PAPER – OPEN ACCESS

Utilization of Big Data Analytics in the Decision-Making Process for Performance Improvement at PT Perkebunan Nusantara IV

Author : Ferryanto BM Situmorang, et al
DOI : 10.32734/lwsa.v9i2.2848
Electronic ISSN : 2654-7066
Print ISSN : 2654-7058

Volume 9 Issue 2 – 2026 TALENTA Conference Series: Local Wisdom, Social, and Arts (LWSA)

LWSA Conference Series 09 (2026)
TALENTA Conference Series
Available online at https://talentaconfseries.usu.ac.id

Utilization of Big Data Analytics in the Decision-Making Process for Performance Improvement at PT Perkebunan Nusantara IV

Ferryanto B. M. Situmorang[1], M. Anggia Muchtar[2], Emerson P. Sinulingga[3]

[a]*Student of Master of Management Study Program, Postgraduate School, Universitas Sumatera Utara, Medan 20155, Indonesia*
[b]*Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan 20155, Indonesia*
[c]*Department of Electrical Engineering, Faculty of Engineering, Universitas Sumatera Utara, 20155, Indonesia*

ferryportal@gmail.com

Abstrak

Studi ini mengatasi kesenjangan 24% dalam produktivitas CPO di PT Perkebunan Nusantara IV (PTPN IV) Regional V, di mana hasil FFB saat ini sebesar 12,31 ton/ha menghasilkan produktivitas CPO sebesar 2,59 ton/ha, yang masih jauh dari target KPI sebesar 3,403 ton/ha. Penelitian ini menunjukkan bagaimana Big Data Analytics (BDA), yang didukung oleh artefak Knowledge Management (KM), meningkatkan pengambilan keputusan strategis untuk meningkatkan kinerja. Dengan menggunakan analitik deskriptif, diagnostik, prediktif, dan preskriptif, studi ini mengidentifikasi pendorong utama produksi FFB dan meramalkan tren masa depan. Temuan studi ini memproyeksikan peningkatan 16,65% dalam produksi FFB pada tahun 2025, mencapai 399.362,63 ton, yang menghasilkan produktivitas CPO sebesar 3,52 ton/ha, sedikit melebihi target KPI. Analisis Pareto mengungkapkan bahwa tiga afdeling yang berkinerja buruk menyumbang 80,59% dari kekurangan produktivitas, menyoroti afdeling untuk intervensi yang lebih terfokus. Analitik preskriptif kemudian memberikan strategi yang dapat ditindaklanjuti untuk alokasi sumber daya dan perencanaan intervensi, memungkinkan manajer untuk memprioritaskan tindakan di area yang akan memberikan peningkatan kinerja terbesar. Studi ini menekankan bagaimana BDA dapat digunakan untuk memprioritaskan intervensi strategis, mengoptimalkan alokasi sumber daya, dan memastikan tujuan kinerja tercapai. Dengan mengintegrasikan output BDA dengan proses KM, pendekatan ini membantu organisasi untuk menutup kesenjangan kinerja dan meningkatkan produktivitas. Studi ini memberikan kontribusi pada literatur manajemen strategis dengan menunjukkan peran BDA dalam meningkatkan pengambilan keputusan dan kinerja dalam lingkungan operasional yang kompleks.

Kata Kunci: Strategi Komunikasi; Brand Awareness; Citra Korporasi; Dashboard Monitoring; PTPN Group.

Abstract

This study addresses a 24% gap in CPO productivity at PT Perkebunan Nusantara IV (PTPN IV) Regional V, where current FFB yields of 12.31 tons/ha result in CPO productivity of 2.59 tons/ha, falling short of the target KPI of 3.403 tons/ha. The research demonstrates how Big Data Analytics (BDA), supported by Knowledge Management (KM) artifacts, enhances strategic decision-making to improve performance. Using descriptive, diagnostic, predictive, and prescriptive analytics, the study identifies key drivers of FFB production and forecasts future trends. The findings project a 16.65% increase in FFB production in 2025, reaching 399,362.63 tons, resulting in CPO productivity of 3.52 tons/ha, slightly exceeding the KPI target. Pareto analysis reveals that three underperforming afdelings contribute to 80.59% of the productivity shortfall, highlighting them for targeted intervention. Prescriptive analytics then provides actionable strategies for resource allocation and intervention planning, allowing managers to prioritize actions in areas that will drive the greatest improvements. This study emphasizes how BDA can be used to prioritize strategic interventions, optimize resource allocation, and ensure that performance goals are met. By integrating BDA outputs with KM processes, this approach helps organizations close performance gaps and improve productivity. The study contributes to strategic management literature by demonstrating the role of BDA in enhancing decision-making and performance in complex operational environments.

Keywords: big data analytics; FFB productivity; extreme gradient boosting; generalized additive model; performance improvement; strategic decision-making

Selection and peer-review under responsibility of The 2nd International Joint Conference on Digitalisation of Sustainable Social Sciences (IJCDSS) 2025
p-ISSN: et al, e-ISSN: 2654-7058, DOI: 10.32734/lwsa.v9i2.2848

1. Introduction

The palm oil industry is a strategic pillar of Indonesia's agricultural sector, and improving plantation productivity remains key to national competitiveness and the performance of state-owned enterprises. PT Perkebunan Nusantara IV (PTPN IV), or PalmCo, has invested heavily in digital platforms such as SAP ERP, DFarming, GIS dashboards, IoT, PalmCo Learning Management System (PALMS) –knowledge platform–, and PalmCo Business Cockpit to enhance operational visibility. However, despite these investments, aggregate productivity has shown only modest increases, from 21.6 tons/ha in 2020 to 22.2 tons/ha in 2023, with a peak of 22.7 tons/ha in 2022. More critically, this stability masks substantial regional disparities, particularly in Regional V (RPC 5), where FFB productivity and CPO productivity are below the KPI target.

In RPC 5, current FFB productivity stands at 12.31 tons/ha in Desember 2024, resulting in CPO productivity of 2.59 tons/ha, about 24% below the regional KPI target of 3.403 tons/ha. The performance shortfall is due not only to lower FFB yields but also the oil extraction rate (OER), which, when combined with FFB yield, determines CPO productivity. By focusing on high-potential afdelings –those with better OER and higher-than-average production potential– RPC 5 can close this gap and potentially exceed the KPI target.

Despite the promise of Big Data Analytics (BDA) in transforming operational data into actionable insights, most studies focus on aggregate outcomes and predictive accuracy, with limited integration of Knowledge Management (KM) for decision-making. This study fills the gap by integrating Generalized Additive Models (GAM), Extreme Gradient Boosting (XGBoost), and Pareto analysis into a BDA–KM pipeline, enabling targeted, actionable decisions for improving productivity in PTPN IV Regional V. By using this combined approach to analyze operational and agronomic data, the study translates complex analytics into reusable decision support, thereby contributing to managerial decision-making and continuous performance improvement.

The integration of prescriptive analytics and KM provides a five-zone operational action map, which translates analytics outputs into actionable insights. This tool helps prioritize interventions, ensuring resources are allocated efficiently and interventions are targeted where they will have the most significant impact. By focusing on the "vital few" high-potential afdelings, Regional V can improve its CPO productivity and exceed the 2025 KPI.

2. Literature Review

2.1 Big Data Analytics in Plantation and Agribusiness Systems

Large-scale plantation and agribusiness systems involve complex interactions among agronomic conditions, climate variability, and operational execution that generate substantial productivity differences across locations and managerial units. The expanding availability of enterprise-system records (e.g., ERP transactions, cost postings, operational logs), remote-sensing products, and environmental datasets has positioned BDA as a central tool for performance monitoring, forecasting, and decision support in these sectors (Corley & Tinker, 2016; Wolfert et al., 2017). Empirical studies show that BDA can improve yield estimation, climate–production analysis, and short-term forecasting in agriculture and oil palm plantations, particularly when agronomic and operational variables are jointly modeled using modern machine-learning approaches (Liakos et al., 2018; Jamshidi et al., 2024). Evidence also suggests that analytics delivers greater decision-making value when embedded in organizational routines rather than treated solely as reporting output (Chatterjee et al., 2023). Despite these advances, a persistent limitation in the literature is the strong emphasis on predictive accuracy at aggregated or coarse spatial units, with limited attention to how analytical outputs are translated into actionable managerial guidance at the operational level where interventions are executed.

2.1 Diagnostic and Predictive Analytics for Productivity Analysis

Descriptive analytics supports monitoring historical production patterns and variability, but it does not explain why productivity gaps persist across operational units. Diagnostic analytics therefore links outcomes to plausible drivers such as rainfall, stand characteristics, and labor execution (Lobell & Burke, 2010). In oil palm, recent field evidence shows that large yield gaps are frequently associated with controllable agronomic and operational practices –especially nutrient management, harvest interval, weed control, and pruning– (Monzon et al., 2023) highlighting the managerial relevance of diagnosing operational drivers rather than attributing outcomes primarily to genetics. Smallholder studies in Indonesia further show that harvest intervals are often longer than recommended, which is associated with higher harvest losses and quality implications (de Vos et al., 2023), making harvesting discipline a concrete diagnostic lever. At plantation scale, yield variability is often dominated by differences between plantations rather than climate averages alone, reinforcing the need for unit-level diagnosis of execution and management routines (Fleiss et al., 2022).

