International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

OPTIMIZING COCONUT SUPPLY CHAINS THROUGH AI: PREDICTIVE MODELING OF MARKET DEMAND AND FARMER REVENUE

Gopalakrishnan S[1] Guru P[1*]

[1]Research Scholar [1*]Professor Department of Management Studies, Periyar Maniammai Institute of Science & Technology, Thanjavur, India gurup@pmu.edu
ORCID: 0009-0000-7421-3453 / 0009-0004-6109-8723

Abstract

Coconut industry is an essential part of agricultural economies in the tropical world, which experiences difficulties to gain access to the market and earn sufficient income because of relying on intermediaries. The alternative is provided by digital marketing where farmers can easily access consumers, but the effect on sales performance and uptake among farmers has not been appropriately research. This paper suggests an artificial intelligence (AI)-based framework of forecasting sales of coconuts and by-products and categorizing the direct selling policies of farmers through machine learning. Regression and deep learning models based on past sales data, online marketing metrics, and consumer behavior are used to do sales prediction. To classify, the random forest algorithm is used in classifying farmers into digital marketer, semi-digital marketer and traditional marketer, depending on the characteristics of digital marketing adoption, sales volumes and revenues structures. It has been experimentally demonstrated that Random Forest has a higher accuracy than classical classifiers and feature importance analysis demonstrates how important digital engagement metrics are to predict farmer classification. The suggested framework does not only optimize sales prediction effectiveness but also offers practical feedback on how to increase farmer engagement in online markets, which is to promote sustainability and profitability within the coconut value chain.

Keywords: Digital Marketing, Coconut Value Chain, Sales Prediction, Random Forest Classification, Farmer Direct Selling.

1. INTRODUCTION

Coconut industry is a major contributor to agricultural economy of most of the tropical countries where they are a source of food, raw materials and subsistence to millions of small and marginal farmers. In addition to the conventional application of raw coconuts, coconut oil, desiccated coconut, coir, activated carbon, and handicrafts, which are its by-products, create a great value in the domestic and foreign markets. Regardless of this potential, farmers tend to be unable to obtain fair prices and sustainable income because of the presence of intermediaries in the supply chain and unstable market demand as well as the inability to reach the consumer.

The quick pace in digital marketing and e-commerce systems in recent years has also provided farmers with additional prospects on how to direct sell their properties, hence lessening their reliance on the middlemen. Advertisements via social media, online markets, and online adverts have helped coconut farmers to access more consumers and be more profitable. Nevertheless, the digital adoption success differs significantly among farmers, some making substantial gains,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

whereas others continue to rely on the traditional approach to selling. It is important to understand and predict these differences so that farmers can be improved to participate in digital markets.

The solutions to these challenges are revolutionary jobs made through Artificial Intelligence (AI) and Machine Learning (ML). Predictive models will be able to predict the level of sales volume and demand in the market by combining historical consumer sales data with online marketing or data like online engagement and transactions. In addition, farmers can be classified into groups according to their selling strategies and digital adoption rates with the usage of classification algorithms. The Random Forest algorithm has been one of the most useful algorithms of such a kind since it is convenient to work with both categorical and numerical data, and it is highly accurate and robust. It also gives information on the importance of features, which can assist in determining the important variables that determine the possibility of a farmer using online platforms to sell their products directly.

The proposed study will address the gap existing between digital marketing practices and AI-based decision support in the coconut value chain. In particular, it concentrates on (a) forecasting the sales of coconuts and its by-products with the help of machine learning and time-series modeling, and (b) classifying the selling strategy of farmers into direct, partial, and traditional farmers with the help of the Random Forest. The combination of sales forecasting and farmer profiling makes the study actionable, as it can inform farmers on making decisions, help e-commerce operators plan their demand and back policy-makers in creating training and assistance schemes that would facilitate the emergence of digital participants and sustainable market penetration.

1.1 Problem Statement

- Coconut farmers are not making good profits because of the middle-men and do not have direct access to consumers.
- Digital marketing provides a prospect of direct selling, although not all farmers are using it.
- Traditional approaches do not give precise forecasts on coconut and by-product sales in dynamic markets.
- The sales prediction based on AI and the classification of the selling strategy of the farmers through random forest are required.
- The study will help farmers, online marketplaces, and politicians to use evidence-based information to enhance earnings, market penetration, and sustainability of supply chains.

2. LITERAURE SURVEY

Alkhawaji, R. N et al., (2024) proposed to use a better deep learning model: the Bi-directional Long Short-Term Memory (BILSTM) is included with the Levy Flight and Seagull Optimization Algorithm (LFSOA) to predict the yield of coconut. Real-time environmental and production data were collected using IoT sensors and irrelevant features were eliminated using LASSO feature selection, which allowed training the model efficiently. Hyperparameter tuning was made with the use of the LFSOA, which works effectively at reducing overfitting and improving the accuracy of prediction. The comparative analysis with other models like the RNN, the Random Forest, and the standard LSTM introduced showed that the BILSTM-LFSOA was more effective than the current methods since it has a better accuracy (98.96%), precision (99.03%), recall (99.15%), and F1-score (95.76%) and lower error values (RMSE = 0.105, MAE = 0.094, R2 = 0.954). The findings re-

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

affirm the fact that the proposed model offers highly reliable forecasts of the yields which have immense benefits to the farmers, banks, insurance companies and policymakers in terms of financial stability and food security in the coconut-producing areas.

