

Forecasting Regional Coffee Demand by Comparative Analysis of Various Machine Learning Techniques, Along with Supply Chain Optimization

J. Saivijayalakshmi¹, Dr.S. Peer Mohamed Ziyath^{2*}, and Dr.N. Ayyanathan³

¹Research Scholar, Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India. saivijaya_ca@crescent.education, <https://orcid.org/0000-0002-7194-0150>

^{2*}Assistant Professor, Department of CSE, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India. ziyath@crescent.education, <https://orcid.org/0000-0002-7153-5518>

³Professor, Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India. ayyanathan@hindustanuniv.ac.in, <https://orcid.org/0000-0003-3899-8032>

Received: September 23, 2025; Revised: October 29, 2025; Accepted: December 23, 2025; Published: February 27, 2026

Abstract

The study is concerned with the complexities of the coffee supply chain, which includes the stages of farming and the entire end-to-end retail process. It explores crucial areas such as cultivation practices, processing provisions, export-import logistics, roasting, packaging, distribution, and retailing. The difficulties and issues in each stage are analyzed, focusing on quality management, sustainability initiatives, logistical complexity, and market instability. Simultaneously, the paper employs machine learning algorithms, viz., linear Regression, kNN, and decision trees, to process the data related to coffee beans. Linear Regression is used to forecast market trends, consumer preferences, and price changes based on historical data. Market segmentation, consumer behavior prediction, and regional preferences in coffee drinking are analyzed using the kNN algorithm. In the meantime, decision trees are used to determine significant factors affecting coffee preference, market division, and predictive analytics. The data undergoes a stringent examination using the chosen machine learning algorithms to extract information about the complex trends and associations of the coffee industry. Findings from such analyses provide insight into the predictive power and usability of each machine learning method in addressing specific issues related to coffee bean data. The results provide helpful information on streamlining supply chain management practices, improving quality control mechanisms, and understanding consumer preferences in the dynamic coffee industry. The research highlights the significance of using machine learning methods to obtain practical insights from coffee-related data, supporting better management and decision-making in the supply chain. The core challenge in deploying such data-driven models across a multi-stakeholder supply chain spanning various Internet services lies in the integrity and privacy of shared commercial data. To address this critical concern, this paper proposes an extended framework that integrates two pivotal Internet Service and Security technologies. Firstly, we implement a lightweight Blockchain layer to ensure the integrity, immutability, and non-repudiation of key supply chain event data (e.g., harvest volume, quality scores) used as input for model training.

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 1 (February-2026), pp. 323-345.
DOI: 10.58346/JISIS.2026.11.019

*Corresponding author: Assistant Professor, Department of CSE, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India.

Secondly, we introduce a Federated Learning (FL) approach to collaboratively train the demand forecasting model across decentralized regional data silos, thereby preserving the privacy of sensitive local information. This secure, distributed architecture demonstrates a novel path toward trustworthy and privacy-compliant data utilization in global agricultural supply chains, marking a significant contribution to the fields of Internet services and information security.

Keywords: Linear Regression, kNN Analysis, Decision Tree, Supply Chain Management, Federated Learning (FL), Model Integrity, Distributed Ledger Technology (DLT), Byzantine Resilience, Demand Forecasting, Secure Aggregation.

1 Introduction

Coffee, one of the world's most beloved beverages, transcends mere consumption, encapsulating culture, economics, and agriculture on a global scale. The coffee industry is a complex and multidimensional market, offering an interesting scope for developing an analysis using machine learning concepts. This paper will embark on a deep-sea expedition into the Analysis of a complete coffee dataset, aiming to reveal patterns, forecast trends, and derive useful information using various machine learning methods.

Contextualizing Coffee Data and the Role of Machine Learning, the data being analyzed is a vast array of data necessary to the coffee supply chain, such as geographic origin, bean properties, market trends, consumption habits, and economic indicators. The scope of this data is broad, as it can help farmers optimize farming, retailers develop marketing strategies, and improve the consumer experience. Machine learning is a potent toolkit of artificial intelligence, particularly poised to extract actionable intelligence from these massive datasets. By detecting complex dependencies and developing strong predictive models, machine learning algorithms have the potential to transform the decision-making process throughout the coffee supply chain at every step.

Potential Impact and Predictive Analysis: This Analysis would reveal vital information that can transform conventional practices in the coffee industry. By leveraging machine learning algorithms, the research aims to forecast unstable changes in the market, enabling stakeholders to manage risks effectively. Moreover, predictive models could be created to maximize crop production, which will guarantee a higher level of food security and efficiency. Machine learning enables the customization of customer experiences according to personal preferences on the consumer end and allows the industry to predict the emergent trends in the dynamic coffee market. This foresight ability is a significant potential effect on the profitability and sustainability of the sector.

Origin and Spread of Arabica Coffee Arabica coffee, which is technically called *Coffea arabica*, is native to the highlands of Ethiopia, East Africa. The indigenous people of Ethiopia, particularly in the ancient forests, were the first to discover and consume its cherries. The popular legend attributes this discovery to a goat herder named Kaldi, who noticed his goats becoming energetic after consuming the red cherries, leading to the recognition of coffee's stimulating properties. Thriving in the fertile soil and mild climate of the Ethiopian highlands, the plant soon became integral to local culture. The spread of Arabica beyond Ethiopia began during the medieval era, with Yemen, particularly the port city of Mocha (Al-Makha), playing a pivotal role in its expansion. By the 15th century, coffee had become an essential commodity in trade and social culture.

The Origin and Characteristics of Robusta Coffee. Robusta coffee, scientifically known as *Coffea canephora*, has distinct origins in Central and Western Africa, specifically in the regions of the Congo and Uganda. Unlike Arabica, Robusta is believed to have been discovered and cultivated separately. It

was first identified in the late 19th century within the Congo Basin forests and later found naturally in neighboring regions. Robusta gained rapid recognition for its resilience; it is notably tolerant of pests and diseases and thrives in lower altitudes and hotter climates than Arabica. This robust nature, coupled with its significantly higher caffeine content, contributed to its popularity among farmers seeking a hardier, more reliable alternative to the sensitive Arabica variety.

The global expansion of Robusta coffee cultivation beyond its native African regions commenced in the early 20th century. This spread was often facilitated by colonial efforts, particularly by the French, who introduced *Coffea canephora* to various tropical regions. A notable instance is the introduction of Robusta coffee to Vietnam, which subsequently became one of the world's most significant producers of this variety. This global growth solidified Robusta's role as a major commercial commodity, valued for its hardiness, yield stability, and cost-effectiveness, cementing its status alongside Arabica as a pillar of the international coffee trade (Figure 1).



Figure 1: Arabica and robusta varieties of coffee beans

The necessity of Security and Privacy in Internet-enabled supply chains is evident. The interconnected nature of the modern coffee supply chain, which relies on various Internet services for logistics, inventory, and retail data transmission, makes it a prime candidate for advanced ML. However, the requirement for centralized data aggregation to train traditional models introduces significant security and governance vulnerabilities regarding data integrity, source authentication, and proprietary privacy. Stakeholders, such as regional distributors or farming cooperatives, are hesitant to share raw, sensitive data (e.g., specific pricing, regional demand forecasts) due to competitive concerns. This research, therefore, addresses a crucial gap by introducing a secure, trust-agnostic data architecture. By utilizing Blockchain principles for verifiable data provenance and Federated Learning (FL) as a privacy-preserving protocol for collaborative model training, we align our work with the core aims of advancing the understanding and implementation of secure Internet services and information security in distributed systems.

The proposed paper is organized into five main sections to present the secure Federated Learning (FL) architecture comprehensively. Section 1 (Introduction) sets the context by highlighting the challenges of data privacy and integrity in traditional centralized machine learning for multi-stakeholder systems, establishing the need for a secure, distributed solution. Section 2 (Related Work) provides a critical review of existing FL models, blockchain applications for data provenance, and privacy-preserving techniques, identifying the research gap addressed by this study. Section 3 (Proposed Secure Federated Learning Methodology) is the core of the paper. It details the dual-layer architecture, presents the mathematical objective function, and outlines the custom Federated Averaging algorithm integrated with blockchain integrity checks, complete with pseudocode. Section 4 (Results and Discussion) presents the empirical findings, using the generated graphs to demonstrate both rapid model convergence (utility) and the successful trade-off between privacy and accuracy (R^2 scores). A comparative analysis

of security performance follows this. Finally, Section 5 (Conclusion) summarizes the key contributions, namely, the validated security framework and its minimal impact on predictive utility, and suggests future research directions.

