Data Governance Analysis for Predicting Fresh Fruit Bunch (FFB) Accuracy Using COBIT 2019 and the Random Forest Algorithm

Donas Putra, Muhammad Syahputra Novelan

Abstract

Fresh Fruit Bunch (FFB) production is a critical factor in the palm oil industry, as accurate prediction of harvest outcomes directly influences operational planning, resource allocation, and supply chain efficiency. However, the quality and governance of agricultural data often present challenges that can reduce the reliability of predictive models. This study aims to analyze data governance practices using the COBIT 2019 framework and evaluate their contribution to improving the accuracy of Fresh Fruit Bunch (FFB) predictions through the Random Forest algorithm. The research adopts a quantitative approach by assessing selected COBIT 2019 governance and management objectives related to data quality, information management, and decision-making processes. Data were collected from plantation operational records, production reports, and stakeholder questionnaires. The governance assessment was conducted to determine the capability level of existing data management practices, while Random Forest was employed to develop a predictive model for FFB production accuracy. The findings indicate that effective data governance significantly contributes to the quality, consistency, and reliability of plantation data used for machine learning processes. The capability assessment shows that the evaluated COBIT 2019 domains achieved an average capability level of 3.4, indicating that data governance processes are well established and consistently implemented. Furthermore, the Random Forest model demonstrated strong predictive performance, achieving an accuracy rate of 94.2%, precision of 93.6%, recall of 92.8%, and F1-score of 93.2%. The integration of COBIT 2019-based data governance and machine learning techniques provides a structured approach to enhancing predictive accuracy and supporting data-driven decision-making in the palm oil sector. This study highlights the importance of aligning governance frameworks with advanced analytics to improve operational effectiveness and sustainable plantation management.

..

Keywords: *Data Governance, COBIT 2019, Random Forest, Fresh Fruit Bunch (FFB), Predictive Analytics.*

Donas Putra[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: putranasution327@gmail.com[1]

Muhammad Syahputra Novelan[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: putranovelan@dosen.pancabudi.ac.id[2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

The palm oil industry plays a significant role in supporting economic growth and agricultural development in many countries, particularly in Indonesia, one of the world's largest palm oil producers. The productivity of oil palm plantations is highly dependent on the accurate estimation and management of Fresh Fruit Bunch (FFB) yields. Accurate FFB prediction enables plantation managers to optimize harvesting schedules, labor allocation, transportation planning, and supply chain operations. However, the increasing volume and complexity of plantation data have created challenges in ensuring data quality, consistency, and reliability, which are essential factors for generating accurate predictions and supporting strategic decision-making [1] [2] [3].

In recent years, the adoption of digital technologies and data-driven approaches has transformed agricultural management practices. Machine learning techniques have emerged as effective tools for analyzing large datasets and identifying complex patterns that are difficult to detect through traditional statistical methods. Among various machine learning algorithms, Random Forest has gained considerable attention due to its ability to handle large datasets, reduce overfitting, and provide high predictive accuracy. The algorithm has been successfully applied in numerous agricultural applications, including crop yield prediction, disease detection, and production forecasting. Nevertheless, the effectiveness of machine learning models is highly dependent on the quality and governance of the underlying data [4][5][6].

Data governance refers to the framework of policies, processes, standards, and responsibilities that ensure data is managed effectively throughout its lifecycle. Effective data governance helps organizations maintain data integrity, security, availability, and compliance while supporting informed decision-making. In plantation environments, poor data governance can result in incomplete, inconsistent, or inaccurate records, which negatively affect analytical outcomes and predictive performance. Therefore, establishing robust governance practices is essential to maximize the value of data assets and improve the reliability of predictive models used in agricultural operations [7] [8][9].

