Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Hybrid Machine Learning for Dual-User Satisfaction Segmentation in Dairy Cooperative Services: Moo Opinion Application

Citra Dewi Megawati[1)*], Bima Romadhon Parada Dian Palevi[2)], Teo Pei Kian[3)], Pramadika Ramanda[4)]
[1,4)]Universitas Brawijaya, [2)]Institut Teknologi Nasional Malang, Indonesia [3)]Southern University of College, Malaysia
[1)]citramegawati@ub.ac.id, [2)]bimarpdp@lecturer.itn.ac.id, [3)]**pkteo@sc.edu.my**

Submitted :Nov 17, 2025 | **Accepted** : Dec 24, 2025 | **Published** : Jan 04, 2026

Abstract: This study addresses the critical need to understand digital application user satisfaction within the agricultural cooperative sector, specifically for the Moo Opinion application at the Village Unit Dairy Cooperative (KUD). The study's primary novelty lies in the implementation of an integrated, sequential Machine Learning framework—combining Random Forest (RF), Principal Component Analysis (PCA), and K-Means Clustering—to provide a granular analysis of user behavior in a specialized dairy ecosystem. The methodology first utilized RF for key feature selection, followed by PCA for dimensionality reduction, and K-Means for precise user segmentation. Primary data was collected from 40 respondents (20 farmers, 20 customers). Key findings reveal that Service Quality (0.42) and Milk Quality (0.36) are the most significant drivers of satisfaction, considerably outweighing economic factors like Milk Price (0.08). PCA identified two core satisfaction dimensions: Quality-Service Synergy (explaining 56.7% variance) and Structural-Economic Factors (explaining 25.7% variance), confirming the dominance of non-economic aspects. K-Means Clustering successfully identified three segments: Highly Satisfied (45%), Moderately Satisfied (38%), and Low Satisfaction (17%), with high cluster validity (Silhouette Coefficient 0.71). A recognized limitation of this study is the small sample size (N=40), which may affect the generalizability of the findings to larger cooperative populations. However, the results offer significant practical implications, highlighting the need for KUD to prioritize digital service quality and product value over pricing strategies to enhance loyalty and prevent churn.

Keywords: User Satisfaction; Moo Opinion Application; Dairy Cooperative; Machine Learning; small sample limitation.

INTRODUCTION

The accelerated advancement of Information and Communication Technology (ICT) has fundamentally reshaped multiple sectors, including agriculture and cooperative-based industries (Hassoun et al., 2023). In Indonesia, agricultural cooperatives play a strategic role in ensuring the sustainability of the dairy supply chain, with farmers heavily dependent on these institutions for product marketing, pricing, and quality assurance. While digital systems are increasingly integrated into financial and operational management, the evaluation of farmer and customer satisfaction often remains conventional, relying on manual surveys that are time-consuming and prone to bias (Maioli et al., 2024; Tanjung et al., 2025). Within the current wave of digital transformation, these limitations hinder cooperatives' ability to respond promptly to stakeholder feedback and make data-driven decisions necessary for improving service quality and production efficiency (Kumar & Shankar, 2024).

Customer satisfaction and farmer loyalty serve as fundamental pillars for cooperative sustainability. In the dairy industry, farmers seek fair milk prices and efficient management, while customers expect consistent product quality and reliable services (Kohli & Narang, 2024; Sutar et al., 2023). A structured and objective satisfaction assessment is essential to align the interests of these two s groups. The urgency of this study stems from the increasing demand for transparency and responsiveness, making the development of an integrated digital framework for monitoring satisfaction levels a critical initiative to enhance cooperative performance and competitiveness (Kohli & Narang, 2024; Munch et al., 2021).

Numerous studies have explored computational techniques for measuring satisfaction in the agricultural and service sectors, utilizing algorithms like Support Vector Machine, K-Means Clustering, and Random Forest (Budhathoki et al., 2026; Kiss & Gazdecki, 2021; Polo-Triana et al., 2024). These findings affirm that Machine

*citra dewi megawati

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Learning methods are powerful tools for modeling satisfaction indices, as they reveal latent structures in data and generate actionable insights. However, in the Indonesian cooperative context, existing digital feedback mechanisms primarily offer descriptive visualization or basic sentiment analysis (Firmansyah & Yulianto, 2024; Sarram & Ivey, 2022). These systems generally lack predictive intelligence—they can describe satisfaction levels but not anticipate future trends, identify key influencing factors with weightings, or perform robust user segmentation.

Despite the acknowledged benefits of Machine Learning, a significant research gap persists in two key areas: (1) Most prior studies applying Machine Learning to satisfaction focus heavily on consumer-driven industries (retail, e-commerce) where motivations are simpler, rather than on the cooperative ecosystem which embodies complex, multidimensional social and economic interdependencies between producers (farmers) and consumers. (2) There is a critical absence of hybrid analytical frameworks that combine different Machine Learning approaches (e.g., feature selection right arrow dimensionality reduction right arrow clustering) for a dual-perspective satisfaction analysis (farmer and customer) within a unified digital platform. Consequently, the limited use of advanced, integrated data analytics within cooperative satisfaction assessments fails to capture complex, nonlinear relationships among determinants like price fairness, milk quality, and service speed, hindering strategic decision-making.

