Machine learning application for sustainable agri-food supply chain performance: a review

I Santoso1, M Purnomo2, A A Sulianto1 and A Choirun1

1Department of Agro-industrial Technology, Faculty of Agricultural Technology, Universitas Brawijaya, Malang Indonesia
2Department of Socio-Economics, Faculty of Agriculture, Universitas Brawijaya, Malang Indonesia

Email: imamsantoso@ub.ac.id

Abstract. The agri-food supply chain consists of activities in “farm-to-fork” order, including agriculture (i.e., land cultivation and crop production), production processes, packaging, warehousing systems, distribution, transportation, and marketing. Data analytics hold the key to ensuring future food security, food safety, and ecological sustainability. While emerging 'smart' technologies such as the internet of things, machine learning, and cloud computing can change production management practices. The current study presents a systematic review of machine learning (ML) applications in the agri-food supply chain. This framework identifies the role of ML algorithms in providing real-time analytical insights to assist proactive data-driven decision-making processes in the agri-food supply chain. It also guides researchers, practitioners, and policymakers on successful management to increase the productivity and sustainability of agri-food.

1. Introduction
The food industry's success in today's global competition is heavily influenced by using technology, information, and communication. The tougher competition in this global market, the higher customer expectations for products and services, and the shorter product life cycles will force companies to prioritize their supply chains to achieve competitive advantages that can support their business continuity [1]. SCM is a complex activity because it includes all activities ranging from scheduling, procurement of raw materials, production processes, and delivery of products to consumers. All SCM activities require a data management system to manage daily administrative activities, processes, and operations, as well as logistics [2]. This is done because data is an important asset for process design and operation control. Advances in technology and information have led to various changes in SCM practices. SCM actors are required to make improvements and process information from the market quickly. Intelligent use of data has great potential for sustainable SCM practices. Manufacturers are expected to adopt new technologies capable of quickly and intelligently interpreting data. Traditional decision support systems cannot be applied to deal with big data accurately. Therefore, in the era of big data, supply chain professionals are trying to implement smart supply chains to handle big data.

Artificial intelligence (AI) methods are used to address these big data-related challenges. Machine learning, a subset of AI, is widely used to identify hidden patterns in the data [3,4]. Machine learning (ML) algorithms can detect data whose data patterns are unknown and direct researchers to achieve the expected goals. The application of the ML Algorithm in the supply chain has attracted the interest of
researchers. So far, researchers have analyzed and interpreted big data using traditional methods. Traditional methods are not capable of analyzing large and unstructured data. In addition, it also cannot identify and predict the most effective factors, especially in the case of supply chain performance. Therefore, researchers began to replace traditional analytical methods by using machine learning techniques. ML has many advantages, such as analysing big data, solving nonlinear problems, and being able to recognize and predict values.

However, to the best of our knowledge, not many studies have been conducted to specially review the status of machine learning applications in Agri-Food Supply Chain (AFSC) management. In this study, we present a systematic literature review on the application of ML in developing sustainable AFSC. The agricultural sector in the future is expected to adopt machine learning as a decision support tool. In addition, this review can be used as a reference by researchers and practitioners to understand the application of machine learning in AFSC.

2. Material and methods
The literature review was compiled using the systematic literature review (SLR) method. We used selected keywords to specify the conceptual boundaries, such as "decision trees", "artificial neural network", "clustering", "genetic algorithm", "Bayesian network", "random forest", "agri-food supply chain", "food industry", etc. Publications were analyzed for the period between 2010 and 2020. The final result, 20 articles were used to analyze and conduct the review.

3. Results and discussions

3.1. Machine learning (ML)
One of the algorithms developed by many researchers to predict certain data is machine learning. ML was developed to be able to learn on their own without any direction from the user. The basis of ML development is based on statistical science, mathematics, and data mining. It is expected to be able to analyse data without having to do commands or reprogramming. ML builds prediction models referring to historical data [4]. ML used in AFSC is classified into 2 groups, namely supervised learning and unsupervised learning.