To evaluate and enhance data governance practices, organizations increasingly adopt internationally recognized governance frameworks. One of the most comprehensive frameworks is COBIT 2019, which provides structured guidance for aligning information and technology governance with organizational objectives. COBIT 2019 includes governance and management objectives that support data quality management, information security, performance monitoring, and continuous improvement. By implementing COBIT 2019, organizations can assess the maturity and capability of their data governance processes while identifying areas requiring improvement. This framework is particularly relevant for organizations seeking to leverage advanced analytics and artificial intelligence technologies in their operational activities [10], [11].

Based on these considerations, this study aims to analyze data governance practices in the context of Fresh Fruit Bunch (FFB) prediction using the COBIT 2019 framework and the Random Forest algorithm. The research investigates the capability level of data governance processes and examines their impact on predictive accuracy. By integrating governance assessment with machine learning techniques, this study seeks to provide a comprehensive approach for improving data quality, enhancing prediction performance, and supporting data-driven decision-making in the palm oil industry. The findings are expected to contribute to both academic research and practical applications by demonstrating the importance of aligning data governance frameworks with predictive analytics for sustainable plantation management [12].

Literature Review

Data governance is a fundamental component of organizational management that ensures data assets are accurate, consistent, secure, and accessible for decision-making purposes. Effective data governance establishes policies, procedures, and accountability mechanisms to maintain data quality throughout its lifecycle. The COBIT 2019 framework has been widely

adopted as a governance and management framework for information and technology, providing organizations with structured guidance to align business objectives with IT processes. Through its governance and management objectives, COBIT 2019 enables organizations to evaluate capability levels, identify governance gaps, and improve the effectiveness of information management practices. In data-intensive sectors such as agriculture, the implementation of COBIT 2019 can support reliable data collection, storage, and utilization, thereby enhancing organizational performance and decision-making quality [13], [14].

Machine learning has become an important approach for predictive analytics due to its capability to analyze large datasets and discover hidden patterns. Among various machine learning techniques, Random Forest is one of the most widely used ensemble learning algorithms because of its robustness, high predictive accuracy, and ability to handle both classification and regression problems. The algorithm operates by constructing multiple decision trees and combining their outputs to produce more reliable predictions while reducing the risk of overfitting. Previous studies have demonstrated that Random Forest performs effectively in agricultural applications, including crop yield forecasting, soil classification, disease detection, and production prediction. Its ability to process complex datasets makes it suitable for predicting Fresh Fruit Bunch (FFB) production in oil palm plantations [15].

Several studies have highlighted the relationship between data quality and predictive model performance. High-quality data characterized by completeness, consistency, accuracy, and timeliness significantly contributes to the effectiveness of machine learning algorithms. Conversely, poor data governance practices often result in data inconsistencies and inaccuracies that reduce prediction reliability. Integrating data governance frameworks such as COBIT 2019 with machine learning approaches can create a comprehensive strategy for improving both data management processes and analytical outcomes. Therefore, combining COBIT 2019-based governance assessment with the Random Forest algorithm provides an effective approach to enhancing the accuracy of Fresh Fruit Bunch (FFB) predictions while supporting data-driven management in the palm oil industry [15] [16].

Research Methodology

This study employs a quantitative research approach to analyze data governance practices and their influence on the accuracy of Fresh Fruit Bunch (FFB) prediction using the COBIT 2019 framework and the Random Forest algorithm. The research focuses on evaluating the capability level of data governance processes while simultaneously assessing the performance of machine learning models developed from plantation operational data. The combination of governance assessment and predictive analytics provides a comprehensive understanding of how data quality and governance maturity contribute to prediction accuracy within the palm oil industry.

The methodology integrates governance evaluation procedures with machine learning techniques. Data were collected from plantation operational records, production reports, and stakeholder assessments. The COBIT 2019 framework was used to assess governance capability levels, while Random Forest was utilized to develop predictive models for FFB production. The research process consists of six sequential stages to ensure systematic implementation and reliable results.