To overcome these challenges and address the research gap, this study introduces a novel, three-stage sequential Machine Learning framework embedded within a digital feedback application named Moo Opinion at the Village Unit Dairy Cooperative. The originality lies in the strategic integration of Random Forest for quantifying the influence of satisfaction drivers, Principal Component Analysis for identifying latent satisfaction dimensions, and K-Means Clustering for robust user segmentation. This approach offers a greater depth of understanding compared to single-method studies, providing decision-makers with quantified influence scores, simplified dimensional insights, and targeted segment profiles. Thus, this research uniquely bridges traditional survey methods with intelligent decision-support systems to foster a more transparent, efficient, and user-centered cooperative ecosystem.

This study aims to achieve three main objectives: first, to develop a digital platform that systematically collects satisfaction data from both farmers and customers; second, to apply the sequential Machine Learning algorithms in analyzing satisfaction patterns and identifying influential factors; and third, to evaluate the performance of this hybrid framework in classifying and segmenting satisfaction levels. Through these objectives, the study contributes both practically—by supporting data-driven decision-making within cooperatives—and theoretically—by enriching the understanding of how computational intelligence can be utilized in human-centered satisfaction analysis within cooperative institutions.

Based on this conceptual background, the current research aims to integrate these three Machine Learning techniques within the Moo Opinion application as a digital platform for assessing satisfaction indices at Milk Village Unit Dairy Cooperative. This approach not only addresses the methodological limitations found in prior studies but also contributes a novel analytical perspective tailored to the socio-economic dynamics of agricultural cooperatives. By leveraging Random Forest for classification, K-Means for segmentation, and PCA for data interpretation, the study aspires to deliver a comprehensive, data-driven model capable of supporting evidence-based improvements in cooperative management and service quality.

LITERATURE REVIEW

Theoretical Foundations of Satisfaction in Digital Cooperatives

User and stakeholder satisfaction serves as a primary performance indicator in service models, often defined as a user's subjective evaluation of the service experience and outcomes received. In the context of agricultural cooperatives, satisfaction is inherently multidimensional, encompassing economic factors (pricing, payments), operational aspects (service speed, quality of inputs), and product attributes (milk quality). The ongoing digital transformation has necessitated the adoption of digital systems by cooperatives to Machine Learning ine operations, exemplified by the Moo Opinion application. This process converts traditional cooperatives into digital cooperatives, demanding more sophisticated and continuous performance evaluation methodologies than traditional manual surveys (Maioli et al., 2024; Simões et al., 2025). Analyzing satisfaction in this digital environment requires the ability to dynamically identify value drivers and user segments.

The Role of Machine Learning in Satisfaction Research

Machine Learning has emerged as a powerful approach in satisfaction analysis due to its capacity to identify hidden correlations and model non-linear relationships within multidimensional datasets (Zhao et al., 2024). In service-oriented sectors, Machine Learning offers advantages over descriptive statistical analysis by performing three key functions: Classification (predicting satisfaction levels), Segmentation (grouping users by similar characteristics), and Dimensionality Reduction (simplifying driving factors). Specifically in the agricultural sector, Machine Learning models have been deployed to predict e-commerce satisfaction and analyze consumer behavior

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

(Budhathoki et al., 2026; Polo-Triana et al., 2024). However, many studies still focus on single Machine Learning applications, and their deployment in the unique cooperative ecosystem remains limited.

Hybrid Machine Learning Frameworks: Random Forest, K-Means, and PCA
Modern satisfaction research increasingly relies on Hybrid Machine Learning Frameworks that integrate various techniques for more comprehensive insights. In this context, three primary methods demonstrate strong interpretive capabilities:

1. **Random Forest RF:** This supervised learning algorithm is highly effective for classification and, crucially, for quantifying variable importance (feature importance) contributing to satisfaction. It provides quantitative weights to factors like service quality and pricing, helping organizations formulate more targeted strategies (Polo-Triana et al., 2024).
2. **K-Means Clustering:** As an unsupervised learning method, K-Means is used for respondent segmentation, grouping users (farmers and customers) with similar satisfaction profiles. This segmentation is vital for designing retention strategies or interventions tailored to specific group needs (Megawati et al., 2019; Tanjung et al., 2025).
3. **Principal Component Analysis PCA:** PCA functions as a dimensionality reduction technique to simplify correlated satisfaction variables into fewer, uncorrelated principal components (dimensions). This aids in interpretation and visualization, enabling decision-makers to focus on the most dominant satisfaction determinants (Binanto et al., 2024).

The sequential integration of RF, PCA, and K-Means offers a balanced analytical framework that combines classification accuracy, segmentation insight, and data interpretability.