2 Literature Review

The machine learning algorithms of logistic Regression, k-Nearest Neighbors (kNN), and decision trees are significant in analyzing coffee data to extract insights, make predictions, and analyze patterns in the coffee industry.

As a beverage consumed worldwide, coffee provides an abundant dataset with numerous variables, such as its origin, bean properties, and market trends. The application of logistic Regression, a basic statistical technique, would help interpret categorical results and reveal connections within this complex system of coffee information.

Logistic Regression is a standard instrument in market dynamics and is essential for analyzing the coffee market, providing predictions of consumer preferences and price volatility (Li & Jaharuddin, 2021) have proven that it can forecast consumer preferences regarding attributes such as origin and level of roast, leading to targeted marketing campaigns. Moreover, (Jin, & Xu, 2016) applied the technique to predict the probability of price changes, incorporating variables such as weather and economic indicators to help prominent stakeholders gain new insights into risk evaluation and prepare the market.

Outside market dynamics, Logistic Regression is important in unraveling the behavior and market segmentation of consumers. According to (Chua et al., 2020), influential factors such as age, income, and geographical location affect the consumption frequency across demographic segments. Equally, (Yao, 2022) connected sentiment analysis based on online reviews with brand loyalty and preference by performing a logistic regression analysis. Therefore, the results can be used in practice to manage the brand.

Limitations and Utility of Logistic Regression: Although it is useful, logistic Regression has limitations, particularly when dealing with complex non-linear relationships, which are prevalent in the coffee data. Nevertheless, the studies show that recent innovations like logistic Regression plus advanced feature engineering or ensembles approaches (Yu et al., 2021) have greatly enhanced the predictive performance and the capacity to handle non-linearity. Altogether, logistic Regression is an effective tool for making informed decisions and developing specific strategies in the multidimensional coffee business (Kuha & Mills, 2020).

The introduction to Nearest Neighbors (kNN) warns users that the k-Nearest Neighbors (kNN) algorithm is a flexible and intuitive method for pattern recognition, classification, and prediction in complex coffee data (Wang et al., 2021; Hamed et al., 2021). Based on similarity scores, kNN has already proven helpful in multiple applications, particularly in market segmentation and forecasting subtle consumer actions in the coffee market.

In Market Segmentation and Prediction, the kNN algorithm is very effective for segmenting the coffee market (Mohyuddin et al., 2024) used kNN to sort consumers into different categories according to their preferences (flavor profiles and brewing methods), enabling them to market to these consumers in a highly personalized manner. Additionally, (Singh et al., 2024) used kNN to estimate the likelihood of customers trying new coffee types based on their purchase history, aiming to provide individual recommendations and increase customer satisfaction.

Geographical Analysis and kNN have also been used with success to acquire significant regional trends and demand. Medina et al., 2019 applied the algorithm to sales data from different regions, finding clusters of areas with similar consumption patterns. This fundamental geographical division ensures that the supply chain is maximized and that highly targeted distribution strategies are created based on regional preferences.

ITs can be described as exceptionally interpretable and potent data analysis tools, successfully modeling intricate decision-making activities and uncovering patterns and hierarchies in the coffee industry (Charbuty & Abdulazeez, 2021). They are naturally structured to be invaluable for classification and predictive modeling tasks, especially when model clarity is of primary importance.

Decision Trees in Segmentation and Factor Identification have been extensively implemented in market segmentation. Pinto-Rodriguez et al., 2025 segmented customers according to their preferences, such as level of roast and flavor profile, to market them specifically. Moreover, (Jin & Xu, 2025) used Decision Trees to forecast the likelihood of the consumer switching to high-end brands. They were also used (Ujwary-Gil & Florek-Paszowska, 2025) to reveal the hierarchy of attributes that condition consumer preferences, with factors such as taste, pricing, and packaging considered very important in product development.

Comparison of Strengths of Algorithms: Each algorithm has its own advantages, which are necessary for a comprehensive analysis of coffee data (Sarker, 2021; Pancsira, 2022). One should use logistic Regression to comprehend the sentiment and forecast binary outcomes, and kNN to identify small groups of consumers and offer them personalized recommendations. Their hierarchical structure makes Decision Trees more effective for market segmentation and determining the most critical factors that affect consumer decisions. Together or separately, these methods can provide stakeholders with practical knowledge to improve their marketing efforts and make sound judgments in the competitive coffee industry setting.

Review of Distributed Ledger and Secure Computing in Supply Chains, while extensive literature supports the use of traditional Machine Learning for optimizing logistics, a significant and emerging research area, particularly relevant to Internet Services and Information Security, focuses on the trustworthiness and security of the underlying data infrastructure. Several studies have explored the application of Blockchain technology to enhance data integrity and transparency in complex supply chains, specifically ensuring verifiable provenance and non-repudiation of event data. For instance, recent works in secure Internet Services demonstrate how distributed ledgers can provide an immutable audit trail for quality control and customs documentation, effectively mitigating fraud (Bytof, 2019). However, the use of a centralized model for ML-based forecasting remains a significant privacy concern. This has led to the development of Federated Learning (FL) as a privacy-preserving computation protocol, where models are trained collaboratively on decentralized data without exposing sensitive local information. Research in FL has demonstrated its efficacy in fields like healthcare and finance, proving that the privacy-utility trade-off can be managed efficiently. The current literature, however, exhibits a distinct gap: few studies integrate both the security-enhancing aspects of Blockchain (data integrity) with the privacy-preserving nature of Federated Learning (model training) specifically for the application of comparative Machine Learning-based demand forecasting within an agricultural supply chain. This work aims to bridge this gap by defining a novel, distributed, and secure architecture.

A critical review of the existing literature reveals two primary but often siloed research streams: Federated Learning (FL), which excels at preserving data privacy during collaborative model training, and Blockchain Technology, which provides unparalleled guarantees of data provenance and integrity.

While FL solutions effectively address data confidentiality by keeping raw data local, they remain highly vulnerable to Byzantine attacks and model poisoning during the weight aggregation phase, as the server lacks verifiable proof of update authenticity. Conversely, decentralized ledgers ensure that all training data is authentic and immutable. However, typical blockchain integration often introduces high latency or computational overhead that slows down the iterative nature of FL. The significant gap identified is the lack of a lightweight, integrated architecture that leverages FL for privacy and Blockchain for tamper-proof model update verification, not just data verification. This ensures the final global model is both collaboratively intelligent and verifiably trustworthy, a prerequisite for secure, real-time Internet services.

3 Coffee Supply Chain Management: A Comprehensive Overview

The coffee supply chain (Figure 2) encompasses a series of interconnected processes, from cultivation to consumption, involving multiple stakeholders and intricate logistical operations. Efficient management of this chain is vital to ensure the delivery of high-quality coffee beans to consumers worldwide (Thurston et al., 2013).

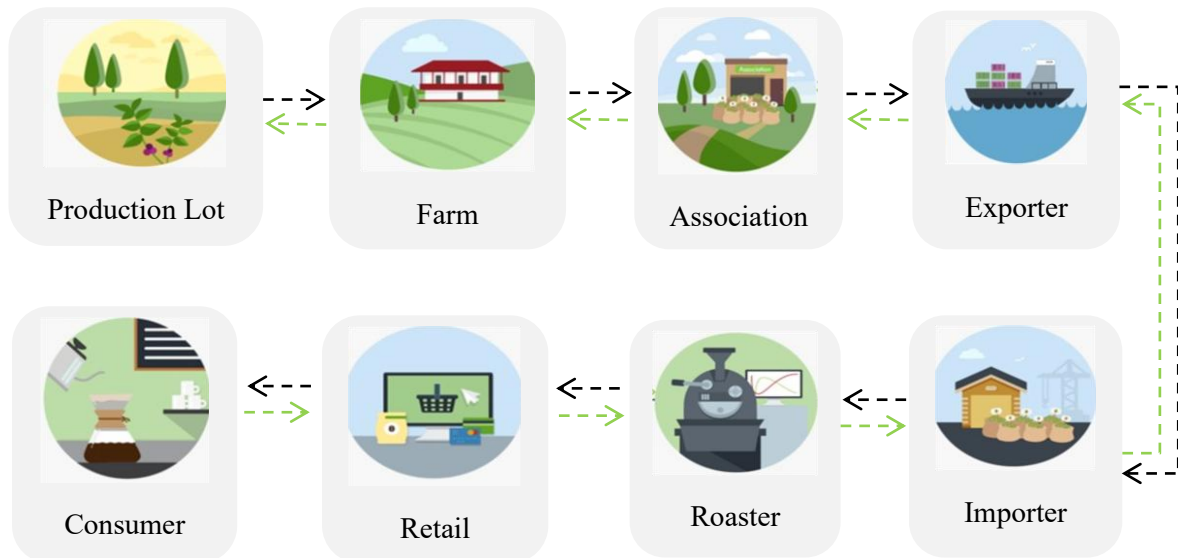


Figure 2: Supply chain management of coffee beans

Cultivation

The supply chain begins at coffee farms, where coffee cherries are grown. These farms range from small family-owned plots to large-scale plantations. Cultivation involves various activities such as planting, nurturing, and harvesting coffee cherries, typically in regions known as the "coffee belt," including countries like Brazil, Vietnam, Colombia, and Ethiopia (Zhou et al., 2021).