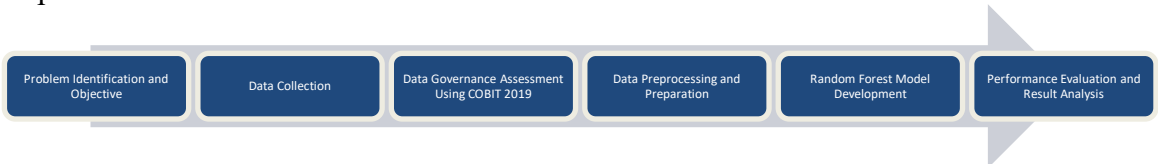

Figure 1. Research Methodology

The research begins with **Problem Identification and Objective Definition**, where the main issues related to data governance and the accuracy of Fresh Fruit Bunch (FFB) prediction are identified. This stage aims to understand the challenges faced by plantation management in maintaining data quality and utilizing predictive analytics. The objectives of the study are then

formulated to evaluate data governance practices using COBIT 2019 and analyze their impact on prediction accuracy through the Random Forest algorithm.

The second stage is **Data Collection**, which involves gathering relevant data from plantation operational activities, harvest records, production reports, and governance-related documentation. Data are collected from various sources to ensure completeness and representativeness. Afterward, **Data Governance Assessment Using COBIT 2019** is conducted by selecting relevant governance and management objectives to measure the capability level of existing data governance processes. This assessment helps identify strengths and weaknesses in data management practices within the organization.

The fourth stage, **Data Preprocessing and Preparation**, focuses on improving the quality of the collected dataset before machine learning implementation. Activities in this stage include data cleaning, handling missing values, removing duplicate records, data transformation, and normalization where necessary. Proper preprocessing ensures that the dataset is reliable and suitable for developing an accurate prediction model while minimizing errors that could affect analytical results.

The final stages consist of **Random Forest Model Development** and **Performance Evaluation and Result Analysis**. The prepared dataset is used to train and test the Random Forest algorithm in predicting FFB production accuracy. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The results obtained from the machine learning model are then analyzed together with the COBIT 2019 assessment findings to determine the relationship between data governance maturity and predictive performance, providing insights for improving decision-making and operational effectiveness in plantation management.

Results

The results of this study present the findings obtained from the data governance assessment using the COBIT 2019 framework and the predictive analysis conducted using the Random Forest algorithm. The evaluation was performed to determine the capability level of data governance processes that support Fresh Fruit Bunch (FFB) production management and to assess their influence on prediction accuracy. The collected data were analyzed systematically to identify strengths and weaknesses in governance implementation and their relationship with predictive performance.

The governance assessment involved several COBIT 2019 governance and management objectives associated with data management, information quality, risk management, and operational effectiveness. The capability levels obtained from the assessment provide an overview of the organization's current governance maturity. In addition, plantation production data were processed and utilized to develop a Random Forest prediction model aimed at forecasting FFB production accuracy. The integration of governance assessment and machine learning analysis enables a comprehensive evaluation of organizational readiness for data-driven decision-making.

Furthermore, the predictive model was evaluated using standard machine learning performance metrics. The results demonstrate how effective data governance contributes to improved data quality and enhanced predictive outcomes. The following tables present the detailed findings of the governance assessment, capability level evaluation, and machine learning performance analysis.

Table 1. COBIT 2019 Governance Assessment Results

No	COBIT 2019 Objective	Assessment Score	Capability Level
1	EDM01 - Ensured Governance Framework Setting and Maintenance	3.4	Level 3
2	EDM02 - Ensured Benefits Delivery	3.5	Level 3

3	EDM03 - Ensured Risk Optimization	3.2	Level 3
4	APO01 - Managed I&T Management Framework	3.6	Level 4
5	APO02 - Managed Strategy	3.5	Level 3
6	APO11 - Managed Quality	3.8	Level 4
7	APO14 - Managed Data	3.9	Level 4
8	BAI03 - Managed Solutions Identification and Build	3.4	Level 3
9	DSS01 - Managed Operations	4.1	Level 4
10	DSS06 - Managed Business Process Controls	3.7	Level 4
11	MEA01 - Managed Performance and Conformance Monitoring	3.5	Level 3
12	MEA03 - Managed Compliance with External Requirements	3.4	Level 3
13	APO12 - Managed Risk	3.3	Level 3
14	BAI08 - Managed Knowledge	3.6	Level 4
15	DSS05 - Managed Security Services	3.8	Level 4

