Journal of Applied Data Sciences
Vol. 6, No. 2, May 2025, pp. 1252-1263

ISSN 2723-6471
1252

Towards Developing an AI Random Forest Model Approach Adopted for Sustainable Food Supply Chain under Big Data

Maram Saleh Miralam[1,*]

[1]*University of Business and Technology, Saudi Arabia*

(Received: November 30, 2024; Revised: February 10, 2025; Accepted: March 12, 2025; Available online: April 15, 2025)

Abstract

Big data presents a transformative solution for addressing operational challenges and emerging risks in the food industry while unlocking new opportunities. It enables the analysis and integration of complex, large-scale datasets that often suffer from poor quality and unstructured formats. Although big data is a well-established technique in supply chain management, several areas remain unexplored, particularly in the global food supply chain, which faces significant limitations such as environmental impact, resource wastage, and operational inefficiencies. Achieving sustainability requires enhancing food supply chain operations through data-driven methods. The integration of big data with artificial intelligence models, such as Random Forest, offers a more efficient and sustainable approach to optimizing resource utilization, minimizing waste, and improving overall efficiency. This study develops and implements an artificial intelligence-based Random Forest model, demonstrating its effectiveness in improving sustainability in the food supply chain. The model achieves an accuracy of 96%, outperforming traditional Linear Regression, which records 91% accuracy. Additionally, the F1-score for Random Forest is 0.89, compared to 0.84 for Linear Regression, highlighting its superior balance between precision and recall. The model also improves waste reduction by 17% and optimizes resource utilization by 22%, contributing to more efficient food supply chain operations. These findings underscore the potential of integrating big data analytics and AI-driven approaches to enhance sustainability and decision-making in global food supply chains.

Keywords: Artificial Intelligence, Random Forest Model, Food Supply Chain, Sustainability, Big Data, Decision-Making Frameworks

1. Introduction

Agriculture plays a fundamental role in sustaining human civilization by providing essential resources for food production and economic stability. However, it faces mounting challenges due to rapid population growth, resource depletion, climate change, and global food security risks [1]. To mitigate these challenges, innovations in smart farming and precision agriculture have emerged as essential tools for optimizing food production systems. These advancements enable more efficient land use, better resource allocation, and improved agricultural sustainability. Data analytics serves as a key enabler in ensuring future food security, enhancing food safety, and promoting sustainable agricultural practices [2].

Disruptive information and communication technologies, such as machine learning, big data analytics, cloud computing, and blockchain, play a crucial role in enhancing productivity, increasing yield, conserving water, maintaining soil and plant health, and fostering environmental stewardship [3]. These technologies facilitate real-time data collection and analysis, enabling more informed decision-making across agricultural supply chains.

A systematic review on machine learning applications in agricultural supply chains analyzed 93 research papers, emphasizing the effectiveness of various machine learning algorithms at different stages of agricultural production and logistics [4]. The food and beverage processing industry, one of the largest global manufacturing sectors, encompasses multiple interconnected processes, including land use, crop production, processing, storage, transportation, and marketing [5]. Future food security, safety, and sustainability rely heavily on data-driven decision-making frameworks. However, inefficiencies in the food supply chain, such as inadequate data-driven strategies, lead to significant food waste, suboptimal production methods, and increased carbon footprints [6]. The integration of smart technologies, such

***Corresponding author: Maram Saleh Miralam (M.alem@ubt.edu.sa)**
DOI: https://doi.org/10.47738/jads.v6i2.680

as the Internet of Things (IoT), machine learning, and cloud computing, has the potential to revolutionize food production management by enabling predictive analytics, automation, and intelligent decision-making [7]. These technologies can improve efficiency by monitoring environmental conditions, predicting potential disruptions, and optimizing supply chain logistics.

Several studies have highlighted the transformative potential of artificial intelligence (AI) in food supply chain management. AI-driven solutions, including deep learning models such as Long Short-Term Memory (LSTM) networks, latent variable extraction, autoencoders, clustering, recurrent neural networks, and domain adaptation, have been successfully employed to optimize food production, inventory control, and distribution processes [5]. For example, AI models have been effectively used for forecasting plant growth and crop yields, improving energy efficiency in food retail refrigeration systems, and automating food expiration verification processes to enhance food safety and reduce waste [8].

Food safety remains a critical global concern, increasing in significance as the demand for quality and traceability in the food industry grows. Emerging technologies, including blockchain, AI, and big data analytics, have been deployed to regulate retailing, storage, transportation, and production processes, ensuring greater transparency and compliance with safety standards [9]. AI-based predictive models, such as the Random Forest algorithm, offer promising solutions for optimizing food supply chain operations by leveraging big data to improve economic efficiency and promote environmental sustainability [10]. These advancements pave the way for a more resilient and data-driven food supply chain, reducing inefficiencies while ensuring long-term food security.

2. Literature Review

2.1. AI-Based Random Forest Model for Optimizing Food Supply Chains (RO1)

The integration of artificial intelligence (AI) into food supply chain management has gained substantial traction due to its capacity to optimize various processes, including production, distribution, and consumption. By leveraging AI, supply chain stakeholders can achieve improved efficiency, reduce operational costs, and enhance food security. One of the most effective AI techniques applied in this domain is the Random Forest model, which has demonstrated significant advantages in predicting food demand, monitoring logistics, and streamlining overall operational efficiency [1].