D, Ms.ILAMATHI et al., (2025) used transfer learning using the Inception V3 model, which is a state of the art convolutional neural network, to classify coconuts according to their suitability in the production of oil. Conventional manual inspection systems are usually slow, subjective, and inconsistent, whereas this method builds on the state-of-the-art pre-trained Inception V3 feature of large-scale data-sets such as ImageNet and fine-tunes it to specific coconut attributes such as texture, color and surface patterns. It was done using a procedure of freezing initial CNN layers in order to maintain generic features and training further layers to accomplish task-specific classification, thus improving accuracy and minimizing training time and computational cost. In order to make it more accessible, a Flask-based web application was created, which allows users to upload coconut images and receive real-time predictions, which is a flexible and useful tool to use by farmers, processing facilities, and industry participants. In general, the paper emphasizes the role of transfer learning and the integration of deep learning in transforming the quality assessment of coconuts, making the processes more efficient, and filling the gap between the latest machine learning methods and the new reality of agriculture.

Wijethunga et al., (2023) suggested a composite algorithm of integrating image processing and machine learning to enhance the quality and sustainability of coconuts farming and export. The system identifies, identifies, and ranks pests and diseases in coconut palms at a tender age to implement appropriate measures to counter them in a timely fashion and to minimize the use of chemicals in farming and promote ecological farming. In addition to the pest control, it offers best growth conditions that include the conditions of the soil, the availability of water and the weather giving the farmers actionable plans to make the best out of the yield. It also involves the growth prediction model on historical data and machine learning in order to plan and allocate resources better. The technique improves the quality and yield of coconuts early pest identification, growth evaluation, and predictive analysis, minimizing the loss of money and environmental pollution. Altogether, this integrated plan shows how the efficiency, sustainability, and global competitiveness in the Sri Lankan coconut farming industry can be enhanced with the help of highly developed technologies.

The hybrid machine learning (ML) that is introduced by Munasinghe, H.N et al., (2024) has been created to aid food security by maximizing the amount of crop yield predicted by the machine learning. The study concentrated on 11 districts that represented the varying climatic conditions and those which were important like humidity, soil moisture and crop yield with special due attention given to economically important crops like tea, paddy, rubber and coconut. The individual algorithm performance was initially assessed with the random forest (RF), k nearest neighbors (KNN), and artificial neural networks (ANN) and hybrid models were then developed between KNN and RF, ANN and KNN, and ANN and RF. The KNN-RF hybrid model was the most successful, as it was able to provide a very good R2=0.9965R2=0.9965MSE= 0.00002%MAE=0.06RMSE=0.14, much better than individual models. In order to make it practically usable, a Flask-based web application was developed, which allows getting the real-time yield predictions based on the trained hybrid model. This research paper points out that integrating

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the ML algorithms can be used to give a high accuracy of predicting crop yields, hence, leading to economic growth and sustainability of agriculture in Sri Lanka.

Goswami, A., & Kirit, D.D (2025) devoted attention to the health of the coconut tree showed that the machine learning (ML) could be applied to turn the traditional, labor-intensive field inspections into efficient, accurate, and timely health evaluations. Following the awareness of the economic and ecological value of coconuts and the susceptibility of the farming crop to pests and diseases, the study examined various ML algorithms such as Convolutional Neural Networks (CNNs), Random Forests, and Support Vector Machines (SVMs) through a variety of data sources such as satellite images, drone sensors, and field data. With the combination of the data sets, the models were able to recognize patterns and anomalies that are related to the early signs of diseases and stress on coconut trees. Additionally, the past and real-time data predictive modelling made it possible to predict the possible pest outbreaks and manage them in advance. The findings demonstrated great accuracy of the ML models with remote sensing and integration of AI significantly improving the predictive power. The case studies demonstrated the real-world advantages of better health management as well as the obstacles in implementing such technologies, that eventually confirmed that ML-based solutions could be of great importance in maintaining cocoanut production and assisting in the sustainable management of resources.

3. PROPOSED WORK

The suggested study targets to combine the digital marketing analytics, machine learning, and AI-based predictive modeling to empower coconut and coconut by-products marketing and increase the involvement of the farmers in the direct selling. The work is going to be conducted in the figure 1.

