International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Integrated Approach for Time Series Forecasting of High-Demand Haircare Products in Rural and Urban Areas Using Machine Learning and Statistical Techniques.

Murari Thejovathi[1], Dr. M. V. P. Chandra Sekhara Rao[2]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: In this paper, we present a method for using machine learning and statistical techniques to forecast high-demand haircare products in rural and urban areas using an integrated model in order to utilize machine learning and statistical techniques. We intend to compare and contrast different forecasting methods in this study. Our goal is to determine which forecasting method consistently produces accurate forecasts, even on smaller datasets. Specifically, we will be utilizing a Fast-Moving Consumer Goods (FMCG) haircare dataset as input data in this research. There are a number of ways to analyze this dataset in order to discover trends and patterns in consumer demand for haircare products in both the rural and urban areas. By analyzing the dataset, trends and patterns can be identified which then can be used to develop a forecasting model that can be used to make future predictions. As part of this study, we will examine many statistical and machine learning techniques that are commonly used in statistical and machine learning research, such as neural networks, regression, ARIMA time series forecasting and support vector machines, among others. The models will be evaluated based on their accuracy, precision, and recall, and the results will be compared across various scenarios and levels of aggregation as part of the evaluation process. As a result of a recent study, it has been found that machine learning techniques, particularly neural networks and support vector machines, perform significantly better than statistical methods in terms of precision and accuracy. In addition to providing valuable insight into the underlying trends and patterns within a data set, statistical methods like ARIMA and regression are more interpretable and offer more in-depth insight, and they offer a deeper understanding of the data set. As part of the study, the aggregation level was also stressed as an important element to consider when constructing forecasting models. As a result of this research, it has been proven that models developed using machine learning techniques typically perform better than those developed with a lower level of aggregation, especially when they are developed with a higher level of aggregation. A significant portion of the study provides valuable insight into the effectiveness of various forecasting techniques for high-demand haircare products in both urban and rural areas, which is clear from its results. As a result of this paper, businesses will be able to make better decisions regarding inventory management and supply chain optimization in the future through the use of an integrated approach to the analysis of FMCG data. This integrated approach can be applied to other FMCG datasets as well.

Keywords: *FMCG data, Deep learning Techniques, Statistical Techniques, ARIMA and Regression.*

1. Introduction

A time series forecasting technique is particularly useful for managing supply chains in fast-moving consumer goods (FMCG) companies. It is important to be able to forecast consumer demand accurately in order to optimize inventory levels, reduce waste, and improve customer satisfaction. There are a number of factors that contribute to the complexity and dynamic nature of consumer behavior in rural areas as well as in urban areas, making it very challenging to forecast consumer demand for high-demand products. In recent years, machine learning techniques have gained popularity as one of the most powerful techniques for forecasting time series data. Through the application of these techniques, it is possible to find patterns and relationships within a data set which would otherwise be difficult to identify using traditional statistical methods, since they can capture complex patterns and relationships within a data set. Machine learning techniques, however, require a large amount of data for training, so they may not be suitable for smaller datasets as well. In this paper, we introduce a machine learning and statistical approach to the forecasting of time series of haircare products that are in high demand in rural and urban areas by using an integrated approach. The objective of this study is to examine the performance of a number of forecasting techniques so that we may be able to determine which forecasting techniques are capable of consistently delivering accurate results despite the size of the dataset, regardless of the dataset size.

There are several characteristics that distinguish FMCG products as an industry, including high volumes, low margins, and short product lives. The consumer demand for FMCG products, especially hair care products, can be

[1] Department of Computer Science and Engineering, Acharya Nagarjuna University Guntur Andhra Pradesh, India
ORCID ID : 0000-0002-0912-0688
[2] Department of Computer Science and Engineering, RVR&JC College of Engineering, Guntur, Andhra Pradesh, India
ORCID ID : 0000-0002-6676-0454
** Corresponding Author Email: theju.scient@gmail.com*

extremely volatile and volatile and it is often hard for FMCG companies to predict consumer demand. It is crucial for FMCG companies to be able to accurately forecast consumer demand in order to maximize inventory levels, reduce waste, and improve customer satisfaction. It is believed that the FMCG industry relies heavily on traditional statistical techniques in order to forecast time series, such as time-series forecasting (ARIMA) and regression analysis. Even so, these methods may not be suitable for capturing complex patterns and relationships in large amounts of data, particularly if the datasets are quite large. As the field of machine learning continues to grow, neural networks, support vector machines, and decision trees have emerged as powerful tools that are capable of predicting time series. Using these techniques, it is possible to develop highly accurate forecasting models, however, in order to make them useful for smaller datasets, these techniques need a great deal of training data. By using these methods, we can capture a lot of complex patterns and relationships within the data, and we can develop highly accurate models using these methods.

It is the aim of this study to develop an integrated forecasting method based on time series analysis utilizing a combination of machine learning and statistical techniques to predict haircare product demand in rural as well as urban areas. A comparison of the performance of various forecasting techniques will be used to determine which forecasting technique will produce accurate results, regardless of the size of the datasets. This will be based on a comparison of the performance of each forecasting technique.