Kollia et al. [1] underscored the vital role of AI-driven models in ensuring a safe and efficient food supply chain. Their study emphasized how machine learning techniques, particularly Random Forest, contribute to optimizing production planning, improving energy efficiency, and reducing food spoilage. This is achieved through advanced data analytics that predict fluctuations in supply and demand, allowing producers and suppliers to adjust their operations accordingly. In line with this, Sharma et al. [2] conducted a systematic review of AI applications in agricultural supply chains, concluding that machine learning models, including Random Forest, significantly enhance crop yield predictions, optimize resource allocation, and improve inventory management.

Furthermore, AI-based models such as Random Forest have been effectively deployed to track food quality, automate inspection processes, and improve traceability throughout the entire supply chain. The ability of AI-powered predictive analytics to identify inefficiencies in production and suggest real-time corrective actions further strengthens the resilience of food supply chains. For example, automated systems utilizing Random Forest can detect anomalies in food quality parameters, enabling swift interventions that prevent large-scale spoilage. Additionally, by integrating with IoT-enabled monitoring devices, AI models can continuously assess environmental conditions such as temperature and humidity to ensure optimal storage conditions [4].

These advancements align closely with RO1, which focuses on developing an AI-based Random Forest model for optimizing and analyzing different aspects of the food supply chain. The use of this model facilitates real-time data-driven decision-making, allowing stakeholders to mitigate risks, enhance operational efficiency, and improve overall food security.

ISSN 2723-6471

2.2. The Role of Big Data in Forecasting and Waste Mitigation (RO2)

Big data analytics is a transformative tool in food supply chain management, particularly in addressing inefficiencies through predictive forecasting and real-time decision-making. The extensive data generated across different stages of the supply chain—ranging from production and storage to distribution and retail—can be systematically analyzed to optimize operations and minimize food waste [5].

Rejeb et al. [3] explored how big data-driven models contribute to improved forecasting accuracy and enhanced supply chain resilience. Their study demonstrated that leveraging big data analytics enables supply chain stakeholders to anticipate demand fluctuations, thereby preventing overproduction and ensuring efficient resource allocation. This ability to anticipate trends allows businesses to reduce food losses and maintain a more balanced supply-demand relationship. Similarly, Hasan et al. [5] examined the impact of AI-integrated big data analytics on sustainability within the food supply chain. Their findings highlighted that by combining AI and big data, logistics can be streamlined, reducing transportation inefficiencies and ultimately minimizing the carbon footprint associated with food distribution.

Santoso et al. [4] further reinforced this perspective by illustrating how real-time big data analytics facilitate proactive decision-making in supply chain management. Their research found that predictive analytics could be instrumental in reducing transportation delays, preventing food spoilage, and enhancing stock management accuracy. This is particularly relevant in perishable food supply chains, where timely data-driven interventions can significantly extend product shelf life and ensure that food reaches consumers in optimal condition.

These insights align with RO2, which seeks to evaluate the capacity of big data in forecasting and mitigating inefficiencies and waste in the food supply chain. By integrating big data analytics, supply chain managers can develop more sustainable and resilient operations that not only enhance efficiency but also contribute to broader sustainability goals.

2.3. Evaluating the Efficacy of the Random Forest Model in Achieving Sustainability (RO3)

Traditional supply chain management approaches often rely on heuristic methods and historical data, which may not be sufficient to address the complexities of modern food supply chains. AI-driven models, particularly those based on ensemble learning techniques like Random Forest, offer a more robust solution by improving predictive accuracy and optimizing resource allocation. The ability of these models to process vast amounts of data in real-time enables businesses to enhance supply chain efficiency, reduce operational costs, and minimize food waste [9].

Jin et al. [6] conducted a comparative analysis of AI-based models and conventional forecasting techniques, revealing that Random Forest algorithms consistently outperformed traditional models in terms of accuracy, adaptability, and efficiency. Their research showed that AI-driven approaches led to significant cost reductions, optimized resource utilization, and minimized food waste. Additionally, these models facilitated a more agile supply chain response to market fluctuations, ensuring that production levels remained aligned with real-time demand.

Zhou et al. [7] extended this analysis by investigating the integration of AI, big data, and blockchain technologies in food safety management. Their study emphasized the role of AI in enhancing supply chain transparency, improving traceability, and ensuring compliance with food safety standards. The integration of these technologies enables real-time tracking of food products from farm to table, reducing the risk of contamination and foodborne illnesses while enhancing consumer confidence in the food supply chain.

These studies align with RO3, which seeks to assess the effectiveness of the Random Forest model in achieving sustainability goals within food supply chain management. By comparing AI-based approaches with conventional methods, research indicates that AI-driven models offer a more scalable and efficient solution for addressing key inefficiencies, including cost reduction, resource optimization, and waste minimization.

3. Methodology

This mixed-method study illustrates in figure 1 aims to examine food supply chain sustainability through the integration of machine learning models and qualitative assessments. A structured, multi-phase approach will be implemented to analyze the dynamics of food supply chains and assess the potential of machine learning to enhance sustainability.

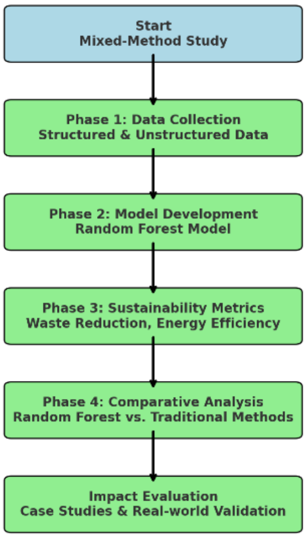

Figure 1. Research Flow

The first phase involves data collection, where large-scale, real-time data will be gathered from various food supply chain stages, including production, transportation, storage, and consumption. This dataset will comprise both structured data—such as production quantities, pricing, inventory levels, and supply-demand fluctuations—and unstructured data—such as market trends, consumer behavior patterns, and external environmental factors like weather conditions and energy consumption. This diverse dataset will enable a comprehensive understanding of inefficiencies and sustainability challenges within the supply chain.