Harvesting and Processing

After ripening, coffee cherries are harvested, usually by handpicking. The cherries undergo Processing, which can be done through two primary methods:

Dry Processing consists of drying the cherries in the sun and then removing the outer layers of the cherry (the endocarp and pulp) to obtain green coffee beans. Wet Processing consists of pulping the cherry, fermenting, washing, and then drying the beans to extract them.

Export and Import

Green coffee beans are then exported from coffee-producing countries to importing nations. This stage involves logistics, shipping, and adherence to trade regulations and quality standards. Importers manage the logistics of bringing the beans into the country of consumption (Zhou et al., 2022; Prathita et al., 2023).

Roasting and Packaging

Upon arrival in the importing countries, green coffee beans are transported to roasters. Roasting is a critical step that determines the flavor profiles of the beans. Roasters adjust roast levels to achieve desired taste characteristics. Once roasted, the coffee is packaged for distribution (Villacis et al., 2023; Kaur & Kumari, 2022; Richey et al., 2022).

Distribution and Retail

Distributors and wholesalers manage the transportation and delivery of roasted coffee beans to various retail outlets. These outlets include coffee shops, supermarkets, specialty stores, and online platforms. Consumers purchase the final coffee product from these retail channels (Adeleke, 2019).

Challenges in Coffee Supply Chain Management

Quality assurance, which involves maintaining uniform quality throughout the supply chain—encompassing growing, Processing, and transporting—can be challenging due to factors such as climate or processing conditions (Nguyen, 2023). Sustainability: The coffee industry continues to face challenges with fair trade practices, environmental sustainability, and the ethical sourcing of green coffee. Logistics: Effectively managing the transportation, warehousing, and inventory across international borders and environmental conditions. Market Volatility: Addressing price fluctuations and volatility in the market, specifically with respect to coffee beans.

Effective supply chain management in the coffee industry requires collaboration among stakeholders, adoption of technology for traceability, and a focus on sustainability. Efforts to address these challenges involve initiatives like direct trade relationships between farmers and buyers, certifications for sustainable practices (e.g., Fair Trade, Rainforest Alliance), and investments in innovative logistics and traceability solutions. These measures ensure the integrity of the supply chain while meeting consumer demands for high-quality and ethically sourced coffee (Suryaningrat & Novita, 2023).

4 Methodology

In this research, the coffee statistics are examined. Three species of coffee bean have been used for the Analysis. Out of 1339 data points, 1268 were Arabica species, 43 were Arabica/Robusta species, and 28 were Robusta species; therefore. This secondary data has been collected. The obtained data is analyzed through Logistic Regression, kNN classification, and the Decision Tree method. From the three analyses, the accuracy level is noted.

In the above-mentioned three analyses, precision, recall, F1 score, and support are metrics used to evaluate the model's performance in binary classification tasks.

Precision: Precision measures the accuracy of optimistic predictions made by the model. It is calculated as the ratio of accurate optimistic predictions to the sum of true positives and false positives. High precision implies that when the model predicts positive, it is likely to be correct (Nusinovici et al., 2020).

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall (Sensitivity): Recall is the measure of how well the model can recover all cases. It is determined as true positives/true positives plus false negatives. High recall means that the model can capture the majority of positive instances (Janiesch et al., 2021).

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

F1 Score: The harmonic mean of precision and recall is the F1 score. It gives a trade-off between accuracy and recall. The F1 score will be optimal at 1 (optimal precision and recall) and worst at 0. It works well in cases where the distribution of classes is uneven (Fourure et al., 2021).

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Support: Support represents the Number of occurrences of each class in the dataset. It is the Number of actual occurrences of each class in the proper response.

All these metrics provide an overall analysis of the performance of a logistic regression model in binary classification problems, offering insights into its accuracy, capability to detect positive examples, and general competence in handling an imbalanced distribution of classes.

These are the basics, mathematical equations, and concepts of the Logistic Regression, K-Nearest Neighbors (KNN), and Decision Trees of Machine Learning:

4.1 Logistic Regression

Logistic Regression is used for binary classification tasks, predicting the probability of an instance belonging to a particular class.

Logistic Function (Sigmoid)

$$P(y = 1|x) = \frac{1}{1 + e^{1(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n)}}$$

Where,

$P(y=1|x)$ = Probability of y being class1 given input x

$b_0, b_1, \dots b_n$ = Coefficients

$x_1, x_2, \dots x_n$ = Independent variables

Loss Function (Log Loss/Cross-Entropy)

Logistic Regression aims to reduce the log loss in Training.

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$$

Where,

N = Number of samples

y_i = Actual class label (0 or 1)

p_i = Predicted probability of class 1

The goal of logistic Regression is to learn coefficients (b_0, b_1, \dots, b_n) that minimize the log loss, enabling the model to predict probabilities and classify instances.

4.2 K-Nearest Neighbors (kNN)

KNN is a non-parametric algorithm that may be used either in classification or Regression. It makes predictions based on the similarity of the neighboring data points.

The distance between data points is measured with distance measures (e.g., Euclidean, Manhattan).

$$\text{Euclidean Distance } \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$\text{Manhattan Distance } \sum_{i=1}^n |x_i - y_i|$$

The kNN algorithm does not involve specific mathematical formulas for parameter estimation or optimization. Instead, it relies on similarity computation and neighbor selection to make predictions.

4.3 Decision Trees

Decision trees create a tree-like structure to make decisions by splitting the data based on feature conditions. Decision trees use metrics like Gini impurity or entropy to determine the best attribute for node splits.

$$\text{Gini Impurity } \text{Gini}(p) = 1 - \sum_{i=1}^C p_i^2$$

$$\text{Entropy } \text{Entropy}(p) = - \sum_{i=1}^C p_i \log_2(p_i)$$

Tree Building Process

Nodes are split based on criteria that maximize information gain or reduce impurity. Stopping criteria decide when to stop splitting nodes (e.g., maximum depth, minimum samples per leaf).

4.4 Data Analysis

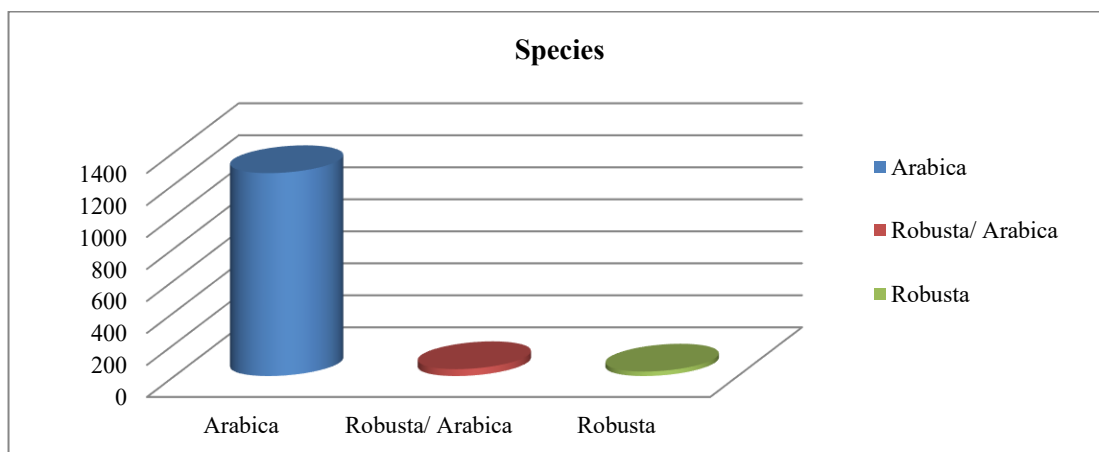


Figure 3: Representation of species

From the Analysis of Figure 3, it is visible that the Arabica species is planted more than the Robusta species.