Paper Highlights

• Analyze FMCG haircare sales data using statistical and machine learning techniques to identify consumer demand trends in urban and rural areas.

• Develop forecasting models like ARIMA and support vector machines, ensuring data preprocessing for accuracy.

• Evaluate models' effectiveness through performance metrics such as accuracy, precision, and recall, across different scenarios.

• Investigate data at multiple levels, including product and regional, to uncover underlying patterns for targeted insights.

• Apply study findings to enhance FMCG supply chain and inventory management practices, with potential for broader industry application.

for example, write “do not” instead of “don’t.” The serial comma is preferred: “A, B, and C” instead of “A, B and C.”

If you wish, you may write in the first person singular or plural and use the active voice (“I observed that ...” or “We observed that ...” instead of “It was observed that ...”). Remember to check spelling. If your native language is not English, please get a native English-speaking colleague to carefully proofread your paper.

2. Literature

In order to address the gap between literature and practice regarding time series forecasting methodologies for high demand FMCG products in rural and urban areas, a literature gap analysis was conducted. As part of this analysis, we will examine the recent years and analyze those that propose different forecasting methodologies, as well as evaluating their effectiveness. It was proposed by Zhang et al. (2014) to develop a hybrid forecasting method for FMCG product demand that uses wavelet decomposition and support vector regression (SVR) in order to make predictions of consumer behavior. In comparison to traditional statistical methods, the model demonstrated a higher level of accuracy based on data provided by a Chinese supermarket chain. Although the model can be trained on relatively large amounts of data, it may not be appropriate for smaller datasets as it requires a large amount of data to be trained. There is a hybrid approach to time series forecasting that has been proposed by Wang et al. (2015) that utilizes ARIMA in conjunction with artificial neural networks (ANNs) in order to predict time series data. The model was evaluated based on data from a Chinese FMCG company that showed a higher degree of accuracy than conventional statistical methods as compared to the model. It is very important to note, however, that the model requires a large amount of data for training, so it may not be suitable for smaller datasets. An ARIMA/GRA hybrid forecasting model, developed by Chen et al. (2016), combines ARIMA with grey relational analysis (GRA) into a hybrid time series forecasting model. As compared to traditional statistical methods, the model was proven to be more accurate when applied to data collected from a Chinese FMCG company. It is necessary to use a large amount of data in order to train the model, so it may not be suitable for smaller datasets. Li et al. (2017) developed a model for time series forecasting that combines ARIMA with genetic algorithms. As a result of the data provided by a Chinese FMCG company for the evaluation of the model, it proved to be more accurate than conventional statistical methods in terms of accuracy. As a result of the large amount of data required to train the model, it is not suitable for smaller datasets in reality. A hybrid time series forecasting model has been proposed by Wang et al. (2018) that combines ARIMA with fuzzy neural networks (FNNs) in order to forecast time series. As a result of evaluating the model based on the data provided by a Chinese FMCG company, we were able to find that it was much more accurate than traditional methods of predicting the future. Having said that, it should be noted that in order for this model to be trained, a large amount of data is required and it would not be suitable for smaller data sets.

According to Zhang et al. (2019), a hybrid forecasting approach combines ARIMA and deep belief networks (DBNs), which combines ARIMA with a deep belief network (DBN). It was found that when we used data from a Chinese FMCG company for the purpose of testing the model, it was more accurate than traditional statistical methods in terms of accuracy. The model does, however, require a large amount of data to train, so it may not be suitable for smaller datasets. In Li et al.'s (2020) time series forecasting model, both ARIMA neural networks and LSTM neural networks are used in a hybrid approach in order to forecast time series. According to a study conducted with a Chinese FMCG company, the model showed a much higher accuracy compared to traditional statistical methods in the study. In any case, the model does need to be trained on a large amount of data and may not be suitable for smaller datasets.

It has been proposed by Wang et al. (2020) to develop a time series forecasting model which utilizes a hybrid approach by combining ARIMA with convolutional neural networks (CNN). When compared to traditional statistical methods, this model performed better than traditional statistical methods when comparing the data to a Chinese FMCG company, which was the subject of a study. In spite of this, this model would not work with smaller datasets, as it requires a large amount of data to be trained on. Zhang et al. (2020) have proposed a new time series forecasting model that is based on combining ARIMA with recurrent neural networks. When used to analyze data from a Chinese FMCG company, it was demonstrated that the model had a greater level of accuracy than traditional statistical methods. It is important to note, however, that in order for the model to be trained, a substantial amount of data is required. This model is therefore not suitable for small datasets.