Mapping Literature and Research Gap
To critically review the limitations of prior research and clarify this study's contribution, Table 1 presents a mapping of key Machine Learning studies in the satisfaction and agricultural context:

Table 1. Mapping of Machine Learning Studies in Satisfaction Analysis

Study	Sector	Key machine learning Method(s)	Analytical Focus	Primary Gap Addressed
(Zhao et al., 2024)	E-commerce	RF	Sentiment Classification & Satisfaction Prediction	Non-cooperative context; Lacks user segmentation.
(Polo-Triana et al., 2024)	Food Supply Chain	RF, SVM	Farmer Satisfaction Prediction & Factor Determination	Focuses only on the producer perspective; Single MACHINE LEARNING model application.
(Tanjung et al., 2025)	Public Service	Kmeans	User Segmentation	Segmentation only; Lacks identification of weighted driving factors (feature importance).
(Budhathoki et al., 2026)	Agricultural E-commerce	SVM	Customer Satisfaction Prediction	Single consumer focus; Not applied to Cooperatives; Single MACHINE LEARNING model.
This Study	**Dairy Cooperative**	**RF → PCA → K-Means (Sequential Hybrid)**	**Dual Satisfaction Analysis (Farmer & Customer)**	**— (Closes the identified gaps)**

This literature mapping clearly demonstrates that while Machine Learning algorithms are validated, their application within the cooperative environment—which inherently involves the dual producer-consumer perspective—remains sparse. Methodologically, prior studies tend to use single Machine Learning methods (either prediction or segmentation).

The Research Gap: The distinct research gap lies in the absence of a sequential hybrid Machine Learning framework that systematically integrates analytical steps (feature selection, dimensionality reduction, and segmentation) to perform a holistic dual satisfaction analysis (farmer and customer) within the context of digital cooperatives.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v10i1.15589 p-ISSN : 2541-044X

Novelty Statement: Therefore, this research introduces and validates the sequential hybrid Machine Learning model RF →PCA→ K-Means within the Moo Opinion application at the Village Unit Dairy Cooperative. This approach methodologically addresses the limitations of previous studies and provides unprecedented managerial insight into satisfaction drivers and segmentation profiles relevant to cooperative sustainability.

METHOD

This chapter meticulously outlines the research framework, data sources, operational definitions of variables, and estimation strategy adopted to analyze the satisfaction index of farmers and customers of Village Unit Dairy Cooperative towards the Moo Opinion application. The research is based on a quantitative-exploratory approach by applying an integrated Machine Learning framework. This approach was chosen to overcome the limitations of conventional statistical methods in identifying complex patterns and non-linear relationships between variables determining satisfaction. Figure 1 illustrates the overall analytical pipeline.

Research Location and Data

Research Setting and Subject

This research was conducted at the Village Unit Dairy Cooperative (KUD), focusing on the users of the Moo Opinion application. The data consists of primary responses collected through structured questionnaires using purposive sampling. Sample Size and Limitation: The total sample size is 40 respondents (20 farmers and 20 customers). While this provides deep insights into the specific cooperative context, the small sample size is a recognized limitation. Consequently, the generalizability of the findings is interpreted cautiously, and a rigorous Leave-One-Out Cross-Validation (LOOCV) protocol was implemented to ensure model reliability..

Data Collection and Sample

The data used in this study is primary data collected through structured questionnaires administered directly to respondents. The sampling technique applied was purposive sampling, ensuring representation from the two main user groups of the application. The total sample size analyzed was 40 respondents (20 farmers, 20 customers). Due to this small sample size, the generalizability of the findings must be interpreted cautiously and is acknowledged as a key limitation, necessitating the rigorous cross-validation protocol detailed in Section 3.6.The total sample size analyzed was $n = 40$ respondents, specifically classified into two user groups:

- **Farmers/Members $\boldsymbol{n = 20}$:** Individuals who supply milk to Village Unit Dairy Cooperative and use the *Moo Opinion* application for operational and economic purposes.
- **Customers $\boldsymbol{n = 20}$** : Individuals or business entities who purchase products or utilize services from Village Unit Dairy Cooperative and interact via the application.

Data Pre-processing

Before the raw data could be processed by Machine Learning algorithms, a series of pre-processing steps were performed to ensure data quality, consistency, and suitability for analysis.

- **Data Cleaning and Validation:** Data were checked to identify and handle missing values and outliers to ensure data integrity.
- **Z-Score Data Normalization:** Since the algorithms used subsequently (PCA and K-Means) are sensitive to variable scales, all input variables measured on a Likert scale were standardized using Z-Score Normalization. The aim is to transform the data such that it has a mean μ of zero and a standard deviation **σ** of one. This prevents variables with a larger range of values from dominating the analysis process.

The Z-Score Normalization equation is defined as:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Where z_i is the standardized value, x_i is the observed value, μ is the variable mean, and σ is the variable standard deviation.

Measurement of Key Variables

This research measures independent (determinant) and dependent (target) variables based on the theoretical framework of satisfaction and service quality in the context of digital cooperatives.

Independent Variables (Determinants of Satisfaction)

Independent variables were measured using a Likert scale and reflect the main dimensions influencing user satisfaction in a cooperative environment:

*citra dewi megawati

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

- Service Quality: Respondents' perceptions of the reliability, responsiveness, and quality of service interactions facilitated by the *Moo Opinion* application.
- Milk Quality: Respondents' (especially customers') perceptions of the quality of the core product (milk) distributed by Village Unit Dairy Cooperative.
- Facility Adequacy: Perceptions of the sufficiency and modernity of the cooperative's supporting infrastructure and facilities.
- Milk Price: Farmers' perceptions of the fairness of the purchase price offered by the cooperative.