Algorithm ML supervised learning uses data labelled. Supervised learning algorithms accept a set of inputs with the right output. This means that the algorithm can modify the model to fit the desired results [5]. Supervised learning is categorized into two groups, namely regression and classification techniques [6]. The ML algorithms that are widely used to analyse SCM performance are decision trees, random forests, Bayesian networks, and regression analysis [7]. Random forest (RF) is an algorithm used for big data classification. The classification process in RF is based on the training of the sample data owned [8]. Decision tree algorithms classify data into smaller subsets, where each set contains a "yes" or "no" response [9,10]. Bayesian networks are widely used to predict outcomes based on Bayes' theorem, namely by calculating the class conditional probabilities and the previous probabilities. Then, regression analysis is widely used to generate equations that show the relationship between input and output parameters.

Unsupervised learning does not use labels in predicting target features/variables. Unsupervised learning uses the similarity of the attributes that are owned. If the attributes and properties of the extracted feature data are similar, they will be grouped (clustering). The number of clusters can be unlimited. The model labels are from these groups, and if new data is to be predicted, it will be matched with groups with similar features [5]. Unsupervised learning consists of artificial neural networks, k-means clustering, principal component analysis, and deep learning. ANN is an information processing technique inspired by the workings of the biological nervous system, namely brain cells in humans [11]. K-means clustering is widely used to divide data into k clusters. In comparison, deep learning is known as deep ANN because it has many hidden layers [4,5,8,11].
3.2. Machine Learning in agri-food supply chain

AFSC covers all “upstream” to “downstream” activities from farm cultivation and crop production, processing, testing, packaging, storage, and marketing [9,10]. Effective AFSC activities require management activities and decision-making processes at strategic, tactical, and operational levels. AFSC has complex characteristics compared to other supply chains due to its perishable, seasonal, fluctuating nature and increased consumer awareness of product quality and safety [12]. AFSC includes four operations: production planning and control at suppliers, food processing at processors (producers), distribution at distributors, and consumption at retailers/customers (see Figure 1).

Figure 1. Agri-food supply chain operations (designed by author).

3.2.1 Production planning and control. Process planning and control include all the activities needed to achieve efficient, effective, and economical operations [13]. At this stage, the activities carried out include production planning, supplier selection, predicting supply chain risks, stock market prediction, supplier segmentation, inventory management, demand forecasting, and performance prediction. Table 1 presents a list of different ML algorithms used for production planning and control.

3.2.1.1 Production planning. The use of ML tools will increase accuracy in production planning because it considers multiple constraints. Machine learning tools are more effective than manual methods, especially for companies that rely on build-to-order and make-to-stock production workflows. ML tools can reduce supply chains latency for components and parts [7].

3.2.1.2 Supplier selection. Suppliers have an important role as it relates to time, quality, and cost [7]. A potential support vector machine (P-SVM) combines DT to solve the supplier selection problem. The P-SVM can be used to construct binary classification and select various features simultaneously.

3.2.1.3 Predicting supply chain risk. The SCRM study that uses Machine Learning is Bruzzone, and Orsoni [14], supply chain risks such as production losses are predicted by Artificial Neural Networks (ANNs). ANN is equipped with specific scenarios such as inputs (production times, capacities, and quantities) and output (cost estimates). Based on the training data, ANNs learn to correlate between input and output to obtain cost estimates for different scenarios [15].

3.2.1.4 Inventory management. The inventory model proposed to manage supply with one supplier and several retailers consists of two groups, namely centralized and non-centralized inventory models. The algorithms are compiled according to customer demand [9]. The algorithms used to solve inventory management problems, especially perishable products, are Q-learning and the Sarsa methods.
Table 1. Machine learning algorithm in agri-food supply chains.

| Operations                              | Authors | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
|-----------------------------------------|---------|----|----|----|----|----|----|----|----|----|
| **Production planning and control**     |         |    |    |    |    |    |    |    |    |    |
| Production planning                     | [7,13]  | ✓  |    |    | ✓  | ✓  | ✓  |    |    |    |
| Supplier selection                      | [5,7,16]| ✓  | ✓  |    | ✓  | ✓  | ✓  | ✓  | ✓  |    |
| Predicting supply chain risks           | [15]    |    | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Stock market prediction                 | [17,18] | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Supplier segmentation                   | [7]     | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Inventory management                    | [7]     | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Demand forecasting                      | [6,7,19]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Lead time forecasting                   | [20,21]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Performance prediction                  | [22,23]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| **Food Processing**                     |         |    |    |    |    |    |    |    |    |    |
| Process control                         | [24]    | ✓  |    |    | ✓  | ✓  | ✓  |    |    |    |
| Scheduling                              | [23–25]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |
| Quality control                         | [2,26]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| **Distribution**                        |         |    |    |    |    |    |    |    |    |    |
| Distribution planning                   | [7]     | ✓  |    |    | ✓  | ✓  | ✓  | ✓  |    |    |
| Transportation                          | [7,27]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Food Delivery                           | [28,29]| ✓  |    |    | ✓  | ✓  | ✓  | ✓  |    |    |
| **Consumption**                         |         |    |    |    |    |    |    |    |    |    |
| Customer Satisfaction                   | [4,30]| ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Consumer perception                     | [30]    | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| Buying behaviour                        | [4]     | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  | ✓  |    |    |
| **Description**                         |         |    |    |    |    |    |    |    |    |    |
| 1: Support Vector Machine               | 2: Genetic Algorithm                  | 3: Clustering                  | 4: Regression                  | 5: Decision Trees             |
| 6: Deep Learning                        | 7: Bayesian Network                   | 8: Random Forest               | 9: Artificial Neural Network    |