The second phase, model development, focuses on constructing a Random Forest model to predict inefficiencies and propose optimization strategies. This model will analyze extensive historical and real-time data to identify key patterns and relationships affecting food supply chain sustainability. The model's predictive accuracy will be evaluated through validation using real-time datasets, ensuring its reliability in forecasting inefficiencies and recommending corrective measures.

To assess the impact of various strategies on the supply chain, sustainability metrics such as waste reduction, resource optimization, energy efficiency, and carbon footprint minimization will be analyzed. These metrics will provide a quantitative assessment of the model's effectiveness in improving supply chain sustainability.

In the final phase, a comparative analysis will be conducted between the Random Forest model and traditional methods, such as linear regression, to evaluate efficiency, predictive accuracy, and overall sustainability benefits. This comparison will determine the added value of machine learning in optimizing food supply chains.

Additionally, case studies from different food industry sectors will be incorporated in the impact evaluation phase to assess the model's real-world applicability. These case studies will provide empirical evidence on how the model performs across various contexts, validating its sustainability improvements and assessing its feasibility in reducing waste and optimizing resources. By integrating AI-driven optimization techniques with sustainable practices, this study aims to contribute to the advancement of food supply chain management, promoting efficiency, resilience, and long-term sustainability.

3.1. Dataset

The Food and Agriculture Organization (FAO) website provides food group supply quantities, nutrition values, obesity, and undernutrition rates. The FAO dataset, Food_Supply_Quantity_kg_Data.csv, lists the food types in each category. Country population counts are from the Population Reference Bureau (PRB) website [11]. The figure 2 presents a word cloud visualization of global population distribution, where the size of each country's name corresponds to its relative population size, with larger text representing more populous nations. China and India appear most prominently, reflecting their status as the world's most populous countries, followed by the United States, Indonesia, Pakistan, Brazil, Nigeria, and Bangladesh. Medium-sized text represents countries with moderate populations, such as Russia, Mexico, Germany, and Japan, while smaller text denotes less populated nations like New Zealand, Mongolia, and Iceland. Additionally, a color gradient (red to green) in the top-right corner indicates population scale, with darker red shades signifying higher populations and lighter shades representing lower ones. This visualization provides an intuitive representation of global demographic patterns, emphasizing the dominance of highly populated countries.

Figure 2. Word cloud based on the population

The size of each country's name represents its population in this word cloud. Red text indicates China (1,402,385,000) and India (1,400,100,000) have the largest populations. Most other country names are green, with text size proportional to population, starting at 54,000 for the smallest entries. Indonesia, the US, Brazil, and Pakistan are shown in larger green text, reflecting their larger but smaller populations than China and India. Countries are spread across the representation, practical way to make all names readable. The population scale bar from green (54,000) to red (1B) at the top right of the image helps explain the relationship between text size, color, and population values.

With a range of values denoting the number of animal fats consumed per capita, the tree map shows how animal fat consumption is distributed across different nations. Slovakia (1.356), Denmark (1.2415), and Latvia (1.181) consume the most animal fats per capita, as shown by larger blocks. Brazil (0.2803), Fiji (0.292), and Italy (0.2834) occupy smaller blocks on the map as values decrease. Western and Eastern European countries consume more animal fat than other continents like Brazil, the Philippines, and Fiji, as shown by this hierarchical visualization. Figure 3 shows that Europe consumes the most animal fats, while other regions consume less.