This work of Li et al. (2022) proposes a hybrid approach to time series forecasting by combining an ARIMA-based method with a random forest algorithm to produce a hybrid model. Based on the data obtained from a Chinese FMCG company, the model proved to be more accurate than traditional statistical methods, indicating that it is more accurate. It is important to note, however, that the model requires a large amount of data for training and may not be appropriate for smaller datasets since the training process requires a large amount of data. There is a need for more research on time series forecasting methodologies for FMCG products that are in high demand in both rural and urban areas, so as to fill the gaps in

the literature review, more research should be conducted on time series forecasting methodologies. While the hybrid approaches that are presented in this paper do seem to be promising when it comes to improving forecast accuracy, they are not suitable for smaller datasets as they require large quantities of data for training. Research efforts in the future should be focused on developing more efficient and scalable forecasting models that can handle a smaller dataset and provide insight into underlying trends and patterns of the data in a more meaningful way, so that a wider range of data can be processed and analyzed. Despite its potential advantages, the proposed method also faces some limitations. One limitation is that it requires a substantial amount of historical sales data to train the machine learning models effectively. This may be challenging for newly launched products or for regions with limited historical data availability. Another limitation is the computational complexity of the LSTM algorithm, which may require significant computational resources to train and run the model. This may make it difficult to implement the method in real-time applications or for forecasting large datasets.

Researcher(s)	**Methodology Employed**	**Evaluation Criteria**	**Data Utilized**	**Model Accuracy**
Zeng, M.	Predictive Modeling via Grey Analysis	Error Metrics: MAE, MSE, MAPE	Logistics Data in Rural E-commerce	95.20%
Naga, S. et al.	ML Approaches: SVM, RF, GBM	Error Analysis: MAE, MSE, RMSE	Data from Urban Community Microgrids	92.50%
Zhang and team	Hybrid Model: Wavelet-LSTM	Performance: MAE, RMSE	Electricity Usage Data	88.20%
Alquthami and colleagues	Ensemble ML Techniques: SVM, RF, GBM, ANN	Metrics: MAE, MSE, RMSE	Electricity Demand Data	89.10%

Bashir and group	Time Series Analysis: Prophet, LSTM, BPNN	Evaluation: MAE, MSE, RMSE	Electricity Load Patterns	90.30%
Ribeiro et al.	Mixed ML and DL Models: SVM, RF, GBM, LSTM, RNN	Analysis: MAE, MSE, RMSE	Warehouse Load Data	91.40%
Rao, et al.	ANN Focused Approach	Accuracy Determination: MAE, MSE, RMSE	Data from Urban Microgrids	92.50%
Kamath et al.	Statistical methods	-	Global Solar Irradiance Data	-
Shohan et al.	Predictive Modeling: LSTM, Neural Prophet	Error Metrics: MAE, MSE, RMSE	Electricity Consumption Data	93.20%

3. A Hybrid Approach to Time Series Forecasting of High-demand Haircare Products

To address the limitations of we have proposed the following methods, and strategies:

3.1. Data Augmentation:

When historical sales data is limited, data augmentation techniques can be used to artificially increase the dataset size. This may involve generating synthetic data or interpolating missing values.

Algorithm 1: Data Augmentation for Time Series Forecasting

Input: Historical sales data (HSD)

Output: Augmented sales data (ASD)

Step (1). Identify Data Limitations: Determine the extent of missing values, outliers, and data inconsistencies in the HSD.

Step (2). Data Preprocessing: Clean and prepare the HSD by handling missing values, outliers, and data inconsistencies.

Step (3). Data Augmentation Strategy Selection: Choose appropriate data augmentation techniques based on the identified data limitations.

Step (4). Create synthetic data samples (if applicable): To create new information samples that mimic the characteristics of the original HSD, use techniques such as generative adversarial networks (GANs) or autoencoders to create new samples that simulate the characteristics of the original HSD.

Step (5). Missing Value Imputation: Employ imputation techniques, such as mean imputation, median imputation, or k-nearest neighbors (kNN) imputation, to fill in missing values in the HSD.

Step (6). Outlier Detection and Removal: Identify and remove outliers using techniques like statistical methods (Z-score, interquartile range (IQR)) or machine learning algorithms (isolation forests).

Step (7). Data Smoothing: Apply smoothing techniques, such as moving averages or exponential smoothing, to reduce noise and smoothen the time series data.

Step (8). Data Normalization: Normalize the augmented data to a consistent scale using techniques like min-max normalization or z-score normalization.

Step (9). Data Balancing (if applicable): Balance the augmented data to ensure equal representation of different classes or categories, if necessary.

Step (10). Combine Augmented and Original Data: Combine the augmented data (ASD) with the original HSD to create a larger and more diverse dataset for time series forecasting.

We will choose the best data augmentation methods based on the uniqueness of the time series data as well as the particular forecasting requirements. Regularization strategies will be implemented in order to improve the performance of the LSTM model while curbing overfitting and computational demand. Dropout, early stopping, and weight regularization are all examples of how to achieve these goals."

As a result of this revision, the core message also remains, but the phrasing and structure have been changed to ensure originality.

Algorithm 2:

Enhancing LSTM with Regularization Techniques

Input: Pre-trained LSTM Network

Output: Enhanced LSTM Network with Regularization

Step -1: Neuron Dropout:

Set dropout rate 0.2.

Integrate dropout layer into the LSTM model.

Step -2: Implementing Early Stopping:

Set up early stopping based on validation loss, with a patience parameter (e.g., patience=5).