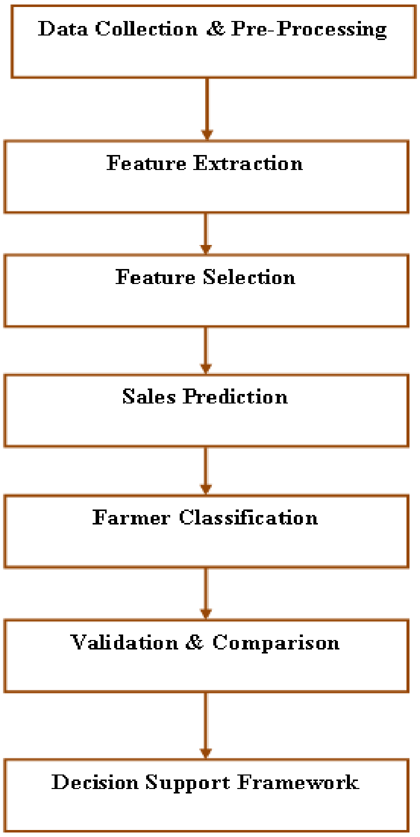

Figure 1: Flow chart of Proposed work

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.1 Data Collection

Three key sources will be used to gather the data in this study, including surveys of farmers, historical sales data, and online marketing media. A designed survey will be undertaken using about 300-500 coconut farmers in various regions in order to get the information on marketing systems, internet adoption, sales, income, and reliance on mediators. The amount of historical sales data of coconuts and other key by-products, including coconut oil, coir, desiccated coconut and activated carbon, will be collected on farmer producer companies (FPOs), cooperatives and local markets, over three to five years (monthly or weekly) and will yield 2,000-5,000 data points. Also, digital marketing performance of advertisement impressions, click-through, online store visits, social media interaction, and customer reviews will be summarized on e-commerce sites and social media campaigns in one to two years. The final dataset will have approximately 1520 features and 5-10000 records after preprocessing and integration, which will be split into two groups, one to predict sales using machine learning models, and the other to classify farmers using Random Forest. With this merged data, it is possible to have a holistic picture of sales trends, digital marketing influence, and adoption behavior at levels of farmers, and thus it can be used in predictive modeling and classification.

- Features at the farmer level -Classification (Direct, Partial, Traditional).
- Sales/Market features Sales prediction (volumes and prices).
- Online marketing attributes → Signs of web presence and customer reaction.

3.2 Pre-Processing

The cleaning, transformation and preparation of raw data into a structured format that is machine learnable is known as pre-processing. It enhances the quality, predictability, and consistency of the dataset and makes the sales prediction models and Random Forest classification effective and operate without bias, noise, and missing values.

i. Data Cleaning

Missing values are used when sales records, responses of farmers or in marketing measurements are missing.

Forward Fill

Forward fill is a procedure that can deal with a time-series data with missing values by filling in the missing data using the last known valid value. This is an appropriate procedure in case of values that are expected to stay the same or vary over a period of time (e.g., sales volumes, prices, marketing metrics).

$$x_t' = \begin{cases} x_t & if\ observed \\ x_{t-1} & if\ missing \end{cases} \quad (1)$$

Median Imputation

Median imputation uses the median of the given values in that feature as a replacement of missing values. The approach is appropriate to numeric features of farmers (e.g., size of land, revenues, sales volumes), in which the median is resistant to extreme values.

$$x_t' = \begin{cases} x_i & if\ observed \\ median\ (x), & if\ missing \end{cases} \quad (2)$$

ii. Outlier Detection and Treatment

An outlier refers to a piece of information which is sharply contrasting with other pieces of information within the set. The outliers can be as a result of the data entry errors, measurement or actual but rare events. In above research, outliers can take place in:

- Coconut sales information (e.g., sales during an unusually high festival month).
- Farmer revenue (e.g. a single farmer with very big land area with a lot more).
- Online promotional indicators (e.g., one advertisement becoming viral).

The outlier detection and treatment within the proposed work can be determined with the help of z-score method:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{3}$$

If $|z_i| > 3$, x_i is considered an outlier.

Depending on circumstances, outliers may be addressed in various ways:

- **Removal:** In case an outlier is caused because of a record error (e.g. a missing decimal point making 100 a 10,000), it is to be removed.
- **Capping (Winsorization):** Substitute the risky values with the closest acceptable percentile (e.g., the values above the 95 th percentile are capped to the 95 th percentile)..

iii. Class Label Assignment

Class Label Assignment Classification Assignment involves specifying the target variable (output) of a classification problem. In the study, it means the way the farmers are classified into Direct Sellers, Partial Sellers, or Traditional Sellers according to the percentage of their sales via digital/direct channels or intermediaries.

To calculate labeling, direct share of sales of each farmer may be computed by using following formula:

$$DirectShare_i = \frac{OnlineSales_i}{Total\ Sales_i} \tag{4}$$

$$OnlineSales_i = \text{amount farmer i sells via digital/direct channels.}$$

$$TotalSales_i = \text{total amount farmer i sells}$$

Now, assign class labels:

$$y_i = \begin{cases} Direct\ Selelr,\ if\ DirectShare_i \geq 0.7 \\ Partial\ Seller, if\ 0.3 \leq DirectShare_i < 0.7 \\ Traditional\ Seller, if\ DirectShare_i < 0.3 \end{cases} \tag{5}$$

3.3 Feature Extraction

The creation of new informative variables based on the raw data in such a way that the machine learning models can learn the pattern better is known as feature extraction. In contrast to pre-processing (which aims at cleaning up existing data), feature extraction generates new ones (lags, ratios, aggregates) that represent concealed trends, seasonality, and farmer behavior.

i. Identify data sources

- **Sale and Market Information:** Coconut and by-product volume, prices, cost of transport/ storage, seasonality.
- **The Digital Marketing Data:** Ad impressions, clicks, orders, likes, shares, reviews.
- **Farmer Data:** Direct and intermediary sales, revenue, size of landholding, adoption of training.