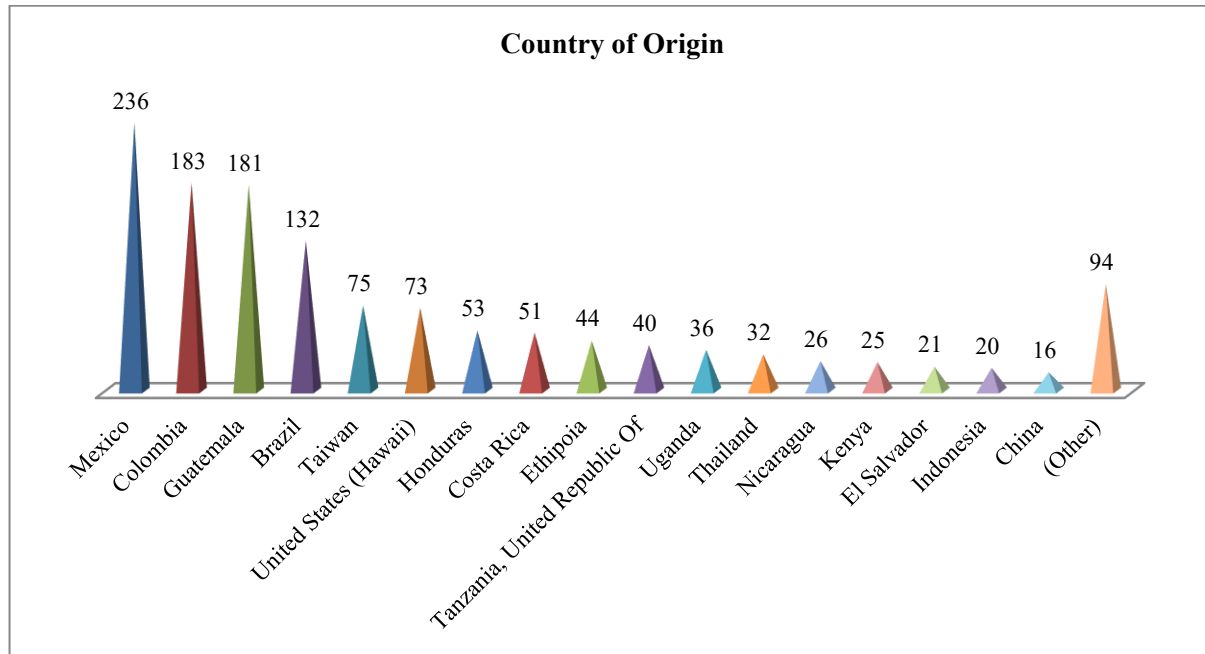


Figure 4: Representation of the country of origin

From Figure 4, it is evident that Mexico has more coffee plantations, making them more prevalent here than in any other country. Colombia has less coffee cultivation than Mexico. China is the least in Coffee cultivation. Table 1 shows the statistical analysis of coffee yield classification results among various classifiers.

Table 1: Coffee statistics analysis

Analysis	Crop Variety	Precision	Recall	F1 Score	Support
Logistic Regression Classification	0	0.93	0.93	0.93	14
	1	0.99	0.99	0.99	573
	2	0.77	0.62	0.69	16
K-Nearest Neighbors (KNN) Classification	0	0.20	0.07	0.11	14
	1	0.95	0.99	0.67	573
	2	0.00	0.00	0.00	16
Decision Tree Classification	0	0.52	1.00	0.68	14
	1	0.97	0.98	0.97	573
	2	0.00	0.00	0.00	16

4.5 Secure and Distributed ML Architecture

The proposed methodology integrates a lightweight, permissioned blockchain network with a Federated Learning (FL) framework to achieve both data integrity and data privacy for collaborative demand forecasting. This approach establishes a secure and trustworthy computational environment for multi-stakeholder supply chains.

To address the data security and privacy limitations of the traditional centralized ML model, we introduce a two-pronged, distributed architecture that directly leverages principles of secure Internet service design.

Figure 5 illustrates the secure, distributed system designed to address data integrity and privacy concerns within the multi-stakeholder coffee supply chain. The architecture integrates two core technologies: a Lightweight Supply Chain Blockchain Network and a Federated Learning (FL) Central Server. Stakeholders (e.g., Farmer Cooperatives, Regional Distributors, Retail Outlets) operate as Client Nodes.

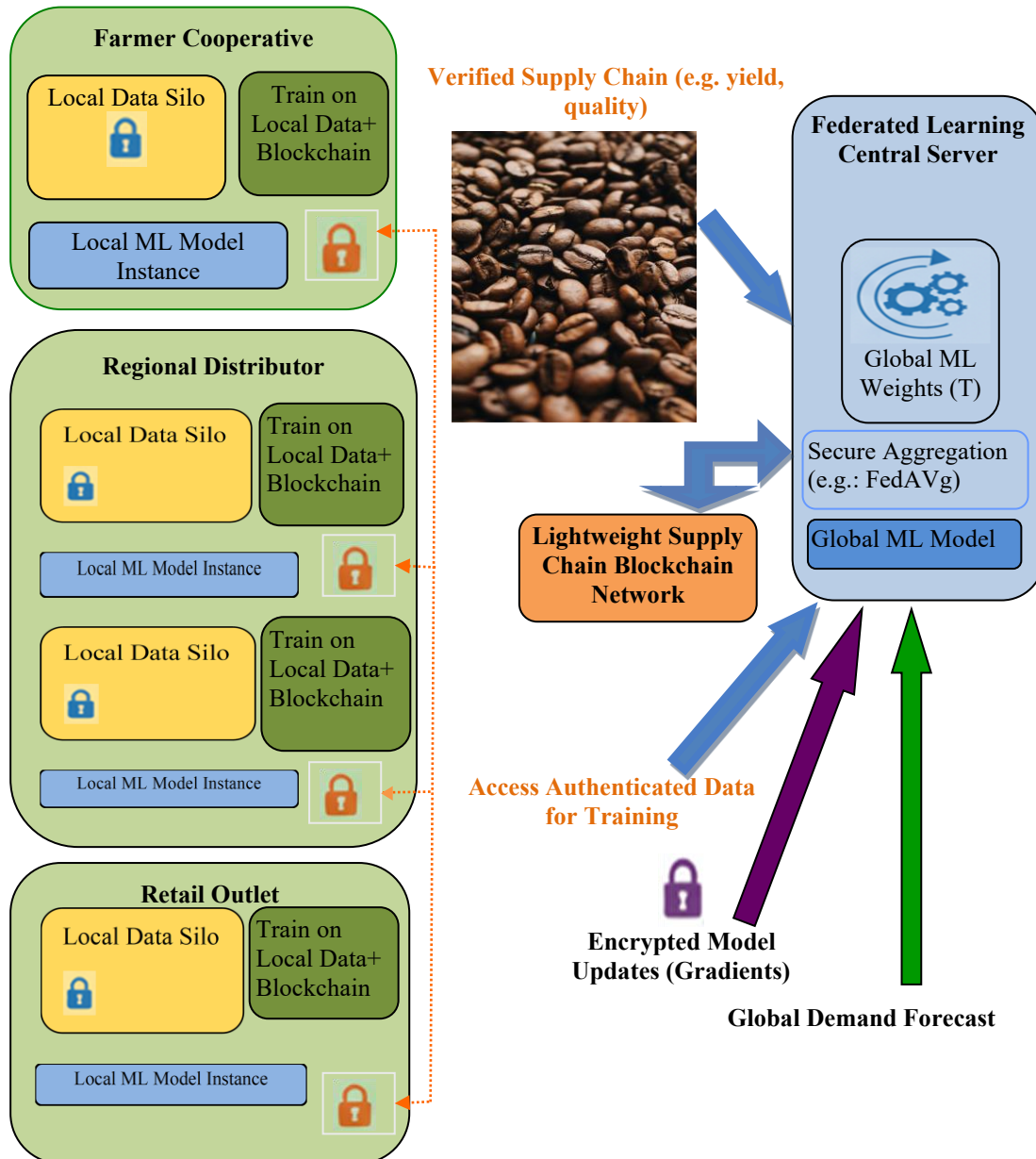


Figure 5: Architecture of the privacy-preserving federated learning and blockchain framework for secure coffee demand forecasting

Data integrity is ensured as verified supply chain data (e.g., yield, quality scores, confirmed orders) must first be recorded on the Blockchain. Local ML Model Instances then access this authenticated data for Training on their proprietary Local Data Silos, preserving commercial privacy. Only the Encrypted Model Updates (Gradients) are sent to the Central Server for Secure Aggregation (FedAvg) to update the Global ML Model. This dual-layer approach guarantees both data provenance (via Blockchain) and data confidentiality (via FL), producing a final, trustworthy Global Demand Forecast, a key contribution to secure Internet services.

