

Data-Driven Governance in Agriculture: Leveraging MyAgrilytics for Predictive Analytics and National Food Security

Theenarachigee Muniandy¹, Mohd Kamarul Othman², Johari Mat³, Azrul Fazwan Kharuddin⁴, Surendran Elangoven⁵

¹ Ministry of Agriculture and Food Security, Malaysia.

Email: theenarachigee@motac.gov.my

² Human Capital Development Division, Public Service Department of Malaysia, Malaysia.

³ Universiti Tun Abdul Razak, Malaysia.

Email: johari.mat@unirazak.edu.my

⁴ Universiti Tun Abdul Razak, Malaysia.

Email: azrul.fazwan@ur.unirazak.edu.my

⁵ Intelli Education Group Sdn Bhd (IEG Campus), Malaysia.

Abstract

This study investigates the development, implementation, and performance of an artificial intelligence (AI)-enabled intelligence analytics platform; Project Sentinel designed to address fundamental challenges in anti corruption investigations: unstructured data overload, slow analytical workflows, and poor relational insight. Drawing on the growing literature that demonstrates the potential of machine learning and natural language processing (NLP) to support criminal investigation, this research examined how AI can automate the structuring of free text, improve entity recognition, reveal underlying linkages through network analysis, and enhance anomaly detection. A mixed methods approach was adopted, integrating quantitative performance metrics (e.g., processing speed, precision, recall, F score) with qualitative feedback from intelligence practitioners. Results indicate that AI automation reduced processing time by more than 50%, significantly improved entity extraction accuracy, and enhanced relational detection across cases. These findings align with extant research on AI in policing and intelligence analytics, which highlights both performance gains and ethical considerations inherent in such systems. Project Sentinel's outcomes also support Sustainable Development Goal (SDG) 16, which calls for accountable institutions and reduced corruption, by enabling faster, more transparent, and data driven investigation processes. In conclusion, Project Sentinel demonstrates that AI can transform anti corruption workflows, but sustainable adoption requires ethical governance, transparency, and interdisciplinary collaboration to mitigate bias and uphold rights. Recommendations include structured ethical oversight, training, and longitudinal evaluation to guide AI integration in public integrity fun The increasing complexity of the agrofood sector, driven by climate variability, geopolitical disruptions, and social media misinformation, underscores the need for integrated, high-quality data to support evidence-based decision-making. This study examines MyAgrilytics, a national initiative by the Ministry of Agriculture and Food Security, designed to unify crop, livestock, fisheries, market, research, and enforcement data into a centralized, analytics-ready data lake. Drawing on literature that emphasizes the importance of big data integration and predictive analytics in agriculture, the platform applies standardized data parsing, validation, and master-data management to improve reliability and reduce duplication. The study adopts a descriptive-exploratory research design, combining purposive sampling of cross-agency datasets, data lifecycle analysis, and predictive modeling to assess data quality, time-to-insight, and analytical capability. Findings indicate that MyAgrilytics achieves high data quality (DQI >93%), dramatically reduces time-to-insight for officers (by ~98%), and enables reliable predictive analytics for crop production, market price monitoring, and disaster impact simulations (MAPE <6%). These results demonstrate the platform's capacity to enhance operational efficiency, proactive decision-making, and cross-sectoral coordination. The initiative aligns directly with the United Nations Sustainable Development Goals (SDG 2: Zero Hunger, SDG 12: Responsible Consumption and Production, SDG 13: Climate Action) by supporting timely interventions, sustainable resource management, and resilience against agrofood disruptions. In conclusion, MyAgrilytics represents a scalable, AI-ready infrastructure that strengthens national food security while enhancing sector competitiveness. The study recommends continued integration of real-time data sources, training for officers, and robust cross-agency governance to sustain long-term impact.ctions.....

Keywords: Agrofood data integration, Data lake, Predictive analytics, Food security, AI in agriculture, MyAgrilytics.

INTRODUCTION

The contemporary agrofood sector is increasingly complex, facing multifaceted challenges that range from climate variability and natural disasters to geopolitical disruptions and misinformation propagated through social media. Effective decision-making in this context requires timely, accurate, and integrated information. Governments and ministries of agriculture are under growing pressure to ensure food security, maintain market stability, and enhance sector competitiveness. In many countries, however, these objectives are constrained by the fragmentation of agrofood data across multiple agencies, ministries, and research institutions. Fragmented data not only delays the generation of actionable insights but also compromises the quality of policy decisions, leading to inefficiencies in production forecasting, disaster response, and market regulation (FAO, 2022; Inmon, 2005).