Because biological production systems exhibit nonlinear responses, thresholds, and diminishing returns, purely linear diagnostics can be inadequate. GAM-type approaches are widely used in agro-environmental research to model interpretable nonlinear mechanisms (Hastie & Tibshirani, 1990; Wood, 2017), and recent oil palm evidence indicates that climate anomalies can affect estate-level yields through nonlinear patterns across multiple pre-harvest lag windows –consistent with threshold-like risk dynamics (Fleiss et al., 2022; Kamil et al., 2024). In parallel, predictive analytics increasingly applies machine-learning models; a recent oil palm study benchmarked 17 ML/DL models using a comprehensive agronomy dataset, illustrating both predictive potential and sensitivity to data completeness and quality (Jamshidi et al., 2024). However, across digital-agriculture research, a recurring limitation is that analytics often stops at monitoring/forecasting rather than being embedded into routines that specify where to intervene first and how insights are codified, shared, and reused –a key requirement for decision-making value creation and knowledge institutionalization in organizations (Makhloufi et al., 2023; Uyar et al., 2024; Tumwebaze et al., 2025).

Predictive analytics increasingly uses machine learning to capture complex nonlinearities. Recent oil palm work benchmarked 17 ML/DL models using a comprehensive agronomy dataset, highlighting both predictive potential and sensitivity to data quality (Jamshidi et al., 2024). To improve managerial usability, explainable ML (e.g., XGBoost + SHAP) can rank key yield drivers and translate predictions into operational levers (Elwirehardja et al., 2023), aligning with broader evidence that interpretability is crucial for adoption in decision support (Paudel et al., 2023; Petropoulos et al., 2025). Combined with prioritization logic (e.g., Pareto-style ranking), interpretable diagnostics and explainable prediction can bridge analytics outputs to planning, budgeting, and field-level control –especially at fine units such as blocks/afdelings.

2.2 Prescriptive Analytics, Knowledge Management, and Decision-Making

Compared with descriptive, diagnostic, and predictive analytics, prescriptive analytics –i.e., analytics that recommends actions and informs resource allocation– remains relatively underrepresented in the oil palm digitalization literature. A recent review of Industry 4.0 technologies in the oil palm sector shows that much of the work concentrates on plantation mapping, disease detection, and fruit grading –valuable for monitoring and classification, but less explicit in translating analytical results into prioritized field interventions and budget decisions (Akmal et al., 2025). Prescriptive analytics is inherently socio-technical: it must specify how recommendations are generated, governed, and enacted with human decision-makers (e.g., advisory versus adaptive/executive use), otherwise outputs remain detached from operational routines (Wissuchek & Zschech, 2024).

In agriculture more broadly, research on data value creation emphasizes that analytics generates value through mechanisms such as prediction/optimization and monitoring/adaptation aimed at business process improvement and organizational performance –yet implementation frequently breaks at the step of converting results into repeatable operating procedures (Uyar et al., 2024). Knowledge management (KM) is therefore the bridge that turns analytical outputs into reusable decision assets: recent empirical studies show that KM processes (capturing, disseminating, and applying knowledge) strengthen how BDA contributes to organizational performance (Makhloufi et al., 2023; Aljehani et al., 2024). Consistent with this view, this study positions model-based thresholds and priority-unit lists as KM artifacts (e.g., playbooks, training modules, and planning templates) to support operational decision-making in PTPN IV, while not evaluating KM platform adoption or usage.

2.3 *Research Gap and Positioning of the Present Study*

As summarized in Table 1, prior work on BDA for agriculture and plantations is often fragmented across the analytics lifecycle: many studies emphasize descriptive monitoring or standalone prediction, are implemented at coarse spatial scales (e.g., region/estate aggregates), and rarely show how analytical outputs are converted into repeatable managerial routines (planning, budgeting, and field control). Meanwhile, the BDA–KM literature consistently finds that analytics contributes to performance more reliably when insights are captured, shared, and institutionalized through KM processes –yet this linkage is usually demonstrated with firm-level perceptual measures rather than operational KPI gap-closure in large-scale plantations.

This study addresses these gaps by proposing an integrated BDA–KM decision framework that (i) conducts diagnostic analytics at fine operational granularity (afdeling–month), (ii) combines diagnostic, predictive, and prescriptive steps in one empirical workflow, and (iii) explicitly translates model outputs into KM artifacts and decision routines (e.g., threshold guides, priority-unit lists, and action planning templates) that can be reused in performance reviews and resource allocation. This moves beyond “insight generation” toward KPI-oriented decision-making and prioritized intervention.

Table 1: Summary of prior studies and positioning of the present study

No	Study	Focus & context	Method/Data	Key findings (from the study)	Present study (this paper): relevance & extension
1	Li et al. (2022)	BDA usage and decision-making quality	Empirical study grounded in dynamic capability theory	BDA usage improves decision-making quality; effects operate through analytics capabilities	Reinforces that decision relevance matters; extended here by specifying *how* decision quality is achieved via reusable thresholds + priority lists embedded in routines.
2	Makhloufi et al. (2023)	BDA–KM–performance link (organizational level)	Survey; PLS-SEM	BDA relates positively to performance; KM acts as a pathway/mediator in value creation	Provides core theoretical grounding for BDA→KM→performance; extended here with *operational* KPI gap-closure at afdeling level (not perceptual firm performance only).
3	Monzon et al. (2023)	Oil palm yield gaps and controllable drivers (action levers)	Field/survey + modeling (smallholders)	Yield gaps are strongly associated with controllable practices (e.g., nutrient management, harvest interval, pruning, weed control).	Supplies *actionable* agronomic levers; extended here by ranking operational units and linking those levers to prioritized interventions and budgeting via KM artifacts.
4	Aljehani et al. (2024)	BDA and organizational performance with KM mechanisms	Survey (telecom); mediation/moderation model	Performance effects depend on complementary capabilities and mediating mechanisms including KM.	Supports the "capabilities + KM" logic; extended here by operationalizing KM artifacts/routines tied to plantation performance management.
5	Uyar et al. (2024)	How agricultural data creates value (sector-wide)	Structured literature review	Maps value-creation mechanisms/targets/impacts; highlights gaps in turning data into sustained performance improvement.	Provides sector-level justification; extended here by implementing a concrete end-to-end pathway from analysis → prioritization → repeatable operating procedures
6	Tumwebaze et al. (2025)	KM in the agriculture sector	Systematic literature review	KM research in agriculture is still developing; emphasizes the need to embed knowledge into processes and practice	Strengthens the argument that KM is the missing "glue"; extended here by treating thresholds/priority lists as KM assets for operational decisions.

2.4 Proposed Research Framework

Building on gaps in prior plantation analytics research –particularly the limited integration of diagnostic, predictive, and prescriptive analytics with KM-enabled decision support– this study proposes an integrated BDA–KM decision-support framework summarized in Figure 1. Grounded in Simon's Intelligence–Design–Choice (IDC) model under bounded rationality, the framework is situated in PTPN IV Regional V, where heterogeneous field conditions and KPI pressure create predominantly semi-structured decisions. Multi-source inputs (production, cost, agronomic, operational, and external rainfall data) feed a four-layer BDA pipeline: descriptive analytics summarizes historical patterns; diagnostic analytics (GAM) identifies key drivers and interpretable nonlinear or threshold effects; predictive analytics (XGBoost) forecasts future productivity; and prescriptive analytics (gap analysis and Pareto ranking) translates results into prioritized intervention targets. Consistent with IDC, descriptive–diagnostic outputs support intelligence, forecasting supports design, and prioritization supports choice. The resulting outputs are codified into KM artifacts (dashboards, SOPs, and planning templates) and embedded into decision routines, with feedback loops enabling reuse and organizational learning for continuous performance improvement.

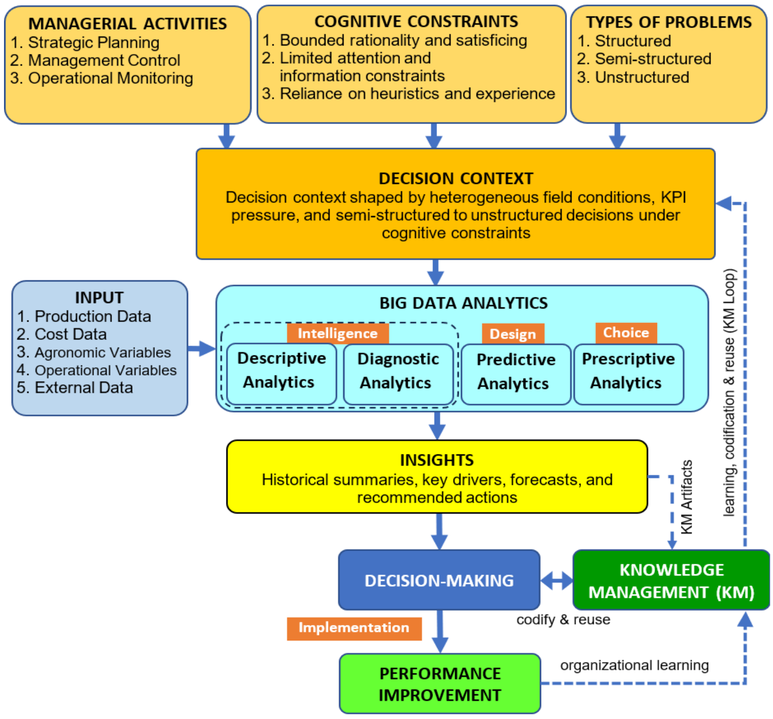

Figure 1. Proposed Research Framework Grounded in Simon's IDC Model

3. Methods

3.1 Research Design and Framework

This study adopts a quantitative design within a Big Data Analytics (BDA) paradigm to support KPI-oriented decision-making in large-scale oil palm plantation management. The analytical workflow follows the proposed BDA–KM decision-support framework (Figure 1), ensuring traceability from the managerial problem formulation to descriptive, diagnostic, predictive, and prescriptive analytics, and ultimately to actionable recommendations codified as knowledge management (KM) artifacts.