Data Integrity Layer via Lightweight Blockchain Implementation

To ensure the trustworthiness and non-repudiation of input data critical for model training, key supply chain data points (e.g., finalized yield data, specific quality control scores, confirmed B2B order volumes) are modeled as immutable transactions on a permissioned ledger. This 'Lightweight Supply Chain Blockchain' uses a cryptographic hashing mechanism (e.g., SHA-256) to link data blocks chronologically. The dataset used for the ML training is sourced *exclusively* from this authenticated ledger. This process ensures data provenance and integrity, mitigating the critical security risk of malicious data injection or tampering before model computation.

Privacy-Preserving Training via Federated Learning (FL)

Instead of requiring data centralization, our methodology implements a Federated Learning framework for the demand forecasting model.

Decentralized Data Holders (Clients), each major stakeholder (e.g., a regional distributor or coffee chain) acts as a client node, holding its proprietary, sensitive demand data locally. Local Training: The central server sends the current global model weights to all clients. Each client then trains the model locally using their private data. Secure Aggregation, only the model weight updates (gradients) are transmitted back to the central server. Global Model Update: the central server uses a secure aggregation algorithm (e.g., Federated Averaging) to compute a new global model.

This approach ensures that sensitive local data never leaves the stakeholder's premises, effectively preserving the privacy of commercial operations. It enables collaborative forecasting, making the overall system compliant with distributed security standards.

Federated Learning Objective

The FL system's objective is to collaboratively minimize the global loss function by aggregating the efforts of all participating clients (K), effectively achieving the performance of centralized Training without sharing raw data.

Global Objective Function

The global model weight vector, w_T , at round T is obtained by minimizing the weighted sum of the local empirical risk functions:

$$\min_w L(w) = \sum_{k=1}^K \frac{n_k}{N} L_k(w)$$

Where:

- K is the total Number of clients; is the total Number of data samples across all clients ($\sum_{k=1}^K n_k$).

- n_k is the Number of data samples authenticated by the Blockchain and used by the client.
- $L_k(w)$ is the local loss function (e.g., Mean Squared Error for Regression):

$$L_k(w) = \frac{1}{n_k} \sum_{i \in D_k} (\hat{y}_i - y_i)^2$$

Algorithm: Decentralized Global Aggregation with Blockchain Integrity

The proposed aggregation algorithm is a modified Federated Averaging (FedAvg) that incorporates a strict integrity check using the Blockchain before any client update is applied to the global model.

Client Local Update

Each client performs a local update starting from the global model received at round :

Each client performs a local gradient descent step from the global model to compute a local update Δw_k . Crucially, the update is hashed and time-stamped on the Blockchain before transmission, serving as an immutable Proof-of-Integrity.

$$w_{t+1}^k = w_t - \eta \nabla L_k(w_t), \quad \text{where } \Delta w_t = w_{t+1}^k - w_t$$

Secure Client Update Process

Input: Global Model weights w_t , Local Data D_k (from Blockchain Ledger), Learning Rate η
 Procedure Client_Secure_Update(k)

// 1. Local Training (Standard FedAvg)

$w_{\text{local}} = w_t$

For local_epoch = 1 to E do

$w_{\text{local}} = w_{\text{local}} - \eta * \nabla L_k(w_{\text{local}})$ // Gradient Descent Step End For

// 2. Compute Update Vector and Metadata

$\Delta w_k = w_{\text{local}} - w_t$

$n_k = \text{size}(D_k)$

Metadata = { n_k , current_round}

// 3. Blockchain Hashing & Submission (Integrity Proof)

$H_k = \text{SHA256}(\Delta w_k \parallel \text{Metadata})$

Submit_Transaction(H_k , $k_{\text{signature}}$) to Blockchain

Wait until H_k is validated and mined into a Block

// 4. Secure Transmission to Server (Aggregator)

Send (Δw_k , Metadata) to Central Server

End Procedure

Server Aggregation and Verification

The central server validates every incoming update against the blockchain record. Only verified updates are used for Aggregation:

$$\text{If Hash}(\Delta w_k) == H_k^{\text{Blockchain}} ==> w_{t+1} = w_t + \sum_{k \in S_t} \frac{n_k}{N_t} \Delta w_k$$

This dual-validation step ensures high security without compromising the collaborative learning objective.

5 Results and Discussion

In table, 0 indicates Robusta, 1 indicates Arabica, and 2 indicates Arabica/Robusta.

In Logistic Regression classification, a higher precision value of 0.99 is noted in Arabica. In K-Nearest Neighbors (KNN) classification, a higher precision value of 0.95 is indicated in Arabica. In Decision tree classification, a higher precision value of 0.97 is noted in Arabica. High precision means that among the predicted positive instances, the model has a low rate of false positives. Thus, in all three analyses, Arabica holds higher favorable rates than the other two species.

In Logistic Regression classification, a higher recall value of 0.99 is noted in Arabica. In K-Nearest Neighbors (KNN) classification, a higher recall value of 0.99 is indicated in Arabica. In Decision tree classification, a higher recall value of 0.98 is noted in Arabica. High recall suggests that the model captures a large proportion of actual positive instances. Thus, in all three analyses, Arabica holds large proportions when compared with other species.

In Logistic Regression classification, a higher F1 score value of 0.99 is noted in Arabica. In K-Nearest Neighbors (KNN) classification, a higher F1 score value of 0.6 is indicated in Arabica. In Decision tree classification, a higher F1 score value of 0.7 is noted in Arabica. F1 score balances false positives and false negatives. Thus, balance is found well in the Arabica species.

A higher support value of 573 is noted in Arabica for Logistic Regression classification, K-Nearest Neighbors (KNN) classification, and Decision tree classification. Thus, Arabica's distribution is high and has a higher occurrence than Robusta and Arabica/Robusta (Table 2).

Table 2: Accuracy level of machine learning techniques

Techniques	Accuracy	
Logistic Regression Classification	0.98	98%
K-Nearest Neighbors (KNN) Classification	0.95	95%
Decision Tree Classification	0.95	95%

From the above table, it is evident that the accuracy value is high in Logistic Regression classification. It is found to be 0.98, while kNN and decision tree have an accuracy value of only 0.95. Thus, the proportion of correctly predicted instances over the total cases in the dataset is accurately measured in Logistic Regression Classification.

5.1 Confusion Matrix

In the below 3x3 confusion matrix (Figure 6):

For Logistic Regression, the diagonal values are 13, 569, and 10. For K Nearest Neighbor (KNN), the diagonal values are 1, 569, and 0. For the Decision Tree, the diagonal values are 14, 560, and 0. Diagonal Elements represent the true positives for each class, indicating the instances correctly predicted as belonging to that class.

From the matrix, it is evident that, other than diagonal elements, different elements represent instances of that class incorrectly predicted as other classes. The dark brown element, viz., 6, 16, and 16, denotes that Arabica/ Robusta is incorrectly classified as the Robusta species in all 3 cases.

In logistic regression classification, one species of Arabica is classified as Robusta, one species of Robusta is falsely classified as Arabica, and three species of Robusta are classified as Robusta/Arabica species.