In response to these challenges, the Ministry of Agriculture and Food Security has initiated MyAgrilytics, a national platform designed to unify diverse agrofood datasets into a single, authoritative, analytics-ready data lake. The platform embodies the vision of “Harvesting Insight” by integrating data from crop production, livestock, fisheries, market facilitation, research and development, regulatory enforcement, and overseas representation. By centralizing data and standardizing its quality, MyAgrilytics seeks to provide a reliable foundation for AI-driven analytics and decision-support tools. This integration is essential in addressing the urgent need for proactive, evidence-based policy responses to both routine and crisis events, including sudden price spikes, natural disasters, and trade disruptions (Shearer, 2000).

Despite the critical importance of data for policy and operational decisions, current data management practices within the agrofood sector exhibit several limitations. First, datasets are often siloed across different divisions and agencies, resulting in duplication and inconsistent reporting. Multiple divisions may request similar data from the same sources, creating redundant workflows and administrative inefficiencies. Second, the lack of standardized data parsing, validation, and master-data management undermines data quality, leading to errors, mismatches, and incomplete information. Third, the absence of a centralized, analytics-ready repository hinders the deployment of advanced AI and machine learning models, which rely on clean, structured data for predictive forecasting and simulation (FAO, 2022; Inmon, 2005). These challenges collectively reduce the Ministry’s agility in responding to crises, limit foresight into market and production trends, and impede effective planning for national food security.

The fragmented nature of agrofood data also affects policy alignment and implementation. Ministries are tasked with a dual mandate: to enhance the competitiveness of the agrofood sector while ensuring the continuous supply of safe, nutritious, and affordable food. Inconsistent or delayed data undermines both objectives by preventing timely interventions in production cycles, supply chain management, and market stabilization. Moreover, social media-driven misinformation about food availability or safety can exacerbate public anxiety and disrupt markets, further highlighting the need for a trusted, unified source of truth (FAO, 2022).

The purpose of this study is to examine the design, implementation, and anticipated impact of the MyAgrilytics platform in addressing these challenges. Specifically, it seeks to explore how the centralization of agrofood data into a validated, AI-ready data lake can reduce duplication, improve data quality, enable predictive analytics, and shorten the time-to-insight for policymakers and analysts. By doing so, the study aims to provide a comprehensive understanding of how integrated data management can enhance the Ministry’s capacity for evidence-based decision-making and proactive response in a volatile agrofood environment.

Literature Review

Several studies have explored the role of centralized data systems and analytics platforms in enhancing decision-making within agriculture and food security. Big data initiatives have demonstrated significant potential in improving crop yield forecasting, livestock management, fisheries monitoring, and market regulation by providing integrated, real-time insights (Kamilaris et al., 2017; Wolfert et al., 2017). For instance, Wolfert et al. (2017) highlighted that agricultural big data platforms reduce information asymmetry across stakeholders, enabling timely interventions that prevent post-harvest losses and stabilize market prices. Similarly, Kamilaris et al. (2017)

emphasized the role of predictive analytics and machine learning in anticipating climatic impacts on crop and livestock productivity, demonstrating that data-driven foresight contributes to resilience in agrofood systems. These findings reinforce the importance of unifying disparate agrofood datasets, as fragmentation hinders the effective deployment of advanced analytics tools and limits actionable insights.

In the domain of market facilitation and trade analytics, previous research indicates that integrated data lakes improve the speed and accuracy of price monitoring and anomaly detection. Kshetri (2018) argued that centralized repositories enable policymakers to identify emerging market disruptions and respond to sudden price fluctuations more effectively. Likewise, studies on fisheries and livestock management have shown that data standardization and cross-agency sharing facilitate compliance monitoring, resource allocation, and sustainability assessments (Li et al., 2020). Collectively, these studies suggest that integrating diverse agrofood datasets across production, market, and enforcement domains can reduce duplication, improve trust in official data, and empower decision-makers with near real-time analytical capabilities.

The theoretical foundation of MyAgrilytics is grounded in the Data-Information-Knowledge-Wisdom (DIKW) hierarchy, which postulates that raw data, when structured, validated, and analyzed, generates actionable knowledge that informs strategic decisions (Rowley, 2007). By consolidating heterogeneous datasets into a single, authoritative repository, MyAgrilytics operationalizes this theory by converting fragmented data into actionable insights for policymakers. Complementing DIKW, the Socio-Technical Systems Theory provides a lens for understanding the interaction between technological infrastructure and organizational processes. The theory underscores that data platforms alone are insufficient unless accompanied by governance frameworks, user-centric interfaces, and cross-agency collaboration (Baxter & Sommerville, 2011). These theoretical perspectives justify the emphasis on both technological and institutional integration in the design of MyAgrilytics.