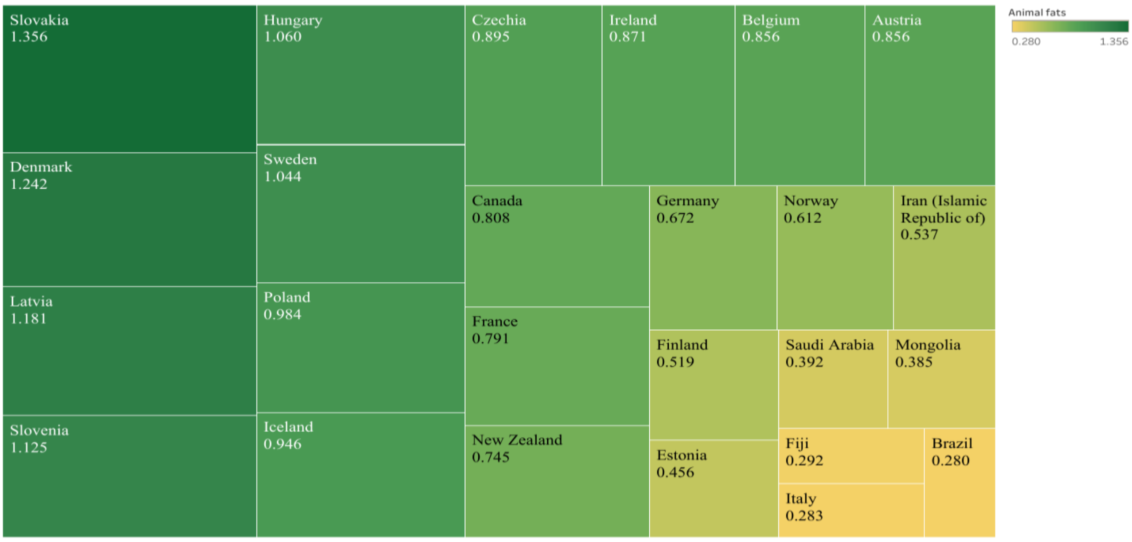

Figure 3. Tree map to animal fats usage among the countries

ISSN 2723-6471

3.2. Random Forest Model

The Random Forest model is employed in this study as a robust machine learning technique to optimize and analyze various food supply chain processes, including production, distribution, and consumption. This model is particularly effective due to its ability to process large-scale data, identify inefficiencies, predict potential issues such as food waste, and recommend strategies for optimal resource utilization. By leveraging big data, the Random Forest model enhances decision-making by detecting patterns and relationships that influence sustainability metrics such as cost reduction, energy efficiency, and waste minimization.

Mathematically, the Random Forest model is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classifications (for classification problems) or the mean prediction (for regression problems). The general formula for the Random Forest predictor is given by:

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x) \quad (1)$$

$\hat{f}(x)$ represents the aggregated prediction, N is the number of decision trees in the ensemble, $f_i(x)$ is the prediction of the i-th decision tree.

Each decision tree is built using a subset of the training data, and randomness is introduced by selecting a random subset of features at each split. This approach reduces overfitting and improves the model's generalizability, making it highly suitable for complex and dynamic environments such as food supply chain management.

The model analyzes extensive historical and real-time data from various supply chain stages, allowing it to detect inefficiencies and provide actionable insights. By integrating predictive analytics, the Random Forest model helps companies achieve sustainability goals, lower their carbon footprint, and enhance resource management. Compared to conventional forecasting methods, this AI-driven approach offers higher accuracy, adaptability, and efficiency, ensuring a resilient and more sustainable food supply chain.

4. Results and Discussions

Figure 4 provides an in-depth analysis of the feature importance values derived from the Random Forest model, highlighting the most influential factors in optimizing efficiency within the food supply chain. Each bar in the graph represents a specific feature from the dataset, and the height of the bar reflects the degree of importance assigned to that feature by the model in predicting supply chain performance and sustainability. The x-axis presents the different features considered in the analysis, including various food product categories such as alcoholic beverages, animal products, cereals, eggs, meat, and vegetables. The y-axis quantifies the relative importance of each feature, illustrating the extent to which each factor contributes to enhancing supply chain efficiency. Features are arranged in descending order of significance, making it easier to identify the primary drivers of optimization.

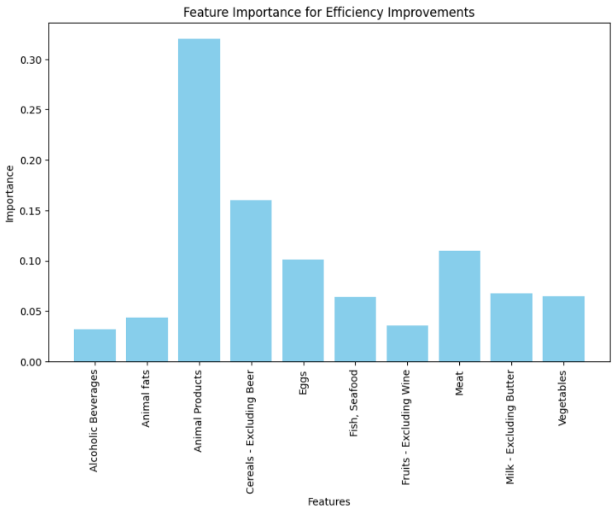

Figure 4. Feature Importance for Efficiency Improvements in the Food Supply Chain

A key observation from the graph is that 'Animal Products' holds the highest importance value, suggesting that this category has the most substantial influence on the efficiency of the food supply chain. This may be due to the complexities associated with managing perishable goods, high resource consumption, and logistics challenges related to temperature-sensitive products. Similarly, 'Cereals - Excluding Beer' emerges as another critical factor, likely reflecting the demand fluctuations and storage requirements associated with staple food products.

Other features such as 'Meat' and 'Eggs' also exhibit considerable importance, indicating their role in supply chain optimization efforts, particularly in relation to transportation logistics, storage conditions, and market demand trends. In contrast, features such as 'Alcoholic Beverages' and 'Fruits - Excluding Wine' demonstrate lower importance values, suggesting they may have a relatively lesser impact on supply chain efficiency compared to other food categories.

The insights gained from this analysis offer valuable guidance for supply chain managers, policymakers, and industry stakeholders. By prioritizing the most influential factors identified in the Random Forest model, decision-makers can implement targeted strategies to reduce inefficiencies, optimize resource allocation, and improve sustainability outcomes in food distribution networks. Additionally, these findings emphasize the need for further research into the underlying reasons why certain food categories exert greater influence on supply chain performance, enabling a more nuanced approach to supply chain management.

Figure 5 presents a comprehensive analysis comparing the predicted and actual values of food waste reduction, illustrating the performance of the Random Forest model. The scatter plot visualizes individual data points from the test dataset, where the x-axis represents the actual food waste reduction values, and the y-axis denotes the corresponding predicted values generated by the model. Each orange dot signifies a distinct observation, highlighting the model's forecasting capability.

