# Blockchain-Enabled Decentralized Clustering for Enhanced Decision Support in the Coffee Supply Chain

Keo Ratanak[1], Muhammad Firdaus[2], Kyung-Hyune Rhee[3]
[1]Dept. of IT Convergence and Application Engineering, Pukyong National University
[2]Dept. of AI Convergence, Pukyong National University
[3]Dept. of Computer Engineering and AI, Pukyong National University

**Abstract**

Considering the growth of blockchain technology, the research aims to transform the efficiency of recommending optimal coffee suppliers within the complex supply chain network. This transformation relies on the extraction of vital transactional data and insights from stakeholders, facilitated by the dynamic interaction between the application interface (e.g., Rest API) and the blockchain network. These extracted data are then subjected to advanced data processing techniques and harnessed through machine learning methodologies to establish a robust recommendation system. This innovative approach seeks to empower users with informed decision-making abilities, thereby enhancing operational efficiency in identifying the most suitable coffee supplier for each customer. Furthermore, the research employs data visualization techniques to illustrate intricate clustering patterns generated by the K-Means algorithm, providing a visual dimension to the study's evaluation.

## 1. Introduction

The coffee industry stands at the intersection of tradition and innovation, deeply rooted in centuries-old practices while adapting to the demands of a rapidly evolving global market. In today's evolving coffee supply chain, blockchain technology offers a chance to tackle complex challenges and transform how we recommend optimal coffee sources. In this sense, blockchain as a decentralized and immutable ledger, offers a paradigm shift in how transactions and data are managed and secured. Its integration into the coffee supply chain promises unprecedented transparency, traceability, and trust among participants.

Inspired by the advantages of blockchain, our research's core objective is to enhance decision-making in the coffee supply chain, boosting operational efficiency and customer satisfaction. The process hinges on extracting transactional data and insights from diverse stakeholders using a Rest API and blockchain as a secure ledger. Blockchain empowers us to create a robust data-sharing platform among supply chain actors. With this data, we employ advanced data processing and machine learning to build a highly accurate recommendation system.

Moreover, this approach empowers coffee supply chain stakeholders with informed decision-making abilities. It surpasses traditional supplier selection by considering multiple variables. This ensures individuals and organizations involved can confidently and efficiently identify ideal coffee suppliers for each customer. Additionally, beyond technical aspects, our study emphasizes data visualization. We use this to effectively communicate complex findings, showcasing clustering patterns from the K-Means algorithm. This visual representation serves to enhance the clarity of our study and equips stakeholders with valuable insights into the structure and dynamics of the coffee supply chain.

## 2. Literature Review

The In this section, we review the relevant literature in the fields of blockchain, coffee supply, machine learning, and recommendation systems to establish the context for our research and identify gaps where our work contributes significantly.

### 2.1. Blockchain

Blockchain on the other hand is a revolutionary digital ledger technology that underpins cryptocurrencies like Bitcoin and Ethereum. It serves as a decentralized and tampered-proof database where transactions are recorded in a chain of interconnected blocks. Each block contains a group of transactions, and once added to the chain, it becomes immutable, ensuring transparency and security. Blockchain eliminates the need for intermediaries in various industries, offering efficient and trustless systems for financial transactions, supply chain management, and beyond. Inspired by [1].

### 2.2. Coffee Supply Chain

The coffee supply chain is characterized by its intricacy, involving numerous stakeholders from coffee growers to distributors. Challenges such as fluctuating climate conditions, pest infestations, and quality control demand precise management. Recently studies have delved into the intricacies of coffee bean processing and quality assurance techniques [2]. However, optimizing the efficiency and transparency of this multifaceted supply chain remains a