Conceptually, MyAgrilytics is structured around three interrelated variables: data centralization, analytics capability, and decision-making effectiveness. Data centralization involves consolidating heterogeneous agrofood datasets, implementing standardization, and removing duplication. Analytics capability refers to the platform's ability to apply AI, machine learning, and predictive models to generate forecasts, anomaly detection, and scenario simulations. Decision-making effectiveness is defined as the timeliness, accuracy, and proactivity of interventions in production, market regulation, and disaster management. The conceptual framework posits that centralized, validated data serves as a foundation for advanced analytics, which in turn enhances the Ministry's decision-making capacity. This causal pathway aligns with findings from prior research on data-driven governance and integrated agricultural information systems (Wolfert et al., 2017; Kamilaris et al., 2017).

Research Methodology

This study adopts a descriptive and exploratory research design to examine the implementation and effectiveness of the MyAgrilytics platform in enhancing decision-making within the agrofood sector. Descriptive research is appropriate for documenting existing processes, identifying data management challenges, and evaluating the preliminary impact of centralized data systems (Creswell & Creswell, 2018). The exploratory component allows the investigation of emerging patterns, use cases, and potential applications of integrated analytics for production forecasting, market stabilization, and disaster preparedness. This mixed approach ensures both qualitative insights into organizational processes and quantitative evaluation of data integration and analytical outcomes (Kharuddin & Ismail, 2017).

The research procedure follows a sequential phased approach aligned with the agile delivery methodology employed in MyAgrilytics. Initially, an assessment of existing datasets across crop, livestock, fisheries, market, R&D, and enforcement domains was conducted to map fragmentation, duplication, and inconsistencies. This was followed by the design and validation of a reference data pipeline incorporating collection, extraction, staging, cleaning/transformation, validation, and analytics/visualization stages (Inmon, 2005; Wolfert et al., 2017). A pilot integration window was implemented between June and September 2025, during which cross-agency collaboration tested the feasibility of consolidating heterogeneous data into the centralized data lake (Kharuddin et al., 2018).

Sampling and data collection were guided by purposive sampling, targeting datasets most relevant to Ministry decision-making processes. These included historical crop yield records, livestock census data, fisheries

production reports, market price indices, and enforcement and R&D reports. Purposive sampling ensures that the selected data are representative of high-impact decision domains while reflecting diversity across structured, semi-structured, and unstructured formats. Data were collected through both automated extraction from digital repositories and manual entry from legacy systems, ensuring completeness and traceability. Metadata standards were applied to document sources, collection dates, and data owners to facilitate future validation and governance.

For data analysis, the study employed a combination of descriptive statistics, data quality assessment metrics, and predictive modeling. Descriptive statistics including mean, median, standard deviation, and trend analysis were used to evaluate data completeness, consistency, and redundancy. Data quality was measured using the Data Quality Index (DQI), calculated as:

$$DQI = \frac{V}{T} \times 100$$

where V represents the number of validated, error-free records, and T represents the total records ingested (Batini et al., 2016). Higher DQI scores indicate improved reliability of the integrated data repository. Predictive modeling, including time-series forecasting and anomaly detection using machine learning algorithms, was applied to assess the analytical readiness of the platform and its capacity to generate actionable insights. Visual dashboards were also developed to evaluate usability and decision-making efficiency, capturing metrics such as query response time and the number of self-service analytical operations performed by officers.

The methodology emphasizes iterative evaluation and continuous feedback. Each integration sprint was assessed for data quality, processing efficiency, and user experience, aligning with best practices in agile data platform implementation (Kamilaris et al., 2017). This approach ensures that the MyAgrilytics platform is both technically robust and operationally relevant to the Ministry's mandate of ensuring food security and sector competitiveness.

Data Analysis, Results, and Findings

Data analysis in this study was conducted to assess the effectiveness of the MyAgrilytics platform in integrating diverse agrofood datasets, improving data quality, and supporting analytics-driven decision-making. The analysis focused on three primary dimensions: data quality improvement, time-to-insight for officers, and analytical capability for predictive use cases. Data from crop, livestock, fisheries, market, and R&D domains were analyzed following the platform's established data pipeline, including extraction, cleaning, validation, and analytics stages. Descriptive statistics, quality indices, and exploratory analytics were employed to evaluate platform performance and derive insights (Batini et al., 2016; Kamilaris et al., 2017).

Data Quality Assessment

The first step involved evaluating the completeness, consistency, and accuracy of datasets integrated into MyAgrilytics. Data quality was measured using the Data Quality Index (DQI), calculated as the percentage of validated, error-free records relative to the total records ingested. Table 1 summarizes the quality assessment results across major domains.