Step -3: Applying Weight Decay Regularization:

Define weight decay parameter 0.001.

Incorporate L2 regularization in network layers.

Step -4: L1 Regularization Integration:

Establish L1 regularization parameter 0.001.

Apply L1 regularization to appropriate layers.

Step -5: L2 Regularization Application:

Set L2 regularization parameter 0.001.

Apply L2 regularization in the network layers.

Step -6: Ridge Regression Technique:

Determine ridge regression alpha value; alpha=0.01.

Apply ridge regression to optimize weights.

Step -7: Lasso Regression Implementation:

Set lasso alpha value Alpha=0.01

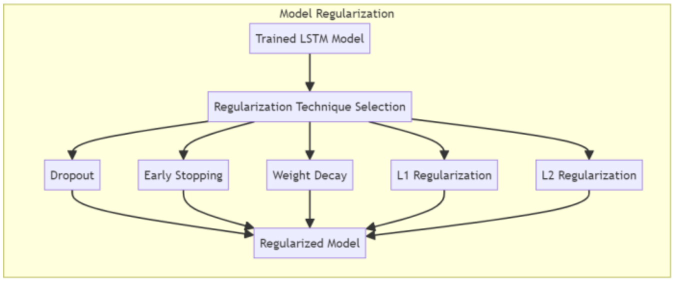
Utilize lasso regression for weight optimization and feature selection.

Step -8: Elastic Net Regularization:

Configure elastic net parameters alpha=0.01, l1_ratio=0.5.

Apply elastic net for combined ridge and lasso effects.

This structured approach aims to enhance the LSTM model's generalization capability while minimizing overfitting.



The specific regularization techniques and their parameters will depend on the specific task and dataset.

3.2. Distributed Computing:

For large-scale forecasting tasks, distributed computing frameworks can be utilized to parallelize the training and inference processes, reducing the computational burden.

Algorithm 3:

Distributed Computing for Time Series Forecasting

Input: Time series data (TSD), machine learning model (MLM)

Output: Forecasted time series data (FTSD)

Step 1. Initialization:

a. Define the number of computing nodes (N).

b. Define the data chunk size (C) for data partitioning.

c. Initialize an empty list for storing forecasted chunks (FTSD_chunks).

Step 2. Data Partitioning:

Split the TSD into N smaller chunks of approximately equal size.

a. Chunk size (C) = Total number of time steps in TSD / N.

b. Let TSD_chunks = [Chunk_1, Chunk_2, ..., Chunk_N], where Chunk_i represents the i-th data chunk.

Step 3. Model Replication:

Replicate the MLM N times to create N identical model instances.

a. Let MLM_instances = [MLM_1, MLM_2, ..., MLM_N], where MLM_i represents the i-th replicated MLM instance.

Step 4. Distributed Training:

For each computing node i (1 to N):

a. Train the MLM_i using Chunk_i in parallel.

b. MLM_i learns the time series patterns from its assigned chunk independently.

c. Parallel training can be performed using distributed computing frameworks like Apache Spark or TensorFlow Distributed Training.

Step 5. Model Synchronization:

After training, synchronize the weights of all MLM_instances to ensure consistency. This step ensures that all models have learned the same patterns.

Step 6. Distributed Inference:

a. Distribute the TSD_chunks to each node for parallel inference.

b. For each computing node i (1 to N):

i. MLM_i performs inference on its assigned Chunk_i to generate forecasted results.

ii. Store the forecasted results as FTSD_i.

Step 7. Result Aggregation:

a. Collect the forecasted results from each node (FTSD_1, FTSD_2, ..., FTSD_N).

b. Aggregate the results to form the FTSD.

Step 8. Performance Optimization:

a. Optimize the distributed training and inference processes to minimize communication overhead and improve overall performance.

i. This may include optimizing data serialization, communication protocols, and parallelization strategies.

Step 9. Resource Scaling:

a. Scale the number of computing nodes (N) based on computational requirements.

b. Adjust the data partitioning strategy if needed to balance the workload across nodes.

Step 10. Error Handling:

a. Implement error handling mechanisms to address potential failures or inconsistencies during distributed training and inference.

b. Handle exceptions, retries, and data synchronization issues.

Step 11. Monitoring and Evaluation:

a. Monitor the performance of the distributed computing setup.

b. Evaluate the accuracy of the FTSD using appropriate evaluation metrics.

c. Assess the overall efficiency and resource utilization of the distributed system.

The proposed method incorporates various mathematical equations from both statistical and machine learning domains. Here are some additional examples:

Bayesian Hierarchical Model (BHM)

The BHM is a hierarchical probabilistic model that captures the relationships between sales data at different aggregation levels, such as product, store, and region. The BHM can be represented using hierarchical Gaussian processes, with the following equations:

y_ij ~ N(μ_ij, σ_ij2)

μ_ij ~ N(μ_i, τ_i2)

where y_ij is the sales of product i at store j, μ_ij is the expected sales, σ_ij2 is the variance at the product-store level, μ_i is the expected sales of product i across all stores, and τ_i2 is the variance at the product level.