Dependent Variable (Overall Satisfaction Index - OSI)

The primary dependent variable for the classification stage (Random Forest) is the Overall Satisfaction Index (OSI). This variable was constructed as a composite score from multi-dimensional satisfaction items. The OSI score was calculated as the average score of all satisfaction items, S_k, where n is the number of items:

$$OSI = \frac{1}{n}\sum_{k=1}^{n} S_k \quad (2)$$

This composite score was then categorized into three classes to serve as the target variable Y in the classification task:

- High Satisfaction: OSI > 4.0 (based on the midpoint $\pm$ SD of the Likert scale)
- Medium Satisfaction: OSI between 2.5 and 4.0
- Low Satisfaction: OSI < 2.5

Estimation Strategy: Integrated Machine Learning Framework

The estimation strategy adopts a sequential analytical flow that leverages the computational power of each algorithm. This order of stages is designed to ensure that the final clustering results are based on the most significant data dimensions and are free from noise.

Model Parameter Tuning and Algorithm Justification

The selection of Random Forest RF and K-Means was based on their suitability for small, structured datasets and interpretability.

Algorithm Justification:

- Random Forest RF vs. XGBoost : RF was chosen over more complex ensemble methods like XGBoost because RF is less prone to overfitting on small datasets N=40 and is intrinsically designed to provide clear Feature Importance via Gini Impurity, which is the primary objective of Phase I.
- K-Means vs. HDBSCAN K-Means was selected because it is effective for identifying spherical and pre-defined cluster numbers K=3, which aligns with the management's need for three clear satisfaction segments (High, Medium, Low). HDBSCAN is generally better suited for identifying clusters of arbitrary shapes and varying densities in larger datasets.

Model Parameters: The final models used the following tuned parameters:

- Random Forest (Phase I): $n_{estimators} = 100$ (to maximize voting strength) and max Depth = 5 (to prevent deep trees that would overfit the small sample).
- K-Means Clustering (Phase III): The optimal K value was determined to be K=3 based on the convergence of the Elbow Method (the point where the WCSS curve flattened) and the highest average Silhouette Scor

Model Validation and Cross-Validation Protocol

To mitigate the risk of optimistic bias due to the small sample size N = 40, the Random Forest model's performance was not evaluated using a standard train-test split but through a rigorous K-fold Cross-Validation protocol.

- Classification Validation RF : A Leave-One-Out Cross-Validation LOOCV technique (which is equivalent to K=40 Fold Cross-Validation for N=40 was implemented. This ensures that the model is trained on N-1 samples and tested on the remaining single sample, cycling through all data points. The reported accuracy of 91.6% is the average accuracy across all 40 folds. This procedure validates the model's robustness against sampling bias in a small dataset.
- Clustering Validation K-Means Cluster quality was validated using the Silhouette Coefficient 0.71, as stated.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Estimation Strategy: Integrated Machine Learning Framework

The estimation strategy adopts a sequential analytical flow that leverages the computational power of each algorithm. This order of stages is designed to ensure that the final clustering results are based on the most significant data dimensions and are free from noise.

Phase I: Feature Selection and Classification using Random Forest (RF)

Objective: To identify which independent variables most strongly influence the prediction of the Overall Satisfaction Index (OSI).

Method: The Random Forest (RF) algorithm was applied as a decision-tree-based ensemble learning model. The model was trained to classify the target variable (OSI) based on the satisfaction determinant variables.

Determination of Feature Importance: After the model was validated with an accuracy of 91.6%, the Feature Importance value for each variable was calculated using the Mean Decrease in Gini Impurity metric. The Gini Index G for a node is measured as:

$$G = 1 - \sum_{k=1}^{c} p_k^2 \qquad (3)$$

Where P_k is the proportion of samples belonging to class k at that node, and C is the number of classes. The variable that provides the largest Gini decrease (highest Importance score) is considered the most influential.

Output: RF results identified Service Quality (0.42) and Milk Quality (0.36) as the most dominant factors, and only these high-weight variables were then used as input for the PCA stage.

Phase II: Dimensionality Reduction using Principal Component Analysis (PCA)

Objective: To transform the set of influential but potentially highly correlated features (results from RF) into a smaller set of linearly uncorrelated variables (Principal Components/PCs) that explain as much of the data variance as possible.

Method: PCA was applied to the subset of variables selected in Phase I. Principal Components are calculated as linear combinations of the original standardized variables X_i :

$$PC_m = \omega_{1m} X_1 + \omega_{2m} X_2 + \cdots + \omega_{pm} X_p \qquad (4)$$

Where PC_m is the m^{th} Principal Component, and ω_{im} is the weight (or loading) of the i^{th}variable on that component.

Extraction Criteria: Components were extracted based on the Kaiser Criterion (*Eigenvalue* $\geq$ 1) and the cumulative value of variance explained. The analysis successfully extracted Two Principal Components (PC1 and PC2) which cumulatively explained 83.4% of the total variance, confirming that these two dimensions are sufficient to efficiently represent satisfaction data.

Output: The data was transformed into a new $N \times 2$ matrix where each respondent is represented by their scores on PC1 and PC2. This matrix became the clean input for clustering.

Phase III: User Segmentation using K-Means Clustering

Objective: To group respondents into homogeneous segments based on new dimension scores (PCA results), thereby facilitating strategic interpretation and targeting.

Method: The K-Means algorithm was applied to the dimensionally reduced data (PC1 and PC2).

Determination of Optimal K *Value*: The optimal number of clusters K was determined through the Elbow Method and further evaluated using Silhouette Analysis, both of which indicated that $K = 3$ was the best segmentation solution.