3.2.1.5 *Demand forecasting*. Single feature time series employ many ML techniques. Multidimensional time series can improve ML performance. One technique that is often used is SVM to choose the best forecasting model from several model predictions [22].

3.2.1.6 *Lead time*. Purchasing lead time is the time interval between the delivery of the order to the company and the delivery to the supplier. Forecasting the purchase lead time is an important task for the procurement department to plan, manage, and control the production process. ML regression is commonly used to estimate the lead time of purchases in the supply chain using real industry databases [21].

3.2.1.7 *Performance prediction*. Tree-based classification is widely used to maximize profits. It works to accept or reject incoming orders based on production capacity and order needs [22]. The most widely used algorithm methods for optimizing the prediction mix are the hybrid model and simulated annealing [23].

3.2.2 *Food processing*. Food processing is the activity of physically or chemically processing raw materials into finished or semi-finished food. Food processing combines raw materials to produce food products that can be marketed and consumed by consumers. The ML developed for the food processing stage is process control, scheduling, and quality control [24].
3.2.2.1 Process control. Process Control aims to regulate and improve operating processes to minimize non-conforming situations. Deep NN and Linear Regression can be used to predict throughput time. The system is regulated by workload control to determine reliable due dates to reduce the percentage of tardy jobs [29].

3.2.2.2 Scheduling. ML technique has a high effectiveness in predicting scheduling accuracy by balancing build to order and make to stock production workflows. Genetic algorithms are used flexibly to solve job shop scheduling problems, while 2 NNs are used for machine allocation and priority assignment [15]. A real-time scheduling system such as NNs, SVM, and Decision Tree (based on Bagging) are integrated, and a Genetic algorithm can be used for supplier selection in AFSC [25].

3.2.2.3 Quality control. Quality control (QC) aims to ensure that product quality can be maintained and improved. Quality control is carried out by testing the final product specifications regularly. ML tools that can be used are hierarchical clustering and SVM [26]. In addition, convolutional NN can classify failure map patterns on wafer products; the tool is integrated into an automatic monitoring system in real-time [2].

3.2.3 Distribution. The distribution is the stage of distributing goods or services from producers to consumers. ML can be used at the distribution stage for distribution planning, transportation, and food delivery. Distribution and transportation are an important activity in SCM that loads heavy costs on organizations. The results of implementing ML in distribution and transportation systems can lead to the delivery time of a product to the appropriate customers by generating better delivery routes and exploring consumer behaviour [27]. Machine Learning (ML) can help decide delivery routes, predict food demand, supply raw materials, and plan logistics [28]. The delivery route problem can be solved with ML by optimizing the location of the shipping agent. Information about the current upcoming traffic situation is sent to inform them of the best route synchronously. Ensuring efficiency and on-time delivery makes it easier to deliver constant orders and even deal with problems such as running out of delivery agents or late deliveries.

3.2.4 Consumption. ML techniques such as deep learning and ANN predict consumer demand, buying behaviour, and consumer perception [30]. The Bayesian network was found to be forecasting the consumers buying behaviour of different food products and performing the quality checks [4]. Factors that influence consumer purchasing behaviour for imported ready-to-eat food can be analysed using ANNs and logistic regression techniques.

