine learning, data handling, and real-world applications. Actively engaged in scholarly forums, she advances AI research with both theoretical rigor and societal relevance.

Amna Abbas is a Master's student in Artificial Intelligence at Sebha University's Faculty of Science. Having earned her B.Sc. in Computer Science from the same institution, she has built a strong foundation in AI, machine learning, data analysis, and algorithm development. At the graduate level, her research focuses on advancing AI-driven solutions—such as predictive modeling, natural language processing, or intelligent systems—to address regional challenges. An active participant in academic life, Amna has contributed to departmental seminars, collaborated on research initiatives, Dedicated and driven, she is poised to emerge as an impactful researcher in the field of AI.

Rabia Al Mamlook is a data scientist with a robust background in industrial engineering and engineering management. She earned her Ph.D. in Industrial Engineering and Engineering Management from Western Michigan University (WMU), USA. She also holds a Master's in Applied Statistics \& Biostatistics from WMU and a Master's in Engineering Management from the University of Tripoli, Libya. Her expertise encompasses machine learning, data mining, statistical process quality control, and data visualization, applied to large datasets in industries and healthcare engineering. Proficient in programming languages such as R, SAS, Minitab, and Python, her research interests include smart manufacturing, data models, and deep learning.

How to cite this paper: Mansour Essgaer, Asma Agaal, Amna Abbas, Rabia Al Mamlook, "Profit Forecasting for Daily Pharmaceutical Sales Using Traditional, Shallow, and Deep Neural Networks: A Case Study from Sabha City, Libya", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.18, No.1, pp. 126-144, 2026. DOI:10.5815/ijieeb.2026.01.08