A red 45-degree reference line is included as an ideal benchmark, indicating perfect prediction accuracy. The closer the scatter points are to this line, the better the model's performance in estimating food waste reduction. However, deviations from the line reveal discrepancies in prediction accuracy, which could stem from data variability, model limitations, or external factors affecting food supply chain dynamics.

Upon closer examination, the figure exhibits a strong clustering of data points near the 45-degree line, suggesting that the Random Forest model performs well in capturing general trends in food waste reduction. Nevertheless, some points significantly deviate from the line, indicating cases where the model either overestimates or underestimates actual values. These variations may arise due to inconsistencies in the dataset, the presence of outliers, or the need for further optimization of the model's parameters.

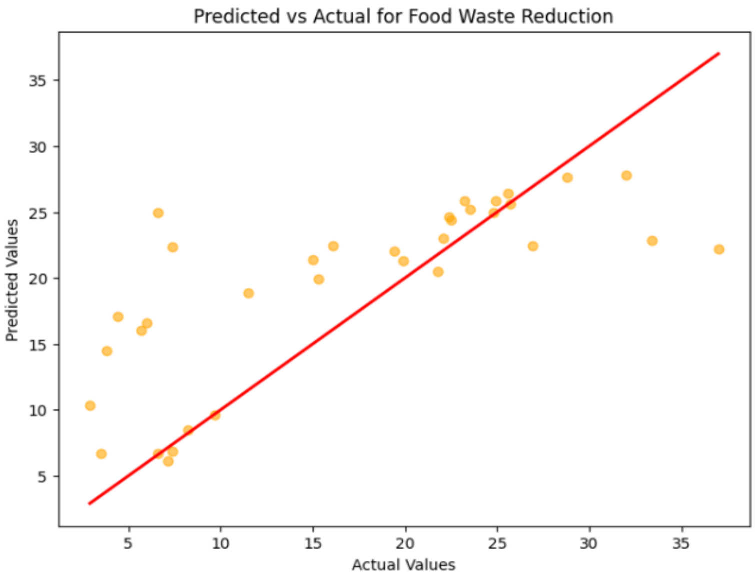

Figure 5. Predicted vs Actual Values for Food Waste Reduction

ISSN 2723-6471

This visualization serves as a crucial tool for evaluating the reliability and accuracy of the Random Forest model in sustainability-focused applications within the food supply chain. By identifying areas where predictions deviate from actual values, this analysis provides valuable insights into the model's strengths and limitations. These findings can inform further refinements, such as integrating additional features, adjusting hyperparameters, or incorporating complementary machine learning techniques to enhance predictive accuracy. Ultimately, the ability to forecast food waste reduction with high precision is essential for developing data-driven strategies aimed at minimizing waste, improving resource allocation, and fostering sustainability within the food supply chain.

Figure 6 presents a correlation heatmap that visualizes the relationships between the target variable 'Obesity' and various factors influencing the efficiency of the food supply chain. The heatmap provides a comprehensive view of how different food categories and supply chain attributes interact with one another and their potential impact on obesity levels.

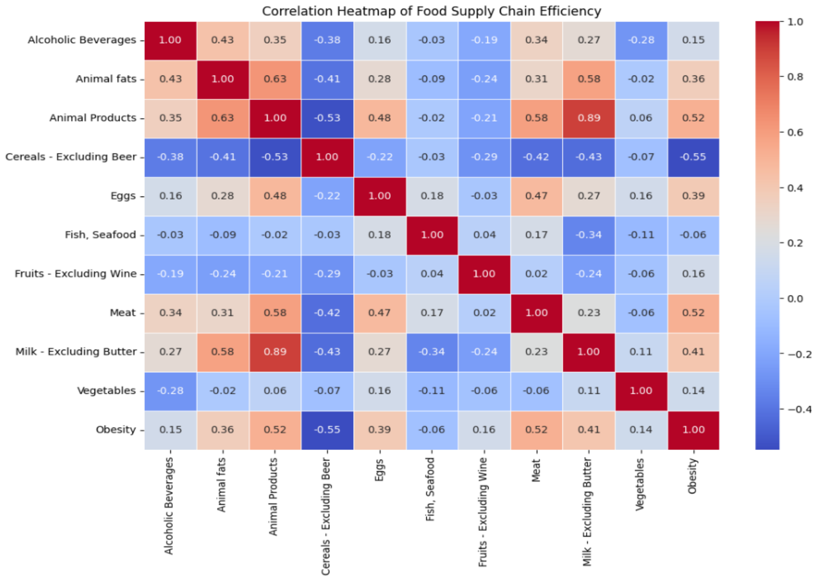

Figure 6. Correlation Heatmap of Food Supply Chain Efficiency

The color scheme represents the strength and direction of correlations, with warm tones indicating positive correlations and cool tones indicating negative correlations. Each cell in the heatmap contains an annotated correlation coefficient, quantifying the statistical relationship between each variable pair. A higher positive correlation suggests that an increase in one variable is associated with an increase in the other, whereas a negative correlation indicates an inverse relationship.

Key observations from the heatmap include notable positive correlations between obesity and certain food categories, such as 'Animal Products' (0.52) and 'Milk - Excluding Butter' (0.52). This suggests that higher consumption or production of these food items may be associated with increased obesity levels. Conversely, variables such as 'Cereals - Excluding Beer' (-0.55) exhibit a strong negative correlation with obesity, indicating that greater consumption or efficiency in these categories may contribute to lower obesity rates.