3.3 Machine learning for AFSC sustainability
The review purpose of analysing agri-food data using ML algorithm is to develop efficient AFSCs. The use of ML in AFSC not only to economic benefits but also to and environmental and social performance (see Figure 2). Many companies struggle to manage their production systems due to increasing market uncertainty. While emerging ‘smart’ technologies such as the internet of things, cloud computing, and machine learning are solutions to transforming traditional production management systems into real-time-based.

The model presented adopts an incremental approach that companies with limited resources can apply to improve their production planning and control process in the context of industry 4.0 and sustainability. The results obtained are that make-to-order companies can obtain greater benefits from a smart product strategy, while make-to-stock companies gain greater profits by implementing a smart process strategy [13]. At the food processing stage, ML algorithm is used for process control, scheduling, and process control. The use of ML provides benefits for social sustainability such as reducing poverty, increasing economic growth while improving public health [31,32]. Thus, increasing social sustainability through the development of active local markets by meeting the customer and market demands. At the distribution stage, the use of ML has resulted in sustainable social and economic
performance improvements through improved food safety, quality, economic savings, and customer satisfaction. The use of ANN in demand forecasting leads to economic savings and increased customer satisfaction. Similarly, optimized vehicle routing and fleet management contribute to fuel economy that improves environmental and economic performance in last-mile deliveries.

Figure 2. Frameworks of the potentials of machine learning for AFSC sustainability (designed by author).

Figure 2 presents a summary outline of the great potentials of the ML algorithm for developing sustainable AFSCs. The findings of the review illustrate AFSC span the three foundational elements of sustainability. Practitioners can use the framework developed in the current study as it explains how the ML algorithm contributes to the overall efficiency of AFSCs and mitigates some of the challenges faced by the agri-food industry. The proposed framework will guide these practitioners in deciding the appropriate AFSC resources and AFSC management to improve sustainable performance.

4. Conclusions
The ML algorithms used to develop sustainable AFSC are grouped into two categories, namely supervised learning (Bayesian network, regression, and decision trees) and unsupervised (ANN, deep learning, GA, and clustering). The main contribution of the study is the ML-AFSC performance applications framework that will further guide academics and practitioners to understand the current states of literature in this agri-food. This framework is based on a literature review that has not been tested empirically. Therefore, future researchers can validate this framework empirically. In addition, researchers can also explore the application of ML to AFSC in different regions.

References
[1] Lezoche M, Hernandez J, Alemany M, Lezoche M, Hernandez J, Alemany M, Panetto H, Kacprzyk J, Lezoche M, Hernández J E, Eva M, Diaz A, Panetto H and Kacprzyk J 2020 Agri-food 4.0: a survey of the supply chains and technologies for the future agriculture To cite this version: HAL Id: hal-02395411 Future Agriculture Comput. Ind. 117 1–15
[2] Nakata K, Orihara R, Mizuoka Y and Takagi K 2017 A Comprehensive Big-Data-Based Monitoring System for Yield Enhancement in Semiconductor Manufacturing IEEE Trans. Semicond. Manuf. 30 339–44
[3] Di Vaio A, Boccia F, Landriani L and Palladino R 2020 Artificial intelligence in the agri-food system: Rethinking sustainable business models in the COVID-19 scenario Sustain. 12
[4] Sharma R, Kamble S S, Gunasekaran A, Kumar V and Kumar A 2020 A systematic literature review on machine learning applications for sustainable agriculture supply chain performance Comput. Oper. Res. 119 104926

[5] Chibani S and Coudert F X 2020 Machine learning approaches for the prediction of materials properties APL Mater. 8

[6] Moroff N U and Sardesai S 2019 Machine Learning in Demand Planning: Cross-industry Overview Proc. Hambg. Int. Conf. Logist. – 27 355–83

[7] Tirkolaee E B, Sadeghi S, Mooseloo F M, Vanchali H R and Aeini S 2021 Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas Math. Probl. Eng. 2021 1–14

[8] Rejeb A, Rejeb K and Zailani S 2021 Big data for sustainable agri-food supply chains: a review and future research perspectives J. Data, Inf. Manag.