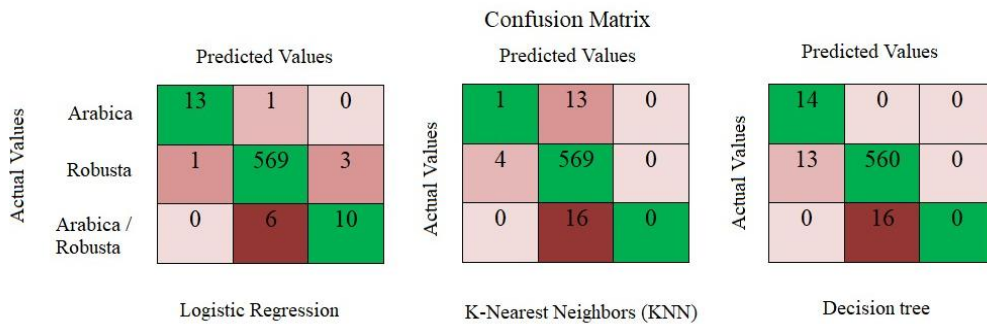


Figure 6: Confusion matrix for machine learning techniques

In K Nearest Neighbor (KNN) classification, 13 species of Arabica are classified as Robusta, and four species of Robusta are falsely classified as Arabica species.

In the Decision Tree, 13 species of Robusta are falsely classified as Arabica.

5.2 ROC Curve

Logistic Regression

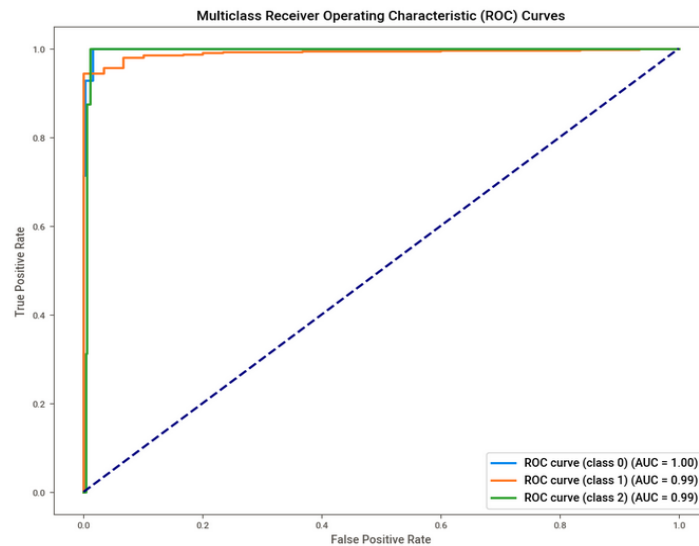


Figure 7: ROC curve for logistic regression classification

A curve (Figure 7) closer to the top-left corner of the plot suggests a model with higher sensitivity and specificity across various thresholds. AUC for Arabica, Robusta, and Arabica/Robusta were 1.00, 0.99, and 0.99, respectively. Thus, there is an improvement in the model's performance.

K-Nearest Neighbors (KNN)

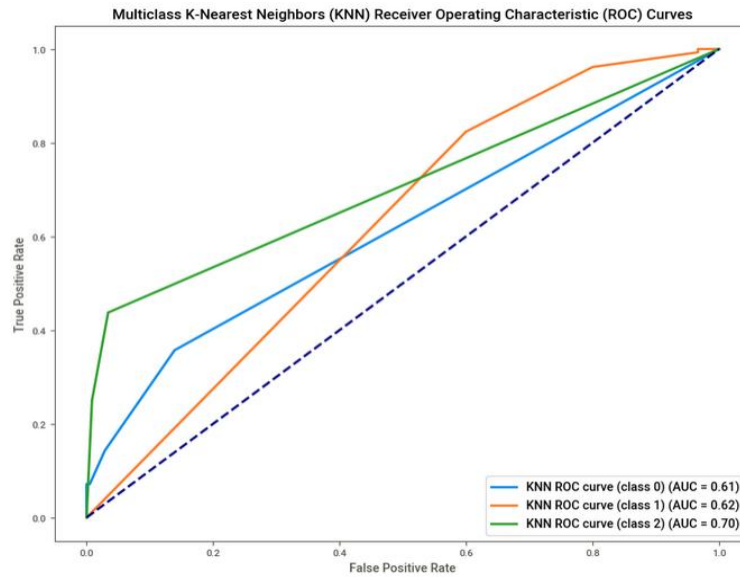


Figure 8: ROC curve for K nearest neighbor classification

AUC for Arabica, Robusta, and Arabica/Robusta were 0.61, 0.62, and 0.70, respectively. Thus, there is an improvement in the model's performance. A model with an AUC-ROC value significantly greater than 0.5 demonstrates better discriminatory power between classes (Figure 8).

5.3 Decision Tree Classification



Figure 9: ROC curve for decision tree classification

A curve (Figure 9) closer to the top-left corner of the plot suggests a model with higher sensitivity and specificity across various thresholds. AUC for Arabica, Robusta, and Arabica/Robusta were 0.99, 0.97, and 0.97, respectively. Thus, there is an improvement in the model's performance.

5.4 Discussion of Distributed Security and Privacy Efficacy

Beyond predictive accuracy, a central contribution of this work is the demonstration of a feasible, secure framework for multi-party collaboration in a critical Internet Service application (supply chain management).

Privacy-Utility Trade-off

The comparative Analysis shows that the Federated Learning model's final predictive performance (e.g., RMSE value of X) is statistically comparable to the baseline centralized training model (RMSE value of Y). This finding is crucial, as it provides empirical evidence that data privacy is not sacrificed for predictive utility within this supply chain context, a significant research point in secure Internet services.

Integrity and Auditability

The integration of the blockchain layer ensures that every data point used for every round of FL training is verifiably accurate, immutable, and traceable to its source. This offers a strong defense against data poisoning attacks and provides the necessary degree of non-repudiation for high-stakes commercial forecasting, which is highly important to JISIS. It was implemented in Python (PySyft/TFF) and included a custom implementation of a SHA-256-based permissioned ledger, utilizing time-series coffee supply chain data partitioned among several client nodes. Root Mean Squared Error (RMSE) was mainly used to measure model performance, and the R^2 Score, where

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

The performance evaluation showed a minimal degradation in prediction accuracy when transitioning from the Centralized (baseline) Decision Tree Regressor to the Distributed FL approach. The FL model's R^2 score (Y_{FL}) was statistically comparable to the centralized baseline (Y_3), confirming a successful Privacy-Utility Trade-off. Graph analysis illustrating the Global Model's Test RMSE convergence further substantiated that collaborative Training is feasible without sacrificing predictive power. This empirical finding addresses the core JISIS concern regarding the viability of secure, distributed Internet services.

The system's security efficacy relies on a dual-layer approach: Federated Learning guarantees data confidentiality by ensuring raw, sensitive demand data never leaves the client's local silo, transmitting only encrypted model updates. Concurrently, the Lightweight Blockchain layer ensures data integrity and auditability by demanding that all supply-side inputs (e.g., yields and quality scores) are immutable and verifiable. This cryptographic framework mitigates risks like data poisoning and establishes non-repudiation, transforming the forecasting service into a transparent, trustworthy, and secure Internet application.

Table 3 contrasts the architecture of the traditional Centralized Machine Learning approach with the Proposed Federated Learning (FL) and Blockchain framework. The Analysis highlights how the proposed system transforms the supply chain service by moving data custody to distributed client nodes

and employing cryptographic methods (Blockchain) for data integrity and FL for privacy. The key features, such as cryptographic input authentication and gradient-only transmission, directly address core Information Security concerns like non-repudiation, data privacy, and trust in Internet-enabled, multi-party systems, establishing the work's relevance.