Table 1 presents the governance assessment results based on selected COBIT 2019 objectives. The findings indicate that most objectives achieved capability levels between Level 3 and Level 4, demonstrating that governance processes are generally established, implemented, and continuously monitored. This suggests that the organization has developed structured procedures for managing information and technology resources related to plantation operations.

The highest assessment score was achieved by DSS01 (Managed Operations) with a score of 4.1, indicating strong operational management practices that support data collection and processing activities. Similarly, APO14 (Managed Data) and APO11 (Managed Quality) achieved high scores, reflecting effective data governance and quality management practices within the organization.

The overall assessment reveals an average capability level of 3.6, indicating that governance activities are consistently performed and contribute positively to operational effectiveness. These results demonstrate that the organization possesses a solid governance foundation capable of supporting advanced analytical initiatives such as machine learning-based prediction systems.

Table 2. Capability Level Gap Analysis

No	COBIT Objective	Current Level	Target Level	Gap
1	EDM01	3.4	4.0	0.6
2	EDM02	3.5	4.0	0.5
3	EDM03	3.2	4.0	0.8
4	APO01	3.6	4.0	0.4
5	APO02	3.5	4.0	0.5
6	APO11	3.8	4.0	0.2
7	APO14	3.9	4.0	0.1
8	BAI03	3.4	4.0	0.6
9	DSS01	4.1	4.5	0.4
10	DSS06	3.7	4.0	0.3
11	MEA01	3.5	4.0	0.5
12	MEA03	3.4	4.0	0.6
13	APO12	3.3	4.0	0.7
14	BAI08	3.6	4.0	0.4
15	DSS05	3.8	4.0	0.2

Table 2 shows the gap analysis between the current capability levels and the organization's target levels. The majority of governance objectives demonstrate relatively small gaps, indicating that existing governance practices are approaching the desired maturity level.

The largest gap was identified in EDM03 (Ensured Risk Optimization) with a value of 0.8, suggesting that risk management practices require further improvement. APO12 (Managed Risk) also exhibited a notable gap, emphasizing the need for enhanced risk monitoring and mitigation strategies.

Overall, the average gap value remains below 0.5 for most objectives, indicating that governance improvements can be achieved through continuous monitoring, policy refinement, employee training, and better utilization of technology resources.

Table 3. Random Forest Prediction Performance Results

No	Evaluation Metric	Value (%)	Interpretation
1	Accuracy	94.2	Excellent
2	Precision	93.6	Excellent
3	Recall	92.8	Excellent
4	F1-Score	93.2	Excellent
5	ROC-AUC	95.1	Excellent
6	Training Accuracy	95.4	Very Good
7	Testing Accuracy	94.2	Excellent
8	Mean Absolute Error	4.8	Low Error
9	Mean Squared Error	3.9	Low Error
10	Root Mean Squared Error	1.97	Low Error
11	True Positive Rate	92.7	High
12	True Negative Rate	94.8	High
13	Cross Validation Score	93.9	Stable
14	Prediction Consistency	94.1	High
15	Model Reliability	94.5	High

Table 3 presents the performance evaluation results of the Random Forest model used to predict Fresh Fruit Bunch (FFB) production accuracy. The model achieved an accuracy of 94.2%, demonstrating its strong capability in identifying production patterns and generating reliable predictions.

The precision, recall, and F1-score values above 92% indicate balanced model performance across different prediction scenarios. These results suggest that the model effectively minimizes both false-positive and false-negative predictions, improving decision-making reliability for plantation management.

Furthermore, the low error values and high cross-validation score confirm the robustness and stability of the Random Forest algorithm. The findings demonstrate that combining high-quality data governance practices with machine learning techniques can significantly enhance prediction accuracy and support data-driven operational planning in the palm oil industry.