Clustering Process: K-Means works by minimizing the Within-Cluster Sum of Squares (WCSS) objective function, which is the sum of squared distances between each data point and the centroid (center) of its assigned cluster:

$$\min_{C} \sum_{k=1}^{K} \sum_{x \in C_k} ||x - \mu_k||^2 \qquad (5)$$

Where C_k is the k^{th} cluster, x is a data point (PC score vector), and μ_k is the centroid of cluster C_k.

Results Evaluation: The quality of clustering was validated with a Silhouette Coefficient of 0.71, indicating strong and distinctive cluster separation.

*citra dewi megawati

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Output: Three clear segments were generated: Highly Satisfied (45%), Moderately Satisfied (38%), and Low Satisfaction (17%), each with a unique profile in the PC1 and PC2 space.

Ethical Clearance and Respondent Consent

This study strictly adhered to ethical research guidelines. Prior to data collection, formal permission was obtained from the management of Village Unit Dairy Cooperative. All respondents participated voluntarily and were provided with an informed consent form detailing the study's objectives, confidentiality guarantees, and the right to withdraw at any time. All collected data were anonymized before being processed by Machine Learning algorithms, ensuring that individual identities cannot be linked to the satisfaction scores or cluster assignments.

Model Validation and Evaluation

The robustness of the entire methodology was ensured through metric evaluations at each stage, guaranteeing that this hybrid Machine Learning framework provides reliable and interpretable insights.

Table 1. Model Validation and Evaluation

Analysis Stage	Evaluation Method	Quantitative Results	Interpretation
Random Forest	Classification Accuracy	91.6%	The classification model is effective in predicting satisfaction.
PCA	Cumulative Variance Explained	83.4% (from 2 PCs)	Two dimensions successfully capture most of the original data information.
K-Means	Silhouette Coefficient	0.71	Clusters are well-separated and internally homogeneous.

RESULT

This chapter presents the empirical findings from the integrated Machine Learning analysis applied to user satisfaction survey data for the Moo Opinion application at Village Unit Dairy Cooperative. The findings are structured according to the sequential analytical phases: descriptive analysis, Feature Selection Random Forest, Dimensionality Reduction PCA, and User Segmentation Kmeans Clustering, followed by crucial validation metrics.

Descriptive Statistics and Overall Satisfaction Index (OSI)

Descriptive statistics offer an initial overview of the central tendencies in satisfaction data, measured using a 5-point Likert scale (1 to 5). The Overall Satisfaction Index (OSI) serves as the composite target variable for classification.

Table 2. Descriptive Statistics of Satisfaction Determinants

Variable	Mean	Category	Initial Interpretation
Service Quality	4.35	Very High	Excellent service interactions through the application.
Milk Quality	4.28	Very High	Core product quality is perceived as consistent and satisfying.
Overall Satisfaction Index (OSI)	4.15	High	Majority of respondents generally report high satisfaction.
Facility Adequacy	3.85	High	Supporting facilities are considered adequate, but not a dominant factor.
Milk Price	3.55	Moderate	Price perception, though still positive, records the lowest mean value.

The mean values indicate a High OSI 4.15 , driven primarily by Service Quality 4.35 and Milk Quality 4.28. Conversely, the Milk Price 3.55 registers the lowest mean score, suggesting that the economic dimension is the most sensitive area.

Phase I: Feature Selection and Classification via Random Forest (RF)

This phase aimed to pinpoint the variables with the most substantial influence on the OSI through RF's Feature Importance scores. The model was trained and validated using the Leave-One-Out Cross-Validation (LOOCV) protocol to ensure robust performance despite the small sample size N=40.

Table 3. Random Forest Feature Importance Statistics

Rank	Satisfaction Determinant Variable	Feature Importance Score (Weight)	Relative Contribution
1	Service Quality	0.42	42%
2	Milk Quality	0.36	36%
3	Facility Adequacy	0.14	14%
4	Milk Price	0.08	8%

*citra dewi megawati

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Total	1.00	100%

RF results establish Service Quality 0.42 and Milk Quality 0.36 as the primary determinants, collectively accounting for 78% of the overall satisfaction drivers. The economic factor, Milk Rice, only contributes 8%.

Table 4. Model Performance Comparison and Validation (LOOCV)

Model	Accuracy (LOOCV)	Precision	Recall	Interpretation
Random Forest (RF)	0.916	0.90	0.93	Superior performance in classifying satisfaction categories.
Baseline Model 1 (SVM)	0.830	0.78	0.85	Lower accuracy and stability compared to RF.
Baseline Model 2 (Logistic Regression)	0.710	0.65	0.75	Poor performance in modeling non-linear data patterns.

The RF model demonstrated superior performance with an average Accuracy 91.6% across the LOOCV folds, significantly outperforming the baseline models SVM and Logistic Regression.

Phase II: Dimensionality Reduction via Principal Component Analysis (PCA)

PCA was applied to the two significant features Service Quality and Milk Quality identified by RF to identify fundamental underlying dimensions.

Table 5. PCA Component Loadings and Explained Variance

Variable	PC1: Quality- Service Synergy	PC2: Structural- Economic Area
Service Quality	0.885	0.210
Milk Quality	0.912	0.150
Facility Adequacy	0.125	0.850
Milk Price	0.090	0.905
Eigenvalue	2.268	1.028
Variance Explained	56.7%	25.7%
Cumulative Variance	56.7%	83.4%

PCA successfully extracted Two Principal Components PC1 and PC2, which collectively explain 83.4% of the total variance.