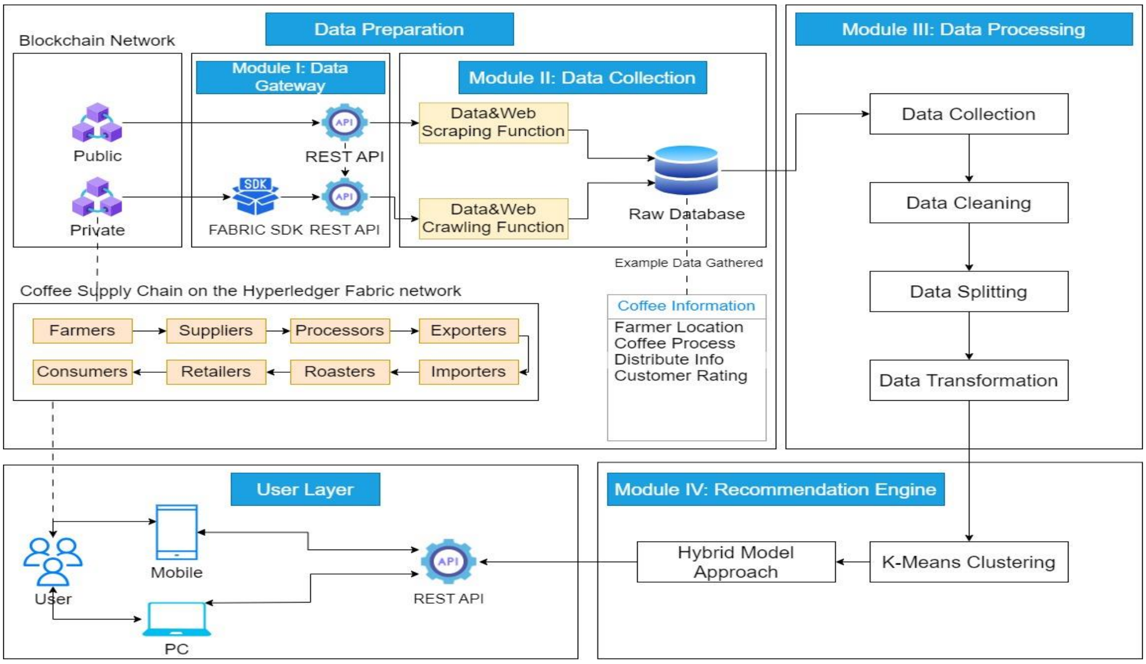


**Figure 1 Overview of BC-Based Coffee SCM Recommendations System**

technical challenge that necessitates innovative approaches, a gap our research endeavors to address.

### 2.3. Machine Learning and Recommendation System

Machine learning and recommendation systems have garnered substantial attention in recent years due to their pivotal role in enhancing user experiences across various domains, including e-commerce, entertainment, and information retrieval. Machine learning techniques, particularly collaborative filtering, content-based filtering, and hybrid methods have been instrumental in developing these systems. Collaborative filtering leverages user behavior and preferences to make personalized recommendations, while content-based filtering relies on item attributes to establish connections between users and items. Hybrid approaches amalgamate both strategies to optimize recommendation accuracy and address cold start problems. Moreover, advancements in deep learning have propelled recommendation systems to new heights by enabling the extraction of intricate patterns and representations from vast datasets. This literature review explores the evolution and integration of machine learning techniques within recommendation systems, highlighting their pivotal role in shaping modern digital experiences. Inspired by [3].

### 2.4. Data Collection and Data Processing

Data collection and processing are vital in supply chain management (SCM), as they provide the foundation for informed decision-making and operational efficiency. Data collection involves gathering information from various supply chain sources, including inventory, demand, logistics, and production schedules. High-quality, timely data collection is essential for accurate SCM strategies. Data processing transforms raw data into valuable insights, employing techniques like analytics and AI to identify patterns and trends. This analytical capability empowers SCM professionals to optimize inventory, streamline distribution, and enhance resilience [4]. The synergy between data collection and processing enables real-time visibility and agility, fostering competitive advantage and operational excellence in modern SCM.

## 3. Proposed System

Figure 1 provides an in-depth view of our proposed BC-Based Coffee Supply Chain Management (SCM) Recommendations System, offering insights into the technological components driving its functionality. At its foundation lies the utilization of blockchain technology, featuring both a public blockchain for transparency and a private blockchain based on Hyperledger Fabric, custom-tailored for coffee supply chain intricacies. The public blockchain serves as a transparent ledger accessible through a REST API link, while the private blockchain relies on both REST API and Fabric SDK to ensure secure and permissioned access to sensitive supply chain data. Our system comprises several modules, each with its specific function [5],[6]. The details are as follows:

- Module 1, known as the Data Gateway, acts as the pivotal bridge between our system and the blockchain networks. It comprises two crucial components: the first, furnishing a REST API link to the public blockchain for real-time access to public coffee supply chain data, and the second, utilizing REST API and Fabric SDK to establish a secure connection with the private blockchain network for retrieving essential supply chain insights.
- Module 2, the Data Collection engine, plays a critical role in aggregating coffee-related data from diverse

sources. It encompasses two primary functions: the Data & Web Scraping Function adept at extracting data from the public blockchain via REST API, and the Data & Web Crawling Function, proficient in accessing the private blockchain network using REST API and Fabric SDK. These functions feed raw data into the Raw Database, where valuable information such as farmer locations, coffee processing details, distribution insights, and customer ratings is meticulously stored.