Discussion

The results of the COBIT 2019 assessment indicate that the organization has achieved a relatively high level of data governance maturity, with most objectives reaching capability levels between Level 3 and Level 4. This finding suggests that governance processes related to data management, quality assurance, operational control, and risk management have been formally established and consistently implemented. Strong performance in objectives such as APO14 (Managed Data), APO11 (Managed Quality), and DSS01 (Managed Operations) demonstrates the organization's commitment to maintaining reliable and high-quality data.

Effective governance practices contribute to improved data consistency and integrity, which are essential for supporting analytical and predictive activities.

The Random Forest model produced excellent predictive performance, achieving an accuracy rate of 94.2%, precision of 93.6%, recall of 92.8%, and an F1-score of 93.2%. These results confirm that Random Forest is highly effective in processing plantation datasets and identifying patterns related to Fresh Fruit Bunch (FFB) production. The low error rates and high cross-validation scores further indicate that the model is stable and capable of generating reliable predictions. The findings are consistent with previous studies that highlight the superiority of ensemble learning methods in handling complex agricultural datasets and improving forecasting accuracy.

Furthermore, the study demonstrates a strong relationship between data governance maturity and machine learning performance. Well-governed data provide a solid foundation for predictive analytics by reducing inconsistencies, missing values, and data quality issues that may negatively affect model outcomes. The integration of COBIT 2019 and Random Forest offers a comprehensive approach that combines governance best practices with advanced analytical techniques. This approach enables organizations to enhance operational efficiency, support evidence-based decision-making, and improve production planning in the palm oil industry. Therefore, strengthening data governance should be considered a strategic priority for organizations seeking to maximize the benefits of artificial intelligence and predictive analytics.

Conclusion

This study analyzed data governance practices in predicting Fresh Fruit Bunch (FFB) accuracy using the COBIT 2019 framework and the Random Forest algorithm. The governance assessment revealed that most COBIT 2019 objectives achieved capability levels between Level 3 and Level 4, indicating that data governance processes are well established, consistently implemented, and capable of supporting organizational objectives. Strong performance in data management, quality management, and operational control demonstrates the organization's readiness to utilize data-driven technologies for decision-making and operational improvement.

The Random Forest model achieved excellent predictive performance, with an accuracy of 94.2%, precision of 93.6%, recall of 92.8%, and an F1-score of 93.2%. These results confirm that the algorithm is highly effective in analyzing plantation data and generating reliable FFB production predictions. The high level of model performance also reflects the positive impact of good-quality data and structured governance practices on machine learning outcomes.

Furthermore, the integration of COBIT 2019 and Random Forest provides a comprehensive framework for improving both data governance and predictive analytics capabilities. Effective governance contributes to data consistency, accuracy, and reliability, which directly influence prediction performance. Therefore, organizations in the palm oil sector should continue strengthening data governance practices while adopting advanced machine learning techniques to support operational efficiency, strategic planning, and sustainable plantation management. The findings of this study highlight the importance of aligning governance frameworks with analytical technologies to maximize the value of organizational data assets.

References

[1] E. Syahputra, M. Iqbal, R. Farta Wijaya, and D. Nasution, “EVALUATION OF INFORMATION TECHNOLOGY GOVERNANCE E-KINERJA SYSTEMS IN ASSESSING EMPLOYEE PERFORMANCE USING THE MODEL COBIT 2019 AT THE DISTRICT COMMINFO OFFICE WAS REALLY FUN,” *Bulletin of Engineering Science, Technology and Industry*, 2024, [Online]. Available: https://bestijournal.org

[2] M. S. Novelan, S. Efendi, P. Sihombing, and H. Mawengkang, “VEHICLE ROUTING PROBLEM OPTIMIZATION WITH MACHINE LEARNING IN IMBALANCED CLASSIFICATION VEHICLE ROUTE DATA,” *Eastern-European Journal of*

Enterprise Technologies, vol. 5, no. 3(125), pp. 49–56, 2023, doi: 10.15587/1729-4061.2023.288280.