Table 1: Data Quality Index (DQI) by Agrofood Domain

Domain	Records Ingested	Valid Records	DQI (%)
Crop	15,000	14,100	94.0
Livestock	8,500	7,960	93.8
Fisheries	6,200	5,880	94.8
Market	12,400	11,780	95.0

Domain	Records Ingested	Valid Records	DQI (%)
R&D	4,700	4,400	93.6

The results demonstrate that the MyAgrilytics platform significantly improves data quality across all domains. The DQI values exceed 93% for all datasets, indicating high reliability for analytics applications. Crop and market data exhibited slightly lower validation errors than other domains, reflecting the standardized reporting procedures and automated ingestion mechanisms employed for these datasets. The integration process reduced duplication and inconsistencies, which were previously noted as significant barriers to timely decision-making.

Time-to-Insight Analysis

A critical measure of the platform's effectiveness is its ability to shorten the time required for officers and analysts to generate actionable insights. Prior to MyAgrilytics implementation, generating a typical report required multiple requests across agencies and manual aggregation of data, often taking 5–7 days. Post-integration, the platform enables self-service analytics through dashboards and query interfaces. Table 2 presents the comparative analysis of average time-to-insight before and after platform deployment.

Table 2: Average Time-to-Insight (Hours) Across Departments

Department	Before MyAgrilytics	After MyAgrilytics	Reduction (%)
Crop Production	120	2	98.3
Livestock	96	1.5	98.4
Fisheries	108	2.2	97.9
Market Monitoring	72	1.0	98.6
Research & R&D	84	2.5	97.0

The results reveal a dramatic reduction in time-to-insight, averaging 98% across all departments. This demonstrates that officers are now able to access integrated, validated datasets and generate multiple analytical views within minutes rather than days. The platform's user-friendly dashboards further support non-technical officers in conducting queries and visualizing trends without specialized training.

Analytical Capability and Use Cases

The MyAgrilytics platform also enables predictive analytics and scenario modeling, including production forecasting, price anomaly detection, and climate/disaster impact simulation. Table 3 illustrates the accuracy of predictive models for selected use cases, measured by mean absolute percentage error (MAPE).

Table 3: Predictive Analytics Accuracy by Use Case

Use Case	MAPE (%)	Interpretation
Crop Production Forecast	4.2	High predictive reliability
Livestock Population Estimation	5.1	Accurate for planning interventions
Fisheries Yield Forecast	4.8	Reliable for resource allocation
Market Price Anomaly Detection	3.9	Early identification of price spikes

Use Case	MAPE (%) Interpretation
Climate/Disaster Impact Modeling 5.5	Suitable for proactive mitigation planning

Predictive models exhibit high accuracy across all domains, with MAPE values below 6%, indicating strong reliability for decision-making. Market price anomaly detection achieved the highest accuracy, reflecting the platform's ability to integrate timely, structured market data. Crop and fisheries forecasting demonstrated consistent performance, supporting proactive interventions in production planning and resource allocation. Climate and disaster impact simulations provide actionable foresight, enabling the Ministry to anticipate disruptions and implement mitigation measures before crises escalate.

The findings collectively indicate that MyAgrilytics effectively addresses prior challenges of fragmented, inconsistent, and duplicated data. The high DQI values demonstrate that data standardization and validation processes successfully improved dataset integrity. The dramatic reduction in time-to-insight shows that the platform empowers officers to access actionable insights rapidly, enhancing operational agility. Predictive analytics capabilities reinforce the platform's strategic value, enabling scenario-based planning and early interventions in production, markets, and disaster response. The integration of structured, semi-structured, and unstructured datasets supports cross-domain analysis, highlighting opportunities for coordinated interventions across crop, livestock, fisheries, and market sectors. Overall, the results confirm that MyAgrilytics provides a robust, analytics-ready environment that enhances evidence-based decision-making, improves efficiency, and strengthens national food security planning.

Discussion and Recommendations

The findings of this study demonstrate that the MyAgrilytics platform successfully addresses the primary objectives of creating a centralized agrofood data lake, improving data quality, enabling predictive analytics, and reducing the time-to-insight for officers and decision-makers. First, the establishment of a centralized data repository directly responds to the challenge of fragmented datasets across crop, livestock, fisheries, market, and R&D domains. The high Data Quality Index (DQI) values observed across all domains (ranging from 93.6% to 95%) confirm that standardized parsing, validation, and master-data management effectively reduce duplication and inconsistencies, thereby enhancing the reliability of datasets for analytics (Batini et al., 2016). This centralization aligns with previous findings emphasizing that integrated agricultural data platforms enhance cross-sectoral coordination, reduce redundancies, and improve overall operational efficiency (Wolfert et al., 2017; Kamilaris et al., 2017).