- Module 3, the Data Processing module, is instrumental in refining and structuring the collected data, although specific data processing steps are not detailed in Figure 1. This preparatory stage readies the data for its journey through the recommendation engine.
- At the core of the system, Module 4, the Recommendation Engine, harnesses the processed data to craft tailored recommendations. It is poised to employ machine learning algorithms and data analysis techniques, offering valuable insights and improvements within the coffee supply chain, to the advantage of both producers and consumers.

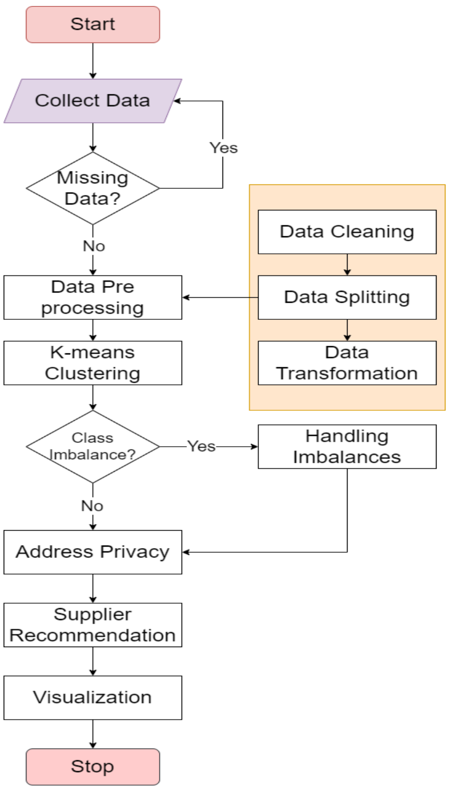


**Figure 2 Data Collection and Data Processing Architecture**

- The User Layer, serving as the interface and interaction hub, enables seamless engagement for coffee producers, distributors, and consumers. Accessible through platforms such as mobile applications and web interfaces, this user-friendly interface ensures effortless communication with the recommendations system via REST API endpoints.

**Algorithm 1** Supplier Recommendation Engine using K-means and Hybrid Model

```
procedure INITIALIZE_CLUSTERS(data, K)
    Initialize cluster centroids randomly or using a heuristic
end procedure
procedure FILTER_CLUSTERS(clusters, filtering_criteria)
    Initialize an empty list filtered_clusters
    for C_i in clusters do
        Compute cluster characteristics (e.g., mean, variance) for C_i
        if Cluster characteristics meet filtering criteria F then
            Add C_i to filtered_clusters
        end if
    end for
    Return filtered_clusters
end procedure
procedure RECOMMEND_SUPPLIERS(user_preferences, filtered_clusters)
    Initialize an empty list recommended_suppliers
    for C_i in filtered_clusters do
        for supplier in C_i do
            if Supplier matches user preferences then
                Add supplier to recommended_suppliers
            end if
        end for
    end for
    Return recommended_suppliers
end procedure
procedure COFFEE_SUPPLY_CHAIN_HYBRID_MODEL(data, K, filtering_criteria, user_preferences)
    Initialize cluster centroids using INITIALIZE_CLUSTERS(data, K)
    prev_centroid ← None
    converged ← False
    while not converged do
        assign, centroid ← K-means clustering(data, cluster centroids)
        if centroids_converged(centroid, prev_centroid) then
            converged ← True
        end if
        prev_centroid ← centroid
    end while
    clusters ← cluster_points(data, assign, K)
    filtered_clusters ← FILTER_CLUSTERS(clusters, filtering_criteria)
    recommended_suppliers ← RECOMMEND_SUPPLIERS(user_preferences, )
    filtered_clusters
    Return recommended_suppliers
end procedure
```

Figure 2 shows the architecture within the proposed system, presenting a technologically detailed perspective. This architecture unfolds as a structured sequence of interrelated components, each with a distinct role in the data processing pipeline. The journey commences with the "Collect Data" phase, where raw data is acquired from diverse sources, initiating the process. Following this, the "Missing Data Detection" decision point assesses the presence of missing values, directing the flow either towards data cleaning to rectify the absence or onward if no gaps exist. The "Data Cleaning" phase rigorously refines the data, addressing duplicates, outliers, and inconsistencies, priming it for subsequent analysis. Subsequently, in the "Data Transformation" phase, the data undergoes critical formatting and feature engineering to suit downstream tasks, particularly machine learning or clustering. Next, "Data Splitting" partitions the dataset into training and testing subsets, facilitating model development and assessment. The detailed process can be found in Algorithm 1.