The central business question guiding this research is: How can big data analytics and knowledge management generate actionable recommendations to improve Fresh Fruit Bunch (FFB) productivity in support of the 2025 CPO productivity KPI of 3.403 tons/ha in PTPN IV Regional V?

To address this question, the study integrates descriptive, diagnostic, predictive, and prescriptive analytics within a unified workflow, and translates analytical outputs into reusable KM artifacts to support repeatable decision cycles and organizational learning.

3.2 Study Area, Unit of Analysis, and Period

The empirical setting is PTPN IV Regional V, characterized by heterogeneous agronomic conditions and operational execution. The unit of analysis is afdeling–month, representing the finest managerial level at which production, cost, and labor performance are routinely recorded and reviewed.

The dataset covers January 2024 to February 2025, representing realized (observed) operational records available at the time of analysis. Year 2024 is treated as a baseline reference period for describing typical production patterns and driver–output relationships, while January–February 2025 provides the most recent realized performance used for near-term validation and gap identification. For forecasting, observations are ordered chronologically and partitioned using a time-based split into a training set (earlier months) and a hold-out test set (most recent months) to avoid look-ahead bias. The validated model is then used to generate projections for March–December 2025 to support KPI monitoring and planning.

3.3 Data Sources and Variable Definition

All variables are derived from two primary sources: SAP ERP (operational, agronomic, and cost records) and Google Earth Engine (GEE) (monthly rainfall estimates). SAP ERP records captured at the block level are aggregated to the afdeling–month level to form the analytical panel. Rainfall is assigned at the estate level and applied uniformly to afdelings within the same estate for the corresponding month.

All production values are stored and processed in kilograms (kg), reflecting the smallest unit in the weighing system. Monthly FFB production per afdeling is obtained by aggregating weighed records to the afdeling–month level; no conversion to tons is applied during model estimation. The 2025 KPI is expressed in CPO tons/ha and is evaluated by converting modeled FFB productivity to implied CPO productivity using an assumed OER of 21%, applied only at the reporting stage for KPI assessment.

3.4 Data Sources and Variable Definition

All variables are derived from two primary sources: SAP ERP, which provides detailed operational, agronomic, and cost records, and Google Earth Engine (GEE) which supplies monthly rainfall estimates. Operational records captured at the block level in SAP ERP are aggregated into an afdeling–month panel dataset, while estate-level rainfall values are assigned uniformly to afdelings within the same estate.

3.5 Variables

Dependent variable

1) FFB production: monthly FFB output per afdeling (kg/month), computed by aggregating weighed FFB records.
2) Independent variables

a) Cultivation and environmental variables

- TM_PP (productive palm population): total number of productive palms per afdeling, aggregated from block-level counts.
- TM_SPH (planting density): number of palms per hectare, computed as an area-weighted average across blocks within each afdeling.
- Hujan: monthly rainfall (mm) derived from GEE.

b) Operational and economic variables

- Pemanen (harvesting activity): monthly count of harvesting-activity records per afdeling, representing harvesting intensity and labor deployment.
- BPemPuk (maintenance and fertilization costs): monthly expenditures on maintenance and fertilization at the afdeling level.
- BPanAng (harvesting and transport costs): combined monthly costs of harvesting and transporting FFB from field to mill per afdeling.

The 2025 productivity KPI is expressed in CPO tons/ha and is evaluated by converting modeled FFB productivity to implied CPO productivity using an assumed oil extraction rate (OER) of 21% and standard unit conversion at the reporting stage; unit conversion is applied only for KPI assessment, not during model estimation.

c) Data Preprocessing

SAP ERP and GEE datasets are cleaned and harmonized to ensure temporal and structural consistency. Preprocessing includes: (i) standardization of units (production consistently in kg) and alignment of cost postings to consistent accounting periods, (ii) removal of duplicate records and treatment of missing/erroneous entries using listwise deletion and limited, justified imputation, (iii) plausibility screening for extreme values using descriptive summaries and boxplots, and (iv) aggregation of transaction-level records into afdeling–month observations. The resulting dataset forms an unbalanced panel suitable for nonlinear and multivariate analysis.

d) Diagnostic Analytics: Generalized Additive Model (GAM)

This study applies GAM to analyze the nonlinear relationships between various predictors (e.g., planting density, rainfall, maintenance) and FFB production. GAM uses smooth spline terms to model the flexible, nonlinear effects of predictors on FFB output. The model incorporates a Gaussian error distribution and identity link function, making it suitable for continuous variables like FFB production. By identifying threshold points or saturation effects, GAM helps understand how predictors influence productivity at different levels, such as how planting density may initially increase FFB production before leveling off.

To assess the significance of these relationships, the model tests the smooth terms, focusing on the shape of the smooth functions and their effective degrees of freedom (EDF). While p-values can be sensitive in penalized spline estimation, the primary insight comes from interpreting the smooth functions, which reveal the key nonlinear and threshold-like dynamics. These findings are critical for guiding operational decisions, highlighting which factors should be optimized for improving productivity.

Moreover, GAM allows for a deeper understanding of the underlying mechanisms behind production fluctuations, providing actionable insights for managers. For example, by analyzing how variables like maintenance intensity or rainfall interact with productivity, the model identifies the specific conditions under which production can be enhanced or optimized. This knowledge can then be translated into targeted interventions, such as adjusting maintenance schedules or reallocating resources during specific climatic conditions, contributing to more effective operational decision-making.

e) Predictive Analytics: XGBoost

To forecast FFB production, we utilize XGBoost at the afdeling-month level to model complex relationships among agronomic variables, rainfall, operational execution, and costs. This approach allows us to capture the nonlinearities and interactions between these factors, which are often not addressed in traditional linear models. The model evaluation follows a time-aware approach, where earlier months are used for training, and the most recent months are reserved for out-of-sample testing. This prevents look-ahead bias, ensuring that the model's predictions are based on historical data only, similar to how it would perform in real-world forecasting scenarios.

To optimize model performance, hyperparameters are fine-tuned using a time-series validation strategy, which helps in addressing seasonality and temporal patterns while avoiding overfitting. Regularization is applied to control for model complexity, ensuring a balance between model fit and generalization. Predictive performance is assessed using standard metrics such as coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), ensuring that the model accurately captures the relationship between the predictors and the target variable (FFB production).

The validated model is then used to generate projections for March–December 2025, providing critical insights for KPI monitoring, capacity alignment, and short-term planning in PTPN IV Regional V, guiding future operational strategies and decision-making.

f) Prescriptive Analytics and Translation into KM Artifacts

In this study, prescriptive analytics is used to guide operational decision-making by identifying key areas where interventions will yield the greatest impact. The process begins with Pareto analysis, which highlights the "vital few" afdelings contributing most to the observed performance gap. These afdelings are then categorized into different zones based on their performance and potential for improvement. The zones include the Priority Zone, where immediate interventions are needed, and the Monitoring Zone, where ongoing observation and adjustments are required. The Improvement Zone identifies units that need optimization, while the Optimal Zone represents high-performing units that should be maintained. Finally, the Benchmark Zone highlights units that serve as best practice models for replication.

The results of this prescriptive analysis are translated into actionable insights through the creation of KM artifacts. These artifacts include prioritized action lists, which guide managers on where to focus their resources. These KM artifacts are essential for embedding analytical outputs into the decision-making process, ensuring that recommendations are actionable, reusable, and integrated into routine managerial practices for continuous performance improvement.

4. Results and Discussions

Consistent with the proposed framework (Figure 1), results are reported sequentially as descriptive, diagnostic, predictive, and prescriptive analytics, and then synthesized into decision-support outputs for KPI improvement.

4.1 Descriptive Analytics: Performance and Variability

Figure 2 shows the monthly aggregated FFB production from nucleus estates (Kebun) (top panel) and purchased FFB at mills (PKS) (bottom panel) for January 2024 to February 2025. The estate-origin production (Kebun) exhibits a consistent upward and downward trend in production volumes, with fluctuations that reflect seasonal variations. On the other hand, the purchased FFB at mills shows more volatility, which can be attributed to varying procurement dynamics and operational adjustments.

The inclusion of 3-month moving average and exponential smoothing trends highlights these fluctuations, allowing for a clearer understanding of the shifts in production over time. This variability underscores the need for a more detailed unit-level analysis of the estates and mills. Simply aggregating total FFB production may not capture the nuances of individual unit performance, making it necessary to benchmark against RKAP targets to identify areas requiring strategic interventions and resource optimization.