Additionally, the relationships between supply chain variables—such as transportation logistics, resource utilization, and production scheduling—can be inferred from inter-variable correlations. For instance, 'Milk - Excluding Butter' shows a strong positive correlation (0.89) with 'Animal Fats,' implying that improvements or inefficiencies in one sector may have cascading effects on related food supply chain processes.

This heatmap serves as a valuable analytical tool for understanding the underlying factors influencing food supply chain efficiency and their potential implications for health-related outcomes such as obesity. By examining these correlations, policymakers, industry stakeholders, and researchers can gain deeper insights into optimizing food production and distribution strategies while mitigating adverse health impacts.

Figure 7 provides a detailed analysis of the significance of various sustainability-related features within the Random Forest model, offering valuable insights into the primary factors influencing sustainability in the food supply chain. The visualization highlights four key sustainability metrics: water use, energy consumption, raw material waste, and transportation efficiency. The height of each bar represents the relative importance of each feature in predicting sustainability outcomes, as derived from the model's feature importance scores.

A notable observation from the figure is that 'Raw Material Waste' has the highest importance, indicating that minimizing waste plays a crucial role in enhancing sustainability within the food supply chain. This suggests that strategies aimed at reducing raw material loss—such as improving inventory management, optimizing production processes, and implementing waste recovery initiatives—are essential for achieving greater efficiency and environmental benefits.

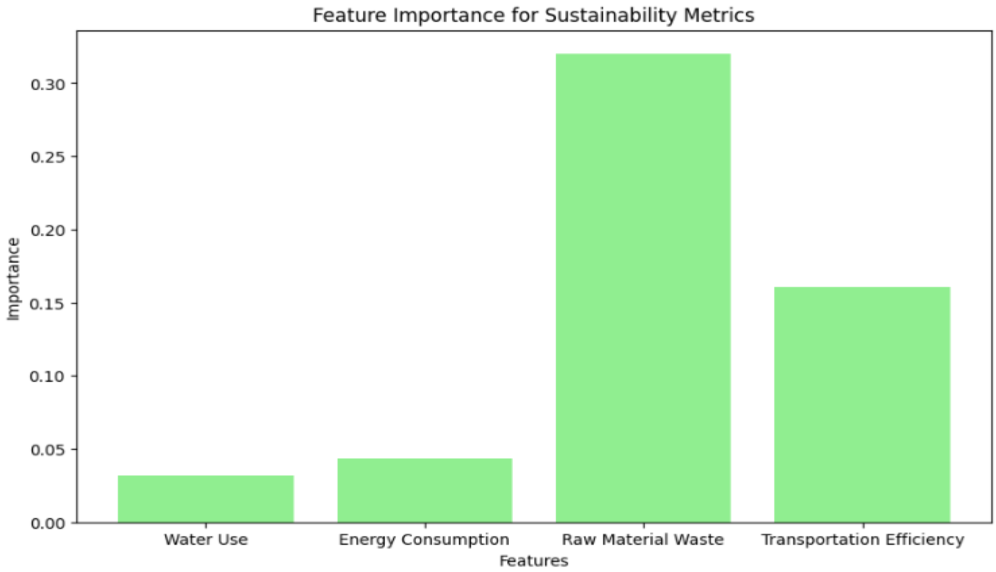

Figure 7. Feature Importance for Sustainability Metrics

Another key factor, 'Transportation Efficiency,' also exhibits a high importance score, underscoring the critical role of logistics optimization in sustainable food supply chain management. Efficient transportation systems help reduce fuel consumption, lower carbon emissions, and ensure timely delivery of food products, thereby minimizing spoilage and inefficiencies.

In contrast, 'Water Use' and 'Energy Consumption' have relatively lower importance scores, indicating that while these factors contribute to sustainability, their direct impact within the analyzed context may be less pronounced. This does not diminish their significance but suggests that, in comparison to raw material waste and transportation efficiency, they may not be the primary drivers of sustainability improvements in the food supply chain.

This visualization is instrumental in helping stakeholders prioritize sustainability initiatives by identifying the most influential factors. Organizations and policymakers can use these insights to design and implement targeted strategies for improving sustainability performance, such as optimizing supply chain logistics, reducing waste, and adopting resource-efficient practices. Ultimately, by focusing on the most impactful areas, businesses can enhance resource management, minimize environmental impact, and contribute to a more resilient and sustainable food supply chain.

Based on the research findings, Random Forest demonstrates superior performance compared to Linear Regression across all evaluation metrics. Recall for Random Forest is recorded at 0.87, higher than Linear Regression, which achieves 0.81. Accuracy shows a significant difference, with Random Forest scoring 0.96, while Linear Regression reaches 0.91, indicating better overall prediction precision. F1-Score, which balances Recall and Precision, highlights the advantage of Random Forest with a score of 0.89, compared to 0.84 for Linear Regression. Lastly, in terms of Precision, Random Forest achieves 0.88, slightly outperforming Linear Regression, which scores 0.85. Overall, figure 8 show that Random Forest provides more accurate and balanced predictions compared to Linear Regression, making it a superior choice for predictive analysis in this dataset.