Second, the platform substantially shortens the time-to-insight for officers, with reductions averaging 98% across departments. These improvements demonstrate that MyAgrilytics allows non-technical officers to self-serve insights through user-friendly dashboards, directly fulfilling the objective of building internal analytical capacity. This is consistent with the theoretical underpinning of the DIKW hierarchy, whereby reliable, structured data is transformed into actionable knowledge to support timely, evidence-based decision-making (Rowley, 2007). The rapid access to data-driven insights supports proactive interventions in production planning, market regulation, and disaster response, which are critical for ensuring continuous food supply and resilience against shocks.

Third, predictive analytics capabilities including crop production forecasting, market price anomaly detection, and climate/disaster simulations highlight the platform's potential to anticipate risks and support strategic planning. MAPE values below 6% across all use cases indicate that AI-driven models are reliable for operational decision-making, confirming the feasibility of integrating machine learning into national agrofood governance (Kamilaris et al., 2017; Li et al., 2020). These analytics capabilities directly contribute to the achievement of the United Nations Sustainable Development Goals (SDGs), particularly SDG 2 (Zero Hunger), SDG 12 (Responsible Consumption and Production), and SDG 13 (Climate Action). By enabling foresight into production and market fluctuations, MyAgrilytics supports policies that enhance food availability, reduce waste, and improve resilience against climate-related risks.

The findings also demonstrate that cross-agency collaboration and adherence to standardized governance rules create a shared data environment that benefits all participating divisions. This supports socio-technical systems theory, which emphasizes that technology and organizational processes must be integrated to maximize performance (Baxter & Sommerville, 2011). The alignment of technological infrastructure with institutional processes ensures that data-driven insights translate into actionable policies, bridging the gap between raw data and national food security objectives.

In conclusion, MyAgrilytics effectively achieves its research objectives by centralizing agrofood data, improving data quality, enabling AI-driven analytics, and significantly reducing time-to-insight. The platform provides a robust foundation for evidence-based policymaking, enhances operational efficiency, and strengthens national food security. These findings are supported by prior literature on big data in agriculture, which emphasizes the benefits of integrated datasets and predictive analytics for resilient agrofood systems (Wolfert et al., 2017; Kamilaris et al., 2017).

Based on these findings, it is recommended that the Ministry of Agriculture and Food Security expand the MyAgrilytics platform to include additional real-time data sources such as IoT-enabled farm sensors, satellite imagery, and socio-economic indicators. Continuous training programs should be implemented to further enhance officers' analytical skills, ensuring maximum utilization of AI-driven insights. Furthermore, a formalized cross-agency governance framework should be maintained to continuously update data standards, enforce validation rules, and monitor predictive model accuracy. These measures will ensure that MyAgrilytics remains a sustainable, reliable, and strategically relevant tool for achieving both sector competitiveness and national food security..

References

1. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2016). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3), 1–52. <https://doi.org/10.1145/1132960.1132963>
2. Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17. <https://doi.org/10.1016/j.intcom.2010.07.003>
3. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
4. FAO. (2022). *Big Data in Food and Agriculture: Opportunities and Challenges*. Food and Agriculture Organization of the United Nations.
5. Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). John Wiley & Sons.
6. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143, 23–37. <https://doi.org/10.1016/j.compag.2017.09.037>
7. Kharuddin, A. F., & Ismail, N. A. (2017). Graphing Calculator Exposure of Mathematics Learning in a Partially Technology Incorporated Environment. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(6), 2529-2537. <https://doi.org/10.12973/eurasia.2017.01238a>
8. Kharuddin, A. F., Ismail, K., Ghazali@Suhaimi, S. S., Abu Bakar, N., & Mokhtar, K. I. (2018). Latent Class Analysis of Dental Treatment: Observed in IIUM Dental Specialist Clinic, Kuantan, Pahang. *IIUM Medical Journal Malaysia*, 17(2). <https://doi.org/10.31436/imjm.v17i2.976>
9. Kshetri, N. (2018). The emerging role of big data in key development issues: Opportunities, challenges, and concerns. *Big Data for Development*, 1–23.
10. Li, Y., Li, X., & Li, B. (2020). Integrating multi-source data for sustainable livestock management. *Sustainability*, 12(5), 1987. <https://doi.org/10.3390/su12051987>
11. Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>
12. Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Mining and Knowledge Discovery*, 1(1), 1–12.
13. 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.2017.01.023>

..