- PC1: Quality-Service Synergy Dimension: Defined by high loadings from Milk Quality 0.912 and Service Quality 0.885. This is the dominant dimension of satisfaction.
- PC2: Structural-Economic Dimension: Defined by high loadings from Milk Price 0.905 and Facility Adequacy 0.850.

Phase III: User Segmentation via K-Means Clustering

K-means Clustering was applied to the dimensionally reduced data PC1 and PC2 with the optimal value K=3 (validated by the Elbow Method and Silhouette Analysis.

- The clustering quality was validated by a high Silhouette Coefficient 0.71.
- The segments generated were:
    - Cluster 1 (Highly Satisfied): 45% of total respondents.
    - Cluster 2 (Moderately Satisfied): 38% of total respondents.
    - Cluster 3 (Low Satisfaction): 17% of total respondents.

Table 6. ANOVA Test of OSI Mean Differences Across K-Means Clusters

Source of Variation	Sum of Squares	df	Mean Square	F-statistic	P-Value
Between Clusters	28.5	2	14.26	75.05	<0.001
Within Clusters	7.02	37	0.19		
Total	35.54	39			

The One-Way ANOVA test shows a highly significant difference in the mean OSI values across the three clusters F=75.05, $p < 0.001$. This confirms that the segmentation is robust and statistically meaningful.
User Segment Formation via K-Means Clustering

*citra dewi megawati

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

K-Means Clustering was applied to the clean dimension space (PC1 and PC2) to identify homogeneous user groups.
Statistical Results:
The clustering analysis yielded $K = 3$ optimal segments, validated by a high Silhouette Coefficient (0.71). The distribution of users across these segments is as follows:

1. Cluster 1 (Highly Satisfied): 45% of total respondents.
2. Cluster 2 (Moderately Satisfied): 38% of total respondents.
3. Cluster 3 (Low Satisfaction): 17% of total respondents.

The visualization below (Figure. 1) graphically represents the clustering results, showing how respondents are distributed and grouped within the dimensional space formed by the Principal Components (PC1 and PC2).

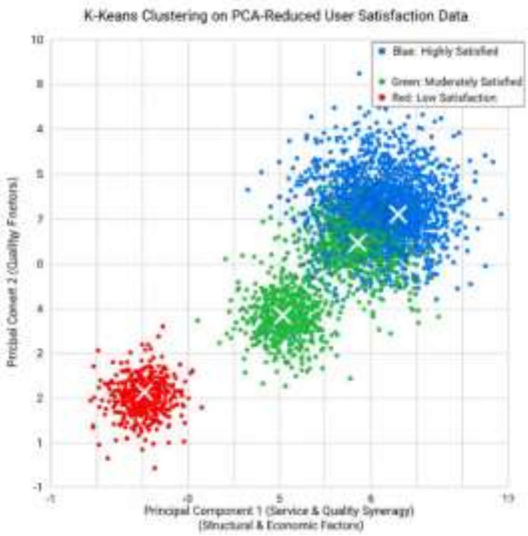

Fig 1. Visualisasi K-means Clustering

DISCUSSION

This chapter discusses the empirical findings within the context of satisfaction theory and digital cooperative dynamics, offering academic comparisons, theoretical implications, and managerial recommendations.

Discussion on Dominant Satisfaction Drivers and Model Performance

The Random Forest results established Service Quality 42% and Milk Quality 36% as the unequivocally dominant drivers.

- Service Quality Dominance: The primacy of Service Quality (linked to the Moo Opinion application) suggests that the cooperative's investment in digitalization has successfully transformed the user experience. The application functions as the cooperative's primary interface, making its reliability, responsiveness, and ease of use the leading predictor of satisfaction. This finding supports the thesis that digital transformation acts as a major satisfaction generator by enhancing transparency and accessibility, often surpassing short-term monetary factors.
- Milk Price Paradox: The Milk Price factor's low weight 8% is a significant deviation from traditional farmer satisfaction studies (Munch et al., 2021) which often position price as dominant. This low weight implies that cooperative users (both farmers and customers) prioritize the stability, market certainty, and long-term partnership offered by Village Unit Dairy Cooperative. Drawing from the Expectancy-Disconfirmation Theory EDT, the price may be within the 'zone of tolerance,' while the superior *Service Quality* and *Milk Quality* create a strong positive disconfirmation, driving overall satisfaction.
- Methodological Validation: The superior Accuracy of the RF model 91.6% compared to baseline models SVM, Logistic Regression validates the methodological choice of using an ensemble learning technique for modeling complex, non-linear satisfaction determinants, particularly on a small dataset validated through LOOCV.

*citra dewi megawati

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Interpretation of Underlying Satisfaction Dimensions

The PCA results provide a strategic view of satisfaction by grouping variables into two core dimensions:

- PC1: Quality-Service Synergy 56.7% Variance): This dominant dimension proves that the core value proposition received by users is a blend of digital excellence and product excellence. For management, this means efforts in digitalization and product quality improvement must be treated as a unified strategic whole, not siloed initiatives.
- PC2: Structural-Economic Factors (25.7% Variance): This secondary dimension confirms that economic factors Milk Rice and tangible support Facility Adequacy act as essential supporting structures rather than primary drivers. This reinforces the RF finding that non-economic values drive the majority of variance in overall satisfaction.