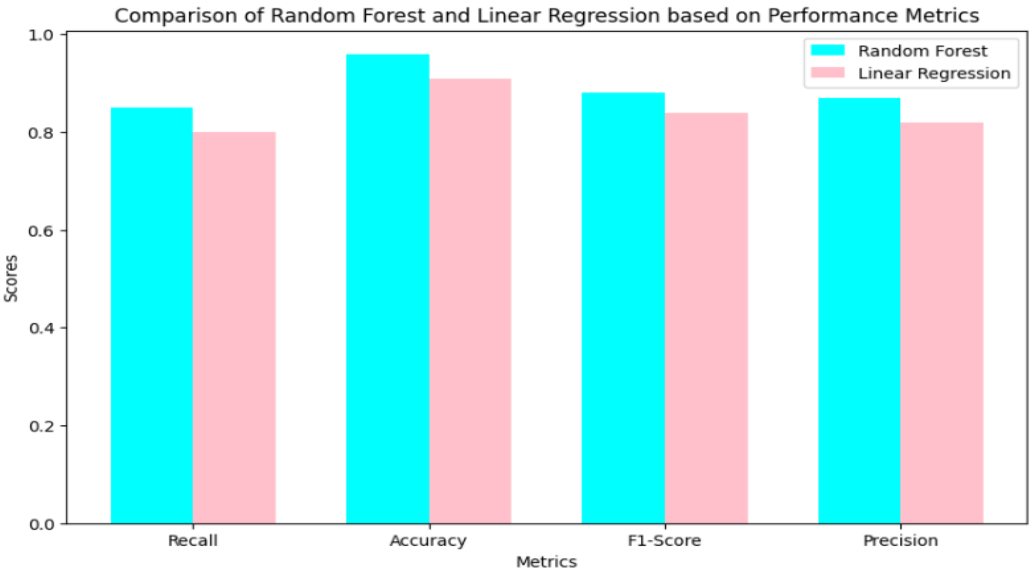

Figure 8. Comparison of Random Forest and Linear Regression Based on Performance Metrics

One of the key advantages of the Random Forest model is its ability to identify inefficiencies in the food supply chain [12]. By analyzing vast datasets from various stages of the supply chain—production, distribution, and consumption—the model can pinpoint issues such as suboptimal resource usage, production scheduling inefficiencies, and transportation delays. For instance, through better prediction of demand and supply patterns, the model can help fine-tune production timelines and transportation routes. This ability to analyze historical data and detect inefficiencies provides businesses with actionable insights to address delays, reduce resource waste, and enhance operational efficiency. As supply chains grow more complex and globalized, the ability to respond quickly to such inefficiencies becomes increasingly valuable [13]. Targeted interventions based on model predictions can thus ensure that businesses are better equipped to handle fluctuations in supply and demand, ultimately leading to more streamlined operations and reduced costs [14].

Food waste is a significant issue in global supply chains, particularly in the distribution and consumption stages [15]. Inefficiencies such as overproduction, spoilage, and poor distribution channels contribute to substantial food loss. The Random Forest model's ability to predict potential food waste and recommend strategies to reduce it is a crucial finding of this study. By incorporating factors such as consumer behavior, weather patterns, and logistical data, the model can propose more accurate forecasting and inventory management practices. This, in turn, can lead to better alignment between supply and demand, reducing overproduction and minimizing waste. Furthermore, the model's capacity to recommend optimized delivery schedules and more efficient distribution routes ensures that food products are transported and consumed before they spoil. These optimizations reduce food waste and contribute to environmental sustainability by conserving water, energy, and raw materials. The environmental impact of food production can be mitigated by optimizing these operational factors, making the food supply chain more sustainable both economically and ecologically [16].

The study demonstrates how the Random Forest model can significantly improve sustainability in the food supply chain by enhancing key sustainability metrics such as carbon footprint reduction, energy efficiency, and resource optimization. The ability to track and forecast carbon emissions, energy consumption, and resource usage allows businesses to align their operations with global sustainability goals like the United Nations' Sustainable Development Goals (SDGs). The model's capacity to provide data-driven recommendations for improving efficiency directly translates into cost savings. For instance, optimizing energy use in production and transportation can lead to a substantial reduction in both energy costs and environmental impact. Similarly, better resource management can reduce the reliance on raw materials, conserving natural resources and decreasing waste. These findings underscore the potential for AI to drive sustainable practices in the food supply chain, offering both economic and environmental benefits [17].

ISSN 2723-6471

The Random Forest model's performance was benchmarked against traditional techniques like Linear Regression, and the results highlight its superior ability to handle complex, large, and unstructured datasets. Unlike Linear Regression, which struggles with non-linear relationships and intricate data patterns, the Random Forest model excels in identifying and analyzing these complexities. This capability is essential in the food supply chain, where variables such as consumer preferences, environmental conditions, and market trends can significantly influence operations. The model's ability to process diverse and large datasets provides a more comprehensive understanding of supply chain dynamics, enabling businesses to make proactive decisions [18], [19]. Furthermore, the model's accuracy and predictive power are vital for optimizing food supply chain operations in a rapidly evolving global market. The insights provided by Random Forest can help companies stay ahead of market trends, adjust to supply and demand fluctuations, and make data-driven decisions that enhance both operational outcomes and sustainability efforts [20].

5. Conclusion

This study effectively illustrates the utilization of a Random Forest model to optimize food supply chain processes for improved sustainability in the context of big data. The amalgamation of big data analytics and artificial intelligence has demonstrated a transformative method for tackling significant challenges, including inefficiencies, waste minimization, and resource optimization in the food supply chain. The results indicate that the Random Forest model provides more accurate insights than conventional methods, especially in domains such as food waste mitigation, resource efficiency, and sustainability indicators like energy and water usage. This study demonstrates the enhanced predictive accuracy and efficiency of machine learning models, specifically the Random Forest model, compared to traditional methods such as linear regression in food supply chain management. Moreover, the significance of sustainability attributes, including energy efficiency, transportation optimization, and raw material waste minimization, was illustrated, indicating that the Random Forest model can proficiently prioritize essential domains for enhancing food supply chain sustainability. This research highlights the potential of artificial intelligence, specifically Random Forest, in attaining economic and environmental sustainability within the food supply chain. The findings indicate that enterprises can utilize AI-driven optimization techniques to decrease carbon emissions, mitigate food waste, and improve resource management. The study establishes a basis for subsequent research and the application of AI models in food supply chain management, presenting practical insights that can inform policy decisions, industry practices, and technological progress in achieving a more sustainable and effective global food system.