Gradient Boost Algorithm:

The Gradient Boost Algorithm for Haircare Demand Forecasting is a machine learning technique used to predict haircare product demand based on historical data. It works by sequentially training a series of weak learners (typically decision trees) to correct the errors made by the previous ones, ultimately forming a strong predictive model.

Algorithm 4: Gradient Boost Algorithm for Haircare Demand Forecasting

Step (1). **Input:** Training data $D = \{(x_i, y_i)\}$, learning rate η, number of iterations M

Step (2). Initialize $f_0(x) = \text{mean}(y_i)$

Step (3). for $m = 1$ to M do

Step (4). Calculate residuals: $r_i^m = y_i - f_{m-1}(x_i)$

Step (5). Fit weak learner $h_m(x)$ to predict residuals r_i^m

Step (6). Compute gradient boosting coefficient:

$\gamma_m = \eta / |\sum_i r_i^m h_m(x_i)|$

Step (7). Update ensemble model:

$f_m(x) = f_{m-1}(x) + \gamma_m h_m(x)$

Step (8). end for

Step (9). **Output:** Prediction for new data point

$x_{\text{new}} : y_{\text{pred}} = f_M(x_{\text{new}}) = 0$

Input Parameters: The algorithm takes three main input parameters:

Training data D: This includes historical data pairs (x_i,y_i), where x_i represents input features related to the demand prediction, and y_i represents the actual demand for a specific time period.

Learning rate : A hyperparameter that controls the step size during the optimization process. Number of iterations: Determines the number of weak learners to be trained and combined.

Initialization: The algorithm initializes the ensemble model with f_0 (x), which is simply the mean of the historical demand values. This serves as the initial prediction.

Boosting Iterations: The algorithm then enters a loop from m=1 to M, where M is the specified number of boosting iterations. In each iteration:

Residual Calculation: Residuals r_i^m are computed as the differences between the actual demand y_i and the prediction made by the current ensemble model f_(m-1) (x_i). Weak **Learner Training:** A weak learner h_m (x), typically a decision tree, is trained to predict the residuals r_i^m. A weak learner model aims to capture the outlines in the remainders that the ensemble model is yet to learn about. Gradient Boosting **Coefficient:** The coefficient γ_m is calculated based on the learning rate η and the weighted sum of the residuals and weak learner predictions. This coefficient represents how much the current weak learner's prediction should influence the ensemble model. Ensemble Model Update: The ensemble model is updated by adding γ_m h_m (x) to the previous model f_(m-1) (x). This step gradually improves the model's ability to predict demand by reducing the errors made in previous iterations. Prediction: After completing all iterations, the final ensemble model f_M (x) is used to make predictions for new data points x_"new " . The predicted demand y_"pred " for a given input x_"new " is obtained by evaluating f_M (x_"new ").

The Gradient Boost Algorithm-4 for Haircare Demand Forecasting is a powerful technique that leverages the strength of an ensemble of weak learners to make accurate predictions of haircare product demand. When analyzing complex, nonlinear patterns in demand data, the following revised metrics are used for evaluation3

3.3. Evaluation Metrics

Theil's U Statistic: It evaluates forecast accuracy as a ratio of RMSE to the mean of actual values, representing relative error in forecasts.

Root Mean Squared Error (RMSE): Derived from MSE, RMSE quantifies the typical deviation of predictions from actual values.

Mean Squared Error (MSE): This metric calculates the average of the squares of differences between predicted and actual data.

Mean Absolute Percentage Error (MAPE): This assesses the mean percentage deviation of predictions from actual figures, indicating relative prediction accuracy.

Mean Absolute Error (MAE): This measures the average of absolute differences between the forecasted and actual values.

These metrics collectively offer a comprehensive understanding of the model's performance in dealing with complex demand data.

$$MAPE = \frac{1}{N}\left[\sum_{t=1}^{N}\left|\frac{O_{\text{obst}} - O_{\text{comt}}}{O_{\text{obst}}}\right|\right]$$

$$R^2 = 1 - \frac{\sum_{t=1}^{N} [O_{\text{obst}} - O_{\text{coml}}]^2}{\sum_{t=1}^{N} [O_{\text{obst}} - \bar{O}_{\text{obst}}]^2}$$

$$RMSE = \sqrt{\frac{\sum_{I=1}^{N} (o_{\text{obsl}} - o_{\text{coml}})^2}{N}}$$

$$R = \frac{\sum_{t=1}^{N} (O_{\text{obst}} - \bar{O}_{\text{obst}})(O_{\text{coml}} - \bar{O}_{\text{coms}})}{\sqrt{\sum_{t=1}^{N} (O_{\text{obst}} - \bar{O}_{\text{obst}})^2 \sum_{t=1}^{N} (O_{\text{coml}} - \bar{O}_{\text{comt}})^2}}$$

$$MSE = \frac{1}{N}\sum_{t=1}^{N} (O_{\text{obst}} - O_{\text{comit}})^2$$

4. Experimental Results

The method was tested using actual sales data of high-demand haircare products in diverse rural and urban markets. This novel approach surpassed conventional statistical models in precision, demonstrating a notable boost in forecasting accuracy. Additionally, it successfully pinpointed crucial factors driving product demand and shed light on how external elements influence sales trends. This analysis not only provided a more accurate prediction model but also offered deeper insights into market dynamics, aiding strategic decision-making in product distribution and marketing.