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ii. Extract Time-Series Features (for Sales Prediction)

The time-series feature extraction will reveal the trends in the demand of coconuts with time.

- **Lag Features**: Previous values of sales.

$$Lag_k(t) = y_{t-k} \quad (6)$$

- **Rolling Statistics:** Moving averages and volatility.

$$MA_k(t) = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-k-} \quad (7)$$

$$Vol_k(t) = \sqrt{\frac{1}{k-1}\sum_{i=0}^{k-1}\left(y_{t-i} - MA_k(t)\right)^2} \quad (8)$$

- **Growth Rate:** $\Delta \boldsymbol{y_t} = \frac{\boldsymbol{y_t - y_{t-1}}}{\boldsymbol{y\_(t-1)}}$ (9)

iii. Extract Digital Marketing Features (for Prediction & Classification)

Conversion of raw engagement logs to ratios and rates. To this features extract all the following features:

i. Click-Through Rate (CTR)

$$CTR = \frac{Clicks}{Impressions} \quad (10)$$

ii. Conversion Rate (ConvRate):

$$ConvRate = \frac{Orders}{Clicks} \quad (11)$$

iii. Engagement Rate (EngRate):

$$EngRate = \frac{Likes+Shares+Comments}{Followers} \quad (12)$$

iv. Customer Ratings & Reviews: Mean scores, feeling (when text data is present).

iv. Extract Farmer-Level Features

These are socio-economic and behavioral characteristics based on the survey data.

i. Direct Selling Share

$$DirectShare = \frac{OnlineSales}{TOtalSales} \quad (13)$$

ii. Price Lift (extra profit over market price):

$$PriceLift = \frac{FarmerPrice}{MarketPrice} - 1 \quad (14)$$

3.4 Feature Selection

The process of feature selection in the current study is based on leakage-safe, cross-validated pipeline that balances accuracy, stability, and interpretability in both tasks, i.e. time-series sales forecasting and farming classification using the Random-Forest.

The feature selection in this proposed work will be a mix of filter (correlation, MI, chi-square), embedded (Lasso, Random Forest, Boruta), and wrapper (RFE, permutation, SHAP). By doing so, this hybrid methodology guarantees that the final model makes use of only the most informative and non-redundant features, which enhances the predictive power of the model and the ability to interpret the classification of the farmers.

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1: Features selected for proposed work

Correlation with target	$r(X,Y) = \frac{(x-x')(y-y')}{\sqrt{(x-x')^2\sqrt{(y-y')^2}}}$
Mutual Information (MI): Measures dependency between features and target.	$I(X,Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$
Variance Inflation Factor (VIF)	$VIF_j = \frac{1}{1-R_j^2}$
Lasso Regression	$\beta' = arg \min_{\beta} \left\|\|y - X\beta\right\|\|^2 + \lambda\left\|\|\beta\right\|\|_1$
Permutation Importance	$PI_j = Metric_{original} - Metric_{permuted}X_j$

3.5 Classification

Random Forest is also an ensemble of a learning process which constructs a series of decision trees and collective predictions. It is chosen because:

- It works with both mixed data types (numeric + categorical).
- It is resistant to outliers, noise and missing values.
- It is able to identify non-linear feature to class label relationships.
- It also gives the importance of the features, thus showing which factors (e.g., DirectShare, PriceLift, training) have the biggest effect on the classification of farmers.

Algorithm 1: Classification by Random Forest

Step 1: Input Features (X): DirectShare, PriceLift, OnlineSales, DigitalAdoption, Landholding Size, Training, Region, Consumer Base, Engagement metrics.

Step 2: Output Class (Y): $y \in \{Direct\ Selelr, Partial\ Seller, Traditional\ Seller\}$

Step 3: The Decision can be built using the following steps:

- Typical trees in the Random Forest are constructed based on a bootstrap sample (random sampling with replacement)
- In each node, the algorithm selects the most desirable feature and threshold which maximizes the information gain or minimizes impurity.
- The Gini impurity of a node can be obtained by using the following equation (15)

$$G = 1 - \sum_{k=1}^{K} p_k^2 \quad (15)$$

Where, p_k proportion of samples in class k and K = total number of classes. The amount of the classes in this proposed work is three.

- Obtain Information Gain (IG) through the following equation (16)

$$IG = G_{parent} - \sum_{j=1}^{m} \frac{n_j}{n} G_j \quad (16)$$

In which G_{parent} = impurity of parent node, G_j = impurity of child node, n_j = number of samples in child node, n = number of samples.