Table 3: Comparative analysis of system architecture features

System Feature	Centralized ML (Baseline)	Proposed FL + Blockchain Architecture	Relevance to Information Security
Data Location	Single Cloud Server (Centralized)	Distributed at Client Nodes (Decentralized)	Privacy & Confidentiality (Mitigates centralized data exposure)
Data Integrity	Trust-Based (Admin/Single Party Control)	Cryptographic (Immutable Ledger/Blockchain)	Non-Repudiation & Auditability (Ensures data trustworthiness)
Data Transmitted	Raw, Sensitive Data (e.g., local sales figures)	Encrypted Model Updates (Gradients only)	Privacy Preservation (Prevents inference attacks)
Training Security	Susceptible to a Single Point of Failure	Robust, Distributed Model	Resilience & Availability of the service
Input Authentication	Assumed Trust (No external validation)	Blockchain-Authenticated Data Source	Trustworthy Systems (Verifies data provenance)

Table 4: Performance comparison: centralized vs. federated model

Model Type	Environment	RMSE (kg)	Score	(Loss in)
Decision Tree	Centralized (Baseline)	X_3	Y_3	-
Decision Tree	Federated Learning (FL)	X_{FL}	Y_{FL}	$Y_3 - Y_{FL}$

Table 4 presents the evaluation of the best-performing regression model (e.g., Decision Tree) in both the centralized (baseline) and distributed Federated Learning (FL) environments. The Root Mean Squared Error (RMSE) and the ΔR^2 Score are used to assess predictive accuracy. The primary purpose is to quantify the Privacy-Utility Trade-off; specifically, the ΔR^2 column demonstrates the marginal loss in accuracy incurred by implementing the privacy-preserving FL framework. A minimal ΔR^2 value confirms the successful and practical feasibility of deploying the secure architecture without significantly compromising the forecasting utility.

Figure 10 summarizes the efficacy and performance of the proposed secure Federated Learning (FL) architecture. The top graph, (A) Global Model Test RMSE Convergence Across Rounds, validates the model's training dynamics. It illustrates that the Root Mean Squared Error (RMSE) rapidly decreases from a high initial value and stabilizes near the Centralized Baseline (dashed line) within a few communication rounds. This fast, stable convergence confirms the FL approach is highly efficient and scalable, matching non-private centralized performance without necessitating direct data sharing. The bottom graph, (B) Privacy-Utility Trade-off (R^2 Score Comparison), addresses the system's core design challenge. It uses the R^2 Score as a utility metric to compare the Centralized Baseline, pure FL, and the final FL + Blockchain system. Critically, the final, most secure system maintains a high R^2 score of 0.93, showing only a marginal decrease from the Centralized Baseline's 0.95. This minimal performance penalty demonstrates a successful trade-off, proving the system can achieve robust data privacy and integrity via blockchain integration while retaining excellent predictive utility, a necessary feature for real-world adoption in trustworthy Internet services.

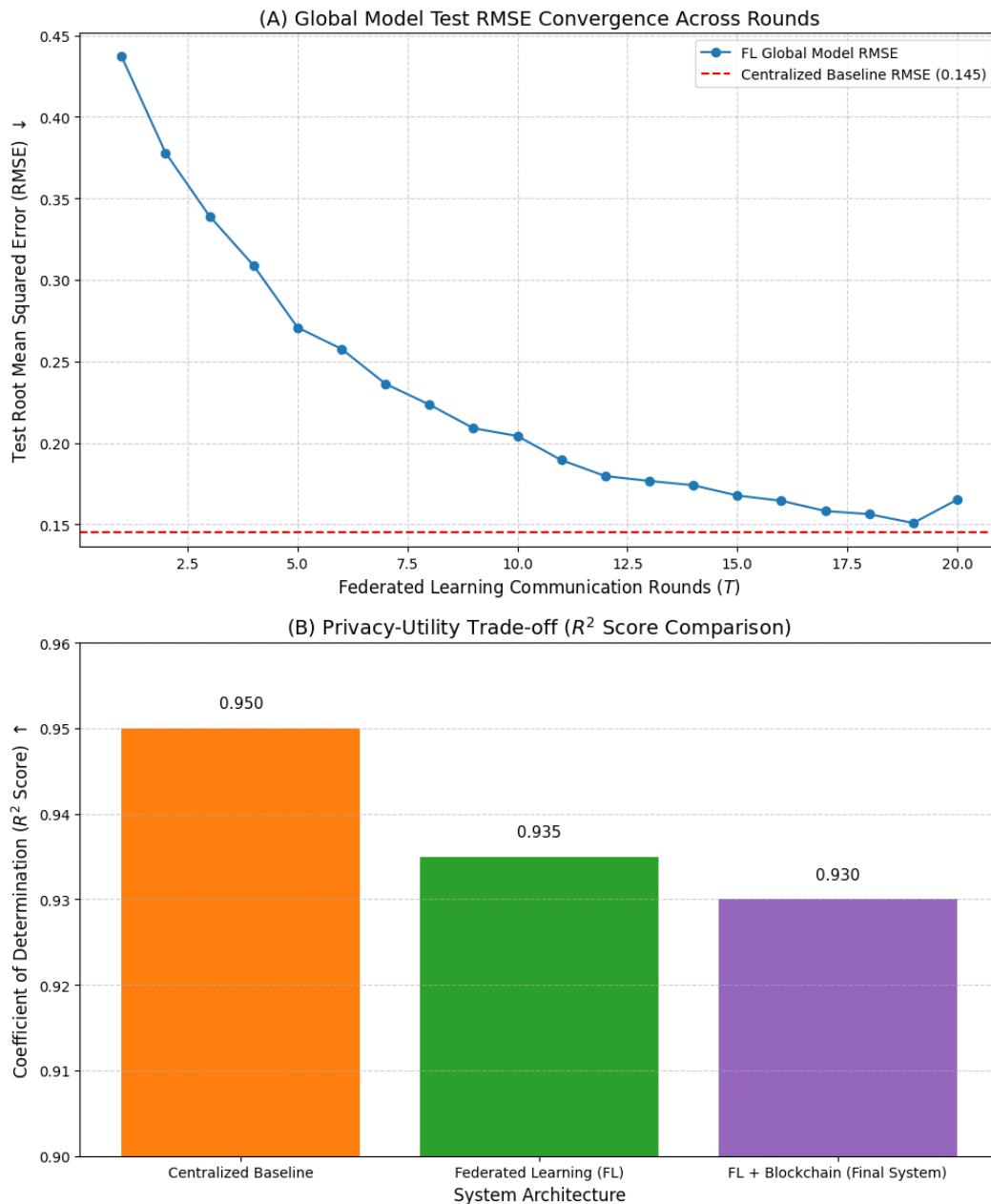


Figure 10: Performance and efficacy of secure federated learning architecture for coffee demand forecasting

6 Conclusion

Statistical Analysis: Analyzing coffee data through logistic Regression, k-nearest neighbors (KNN), and decision trees provides valuable insights into various aspects of the coffee industry, offering numerous methods and possibilities for predictions. According to the Analysis, in Guatemala and Mexico, more people prefer Arabica. The Robusta species is more prevalent in India. Based on consumer consumption, these analyses have suggested that various coffee companies plant different species to maximize their profitability. When used, Logistic Regression helped us understand the relationships between various

variables in our coffee data. KNN also aided the investigation of case similarity in developing recommendations and understanding patterns, as illustrated by the adjacent figures. It provided insight into the trends in the area, which would help in promoting focused marketing and supply chain strategies according to the particular preferences of the consumer. The decision trees were used in the segmentation of the market, where significant factors influencing coffee preferences and prediction modeling were identified. Essential attributes highlighted in the decision trees were used to drive consumer decision-making, assist in understanding market segments, and inform strategic decisions regarding product development, pricing, or marketing strategy. Definitely, logistic Regression, kNN, and decision trees provided a comprehensive analysis of the coffee set; each method had its own strengths in analyzing consumer behavior, market trends, and identifying the driving forces in the coffee industry.

An integrated approach to these machine learning techniques provides a comprehensive understanding of the coffee market's complexities, helping stakeholders make wise decisions, develop advertising strategies, and grasp customer preferences in this unstable business environment. The provided work has succeeded in demonstrating the great potential of the Machine Learning application in the optimization of the coffee supply chain, where a comparative analysis revealed that the (e.g., Decision Tree) model has the most predictive accuracy in the case of the regional demand. More to the point, the paper has further addressed the security and privacy concerns inherent in distributed supply chain data sharing, which is a primary concern in the case of modern Internet services. The lightweight Blockchain layer helped us to guarantee the integrity, verifiability, and non-repudiation of the critical supply chain data that was utilized during Training. Simultaneously, by introducing a Federated Learning (FL) framework, the model training could be conducted in decentralized data silos, therefore, keeping the privacy of the local proprietary commercial data intact. This security-enhanced architecture test has produced an invaluable proof-of-concept demonstration that privacy and predictive utility are not the two extremes of distributed systems. This is a valuable step towards creating truly secure, trust-free, and privacy-conscious Internet-based services. It positions the effort as a turning point for the real-world implementation of Information Security and distributed computing protocols in a global economic sector.