[9] Tsolakis N K, Keramidas C A, Toika A K, Aidonis D A and Iakovou E T 2014 Agrifood supply chain management: A comprehensive hierarchical decision-making framework and a critical taxonomy Biosyst. Eng. 120 47–64

[10] Saetta S and Caldarelli V 2020 How to increase the sustainability of the agri-food supply chain through innovations in 4.0 perspective: A first case study analysis Procedia Manuf. 42 333–6

[11] Abdella G M, Kucukvar M, Onat N C, Al-Yafay H M and Bulak M E 2020 Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption J. Clean. Prod. 251 119661

[12] Kamble S S, Gunasekaran A and Gawankar S A 2018 Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives Process Saf. Environ. Prot. 117 408–25

[13] Oluyisola O E, Sgarbossa F and Strandhagen J O 2020 Smart production planning and control: Concept, use-cases and sustainability implications Sustain. 12

[14] Bruzzone A and Orsoni A 2003 AI and simulation-based techniques for the assessment of supply chain logistic performance Proc. - Simul. Symp. 2003-Janua 154–64

[15] Baryannis G, Validi S, Dani S and Antoniou G 2019 Supply chain risk management and artificial intelligence: state of the art and future research directions Int. J. Prod. Res. 57 2179–202

[16] Cavalcante I M, Frazzon E M, Forcellini F A and Ivanov D 2019 A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing Int. J. Inf. Manage. 49 86–97

[17] Patel J, Shah S, Thakkar P and Kotecha K 2015 Predicting stock market index using fusion of machine learning techniques Expert Syst. Appl. 42 2162–72

[18] Ballings M, Van Den Poel D, Hespeels N and Gryp R 2015 Evaluating multiple classifiers for stock price direction prediction Expert Syst. Appl. 42 7046–56

[19] Fu W and Chien C F 2019 UNISON data-driven intermittent demand forecast framework to empower supply chain resilience and an empirical study in electronics distribution Comput. Ind. Eng. 135 940–9

[20] Biazon de Oliveira M, Zucchi G, Lippi M, Cordeiro D, Rosa da Silva N and Iori M 2021 Lead Time Forecasting with Machine Learning Techniques for a Pharmaceutical Supply Chain 1 634–41

[21] Oliveira F N de, Leiras A and Ceryno P 2019 Environmental risk management in supply chains: A taxonomy, a framework and future research avenues J. Clean. Prod. 232 1257–71

[22] Duan Q, Zeng J, Chakrabarty K and Dispoto G 2015 Data-driven optimization of order admission policies in a digital print factory ACM Trans. Des. Autom. Electron. Syst. 20

[23] Delgoshaei A and Gomes C 2016 A multi-layer perceptron for scheduling cellular manufacturing systems in the presence of unreliable machines and uncertain cost Appl. Soft Comput. J. 49 27–55

[24] Bertolini M, Mezzogori D, Neroni M and Zammori F 2021 Machine Learning for industrial applications: A comprehensive literature review Expert Syst. Appl. 175 114820
[25] Shiue Y R, Guh R and Lee K 2012 Development of machine learning-based real time scheduling systems: Using ensemble based on wrapper feature selection approach *Int. J. Prod. Res.* **50** 5887–905

[26] Wuest T, Weimer D, Irgens C and Thoben K D 2016 Machine learning in manufacturing: Advantages, challenges, and applications *Prod. Manuf. Res.* **4** 23–45

[27] Liu C, Shu T, Chen S, Wang S, Lai K K and Gan L 2016 An improved grey neural network model for predicting transportation disruptions *Expert Syst. Appl.* **45** 331–40

[28] Alfi R and Triyani W 2019 Perancangan Sistem Database Stock Location untuk Optimalisasi Stock Accuracy pada Gudang Pengemasan Suatu Perusahaan Minuman Design of Stock Location Database System for Optimizing the Stock Accuracy of Beverage Company Packaging Warehouse *J. Teknol. dan Manaj. Agroindustri* **8** 117–32

[29] Mezzogori D, Romagnoli G and Zammori F 2020 *Defining accurate delivery dates in make to order job-shops managed by workload control* (Springer US)

[30] Ribeiro F C, Borêm F M, Giomo G S, De Lima R R, Malta M R and Figueiredo L P 2011 Storage of green coffee in hermetic packaging injected with CO2 *J. Stored Prod. Res.* **47** 341–8

[31] Santoso I, Afifa Y N, Astuti R and Deoranto P 2021 Development model on upstream-downstream integration of coffee agroindustry using dynamics modelling approach *IOP Conf. Ser. Earth Environ. Sci.* **733**

[32] Choirun A, Santoso I and Astuti R 2020 Sustainability risk management in the agri-food supply chain: Literature review *IOP Conf. Ser. Earth Environ. Sci.* **475**