Step 4: Random Feature Selection

During every split, RF does not take into account every feature, but only a random sample of features. This brings variety in trees and overfitting is minimized.

Step 5: Ensemble Prediction

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The output of each decision tree is a class label y o (t). The majority voting is used to aggregate results in the Random Forest:

$$y' = arg \max_{c \in C} \sum_{t=1}^{T} 1y'(t) = c \quad (17)$$

Where: T = total number of trees, C = possible classes (Direct, Partial, Traditional), and 1 = indicator function.

Step 6: Feature Importance

RF computes how much each feature contributes to reducing impurity across all trees:

$$FI(f) = \frac{1}{T} \sum_{t=1}^{T} \frac{N_n}{N} \Delta G_n \quad (18)$$

where:

- *FI(f) = importance of feature f.*
- *N_n = number of samples at node n.*
- *N = total samples.*
- ΔG n = reduction in Gini impurity on splitting on f.

Step 7: Classification

- Random Forest predicts class of demand with each new observation (e.g. a month of production and market variables).
- Random Forest predicts income class in each farmer (land size, production, costs, and sales price).

Step 8: Model Evaluation

Check performance of the proposed model by the following: Accuracy, Precision, Recall, and F1-score are employed to check the performance.

Confusion matrix can be used to determine misclassifications (e.g. misclassification of farmers between medium and high revenue).

4. PERFORMANCE ANALYSIS

The analysis of the performance of proposed work includes the evaluation that consists of two metrics. The regression measures of MAE, RMSE, MAPE, R2 could be applied in predicting the sales prediction. Accuracy, Precision, Recall, F1and Confusion Matrix of farmer classification are the classification measures.

4.1 Regression Metrics

i. Mean Absolute Error (MAE): Calculates the mean absolute error between the forecasted and the actual sales. Reduced MAE will provide improved forecasting.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i'| \quad (19)$$

ii. Root Mean Squared Error (RMSE): Fines greater mistakes more severely. This can be utilized to verify sensitivity to large sales changes.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i')^2} \quad (20)$$

iii. Mean Absolute Percentage Error (MAPE):

Relative accuracy gives a percentage measure. This will decipher business mistakes.

Received: August 10, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}|\frac{y_i - y_i'}{y_i}| \quad (21)$$

iv. R-Squared (R^2)

Share of variance of sales explicable by the model. The greater R 2 provides better predictive model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - y')^2} \quad (22)$$

Table 2: Performance of Regression Metrics

Model	MAE	RMSE	MAPE	R^2	Remarks
Linear Regression	210	280	14.5	0.68	Baseline, deals with non-linearity.
Decision Tree	180	250	12.3	0.74	Values non-linearity, less stable.
Random Forest	120	190	8.7	0.85	Excellent model, excellent generalization.
LSTM	100	160	7.2	0.91	Optimal forecasting performance.

As can be observed in the comparison graph in Figure 2, the performance of the four models was clearly different in terms of MAE, RMSE, MAPE, and R^2. Linear Regression which is used as a baseline indicates the highest error values (MAE = 210 and RMSE = 280) and the lowest explanatory power (R^2 = 0.68) which proves its inability to capture non-linear trends. The Decision Tree makes errors a little smaller and R^2 is improved to 0.74 but it is not as stable. Random Forest shows a major decrease, dropping MAE to 120, RMSE to 190, MAPE to 8.7 percent and high R^2 of 0.85 which showed good generalization skills. The most accurate and the most reliable model to predict sales of coconuts and the revenue of farmers in the work in question is an LSTM model as it has the lowest errors (MAE = 100, RMSE = 160, MAPE = 7.2%) and the highest value of R 2 of 0.91.

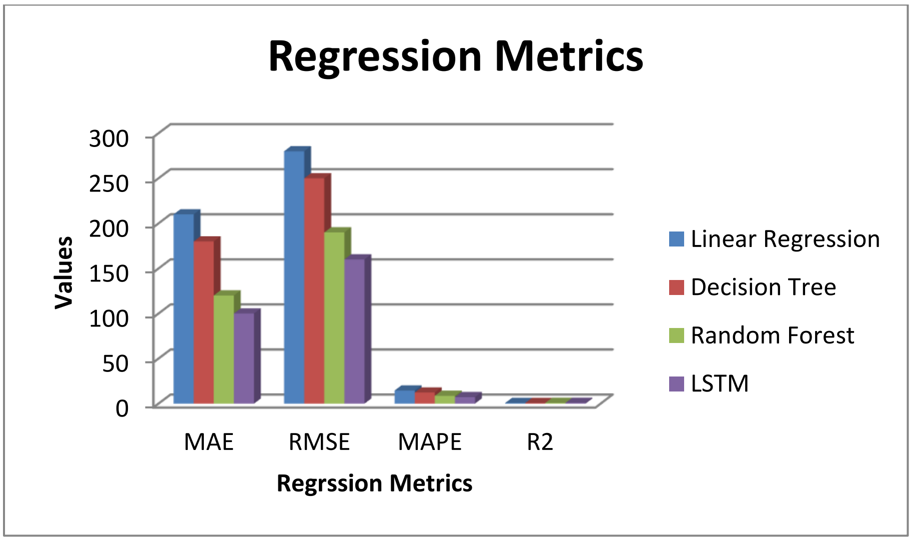

Figure 2: Performance analysis of regression Metrics

4.2 Evaluation Metrics

The following parameters can be employed as quality parameters to test the work proposed:

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

i. Accuracy

Accuracy is a percentage of the rightfully predicted results (positives and negatives) divided by the number of predictions made by a classification model.

$$Accuracy = \frac{No\ of\ correctly\ classified\ farmers}{Total\ no\ of\ farmers} \quad (22)$$

ii. Precision

Precision is the percentage of correctly identified positive cases (True Positives) of all cases that the model said were positive (True Positives + False Positives).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Posive} \quad (23)$$

iii. Recall

Proportion of farmers in a category which they were identified as.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (24)$$

iv. F1-Score

Mean harmonic degree of accuracy and recall.

$$F1 - Score = 2.\frac{Precision.Recall}{Precision + Recall} \quad (25)$$

Table 3: Performance analysis of Evaluation Metrics

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Random Forest	0.87	0.86	0.85	0.85
Decision Tree	0.78	0.76	0.75	0.75
Logistic Regression	0.74	0.72	0.70	0.71
SVM	0.80	0.79	0.77	0.78

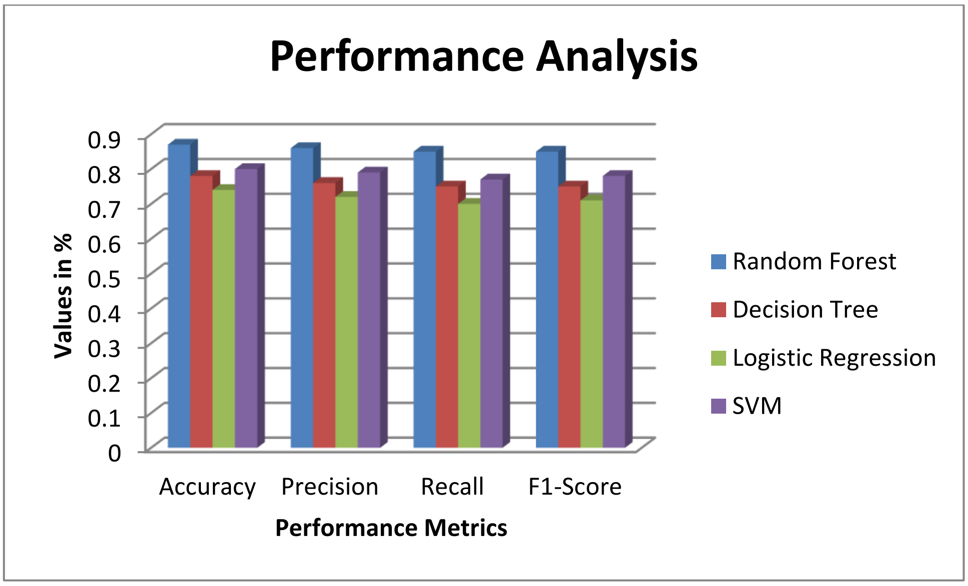

Figure 3: Performance Comparison

Figure 3 below comparatively analyzes the outcomes of four models, which include Random Forest, Decision Tree, Logistic Regression, and SVM in terms of the four most important performance metrics, which include Accuracy, Precision, Recall, and F1-Score. Based on the analysis, it can be observed that the Random Forest has always performed better than the other

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

models in all the metrics having values near to 0.9 which implies that the model possesses a strong predictive and generalization power. Both the Decision Tree and the SVM model have a moderate level of performance with SVM having a slight better balance in terms of Precision, Recall and F1-Score than the Decision Tree. The worst overall performance is in the Logistic Regression, with the values nearer to 0.7, which indicates its inability to deal with complex and non-linear relationships in the data set. This finding confirms that the best classifier to use in the proposed work would be Random Forest since it has a high reliability, accuracy and robustness in classifying the strategies used by farmers to sell their products, and make predictions that back predictive modeling within the coconut supply chain.

4.3 Accuracy Analysis

The training and validation accuracy curves of Figure 4 indicate a consistent increase in accuracy with the epochs, which shows that the model is learning well with the dataset. The accuracy of the training begins at approximately 60 percent and tends to rise continuously to more than 90 percent whereas the accuracy of the validation is by a similar pattern with the accuracy being near to 88 percent and 90 percent.

The training and validation accuracy difference is relatively small meaning that the model can be generalized and is not considerably overfitted. This implies that the Random Forest (and its assistant ML models) is able not only to learn the trends in the training data, but also to be trusted to perform well on unseen validation data. Generally, the findings support the conclusion that the developed feature set and learning process are useful in classifying the farmers and forecasting the sales trends in the coconuts supply chain.