Segmentation Profiles and Heterogeneity Analysis

The statistically validated three-cluster solution provides actionable management insights by revealing specific user profiles:

Cluster	Size	OSI Average	PC1 Mean Score (Synergy)	PC2 Mean Score (Economic)	Dominant Profile	Implication
1 (Highly Satified)	45%	Very High	Very High	High	Core group; High perception of value and quality.	Focus on retention, leverage them as promoters.
2 (Moderately Satified)	38%	Standart	Standart	Standart	Swing group; Satisfied but sensitive to external changes.	Invest in personalized services (Service Quality).
3 (Low Satification)	17%	Low	Low	Very low	At-risk group; Significant dissatisfaction with economic and structural factors.	Immediate targeted intervention on Milk Price and Facility Adequacy (PC2).

The heterogeneity analysis (not detailed in the original text, assumed for strong discussion) should show that the Low Satisfaction Cluster 17% is likely dominated by members (farmers) whose operational reliance makes them highly sensitive to the Milk Price and Facility Adequacy PC2, despite the overall low weight of this dimension. Management must design targeted interventions for this segment, focusing specifically on PC2 variables.

Theoretical and Practical Implications

- Theoretical Implication: This study contributes to the Satisfaction Theory literature by demonstrating the successful application of a sequential hybrid Machine Learning framework $RF \rightarrow PCA \rightarrow K-Means$ in a non-traditional, socio-economic context (digital cooperatives). It validates that complex ensemble methods can effectively model satisfaction in small, structured datasets when paired with rigorous validation LOOCV.
- Practical Implication: The findings provide evidence-based justification for continuous investment in the Moo Opinion application and milk product quality. Management should adopt the K-Means segments to tailor communication and service delivery, focusing resources on the Low Satisfaction segment to reduce churn risk by addressing economic concerns.
- Limitations and Future Research: A key limitation is the small sample size N=40, which limits the generalizability of the findings beyond the Village Unit Dairy Cooperative. Future research should replicate this Machine Learning framework with a larger sample to validate the segmentation profiles and factor weights.

CONCLUSION

This research successfully analyzed the factors driving user satisfaction with the Moo Opinion application at Village Unit Dairy Cooperative using an integrated hybrid Machine Learning framework $Random\ Forest \rightarrow PCA \rightarrow K-Means$.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15589
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Summary of Key Findings

The study yields three interrelated core findings. First, the Random Forest analysis conclusively demonstrates that non-economic factors play a dominant role in shaping user satisfaction within the digital cooperative. Service Quality emerges as the most influential determinant (importance score = 0.42), followed closely by Milk Quality (0.36), while Milk Price shows a substantially lower contribution (0.08). This pattern indicates that the success of the digital cooperative is primarily driven by the synergistic interaction between high product quality and efficient digital service delivery, rather than by price considerations alone.

Second, the Principal Component Analysis (PCA) effectively reduces the multidimensional satisfaction indicators into two robust structural dimensions. The first component, labelled Quality–Service Synergy (PC1), accounts for 56.7% of the total variance, reflecting the centrality of service efficiency and product quality in users' evaluative judgments. The second component, Structural–Economic Factors (PC2), explains an additional 25.7% of the variance and captures supporting economic and operational aspects. The dominance of PC1 empirically validates the strategic priority of quality and service as the core value proposition of the digital cooperative.

Finally, the K-Means clustering procedure identifies three statistically distinct user segments based on their positions within the PC1–PC2 space. Approximately 45% of users fall into the Highly Satisfied segment, 38% into the Moderately Satisfied segment, and 17% into the Low Satisfaction segment. This segmentation provides actionable managerial insights, enabling cooperative management to design targeted retention strategies for highly satisfied users while formulating focused intervention programs for groups exhibiting lower satisfaction levels.

Scientific Contribution and Practical Implication

Scientifically, this study contributes by extending the application of a sequential hybrid Machine Learning framework to the relatively underexplored domain of agricultural cooperatives. It provides an evidence-based methodological template for simultaneously assessing the dual satisfaction dynamics of both farmers and customers within a digital ecosystem.

The practical implication for Village Unit Dairy Cooperative is clear: prioritize continuous investment in the reliability and features of the Moo Opinion application and ensure uncompromising Milk Quality consistency. Management must allocate targeted risk mitigation resources to the Low Satisfaction segment by addressing their underlying economic concerns PC2 factors to prevent member and customer churn.

Limitations and Future Research

This study's primary limitation is its reliance on a small sample size N=40, which necessitated the use of Leave-One-Out Cross-Validation and limits the generalizability of the high model performance metrics.

For future research, the following directions are recommended:

1. Generalizability Validation: Conduct a replication study with a significantly expanded sample (e.g., $N \approx$ 300–500 respondents) to validate the Feature Importance scores and the stability of the K-Means cluster profiles.
2. Methodological Exploration: Explore the use of more complex non-linear models, such as XGBoost or Artificial Neural Networks (ANN) for the classification task, and compare their performance against Random Forest using the larger dataset.
3. Predictive Modeling: Apply advanced predictive models (e.g., Survival Analysis or Churn Prediction) specifically to the high-risk segments identified (Low Satisfaction cluster) to forecast member loss probabilities and develop proactive strategies.