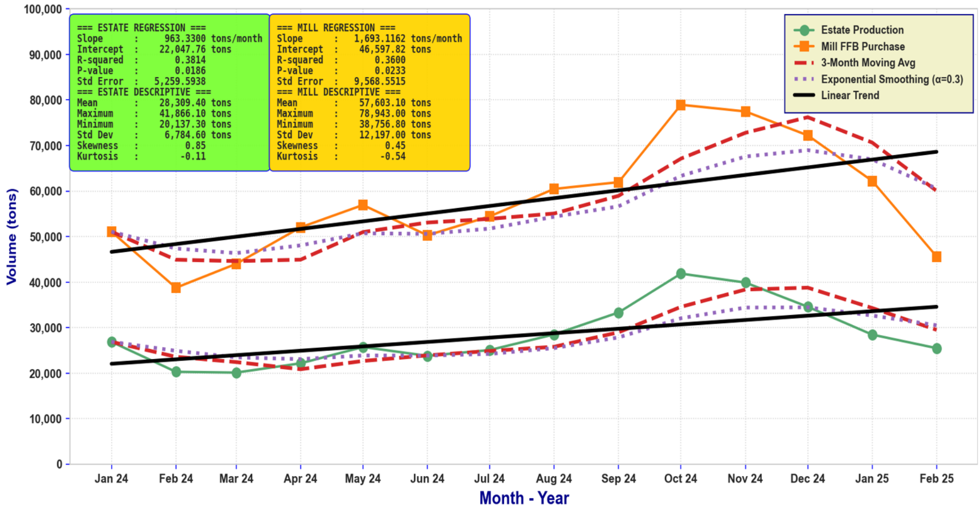

Figure 2: Actual FFB Production in 14 Estates and FFB Purchases in 7 Mills vs. RKAP Targets (Jan–Feb 2025)

4.2 Performance Against RKAP Targets

Figure 3 compares actual FFB production across 14 estates and purchased FFB at 7 mills against RKAP targets for January–February 2025. The results show significant variability: while most estates exceed their targets, some underperform despite favorable totals. Among the 14 estates, 13 estates (green circles) exceeded the RKAP targets for both months, while 1 estate (yellow circle) missed its target in February. This indicates that while most estates perform well, some units need closer monitoring to meet targets.

Comparing estate-origin production and mill purchases highlights the volatility of mill procurement, while estate production is more stable. This emphasizes the need for unit-level analysis. Both estate production and mill purchases fluctuate due to operational and external factors, suggesting the need for tailored interventions to optimize performance and meet RKAP targets consistently.

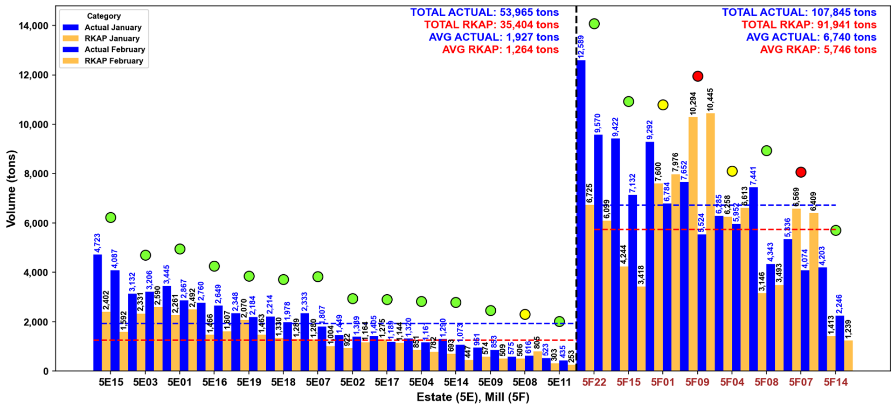

Figure 3: Actual FFB Production in 14 Plantations and FFB Purchases in 7 Mills Versus RKAP Targets in Regional 5, Jan–Feb 2025

Table 2 summarizes the cumulative FFB production for January–February 2025. Estates produced 53,965 tons, achieving 152.41% of the RKAP target, while mills (reported as purchased FFB only) reached 107,845 tons, or 117.30% of the RKAP target. In total, the region achieved 127.06% of the combined RKAP target, highlighting that while the overall performance exceeds the target, it may mask underperformance in specific estates or mills.

Table 1: FFB Production Performance (Jan–Feb 2025)

Description	Actual (tons)	RKAP (tons)	Achievement (%)
Estates (5E)	53,965	35,404	152.41
Mills (5F)	107,845	91,941	117.30
Total	161,810	127,345	127.06

4.3 Descriptive Analytics: Variability and Outlier Analysis

Figure 3 presents boxplots for estate-origin and purchased FFB deliveries to mills for January–February 2025. Estate-origin deliveries exhibit wider interquartile ranges and upper outliers, reflecting substantial variability in harvesting performance across estates. Some estates show exceptionally high production, indicating consistently better-performing units. In contrast, purchased FFB from external suppliers shows even greater dispersion, with higher peaks and more variability. This variability is driven by factors like supplier reliability, logistical constraints, and procurement strategies, which are influenced by external factors such as seasonal demand or transportation issues.

The boxplots highlight that performance challenges are not uniform. Outliers demonstrate how extreme values can skew aggregate results. While some estates and mills consistently deliver high levels of FFB, others show irregular patterns, leading to significant deviations from the average performance.

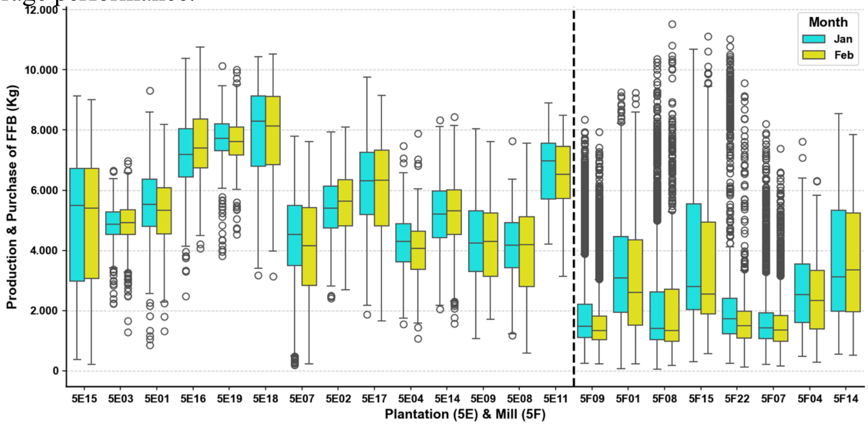

Figure 3: Boxplots of Estate-Origin and Externally Purchased FFB Deliveries to Mills (Jan–Feb 2025)
Source: Research data, 2025

4.4 Correlation and Diagnostic Analysis

The correlation analysis reveals key relationships between operational variables and FFB productivity. Table 2 shows significant correlations, with Spearman's Rank highlighting nonlinear relationships, and Pearson coefficients provided for comparison. Harvesting activity (Pemanen) shows the strongest correlation with FFB production (Spearman $\rho \approx 0.909$), followed by harvesting and transport costs (BPanAng; $\rho \approx 0.706$) and productive palm population (TM_PP; $\rho \approx 0.601$). These results suggest that harvesting intensity and costs are the most influential drivers of FFB production.

However, rainfall (Hujan) demonstrates a weak negative correlation (Spearman $\rho \approx -0.137$), which is statistically insignificant, indicating that linear models may not fully capture its effect on FFB production. This highlights the need for nonlinear models to understand the role of environmental factors more effectively. In conclusion, the analysis suggests that harvesting intensity and cost optimization should be prioritized, while also accounting for nonlinear relationships to enhance FFB productivity

Table 3: Spearman and Pearson Correlation Coefficients Between Variables and FFB Production

No	Variable	Spearman	Pearson	Strength	Significance
1	TM_PP	0.601	0.642	Strong positive	Significant
2	TM_SPH	0.243	0.313	Weak positive	Significant
3	Hujan	-0.137	-0.011	Very weak negative	Not significant
4	Pemanen	0.909	0.895	Very strong positive	Significant
5	BPemPuk	0.246	0.353	Weak positive	Significant
6	BPanAng	0.706	0.616	Strong positive	Significant

Figure 4 displays the estimated smooth functions from the Generalized Additive Model (GAM) for each predictor's influence on FFB production. The plot shows the partial effect of each variable on FFB production while controlling for other factors, with a 95% confidence interval (CI) shaded in light pink.

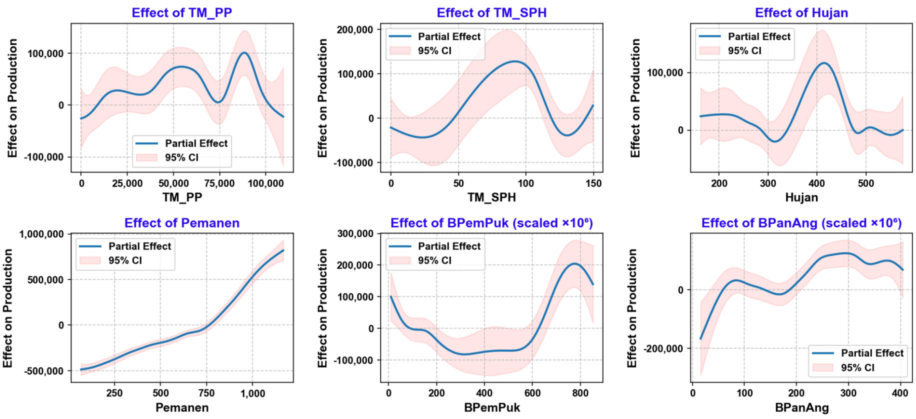

Figure 1: Nonlinear Effects of Key Predictors on FFB Production
Source: Research data, 2025

Figure 4 illustrating the nonlinear effects of key predictors on FFB production. The analysis reveals several important dynamics:

- TM_PP (productive palm population) shows a saturation effect, where further increases in palm population yield diminishing returns.
- TM_SPH (planting density) follows a hump-shaped curve, indicating an optimal density range for maximum production.
- Rainfall has an asymmetric nonlinear effect, where both insufficient and excessive rainfall reduce productivity.
- Pemanen (harvesting activity) has a strong positive effect on production, though with diminishing returns at higher intensities.
- BPemPuk (maintenance and fertilization costs) shows a U-shaped response, indicating budget optimization opportunities.
- BPanAng (harvesting and transport costs) exhibits a positive but nonlinear relationship, reflecting increasing costs as volume rises due to logistical constraints.