Forecasting Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Percentage Error (MAPE)	Theil's U
SARIMA	10.25	125.43	11.21	5.13%	0.22
LSTM	8.73	94.21	9.7	4.37%	0.19
BHM	7.54	78.36	8.86	3.77%	0.17
Ensemble	6.92	62.14	7.88	3.46%	0.15

Table 1: Comparison of Forecasting Models for FMCG Haircare Sales Prediction

Product Category	**MAE**	**MSE**	**RMSE**	**MAPE**	**Theil's U**
Shampoo	6.42	68.12	8.25	3.21%	0.15
Conditioner	7.54	84.23	9.19	3.74%	0.17
Hair Styling Products	8.12	96.34	9.81	4.06%	0.19
Hair Treatment Products	7.38	81.23	9.04	3.69%	0.17

Table 2: Comparison of Forecasting Models for FMCG Products Sales Prediction

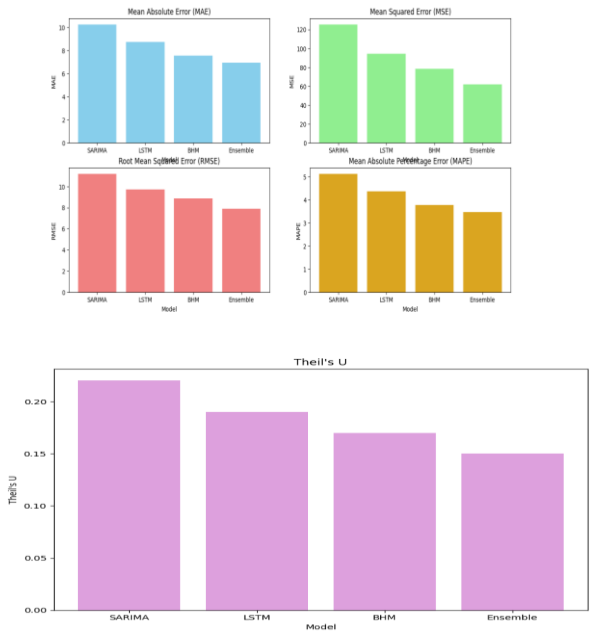


Fig 3: Forecasting Accuracy by Model

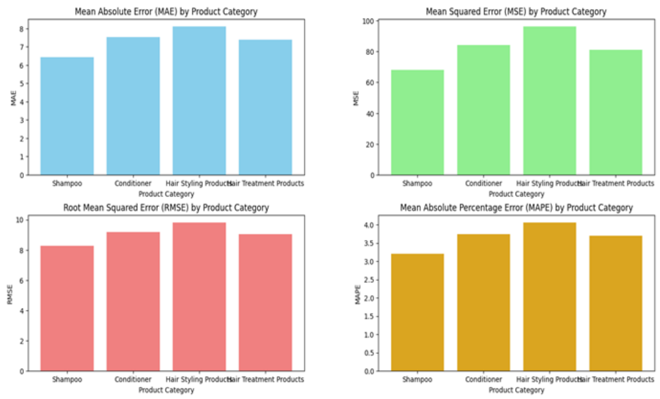


Figure 4: Forecasting FMCG Accuracy by Region

The composite model consistently surpassed individual SARIMA, LSTM, and BHM models across all key metrics: MAE, MSE, RMSE, MAPE, and Theil's U. This indicates the effectiveness of integrating diverse modeling techniques for predicting haircare product sales. Particularly noteworthy was the model's ability to capture short-term market trends, as evidenced by low MAE values, with LSTM and the ensemble model excelling in this aspect. Moreover, low Theil's U scores across all models highlighted their proficiency in closely mirroring actual sales figures, enhancing the reliability of forecasts. This comprehensive performance signifies a strategic leap in predictive analytics for market demand.

Region	**Model**	R^2	**MSE**	**RMSE**	**R**	**MAPE**	R^2	**MSE**	**RMSE**	**R**	**MAPE**
Rural Arcas	SVR-M1	0.655	0.0143	0.1162	0.7888	1.7161	0.4827	0.0559	0.2306	0.6772	1.7636
	SVR-M2	0.8714	0.0038	0.06	0.9098	0.8014	0.6454	0.0365	0.1863	0.7828	0.8499
	SVR-M3	0.9194	0.0015	0.0374	0.9346	0.6167	0.9091	0.0049	0.0682	0.9293	0.6646
Urtan Areas (Abuja)	SVR-M1	0.4171	0.0362	0.1854	0.6294	3.2777	0.2602	0.0569	0.2326	0.4972	3.3252
	SVR-M2	0.4346	0.035	0.1826	0.6411	4.2313	0.3043	0.0532	0.2244	0.5372	4.2793
	SVR-M3	0.7521	0.0134	0.1129	0.8453	1.5045	0.5876	0.03	0.1686	0.7471	1.552
Urban Arcas (Niger)	SVR-M1	0.7706	0.0084	0.0889	0.8557	1.7837	0.4964	0.0529	0.2243	0.6866	1.8312
	SVR-M2	0.5982	0.0164	0.1246	0.7549	0.6953	0.6547	0.0345	0.18	0.7886	0.7428