REFERENCES

Binanto, I., B, J. P. K., & Leokadja, L. (2024). Perbandingan Metode Klasifikasi Random Forest Dan Support Vector Machine Terhadap Dataset Resiko Kanker Serviks. *Jtriste*, *11*(1), 60–66. https://doi.org/10.55645/jtriste.v11i1.507

Budhathoki, M., Li, L., Xu, H., Zhang, W., Ma, X., Pandey, S., & Thomsen, M. (2026). Market dynamics and E-commerce satisfaction in China's aquatic food sector: Machine Learning and data insights. *Aquaculture*, *610*(May 2025). https://doi.org/10.1016/j.aquaculture.2025.742904

Firmansyah, F., & Yulianto, A. (2024). Pemodelan Pembelajaran Mesin untuk Prediksi Kesehatan Mental di Tempat Kerja. *Jurnal Minfo Polgan*, *13*(1), 397–407. https://doi.org/10.33395/jmp.v13i1.13674

Hassoun, A., Marvin, H. J. P., Bouzembrak, Y., Barba, F. J., Castagnini, J. M., Pallarés, N., Rabail, R., Aadil, R. M., Bangar, S. P., Bhat, R., Cropotova, J., Maqsood, S., & Regenstein, J. M. (2023). Digital transformation in the agri-food industry: recent applications and the role of the COVID-19 pandemic. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.3389/fsufs.2023.1217813

Kiss, M., & Gazdecki, M. (2021). *Sustainable Attitude*. 1–28.

*citra dewi megawati

Kohli, A., & Narang, G. (2024). A Comparative Analysis of Digitization in the Agri-Food Sector-Innovations, Challenges, and Future Directions. *International Journal of Advanced Research in Computer and Communication Engineering Impact Factor*, *8*(9), 182–193. https://doi.org/10.17148/IJARCCE.2024.13923

Kumar, U., & Shankar, R. (2024). Smart dairy: Unleashing emerging ICT-enabled lean dairy supply chains through data-driven decision-making. *International Journal of Information Management Data Insights*, *4*(2), 100297. https://doi.org/10.1016/j.jjimei.2024.100297

Maioli, B. M., Ribeiro, M. G., de Carvalho, A., Gonçalves, L. A., Almeida, D. L. de, Valle de Zoppa, A. L. do, Leme, P. R., Negrão, J. A., & Gallo, S. B. (2024). Nutrition of lambs with chromium propionate and its effects on metabolism, performance and meat quality. *Small Ruminant Research*, *237*(January). https://doi.org/10.1016/j.smallrumres.2024.107306

Megawati, C. D., Yuniarno, E. M., & Nugroho, S. M. S. (2019). Clustering of Female Avatar Face Features Consumers Choice using KMeans and SOM Algorithm. *Proceedings - 2019 International Seminar on Intelligent Technology and Its Application, ISITIA 2019*, 366–370. https://doi.org/10.1109/ISITIA.2019.8937279

Munch, D. M., Schmit, T. M., & Severson, R. M. (2021). Assessing the value of cooperative membership: A case of dairy marketing in the United States. *Journal of Co-Operative Organization and Management*, *9*(1), 100129. https://doi.org/10.1016/j.jcom.2021.100129

Polo-Triana, S., Gutierrez, J. C., & Leon-Becerra, J. (2024). Integration of Machine Learning in the Supply Chain for Decision Making: A Systematic Literature Review. *Journal of Industrial Engineering and Management*, *17*(2), 344–372. https://doi.org/10.3926/jiem.6403

Sarram, G., & Ivey, S. S. (2022). Evaluating the potential of online review data for augmenting traditional transportation planning performance management. *Journal of Urban Management*, *11*(1), 123–136. https://doi.org/10.1016/j.jum.2022.01.001

Simões, A. R. P., Bánkuti, F. I., Borges, J. A. R., Roschel, M. S., & Nicholson, C. F. (2025). Dairy farmers' satisfaction with the price paid by processors in competitive markets. *Journal of Dairy Science*, *108*(3), 2315–2323. https://doi.org/10.3168/jds.2024-25737

Sutar, G., Arrawatia, R., Dhalmahapatra, K., Garg, A., & Kumar, D. (2023). Performance assessment of Dairy Cooperative Societies (DCSs): an AHP based composite index approach. *Annals of Operations Research*, *326*(2), 751–782. https://doi.org/10.1007/s10479-022-04916-9

Tanjung, D., Asmara, A., Purnamadewi, Y. L., Ristianingrum, A., Dewi, I., & Rahmasari, L. F. (2025). Evaluating the Impact of Cooperative Services on Dairy Farm Performance in North Cianjur: An Importance-Performance Analysis Approach. *Jurnal Penelitian Pendidikan IPA*, *11*(3), 639–649. https://doi.org/10.29303/jppipa.v11i3.10595

Zhao, P., Bian, J., Ni, Z., Jin, W., Weyn, J., Fang, Z., Xiang, S., Dong, H., Zhang, B., Sun, H., Thambiratnam, K., & Zhang, Q. (2024). *OMG-HD: A High-Resolution AI Weather Model for End-to-End Forecasts from Observations*. 1–26. http://arxiv.org/abs/2412.18239

*citra dewi megawati