The core of the architecture resides in "K-means Clustering," where data undergoes segmentation into distinct clusters based on similarities. To ensure data quality, the system evaluates for "Class Imbalance," and if identified, proceeds to "Handling Imbalances," employing techniques such as oversampling or under sampling to redress this disparity. Maintaining a commitment to data privacy, the "Addressing Privacy Concerns" step implements privacy-preserving measures like data masking or encryption.

Leveraging the clustered data and processed information, the system generates "Supplier Recommendations," providing valuable insights or suggestions pertaining to suppliers or data sources. Finally, the "Visualization" component transforms the data into visual representations that facilitate exploration and decision-making. This holistic architecture orchestrates the entire data processing workflow, meticulously addressing data quality, privacy, and class balance concerns at every stage, ultimately enhancing the system's overall effectiveness and trustworthiness. Inspired by [7],[8].

## 4. Conclusion

In conclusion, the research presented in this study introduces a groundbreaking approach to revolutionize decision-making within the coffee supply chain through the integration of blockchain technology, data processing, and machine learning. Our proposed BC-Based Coffee Supply Chain Management Recommendations System offers a comprehensive solution that leverages the transparency of public blockchains and the security of private blockchains to extract, collect, process, and analyze vital data from diverse stakeholders. This data is then used to create a robust recommendation engine that empowers stakeholders to make informed decisions, optimizing the selection of coffee suppliers for each customer.

The visual representation of clustering patterns generated by the K-Means algorithm adds a valuable dimension to the study, making complex findings accessible and aiding stakeholders in understanding the dynamics of the coffee supply chain. Furthermore, the detailed architecture of the data collection and processing pipeline demonstrates a commitment to data quality, privacy, and addressing class imbalances at every stage.

In a rapidly evolving coffee supply chain landscape, where efficiency and informed decision-making are paramount, our research offers a promising solution. By enhancing operational efficiency and providing stakeholders with actionable insights, this blockchain-enabled clustering approach has the potential to drive positive changes in the coffee industry, benefiting producers, distributors, and consumers alike. As the coffee supply chain continues to evolve, this research serves as a foundation for further advancements and innovations in data-driven decision support.

**Acknowledgement**

This research was supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (201803932), and supported by the MSIT (Ministry of Science and ICT), Korea, under the Special R&D Zone Development Project (R&D) - Development of R&D Innovation Valley support program (2023-DD-RD-0152) supervised by the Innovation Foundation.

## References

[1] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017, June). An overview of blockchain technology: Architecture, consensus, and future trends. In 2017 IEEE international congress on big data (BigData congress) (pp. 557-564). Ieee.

[2] Villalva-Cataño, A., Ramos-Palomino, E., Provost, K., & Casal, E. (2019, October). A model in agri-food supply chain costing using ABC costing: an empirical research for Peruvian coffee supply chain. In 2019 7th International Engineering, Sciences and Technology Conference (IESTEC) (pp. 1-6). IEEE.

[3] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.

[4] Sukati, I., Hamid, A. B., Baharun, R., & Yusoff, R. M. (2012). The study of supply chain management strategy and practices on supply chain performance. Procedia-Social and Behavioral Sciences, 40, 225-233.

[5] Abbas, K., Afaq, M., Ahmed Khan, T., & Song, W. C. (2020). A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. Electronics, 9(5), 852.

[6] Himeur, Y., Sayed, A., Alsalemi, A., Bensaali, F., Amira, A., Varlamis, I., ... & Dimitrakopoulos, G. (2022). Blockchain-based recommender systems: Applications, challenges and future opportunities. Computer Science Review, 43, 100439.

[7] Che, Z. H., & Wang, H. S. (2010). A hybrid approach for supplier cluster analysis. Computers & Mathematics with Applications, 59(2), 745-763.

[8] Che, Z. H. (2010). A two-phase hybrid approach to supplier selection through cluster analysis with multiple dimensions. International Journal of Innovative Computing, Information and Control, 6(9), 4093-4111.