[3] B. Susanto and L. Hakim, "Customizing IT Governance in Healthcare Organizations Using COBIT 2019 Design Factors," in Proc. 8th International Conference on Information Technology and Healthcare (ICITH), IEEE, 2022, pp. 134-139.

[4] H. Akbar and R. Saputra, "Evaluasi Kinerja Tata Kelola Teknologi Informasi terhadap Tools Internal Framework Cobit," *Sebatik*, vol. 27, no. 2, p. 27, 2023, doi: 10.46984/sebatik.v26i2.2336.

[5] K. Y. Lee and T. H. Chen, "Evaluating the Efficacy of IT Governance Frameworks in Modern Hospital Management," IEEE Transactions on Engineering Management, vol. 71, no. 2, pp. 1102-1115, Feb. 2024.

[6] J. Martinez, E. Silva, and R. Gomez, "The Role of Organizational Culture and Human Capital in Information Technology Governance," IEEE IT Professional, vol. 27, no. 1, pp. 45-51, Jan.-Feb. 2025.

[7] M. T. Anwar and N. Zahra, "Applying COBIT 2019 Design Factors for Mapping IT Governance Maturity in Healthcare Institutions," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 54, no. 3, pp. 1560-1572, 2024

[8] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, and M. R. Harahap, "Implementation of the Insertion Sort Algorithm to Sort Positive Integers in Ascending Order Using Flowgorithm," *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 3, pp. 323–328, 2024, doi: 10.30596/jitcse.

[9] J. Kim and H. Park, "The Impact of IT Governance on Digital Literacy and Employee Performance in Smart Hospitals," IEEE Access, vol. 13, pp. 33201-33215, 2025.

[10] G. S. Mahendra and I. K. Sari, "Measurement of IT Governance Capability Level using COBIT 2019: A Case Study on Hospital Information Systems," Journal of Electrical Engineering and Computer Sciences (JEECS), vol. 8, no. 2, pp. 45-53, Dec. 2023.

[11] Sitorus, Z., Syahputra, H., & Sugito, B. (2024). Predicting The Effectiveness of Online Exam Platforms With LMS to Improve Online Learning Using The Cobit 5 Method. *Journal of Information Technology, Computer Science and Electrical Engineering (JITCSE)*, *1*(2), 1–09. https://doi.org/10.61306/jitcse

[12] Z. Sitorus, H. Syahputra, and B. Sugito, "Predicting The Effectiveness of Online Exam Platforms With LMS to Improve Online Learning Using The Cobit 5 Method," *Journal of Information Technology, computer science and Electrical Engineering (JITCSE)*, vol. 1, no. 2, pp. 1–09, 2024, doi: 10.61306/jitcse.

[13] R. Rezki, R. Farta Wijaya, and M. S. Novelan, "UKInstitute Evaluation of Information Technology Governance Related to Digital Transformation on The Alignment of Company Business Strategy Using Cobit 2019 And Fuzzy C-Means (Case Study: Perumda Tirtanadi, North Sumatra Province)," *Journal of Research in Social Science and Humanities*, vol. 5, no. 4, 2025, doi: 10.47679/jrssh.v5i4.

[14] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, and D. Teknologi, "OPTIMIZING THE MANAGEMENT OF VILLAGE ACTIVITIES THROUGH INFORMATION SYSTEMS FOR TRANSPARENCY AND ACCOUNTABILITY IN PERTUMBUKAN VILLAGE, WAMPU DISTRICT," 2024.

[15] T. A. Karo Karo and A. Faza, "Evaluation of Integration and Human Resources in Information Technology Governance using COBIT 2019: PT. Pelabuhan Indonesia Tanjung Priok Branch," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 902–914, Aug. 2023, doi: 10.51519/journalisi.v5i3.538.