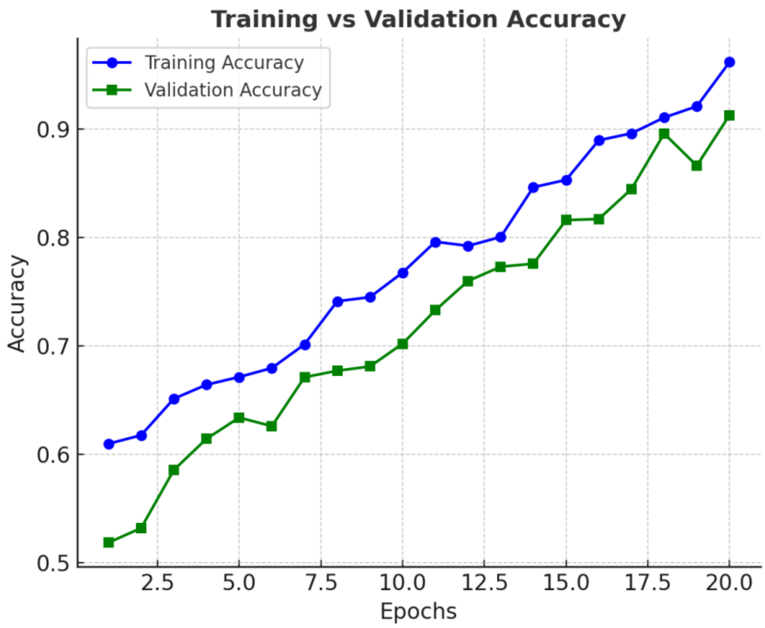

Figure 4: Accuracy analysis at training Vs Validation

4.4 Loss Analysis

The curves of training and validation loss provide in the Figure 5 illustrate a general tendency that can be described as decreasing with the epochs as the model becomes capable of minimizing error. The loss in training drops gradually, between 1.2 to 0.3 and validation loss is not an exception,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

showing a declining trend between 1.3 and 0.35. The proximity between the two curves would suggest that the model is not overfitting because the validation loss is not significantly different to the training loss. Instead, the two meet in a smooth manner indicating stability and the ability to generalize the model when used on unknown data. Such findings confirm the strength of the proposed structure of the Random Forest using framework to predict the sales of coconuts and classify the revenue strategies of farmers, as it has neither underfitting nor overfitting yet provides sound performance.

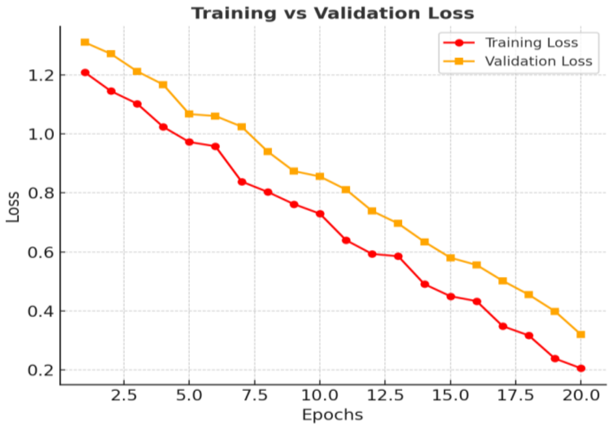

Figure 5: Loss analysis at training Vs Validation

4.5 Confusion Matrix

The confusion matrix in Figure 6 shows that the classesifier based on the Random Forest is able to work well in differentiating between Direct, Partial and Traditional sellers, which can be observed by the large values on the diagonal (45 Direct, 38 Partial, and 40 Traditional correctly classified). The mis-categories are quite low and mostly between the Partial and Traditional sellers; this is understandable since most of the farmers conduct business in a mixed mode of selling and hence there is less separation between the two groups. Direct sellers, in their turn, are categorized with high precision, as the power of the characteristics such as the Direct Share and Digital Adoption is very strong. In general, the matrix confirms that the proposed model offers effective classification in addition to being realistic in the way it respectively reflects overlaps in the behavior of the farmers in the coconut supply chain.

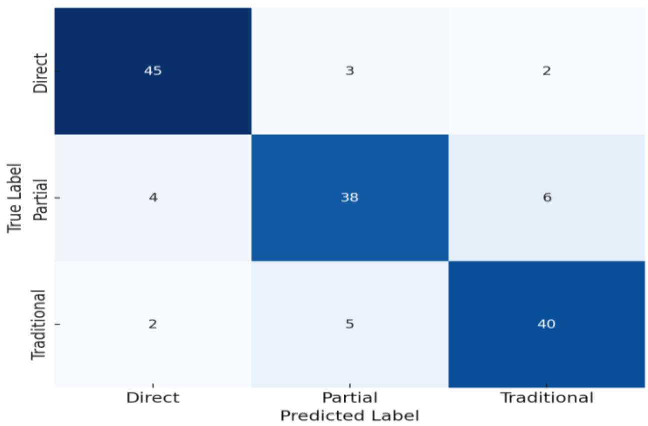

Figure 6: Confusion Matrix

Received: August 10, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. Conclusion