These nonlinear dynamics provide valuable insights for operational decision-making, highlighting key points where interventions could be most effective, such as adjusting palm population or optimizing labor deployment or resource allocation

4.5 Predictive Analytics: Forecasting FFB Production

Predictive analysis using the XGBoost model demonstrates strong performance with R^2 = 0.9154, RMSE = 49,036.24, and MAPE = 30.17% (Table 4), outperforming both GAM and Random Forest models. This highlights the model's ability to capture short-term variations in FFB production across different afdelings. The XGBoost model effectively accounts for factors such as agronomic structure, harvesting intensity, and cost data. The results indicate that XGBoost is the most reliable for forecasting FFB production, helping guide operational decisions and strategic planning in plantation management.

Table 4: Performance Comparison of Predictive Models (GAM, Random Forest, XGBoost)

Variable	R^2	MAE	RMSE	MAPE (%)
XGBoost	0.9154	39,370.20	49,036.24	30.17
Random Forest	0.8466	51,179.89	66,029.30	37.61
GAM	0.8532	50,960.02	64,590.64	39.51

Figure 5 compares the observed versus predicted FFB production for the validation period. The predictions from the XGBoost model closely follow the observed trends, capturing the general seasonal fluctuations in production. However, extreme values in the data are partially smoothed, which reflects the model's tendency to moderate outlier effects. Therefore, while the model provides a solid forecast trajectory, it is important to view these predictions as indicative rather than definitive, with field validation and managerial judgment necessary to account for any unforeseen operational disruptions or abnormal weather conditions.

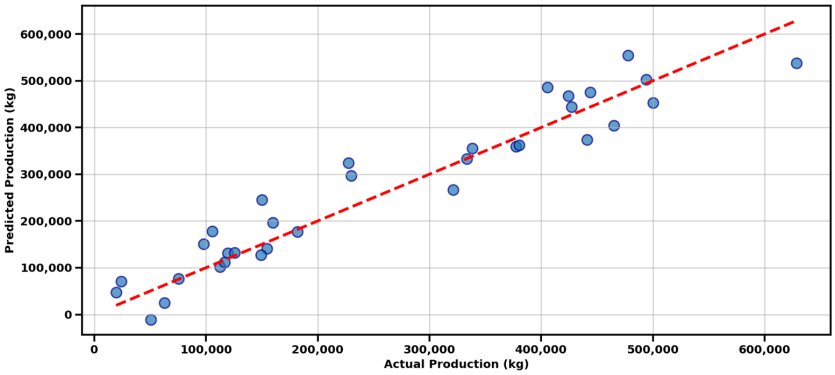

Figure 2: Observed vs. Predicted FFB Production Using XGBoost Model
Source: Research data, 2025

Table 5 compares 2024 production with the 2025 forecast by month, showing total projected FFB production of 399,362.63 tons in 2025, an increase of 56,995.82 tons (16.65%) versus 2024. Under an assumed OER of 21%, the forecast implies an estimated CPO productivity of ~3.52 tons/ha, slightly exceeding the KPI target of 3.403 tons/ha.

Table 5: FFB Production in 2024 and Forecast for 2025 (tons)

Month	2024	2025	Difference	%
January	26,891.44	28,469.39	1,577.95	5.87
February	20,309.19	25,495.53	5,186.34	25.54
March	20,137.28	27,595.50	7,458.22	37.04
April	22,171.98	27,595.50	5,423.52	24.46
May	25,748.93	27,898.08	2,149.15	8.35
June	23,817.50	28,929.39	5,111.89	21.46
July	25,086.63	30,744.73	5,658.10	22.55
August	28,514.11	34,720.56	6,206.45	21.77
September	33,338.04	38,802.51	5,464.47	16.39
October	41,866.12	42,915.97	1,049.85	2.51
November	39,903.10	43,817.40	3,914.30	9.81
December	34,582.49	42,378.07	7,795.58	22.54
Sum	342,366.81	399,362.63	56,995.82	16.65

Assuming an OER of 21%, this level of production translates into an estimated CPO productivity of approximately 3.52 tons/ha, surpassing the KPI target of 3.403 tons/ha. It is important to note that the forecasted FFB production of 399,362.63 tons directly informs the calculation of CPO productivity, which, based on the 21% OER and a plantation area of 23,835.28 hectares, results in an estimated CPO productivity of 3.52 tons per hectare.

Figure 6 illustrates the monthly forecast for 2025, with production levels surpassing 2024 figures in most months, particularly in March, mid-year (June–August), and December. This trend highlights the need for adjustments to RKAP assumptions, ensuring that mill intake capacity, labor scheduling, transport planning, and procurement strategies are aligned with the expected production peaks during these periods.

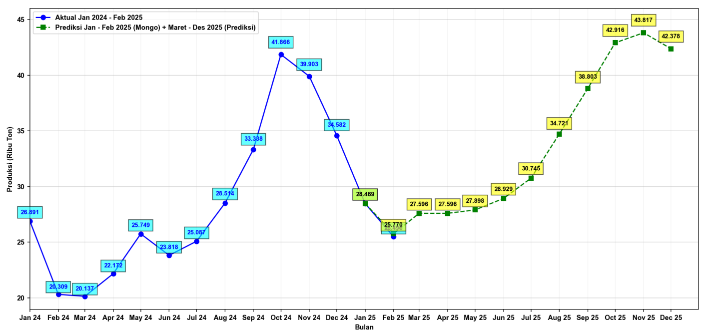

Figure 6: FFB Production Forecast for March-December 2025 (tons)
Source: Research data, 2025

Although the model offers valuable insights, it is essential to use these predictions in conjunction with field validation and managerial experience, considering the inherent uncertainty in forecasted data and the smoothing effects observed in the projections.

4.6 Prescriptive Analytics: Gap and Pareto Analysis

Figure 8 shows the KPI-based gap analysis and Pareto ranking at the afdeling level for January–February 2025. Most afdelings perform near the KPI target, but a subset exhibits significant shortfalls, indicating that underperformance is concentrated in specific units. This suggests that targeted interventions would be more effective than distributing resources evenly across all units.

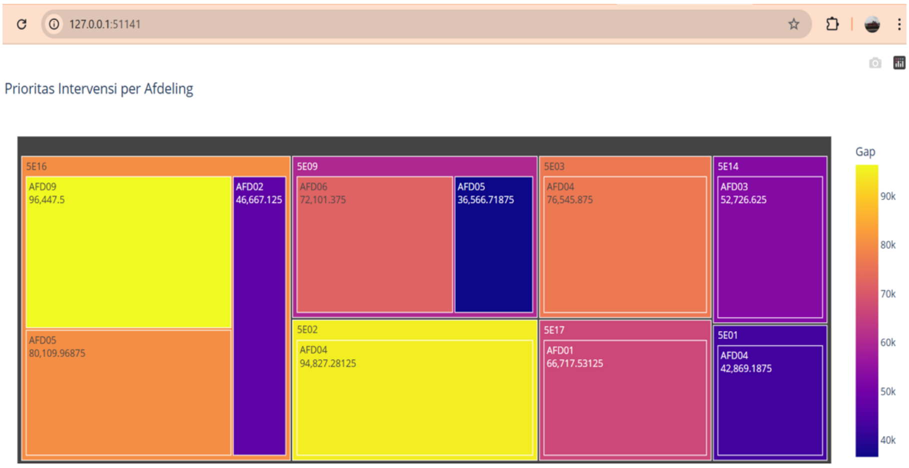

Figure 7: Production Gaps and Pareto Ranking of Afdelings
Source: Research data, 2025

To identify the highest-impact intervention targets, the Pareto analysis in Figure 9 shows a highly concentrated gap profile. The top three afdelings– Afdeling 03 of Pamukan Estate (5E15-03), Afdeling 02 of Gunung Meliau Estate (5E08-02), and Afdeling 05 of Tajati Estate (5E17-05)– account for 80.59% of the total negative gap, underscoring the "vital few" principle. Therefore, these three afdelings are prioritized for intervention, as addressing their combined shortfalls offers the greatest potential to improve Regional V's KPI trajectory in January–February 2025.

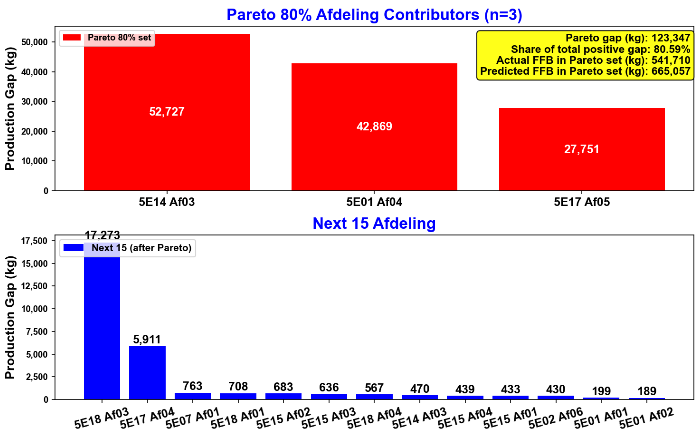

Figure 3: Afdelings with the Highest Production Gap (Pareto Analysis)
Source: Research data, 2025

From a managerial standpoint, combining gap analysis with Pareto ranking operationalizes the principle of focusing on the "vital few" rather than the "trivial many." Instead of spreading resources uniformly across all 81 afdelings, management can prioritize key interventions – such as agronomic audits, maintenance optimization, harvesting supervision, and incentive realignment – in the most impactful units, where improvements are expected to yield the largest performance gains. This targeting approach is particularly valuable under bounded rationality, where managerial focus and budgets are limited.