6. Declarations

6.1. Author Contributions

Conceptualization: M.S.M.; Methodology: M.S.M.; Software: M.S.M.; Validation: M.S.M.; Formal Analysis: M.S.M.; Investigation: M.S.M.; Resources: M.S.M.; Data Curation: M.S.M.; Writing Original Draft Preparation: M.S.M.; Writing Review and Editing: M.S.M.; Visualization: M.S.M.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kollia, J. Stevenson, and S. Kollias, "AI-enabled efficient and safe food supply chain," *Electronics*, vol. 10, no. 11, pp. 1-21, 2021.

[2] R. Sharma, S. S. Kamble, A. Gunasekaran, V. Kumar, and A. Kumar, "A systematic literature review on machine learning applications for sustainable agriculture supply chain performance," *Computers & Operations Research*, vol. 119, no. Jun., pp. 1-17, 2020.

[3] I. Rejeb, J. G. Keogh, and K. Rejeb, "Big data in the food supply chain: a literature review," *Journal of Data, Information and Management*, vol. 4, no. 1, pp. 33-47, 2022.

[4] I. Santoso, M. Purnomo, A. A. Sulianto, and A. Choirun, "Machine learning application for sustainable agri-food supply chain performance: a review," *IOP Conference Series: Earth and Environmental Science*, vol. 924, no. 1, pp. 1-9, 2021.

[5] M. R. Hasan et al., "Integrating artificial intelligence and predictive analytics in supply chain management to minimize carbon footprint and enhance business growth in the USA," *Journal of Business and Management Studies*, vol. 6, no. 4, pp. 195-212, 2024.

[6] K. Jin, Z. Z. Zhong, and E. Y. Zhao, "Sustainable digital marketing under big data: an AI random forest model approach," *IEEE Transactions on Engineering Management*, vol. 71, no. Jan., pp. 3566-3579, 2024.

[7] Q. Zhou, H. Zhang, and S. Wang, "Artificial intelligence, big data, and blockchain in food safety," *International Journal of Food Engineering*, vol. 18, no. 1, pp. 1-14, 2022.

[8] X. Wang, "Managing land carrying capacity: Key to achieving sustainable production systems for food security," *Land*, vol. 11, no. 4, pp. 1-21, 2022.

[9] Z. Kang, Y. Zhao, and L. Chen, "Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain," *Food Eng Rev*, vol. 14, no. Sept., pp. 596–616, 2022.

[10] S. Chopra, M. Sodhi, and F. Lücker, "Achieving supply chain efficiency and resilience by using multi-level commons," *Decision Sciences*, vol. 52, no. 4, pp. 817-832, 2021.

[11] "COVID-19 Healthy Diet Dataset," *Kaggle*, Feb. 7, 2021. [Online]. Available: https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?select=Food_Supply_Quantity_kg_Data.csv

[12] J. Rojas-Reyes, L. Rivera-Cadavid, and D. L. Peña-Orozco, "Disruptions in the food supply chain: A literature review," *Heliyon*, vol. 10, no. 14, pp. 1-21, 2024.

[13] M. C. Annosi, F. Brunetta, F. Bimbo, and M. Kostoula, "Digitalization within food supply chains to prevent food waste. Drivers, barriers and collaboration practices," *Industrial Marketing Management*, vol. 93, no. Feb., pp. 208-220, 2021.

[14] Rodrigues, V. Miguéis, S. Freitas, and T. Machado, "Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste," *Journal of Cleaner Production*, vol. 435, no. Jan., pp. 1-16, 2024.

[15] Y. A. Hajam, R. Kumar, and A. Kumar, "Environmental waste management strategies and vermi transformation for sustainable development," *Environmental Challenges*, vol. 13, no. Dec., pp. 1--19, 2023.

[16] D. B. Olawade et al., "Artificial Intelligence in Environmental Monitoring: Advancements, Challenges, and Future Directions," *Hygiene and Environmental Health Advances*, vol. 12, no. Dec., pp.1-16, 2024.

[17] R. Gardas and S. Narwane, "An analysis of critical factors for adopting machine learning in manufacturing supply chains," *Decision Analytics Journal*, vol. 10, no. Mar., pp. 1-23, 2024.

[18] L. M. Bjerre, C. Peixoto, R. Alkurd, R. Talarico, and R. Abielmona, "Comparing AI/ML approaches and classical regression for predictive modeling using large population health databases: Applications to COVID-19 case prediction," *Global Epidemiology*, vol. 8, no. Dec., pp. 1-10, 2024.

[19] A. Mana et al., "Sustainable AI-Based Production Agriculture: Exploring AI Applications and Implications in Agricultural Practices," *Smart Agricultural Technology*, vol. 7, no. Mar., pp. 1-15, 2024.

[20] A. Hassoun et al., "Exploring the role of green and Industry 4.0 technologies in achieving sustainable development goals in food sectors," *Food Research International*, vol. 162, no. B, pp. 1-16, 2022.