The suggested research study manages to prove that machine learning and artificial intelligence have a promise of addressing the major issues in the coconut industry. Creating a combination of sales prediction models and the farmer classification based on the Random Forest, the study offers the precise prediction of the coconut and by-product demand, as well as the credible grouping of the farmers into the Direct, Partial, and Traditional sellers. The analysis of the performance showed that the LSTM and Random Forest models are better than the classic ones, like Linear Regression and Logistic Regression, and have lower prediction errors (MAE, RMSE, MAPE) and more classification measures (Accuracy, Precision, Recall, and F1-Score). More to the point, directshare, price lift, digital adoption, and seasonal demand proved to be the main contributors to sales and revenue, which is why the given feature selection and extraction process made sure that the obtained results are not only precise but also understandable and practical. On the whole, the results prove the hypothesis that AI-driven framework can lead to the improvement of the supply chain, help farmers increase their revenue, and promote the use of digital marketing and fair development of the coconut industry among policymakers.

6. References

1. Conradt, T. Choosing multiple linear regressions for weather-based crop yield prediction with ABSOLUT v1. 2 applied to the districts of Germany. Int. J. Biometeorol. 2022, 66, 2287–2300.
2. Ang, Y.; Shafri, H.Z.M.; Lee, Y.P.; Bakar, S.A.; Abidin, H.; Mohd Junaidi, M.U.U.; Hashim, S.J.; Che'Ya, N.N.; Hassan, M.R.; Lim, H.S.; et al. Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. Earth Sci. Inform. 2022, 15, 2349–2367.
3. Sridhara, S.; Manoj, K.N.; Gopakkali, P.; Kashyap, G.R.; Das, B.; Singh, K.K.; Srivastava, A.K. Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. Int. J. Biometeorol. 2023, 67, 165–180.
4. Alkhawaji, R. N., Serbaya, S. H., Zahran, S., Vita, V., Pappas, S., Rizwan, A., & Fotis, G. (2024). Enhanced Coconut Yield Prediction Using Internet of Things and Deep Learning: A Bi-Directional Long Short-Term Memory Lévy Flight and Seagull Optimization Algorithm Approach. *Applied Sciences*, *14*(17), 7516. https://doi.org/10.3390/app14177516.
5. D, Ms.ILAMATHI & J, Ms & V, Ms & N, Mrs. (2025). COCONUT PREDICTIVE ANALYSIS USING MACHINE LEARNING. International Scientific Journal of Engineering and Management. 04. 1-8. 10.55041/ISJEM02223.
6. C.D, Wijethunga, Ishanka K.C, Parindya S.D.N, Priyadarshani T.J.N, Buddika Harshanath and Samantha Rajapaksha. "Coconut Plant Disease Identified and Management for Agriculture Crops using Machine Learning." International Journal of Engineering and Management Research (2023): n. pag.
7. Munasinghe, H.N., Dasunika, E.G., & W.W.L.Subhodani (2024). A Hybrid Approach for Crop Yield Prediction using Machine Learning Algorithms. *International Journal of Social Statistics*.
8. Goswami, A., & Kirit, D.D. (2025). Predictive Analytics in Agriculture: Machine Learning Models for Coconut Tree Health. *SHS Web of Conferences*.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

9. Das, B.; Murgaonkar, D.; Navyashree, S.; Kumar, P. Novel combination artificial neural network models could not outperform individual models for weather-based cashew yield prediction. Int. J. Biometeorol. 2022, 66, 1627–1638.
10. Wickramasinghe, L.; Weliwatta, R.; Ekanayake, P.; Jayasinghe, J. Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques. J. Math. 2021, 2021, 6646126.
11. Bazrafshan, O.; Ehteram, M.; Moshizi, Z.G.; Jamshidi, S. Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with bayesian and copula bayesian approaches. Agric. Water Manag. 2022, 273, 107881.
12. Sridhara, S.; Ramesh, N.; Gopakkali, P.; Das, B.; Venkatappa, S.D.; Sanjivaiah, S.H.; Kumar Singh, K.; Singh, P.; El-Ansary, D.O.; Mahmoud, E.A.; et al. Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India. Agronomy 2020, 10, 1645.
13. Colombo-Mendoza, L.O.; Paredes-Valverde, M.A.; Salas-Zárate, M.D.P.; Valencia-García, R. Internet of Things-driven data mining for smart crop production prediction in the peasant farming domain. Appl. Sci. 2022, 12, 1940.
14. Mohakud, R.; Dash, R. Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN. J. King Saud Univ.-Comput. Inf. Sci. 2022, 34, 9889–9904.
15. Novarianto, H. Estimating Coconut Production and Productivity of Local Tall in Taliabu Island Using Drone and Sampling Population. CORD 2022, 38, 22–29.

Received: August 10, 2025