Lastly, the prescriptive layer builds on the earlier analysis stages. Descriptive analytics establish baseline trends, diagnostic analytics uncover nonlinear drivers and thresholds, predictive analytics forecast future production trajectories, and prescriptive analytics translate these insights into ranked, location-specific priorities. When integrated into dashboards, planning briefs, and operational guidance, these priorities can be reused across planning cycles, supporting more consistent and informed decision-making.

4.7 Zoning-Based Operational Action Map (KM Artifact)

Building on the gap and Pareto analysis, this study translates identified productivity shortfalls into a zoning-based operational action map to guide resource allocation and field execution. Notably, the zones are not automatically generated by the predictive model. Instead, zoning is constructed through a transparent, rule-based synthesis of the analytical outputs, primarily focused on (i) the magnitude of KPI gaps and Pareto ranking ("where to act first") and (ii) dominant driver patterns derived from diagnostic evidence ("what to address").

Figure 9 summarizes this process, presenting it as a repeatable decision tool and Knowledge Management (KM) artifact that can be integrated into routine performance reviews, action tracking, and follow-up monitoring. Five zones are defined to support differentiated managerial responses:

a) Priority Zone: Includes the Pareto set of afdelings accounting for approximately 80% of the total negative KPI gap, requiring immediate, intensive intervention.

b) Improvement Zone: Contains underperforming units outside the Pareto set that need targeted optimization, aligned with key drivers like harvesting intensity, cost efficiency, or agronomic structure.

c) Monitoring Zone: Units close to the benchmark that require closer observation and tactical adjustments to prevent slippage.

d) Optimal Zone: Units meeting or exceeding the benchmark, exhibiting stable performance that should be maintained.

e) Benchmark Zone: Strong-performing units whose best practices can be documented and replicated across other areas.

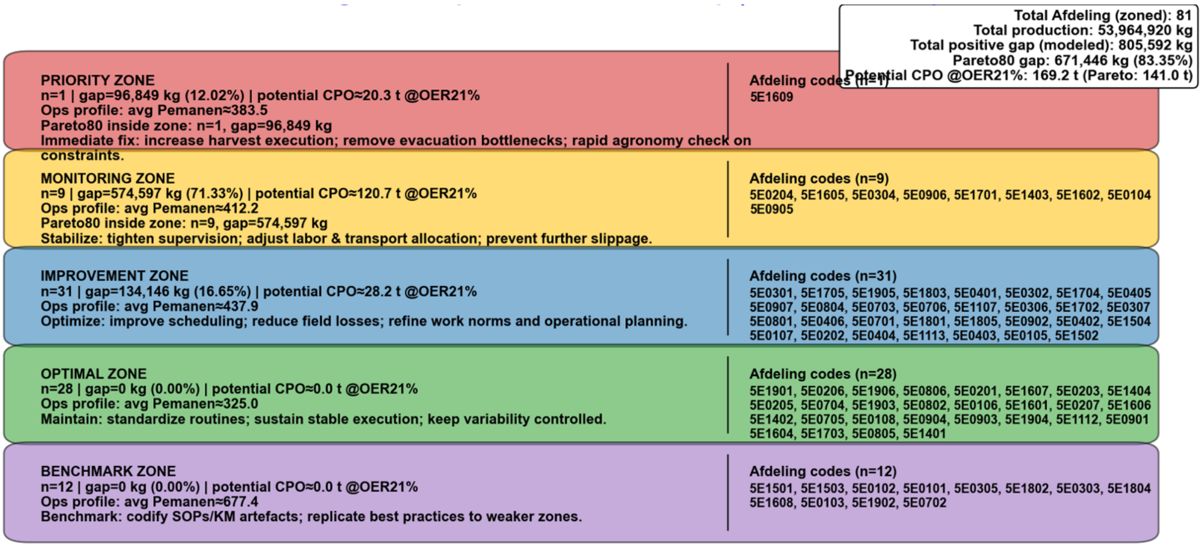

Figure 9: Zoning-based Operational Action Map

4.8 Interpreting Productivity Drivers Through Production Theory

The results indicate that FFB productivity in PTPN IV Regional V is driven by a combination of agronomic, environmental, and operational factors, with nonlinear relationships shaping these dynamics. Structural capacity variables, such as TM_PP (productive palm population) and TM_SPH (planting density), define the feasible production range, establishing the upper limits of productivity based on the available resources. However, the short-term variation in FFB production is more strongly influenced by operational execution –especially Pemanen (harvesting activity) and the associated logistics/cost structure (BPanAng), which play a crucial role in translating potential capacity into actual output. The Generalized Additive Model (GAM) diagnostics reveal key nonlinear dynamics, such as saturation points and diminishing returns, highlighting that production efficiency improves up to a certain threshold of input (e.g., increasing palm population or planting density) before leveling off. This supports the theory of production, which suggests that inputs can result in diminishing returns after reaching an optimal level.

4.9 Predictive Analytics and Planning Relevance

The XGBoost model achieves strong predictive performance ($R^2 \approx 0.92$), demonstrating that monthly afdeling-level production can largely be explained by SAP ERP variables, complemented by rainfall data. The primary contribution of predictive analytics in this study is its relevance for planning: projections for March–December 2025 offer a structured foundation for aligning production expectations with mill intake capacity, labor deployment, and transport planning, thus aiding adjustments to the RKAP (Rencana Kerja dan Anggaran Perusahaan) and improving cross-functional coordination. At the same time, forecast smoothing and residual errors suggest that these projections should be treated as indicative trajectories, subject to field validation, especially under varying weather conditions or operational disruptions. This approach enhances long-term decision-making by helping managers anticipate and manage seasonal and operational fluctuations.

4.10 BDA, Bounded Rationality, and Knowledge Management

From a decision-making perspective, the integrated analytics pipeline demonstrates how BDA can mitigate **bounded rationality** in the context of complex and semi-structured plantation operations (Simon, 1997). The descriptive and diagnostic phases address where performance varies and why, while the forecasting phase offers actionable insights into what is plausible ahead, supporting long-term planning. The prescriptive analysis, through gap–Pareto ranking, identifies where to intervene first, focusing on the "vital few" high-leverage afdelings that contribute most to the performance gap. This approach counters the inefficiency of applying uniform policies across all units. The results emphasize quick wins –units that underperform relative to their potential and can yield faster improvements through operational tightening. Meanwhile, the **Priority Zone** remains the focal point for closing the KPI gap.

These insights are codified into Knowledge Management artifacts such as dashboards, driver/threshold guides, priority lists, zoning maps, and action logs. By embedding these artifacts into decision-making cycles, the framework supports **repeatable review-action-learning cycles**, ensuring that BDA outputs are actionable and continuously integrated into operational practices. This approach empowers managers to focus their cognitive resources on a small set of critical interventions, maximizing operational impact and driving productivity improvements.

4.11 Broader Applicability Beyond Plantation Operations

While this study is grounded in PTPN IV Regional V, the core framework –integrating BDA with KM for operational decision-making– can be adapted to other asset- and operations-intensive industries with distributed units and heterogeneous performance. The logic of this framework is transferable across sectors, where the process remains largely the same: unit-level data is gathered, key drivers are diagnosed (including nonlinearities), short-term forecasts are generated for planning purposes, and prescriptive prioritization is employed. The results are then translated into actionable insights, using KM artifacts like playbooks, dashboards, and action logs, which institutionalize the decision-making process.

In mining, for instance, operational data (e.g., tonnage moved, cycle/queue times, downtime, and maintenance costs) can be used to diagnose controllable drivers (e.g., truck–shovel matching, road conditions, dispatch rules) and prioritize the "vital few" pits or haul routes. Similar to this study's approach, Pareto analysis can be used to highlight the most significant areas requiring intervention. In forestry, stand-level growth, harvest, and road accessibility data can be analyzed to identify nonlinear growth responses and operational constraints. The results can help prioritize stands contributing most to shortfalls in volume or cost overruns. In logistics, operational data (e.g., shipment volume, on-time delivery, route distance) can reveal bottlenecks and their underlying causes. Prescriptive analytics can then identify the most problematic depots or routes that require immediate action, which can be monitored through dashboards and action logs.

This study demonstrates how a structured BDA–KM decision process –integrating descriptive, diagnostic, predictive, and prescriptive analytics– can lead to informed, repeatable decision-making, improving operational efficiency and driving long-term performance improvement in large-scale, complex industries.

5. Conclusion

5.1 Key Findings

This study provides empirical evidence that Big Data Analytics can significantly enhance KPI-oriented decision-making in oil palm plantation management. The findings reveal that 80.59% (123.34 tons of FFB) of the productivity gap in PTPN IV Regional V is concentrated in just three afdelings, emphasizing the need for focused interventions. The study identifies nonlinear relationships between key factors such as productive palm population (TM_PP), planting density (TM_SPH), rainfall, harvesting activity (Pemanen), and logistics/cost structure (BPanAng), which challenge linear planning assumptions and highlight the importance of targeted resource allocation.