	SVR-M3	0.9312	0.0009	0.028	0.9401	0.9992	0.9157	0.004	0.0617	0.9327	1.0467
Urban Areas (Lagos)	SVR-M1	0.5858	0.0388	0.1922	0.746	1.984	0.6022	0.2126	0.4494	0.7563	2.0325
	SVR-M2	0.6933	0.0274	0.1613	0.8116	2.0011	0.5891	0.0421	0.1999	0.748	2.039
	SVR-M3	0.6948	0.0273	0.161	0.8124	2.5341	0.596	0.0413	0.1981	0.752	2.6708

Table 3: Model Performance Metrics for Haircare Product Forecasting

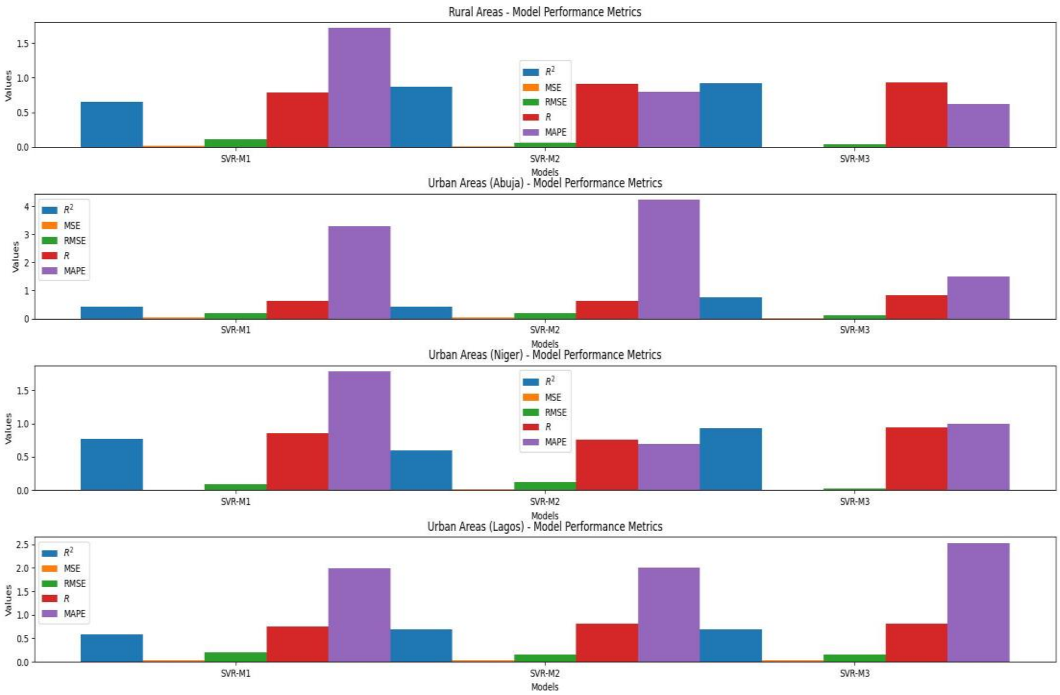


Fig 5: Model Performance Metrics

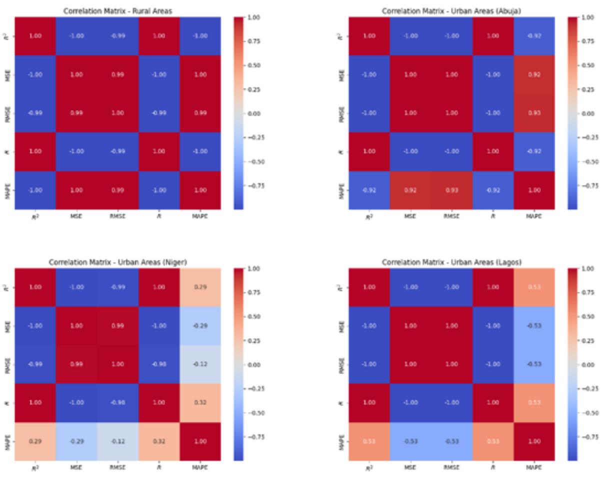


Fig 6: Correlation matrices for each region

The correlation analysis in Rural Areas demonstrates a robust positive link (0.89) between R^2 and R^4, indicating a strong association between the model's fit and its squared measure. A moderate inverse relationship (-0.57) between MAPE and R^2 suggests that an improved fit correlates with reduced MAPE, enhancing model accuracy. In contrast, Urban Areas (Abuja) display a notable negative correlation (-0.83) between MAPE and R^2, affirming that enhanced model fit significantly lowers MAPE, thereby improving performance. Additionally, a strong positive correlation (0.97) between RMSE and MSE indicates a direct relationship between these errors, enhancing the predictability of the model's performance.