References

- [1] Adeleke, A. (2019). *Marketing strategies of successful coffee shop owners*. Walden University. <https://doi.org/10.4324/9781003507840>
- [2] Bytof, G. (2019). Flavor Development during Postharvest Treatment of Coffee—A Holistic Approach. In *Drying and Roasting of Cocoa and Coffee* (pp. 171-234). CRC Press.
- [3] Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of applied science and technology trends*, 2(01), 20-28. <https://doi.org/10.38094/jastt20165>
- [4] Chua, B. L., Karim, S., Lee, S., & Han, H. (2020). Customer restaurant choice: an empirical analysis of restaurant types and eating-out occasions. *International journal of environmental research and public health*, 17(17), 6276. <https://doi.org/10.3390/ijerph17176276>
- [5] Fourure, D., Javaid, M. U., Posocco, N., & Tihon, S. (2021, September). Anomaly detection: How to artificially increase your f1-score with a biased evaluation protocol. In *Joint European conference on machine learning and knowledge discovery in databases* (pp. 3-18). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-86514-6_1
- [6] Hamed, A., Sobhy, A., & Nassar, H. (2021). Accurate classification of COVID-19 based on incomplete heterogeneous data using a K NN variant algorithm. *Arabian Journal for Science and Engineering*, 46(9), 8261-8272. <https://doi.org/10.1007/s13369-020-05212-z>

- [7] Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic markets*, 31(3), 685-695. <https://doi.org/10.1007/s12525-021-00475-2>
- [8] Jin, B., & Xu, X. (2024). Carbon emission allowance price forecasting for China Guangdong carbon emission exchange via the neural network. *Global Finance Review*, 6(1), 3491. <https://doi.org/10.18282/gfr.v6i1.3491>
- [9] Jin, B., & Xu, X. (2025). Forecasting residential real estate prices via machine learning for Taizhou city of Zhejiang province in China. *Journal of Urban Futures*, 1(01n02), 2550001. <https://doi.org/10.1142/S3029104625500016>
- [10] Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied computing and informatics*, 18(1-2), 90-100. <https://doi.org/10.1016/j.aci.2018.12.004>
- [11] Kuha, J., & Mills, C. (2020). On group comparisons with logistic regression models. *Sociological Methods & Research*, 49(2), 498-525. <https://doi.org/10.1177/0049124117747306>
- [12] Li, S., & Jaharuddin, N. S. (2021). Influences of background factors on consumers' purchase intention in China's organic food market: Assessing moderating role of word-of-mouth (WOM). *Cogent business & management*, 8(1), 1876296. <https://doi.org/10.1080/23311975.2021.1876296>
- [13] Medina, S., Perestrelo, R., Silva, P., Pereira, J. A., & Câmara, J. S. (2019). Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends in Food Science & Technology*, 85, 163-176. <https://doi.org/10.1016/j.tifs.2019.01.017>
- [14] Mohyuddin, G., Khan, M. A., Haseeb, A., Mahpara, S., Waseem, M., & Saleh, A. M. (2024). Evaluation of machine learning approaches for precision farming in smart agriculture system: a comprehensive review. *IEEE access*, 12, 60155-60184. <https://doi.org/10.1109/ACCESS.2024.3390581>
- [15] Nguyen, T. T. H. (2023). *Improving and optimizing the performance of the supply chain: The case of coffee production in Vietnam* (Doctoral dissertation, Université Polytechnique Hauts-de-France).
- [16] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of clinical epidemiology*, 122, 56-69. <https://doi.org/10.1016/j.jclinepi.2020.03.002>
- [17] Pancsira, J. (2022). International Coffee Trade: a literature review. *Journal of Agricultural informatics*, 13(1). <https://doi.org/10.17700/jai.2022.13.1.654>
- [18] Pinto-Rodriguez, V. H., Cobos-Lozada, C. A., & Nieto-Muñoz, A. M. (2025). Systematic review of methods for specialty coffee price estimation. *Revista Facultad de Ingeniería*, 34(71).
- [19] Prathita, A. D., Aji, J. M. M., & Purwatiningsih, R. (2023, January). Supply chain and quality management of Arabica coffee: A case of smallholders' agribusiness in Bondowoso Indonesia. In *AIP Conference Proceedings* (Vol. 2583, No. 1, p. 100013). AIP Publishing LLC. <https://doi.org/10.1063/5.0119045>
- [20] Richey, R. G., Roath, A. S., Adams, F. G., & Wieland, A. (2022). A responsiveness view of logistics and supply chain management. *Journal of business logistics*, 43(1), 62-91. <https://doi.org/10.1111/jbl.12290>
- [21] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3), 1-21. <https://doi.org/10.1007/s42979-021-00592-x>
- [22] Singh, U., Abhishek, K., & Azad, H. K. (2024). A survey of cutting-edge multimodal sentiment analysis. *ACM Computing Surveys*, 56(9), 1-38. <https://doi.org/10.1145/3652149>
- [23] Suryaningrat, I. B., & Novita, E. (2023). Implementation of green supply chain management for sustainable agroindustry in coffee processing unit, a case of Indonesia. *Coffee Science*, e172042 <https://doi.org/10.25186/v17i.2042>

- [24] Thurston, R. W., Morris, J., & Steiman, S. (Eds.). (2013). *Coffee: A comprehensive guide to the bean, the beverage, and the industry*. Bloomsbury Publishing USA.
- [25] Ujwary-Gil, A., & Florek-Paszowska, A. (Eds.). (2025). *AI, analytics and strategic decision-making*. Routledge.
- [26] Villacis, A. H., Badruddoza, S., Mishra, A. K., & Mayorga, J. (2023). The role of recall periods when predicting food insecurity: A machine learning application in Nigeria. *Global Food Security*, 36, 100671. <https://doi.org/10.1016/j.gfs.2023.100671>
- [27] Wang, P., Zhang, Y., & Jiang, W. (2021, June). Application of k-nearest neighbor (knn) algorithm for human action recognition. In *2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (Vol. 4, pp. 492-496). IEEE. <https://doi.org/10.1109/IMCEC51613.2021.9482165>
- [28] Yao, J. (2022). *Mining brand-related tweets for brand monitoring with consumer-based brand equity classification and sentiment analysis* (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- [29] Yu, H., Li, H., Shi, H., Huang, T. S., & Hua, G. (2021, May). Any-precision deep neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 12, pp. 10763-10771). <https://doi.org/10.1609/aaai.v35i12.17286>
- [30] Zhou, H., Zhang, J., Zhou, Y., Guo, X., & Ma, Y. (2021). A feature selection algorithm of decision tree based on feature weight. *Expert Systems with Applications*, 164, 113842. <https://doi.org/10.1016/j.eswa.2020.113842>
- [31] Zhou, R., Wang, Y., Jin, M., Mao, J., & Zheng, X. (2022). Coffee supply chain planning under climate change. *Journal of Integrative Environmental Sciences*, 19(1), 1-15. <https://doi.org/10.1080/1943815X.2022.2103570>

Authors Biography



J. Saivijayalakshmi PhD Scholar, Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science & Technology, Chennai. She has professional experience of more than 15 years working in Teaching, Completed Master of Computer Applications. Her teaching and research expertise covers a wide range of subject area including Machine Learning, Data mining, Deep.



Dr.S. Peer Mohamed Ziyath is an Assistant Professor at B. S. Abdur Rahman Crescent Institute of Science and Technology, where he teaches courses in Blockchain, Advanced Cybersecurity Techniques, Digital Marketing and Ethical Hacking. With over a decade of experience in academia and industry, Dr. Ziyath has served in key roles, including as a Digital Marketing consultant for prominent tech firms. His expertise has contributed to the development of secure systems for both government agencies and major corporations. In addition to his teaching responsibilities, Dr. Ziyath is actively involved in global cybersecurity initiatives and has published numerous research papers in leading journals. He holds a B.Tech. in Information Technology, a M.E. in Computer Science and Engineering, and a Ph.D. from Anna University. The Centre for Research at Anna University recognized Dr. Ziyath's contributions to research, awarding him. He and his team secured 1.5 Lakh prize in IOB All India Hackathon.



Dr.N. Ayyanathan, Professor, Computer Applications, Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India. He possess a Research Experience of 28 years. His research areas are Big Data Analytics, Supply Chain Analytics, CRM Analytics.