5.2 Theoretical Implications: BDA–KM Synergy

This study contributes to the literature on the synergy between BDA and Knowledge Management by demonstrating how the integration of BDA and KM processes enhances decision-making in plantation operations. It shows that BDA helps reduce cognitive complexity in complex, semi-structured environments by transforming operational data into actionable insights. These insights are then codified and stored as KM artifacts, supporting continuous performance improvement and structuring decision-making cycles. This study goes beyond treating BDA as a technical tool, positioning outputs –such as driver patterns, thresholds, and Pareto priorities– as reusable knowledge assets. These assets can be integrated into organizational routines, facilitating sustained performance improvement.

5.3 Managerial Implications

For PTPN IV Regional V, the study offers actionable insights for improving performance. Nonlinear diagnostics suggest the limits of uniform policies and highlight the need for customized input allocation tailored to afdeling-specific conditions. Predictive forecasting supports RKAP adjustments, aligning estate production with mill intake capacity and labor/transport planning. Gap and Pareto analysis direct resources to the top three underperforming afdelings, which contribute to 80.59% of the total negative KPI gap, providing a strategic focus for intervention. These insights can be embedded into KM artifacts –such as dashboards, priority lists, and zoning maps– distributed through existing KM channels (e.g., PALMS) to ensure consistent follow-through and continuous improvement across estates.

5.4 Limitations, Future Research, and Scalability

This study is limited by its focus on a single regional unit and a short observation period, which restricts the generalizability of the findings across other regions, years, and agroecological conditions. The study also relies primarily on ERP-based operational and cost data, with other potentially important variables, such as oil properties and terrain, not incorporated. Future research could explore these additional variables to enhance the depth of analysis and improve scenario planning. Additionally, the study does not evaluate how KM mechanisms –such as training completion and reuse rates– impact the adoption of analytics. Further work should investigate the long-term effectiveness of integrating BDA insights into routine operations and whether the benefits persist across repeated review-action-learning cycles. Finally, the results should be interpreted as short-term insights aimed at improving planning and control, with longer-term studies extending the analysis across multiple years and regions to capture broader dynamics.

5.5 Scalability

The BDA–KM decision framework developed in this study is highly scalable and can be extended to other sectors, such as mining, forestry, and logistics. In these industries, similar analytics methods can be used to identify key drivers, forecast future outcomes, and prioritize high-leverage interventions. By applying this integrated approach, organizations in various sectors can address performance gaps, optimize resource allocation, and enhance operational efficiency. Furthermore, the BDA–KM framework can be scaled for enterprise-wide adoption, including across the entire PTPN Holding group, ensuring consistent and sustained performance improvements at a national level. This scalability not only demonstrates the flexibility of the framework but also its potential to drive data-driven decision-making in large, complex organizations beyond the palm oil industry.

References

[1] Abdul, I. (2023). *Merancang kelapa sawit sebagai komoditi unggulan nasional*. PT Literasi Nusantara Abadi Grup.

[2] Abubakar, A. M., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation & Knowledge, 4*(2), 104–114. https://doi.org/10.1016/j.jik.2017.07.003

[3] Aljehani, S. B., Abdo, K. W., Alam, M. N., & Aloufi, E. M. (2024). Big data analytics and organizational performance: Mediating roles of green innovation and knowledge management in telecommunications. *Sustainability, 16*(18). https://doi.org/10.3390/su16187887

[4] Aydin, N., & Yurdakul, G. (2021). Analyzing the efficiency of bank branches via novel weighted stochastic imprecise data envelopment analysis. *RAIRO – Operations Research, 55*(3), 1559–1578. https://doi.org/10.1051/ro/2021067.

[5] Balusamy, B., Abirami. R, N., Kadry, S., & Gandomi, A. H. (2021). *Big data: Concepts, technology, and architecture.* John Wiley & Sons.

[6] Banowati, E., & Sriyanto. (2013). *Geografi pertanian.* Ombak.

[7] Becerra-Fernandez, I., Sabherwal, R., & Richey, R. (2024). *Knowledge management: Systems and processes in the AI era* (3rd ed.). Routledge.

[8] Badan Pusat Statistik. (2024). *Statistik kelapa sawit Indonesia 2023*. Badan Pusat Statistik.

[9] Brodsky, A., Shao, G., & Krishnamoorthy, M. (2015). *Analysis and optimization based on reusable knowledge base of process performance models* (NISTIR 8094). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8094

[10] Chatterjee, S., Chaudhuri, R., Gupta, S., Sivarajah, U., & Bag, S. (2023). Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm. *Technological Forecasting and Social Change, 196*, 122824. https://doi.org/10.1016/j.techfore.2023.122824

[11] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[12] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*, e623. https://doi.org/10.7717/peerj-cs.623

[13] Cote, C. (2021). *4 types of data analytics to improve decision-making*. Harvard Business School Online. https://online.hbs.edu/blog/post/types-of-data-analysis

[14] Dalkir, K. (2011). *Knowledge management in theory and practice* (2nd ed.). The MIT Press.

[15] Dalkir, K. (2023). *Knowledge management in theory and practice* (4th ed.). The MIT Press.

[16] Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business School Press.

[17] de Vos, R. E., Nurfalah, L., Tenorio, F. A., Lim, Y. L., Monzon, J. P., Donough, C. R., Sugianto, H., Dwiyahreni, A. A., Winarni, N. L., Mulani, N., Ramadhan, G., Imran, M. A., Tito, A. P., Sulistiawan, P., Khoirul, M., Farrasati, R., Pradiko, I., Grassini, P., & Slingerland, M. (2023). Shortening harvest interval, reaping benefits? A study on harvest practices in oil palm smallholder farming systems in Indonesia. *Agricultural Systems, 211*, 103753. https://doi.org/10.1016/j.agsy.2023.103753

[18] Eckerson, W. (2011). *Big data analytics: Profiling the use of analytic platforms in user organizations*. BeyeNetwork.

[19] Elwirehardja, G. N., Suparyanto, T., Miftakhurrokhmat, & Pardamean, B. (2023). Determining variables associated with annual oil palm yield: An explainable gradient boosting approach. *Procedia Computer Science, 227*, 262–271. https://doi.org/10.1016/j.procs.2023.10.524

[20] EMC Education Services. (2015). *Data science & big data analytics: Discovering, analyzing, visualizing and presenting data*. John Wiley & Sons.

[21] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General), 120*(3), 253–290. https://doi.org/10.2307/2343100

[22] Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., & Luck, J. (2020). Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning. *Remote Sensing, 12*(12), 2028. https://doi.org/10.3390/rs12122028

[23] Firmansyah, E., Mawandha, H. G., Umami, A., Nurjanah, D., Dinarti, S. I., Puruhito, D. D., & Purwadi. (2022). *Pengelolaan peremajaan kelapa sawit*. Instiper Press.

[24] Fitri, A., Rahim, R., Nurhayati, & Pagiling, A. S. (2023). *Dasar-dasar statistika untuk penelitian*. Yayasan Kita Menulis.

[25] Fleiss, S., McClean, C. J., King, H., & Hill, J. K. (2022). Limited impacts of climatic conditions on commercial oil palm yields in Malaysian plantations. *CABI Agriculture and Bioscience, 3*, 59. https://doi.org/10.1186/s43170-022-00127-1

[26] Franceschini, F., Galetto, M., & Maisano, D. (2018). *Designing performance measurement systems: Theory and practice of key performance indicators*. Springer.

[27] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

[28] Gartner. (2025). *IT glossary*. https://www.gartner.com/en/information-technology/glossary/big-data

[29] Ginting, C. (2024). *Manajemen nutrisi kelapa sawit*. Instiper Press.

[30] Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill/Irwin.

[31] Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics. *Journal of Management Information Systems, 35*(2), 388–423. https://doi.org/10.1080/07421222.2018.1451951

[32] Hajric, E. (2018). *Knowledge management system and practices: A theoretical and practical guide for knowledge management in your organization*. https://anyflip.com/bkmdx/ydhj/basic

[33] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

[34] Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall. https://doi.org/10.1201/9780203753781

[35] Herrmann, J. W. (2015). *Engineering decision making and risk management*. John Wiley & Sons.

[36] Hislop, D. (2005). *Knowledge management in organizations: A critical introduction*. Oxford University Press.

[37] Hwang, K., & Chen, M. (2017). *Big-data analytics for cloud, IoT and cognitive computing*. John Wiley & Sons.

[38] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. https://otexts.com/fpp2

[39] Imran, S., & Indriani, R. (2022). *Ekonomi produksi pertanian*. Ideas Publishing.

[40] International Organization for Standardization. (2019). *ISO/IEC 20546:2019 –Information technology–Big data– Overview and vocabulary*. International Organization for Standardization.

[41] Jain, P., & Sharma, P. (2015). *Behind every good decision: How anyone can use business analytics to turn data into profitable insight*. AMACOM.

[42] Japkowicz, N., & Boukouvalas, Z. (2025). *Machine learning evaluation*. Cambridge University Press. https://doi.org/10.1017/9781009003872

[43] Jamshidi, E. J., Yusup, Y., Hooy, C. W., Kamaruddin, M. A., Mat Hassan, H., Muhammad, S. A., Mohd Shafri, H. Z., Then, K. H., Norizan, M. S., & Tan, C. C. (2024). Predicting oil palm yield using a comprehensive agronomy dataset and 17 machine learning and deep learning models. *Ecological Informatics, 81*, 102595. https://doi.org/10.1016/j.ecoinf.2024.102595

[44] Jennex, M. E. (2006). What is knowledge management? *International Journal of Knowledge Management, 2*(4). https://www.igi-global.com/chapter/knowledge-management/24979

[45] Jones, G. R. (2013). *Organizational theory, design, and change* (7th ed.). Pearson.