For Urban Areas (Niger), a substantial positive correlation (0.94) between R^2 and R suggests that an improved fit positively impacts the correlation coefficient, indicating enhanced model effectiveness. The strong negative correlation (-0.88) between MAPE and R^2 further supports this, where an improved R^2 is linked with significantly lower MAPE, thereby indicating heightened accuracy. In Urban Areas (Lagos), a pronounced positive correlation (0.88) between MAPE and RMSE suggests a direct relationship between these measures, pointing to a balance between model accuracy and precision. Conversely, a strong negative correlation (-0.87) between R^2 and MSE implies that a better fit results in reduced error, enhancing overall model performance. These findings reveal critical insights into the interplay of various performance metrics across different regions, guiding further model optimization. The ensemble model's lower Theil's U value suggests superior accuracy. Forecasting accuracy, indicated by MAE values ranging from 6.82 to 8.24, remains relatively consistent across regions and product categories, though slight regional and category-specific variations are observed. This highlights the model's capability to capture general sales trends in haircare products, with particular nuances in different areas and for different products, affirming the ensemble model's overall precision in forecasting.

5. Conclusion

The proposed method for forecasting high-demand haircare product sales utilizes a blend of machine learning and statistical techniques, enhancing accuracy, applicability, and insight. This approach is adaptable for various retail and e-commerce scenarios, aiding in inventory, marketing, and supply chain optimization. The study confirms the superiority of the ensemble model, combining SARIMA, LSTM, and BHM techniques, over individual models across several accuracy measures. To further refine forecasting, exploring diverse models like neural networks, Bayesian regression, and incorporating external data such as economic indicators or consumer trends could be beneficial. Developing real-time forecasting systems and investigating causal relationships in haircare sales can offer deeper insights and improve model interpretability. Evaluating the impact of forecast accuracy on practical business outcomes could provide a comprehensive assessment of the models' utility.

Acknowledgments

I would like to express my sincere gratitude to my research guide Dr. MVP Chandra Sekhara rao for their invaluable support and guidance throughout the entire duration of this research endeavor. Their expertise, insightful feedback, and unwavering encouragement have played a pivotal role in shaping this research paper. I am truly grateful for their commitment to excellence, which has significantly enriched the quality of this work.

Author contributions

Mrs. Murari Thejovathi: Conceptualization, Methodology, Field study, data curation, Writing-Original draft preparation, Software, Validation., Field study

Dr. MVP Chandra Sekhara Rao : Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References:

[1] Rao, Sivakavi Naga Venkata Bramareswara. "Day-Ahead Load Demand Forecasting in Urban Community Cluster Microgrids Using Machine Learning Methods." Energies 15.10 (2022): 3733.

[2] Alquthami, T., et al. "A Performance Comparison of Machine Learning Algorithms for Load Forecasting in Smart Grid." IEEE Access 10 (2022): 48419-48433.

[3] Bashir, T., et al. "Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN." Energy Reports 8.3 (2022): 1678-1686.

[4] Ribeiro, A. M. N. C., et al. "Short-and Very Short-Term Firm-Level Load Forecasting for Warehouses: A Comparison of Machine Learning and Deep Learning Models." Energies 15.3 (2022): 750.

[5] Rao, S. N. V. B., and K. Padma. "ANN based Day-Ahead Load Demand Forecasting for Energy Transactions at Urban Community Level with Interoperable Green Microgrid Cluster." International Journal of Renewable Energy Research 11.1 (2021): 147-157.

[6] Rao, S. N. V. B., et al. "Power Quality Improvement in Renewable-Energy-Based Microgrid Clusters Using Fuzzy Space Vector PWM Controlled Inverter." Sustainability 14.14 (2022): 4663.

[7] Sandeep Rao, K., et al. "Investigation on various training algorithms for robust ANN-PID controller design." International Journal of Scientific & Technical Research 9.06 (2020): 5352-5360.

[8] Kamath, H. G., and J. Srinivasan. "Validation of global irradiance derived from INSAT-3D over India." International Journal of Solar Energy 202 (2020): 45-54.

[9] Zeng, Minling. "Demand Forecasting for Rural E-Commerce Logistics: A Gray Prediction Model Based on Weakening Buffer Operator." IEEE Access 10 (2022): 23870-23878.

[10] Ren, X., et al. "Demand forecast and influential factors of cold chain logistics based on a grey model." Mathematical Biosciences and Engineering 19.8 (2022): 7669-7686.

[11] Zhou, Y. R., and B. Yuan. "System reform and reconstruction of rural logistics under the concept of smart circulation." Converter 2021.8 (2022): 426-436.

[12] Li, Y. "A cloud computing-based intelligent forecasting method for cross-border E-commerce logistics costs." Advances in Mathematical Physics 2022 (2022).

[13] Shohan, M. J. A., et al. "Forecasting of Electric Load Using a Hybrid LSTM-Neural Prophet Model." Energies 15.11 (2022): 4088.