[46] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

[47] Kamil, N. N., Xiao, S., Syed Salleh, S. N., Xu, H., & Zhuang, C. C. (2024). Nonlinear impacts of climate anomalies on oil palm productivity. *Heliyon, 10*(15), e35798. https://doi.org/10.1016/j.heliyon.2024.e35798

[48] Kampoowale, I. (2025). Linking big data analytics capabilities to organizational learning through knowledge management and data-driven decision-making. *The TQM Journal*. https://doi.org/10.1108/TQM-08-2024-0282

[49] Kementerian Pertanian Republik Indonesia. (2013). *Peraturan Menteri Pertanian Nomor 131/Permentan/OT.140/12/2013 tentang pedoman budidaya kelapa sawit (Elaeis guineensis) yang baik*. Kementerian Pertanian Republik Indonesia.

[50] Kerzner, H. (2017). *Project management metrics, KPIs, and dashboards: A guide to measuring and monitoring project performance* (3rd ed.). John Wiley & Sons.

[51] Khan, A. K., Zaman, S. U., & Alam, S. H. (2024). Impact of big data analytics on organizational performance: The role of business analytics, decision-making quality and sustainability. *Technology Review and Trends*, *3*(1), 389–406. https://doi.org/10.63062/trt/V24.069

[52] Kotulic, R., & Pavelková, J. (2014). The application of the Cobb–Douglas production function in analyzing the effectiveness of productive resources in agricultural enterprises of primary production. *Journal of Central European Agriculture, 15*, 284–301. https://doi.org/10.5513/JCEA01/15.3.1489

[53] Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. META Group.

[54] Lebas, M. J. (1995). Performance measurement and performance management. *International Journal of Production Economics, 41*(1–3), 23–35. https://doi.org/10.1016/0925-5273(95)00081-X

[55] Li, K.-C., Jiang, H., & Yang, L. T. (2015). *Big data algorithms, analytics, and applications*. Chapman & Hall/CRC. https://doi.org/10.1201/b18050

[56] Li, L., Lin, J., Ouyang, Y., & Luo, X. R. (2022). Evaluating the impact of big data analytics usage on the decision-making quality of organizations. *Technological Forecasting and Social Change, 175*, 121355. https://doi.org/10.1016/j.techfore.2021.121355

[57] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), 2674. https://doi.org/10.3390/s18082674

[58] Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology, 150*(11), 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

[59] Makhloufi, L., Vasa, L., Rosak-Szyrocka, J., & Djermani, F. (2023). Understanding the impact of big data analytics and knowledge management on green innovation practices and organizational performance: The moderating effect of government support. *Sustainability*, *15*(11), 8456. https://doi.org/10.3390/su15118456

[60] Monzon, J. P., Lim, Y. L., Tenorio, F. A., Farrasati, R., Pradiko, I., Sugianto, H., Donough, C. R., Rattalino Edreira, J. I., Rahutomo, S., Agus, F., Slingerland, M. A., Zijlstra, M., Saleh, S., Nashr, F., Nurdwiansyah, D., Ulfaria, N., Winarni, N. L., Zulhakim, N., & Grassini, P. (2023). Agronomy explains large yield gaps in smallholder oil palm fields. *Agricultural Systems, 210*, 103689. https://doi.org/10.1016/j.agsy.2023.103689

[61] Malangyoedo, A. (2017). *Sukses pengelolaan perkebunan kelapa sawit produktivitas tinggi*. Lily Publisher. https://web-ikepri.moco.co.id/read-book

[62] McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 60–68. https://hbr.org/2012/10/big-data-the-management-revolution

[63] Meissner, G. (2014). *Correlation risk modeling and management: An applied guide including the Basel III correlation framework–with interactive correlation models in Excel/VBA*. John Wiley & Sons.

[64] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia, 22*(1). https://doi.org/10.4103/aca.aca_157_18
[65] Mohamad Akmal, M. Z., Ooi, J., Ng, W. P. Q., How, B. S., Lam, H. L., Foo, D. C. Y., & Lim, C. H. (2025). Impact of industry 4.0 technologies on the oil palm industry: A literature review. *Smart Agricultural Technology, 10*, 100685. https://doi.org/10.1016/j.atech.2024.100685
[66] Mokhtar, S. M., & Rahman, S. A. (2017). The roles of big data and knowledge management in business decision making process. *International Journal of Academic Research in Business and Social Sciences, 7*. https://doi.org/10.6007/IJARBSS/v7-i12/3623
[67] Monita, C. F., & Zebua, D. D. (2023). Faktor-faktor yang mempengaruhi produktivitas kelapa sawit di PT Mustika Agung Sentosa. *Jurnal Manajemen Agribisnis, 11*, 231–243. https://doi.org/10.24843/JMA.2023.v11.i01
[68] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). John Wiley & Sons.
[69] Neely, A. (Ed.). (2007). *Business performance measurement: Unifying theories and integrating practice*. Cambridge University Press.
[70] Nora, S., & Mual, C. D. (2018). *Budidaya tanaman kelapa sawit*. Kementerian Pertanian.
[71] Nugroho, A. (2019). *Teknologi agroindustri kelapa sawit*. Lambung Mangkurat University Press.
[72] OpenStax. (2019). *Principles of management*. Rice University.
[73] Parmenter, D. (2015). *Key performance indicators: Developing, implementing, and using winning KPIs* (3rd ed.). John Wiley & Sons.
[74] Parwati, W. U., Nadeak, F. H., & Kautsar, V. (2024). Analisis pertumbuhan dan produktivitas kelapa sawit pada variasi kerapatan tanam. *Jurnal Agro Industri Perkebunan*. https://doi.org/10.25181/jaip.v12i2.3535
[75] Pasolong, H. (2023). *Teori pengambilan keputusan*. Alfabeta.
[76] Peruri Venkata Anusha, Anuradha, C., Murty, P. S., & Kiran, C. S. (2019). Detecting outliers in high dimensional datasets using Z-score methodology. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9*. https://doi.org/10.35940/ijitee.A3910.119119
[77] Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
[78] Pyne, S., & Rao, B. P. (2016). *Big data analytics: Methods and applications*. Springer.
[79] Ramasamy, T. (2010). *Principles of management*. Himalaya Publishing House.
[80] Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2020). Reshaping business with artificial intelligence. *MIT Sloan Management Review*. https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/
[81] Rehman, S. U., Elrehail, H., & Taam, A. (2024). Antecedents and consequences of big data knowledge management. *International Journal of Information Management Data Insights, 4*. https://doi.org/10.1016/j.jjimei.2024.100265
[82] Robbins, S. P., & Judge, T. A. (2024). *Organizational behavior* (19th ed.). Pearson Education.
[83] Rusdi, M. I., Bashkeran, T., & Ahmad, Z. (2022). Data-driven technique for palm oil mills boiler water quality monitoring in the state of Perak, Malaysia. *Malaysian Journal of Fundamental and Applied Sciences, 18*, 171–183. https://doi.org/10.11113/mjfas.v18n2.2242
[84] Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations*. The Macmillan Company.
[85] Simon, H. A. (1997). *Administrative behavior: A study of decision-making processes in administrative organizations* (4th ed.). The Free Press.
[86] Sukardi, & Hendrik, B. (2025). Prediksi produktivitas lahan kelapa sawit pasca peremajaan menggunakan algoritma machine learning. *Journal of Science and Social Research, 8*.
[87] Sulardi. (2022). *Budidaya tanaman kelapa sawit*. PT Dewangga Energi Internasional.
[88] Taherdoost, H., & Madanchian, M. (2023). Decision making: Models, processes, techniques. *Current Chinese Economic Development Studies*. https://doi.org/10.37256/ccds.5120233284
[89] Tumwebaze, R. P., Walsh, J. N., & Lannon, J. (2025). Knowledge management in the agriculture sector: A systematic literature review. *Knowledge Management Research & Practice, 23*(2), 131–148. https://doi.org/10.1080/14778238.2024.2359419
[90] Uula, M. M. (2024). *Analisis efisiensi dan produktivitas sektor perkebunan di Sumatra: Data envelopment analysis (DEA) dan Malmquist productivity index (MPI)*. https://sumatranomics.com/wp-content/uploads/2024/03/13387040409_Kategori-Mahasiswa-Juara-3.pdf
[91] Uyar, H., Karvelas, I., Rizou, S., & Fountas, S. (2024). Data value creation in agriculture: A review. *Computers and Electronics in Agriculture, 227*, 109602. https://doi.org/10.1016/j.compag.2024.109602
[92] Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2017). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234–246. https://doi.org/10.1016/j.ijpe.2014.12.031
[93] Wissuchek, C., & Zschech, P. (2025). Prescriptive analytics systems: Revise a SLR from an information systems perspective. *Information Systems and e-Business Management, 23*, 279–353. https://doi.org/10.1007/s10257-024-00688-w
[94] Wixom, B. H., Beath, C. M., & Owe, L. (2023). *Data is everybody's business: The fundamentals of data monetization*. The MIT Press.
[95] Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). Big data in smart farming: A review. *Agricultural Systems, 153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.02
[96] Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman & Hall/CRC. https://doi.org/10.1201/9781315370279
[97] Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association, 111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986
[98] Zahriyah, A., Suprianik, Parmono, A., & Mustofa. (2021). *Ekonometrika: Teknik dan aplikasi dengan SPSS*. Mandala Press.
[99] Zheng, A. (2015). *Evaluating machine learning models: A beginner's guide to key concepts and pitfalls*. O'Reilly Media.