The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains

Mallika Kliangkhlao

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Computer Engineering

Prince of Songkla University

2022

The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains

Mallika Kliangkhlao

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Computer Engineering

Prince of Songkla University

2022

Thesis Title The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains

Author Miss Mallika Kliangkhlao

Major Program Computer Engineering

Major Advisor

....

(Dr. Somchai Limsiroratana)

Examining Committee:

....Chairperson

(Asst. Prof. Dr. Nichnan Kittiphattanabawon)

....Committee

(Assoc. Prof. Dr. Pichaya Tandayya)

....Committee

(Asst. Prof. Dr. Panyayot Chaikan)

....Committee

(Dr. Anant Choksuriwong)

....Committee

(Dr. Somchai Limsiroratana)

The Graduate School, Prince of Songkla University, has approved this thesis as fulfillment of the requirements for the Doctor of Philosophy Degree in Computer Engineering

....

(Asst. Prof. Dr. Thakerng Wongsirichot)

Acting Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate’s own investigations. Due acknowledgment has been made of any assistance received.

....Signature

(Dr. Somchai Limsiroratana)

Major Advisor

....Signature

(Miss Mallika Kliangkhlao)

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

....Signature

(Miss Mallika Kliangkhlao)

Candidate

Thesis Title	The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains
Author	Miss Mallika Kliangkhlao
Major Program	Computer Engineering
Academic Year	2022

Abstract

Agricultural supply chain management depends upon the decision-making to stabilize the market situation. Uncertainties in demand and supply in the market dynamics are the main thread to the management. It then requires product flow and activities to be understood thoroughly and immediately. This task requires comprehensive information, expertise, and processing ability, which are time-consuming and labor-intensive. This research proposes an automatic system framework alongside a Causal Bayesian Networks model for market detection and explanation using streaming data. This research contributes to designing and developing the model by encoding expert knowledge using cause-and-effect assumptions integrating with supply chain ground through. This model can detect the market situation rationally, likewise human logic. The results proved that the proposed model could accurately detect and reasonably explain the event. It illustrates that the model is suitable and ready for application to real-world applications for supporting decision-making in agricultural supply chain management.

ชื่อวิทยานิพนธ์ การออกแบบและพัฒนาแบบจำลองเครือข่ายเบย์เซียนแบบเหตุและผลเพื่ออธิบายห่วงโซ่อุปทานทางการเกษตร

ผู้เขียน นางสาวมัลลิกา เกลี้ยงเคล้า

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2565

บทคัดย่อ

การจัดการห่วงโซ่อุปทานด้านการเกษตรมีวัตถุประสงค์เพื่อดูแลและควบคุมให้กระบวนการผลิตและซื้อขายสินค้าให้มีประสิทธิภาพและมีความสมดุลย์ แต่ทว่าความสมดุลย์ภายในตลาดเป็นภาวะที่เกิดขึ้นค่อนข้างยากเนื่องจากความผันผวนของอุปสงค์และอุปทาน ดังนั้นการจัดการห่วงโซ่อุปทานจึงจำเป็นต้องอาศัยความเข้าใจในกลไกการผลิตสินค้าที่ครอบคลุมตั้งแต่การเพาะปลูก แปรรูป และซื้อขายอย่างละเอียดรอบคอบและทันท่วงที ทว่าการทำความเข้าใจนั้นมีข้อจำกัดคือยังคงต้องอาศัยมนุษย์ในการตีความและประมวลผลซึ่งเป็นงานที่สิ้นเปลืองเวลาและต้องอาศัยทรัพยากรบุคคลที่มีความรู้เฉพาะทาง ดังนั้นงานวิจัยนี้จึงนำเสนอกรอบแนวคิดของระบบอัตโนมัติร่วมกับโมเดลเครือข่ายเบย์เซียนแบบเหตุและผลเพื่อใช้ในการตรวจสอบและอธิบายเหตุการณ์ทางการตลาด งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนาโมเดลดังกล่าวโดยการถอดความรู้เฉพาะทางจากมนุษย์โดยอาศัยแนวคิดสมมติฐานเชิงเหตุและผลบูรณาการร่วมกับองค์ความรู้ด้านห่วงโซ่อุปทาน ผลการทดสอบพบว่าโมเดลที่นำเสนอสามารถตรวจสอบเหตุการณ์ได้ถูกต้องและสามารถอธิบายเหตุการณ์ได้อย่างสมเหตุสมผล ดังนั้นผลลัพธ์บ่งชี้ว่าโมเดลมีความเหมาะสมและสามารถนำไปประยุกต์ใช้ในระบบการจัดการห่วงโซ่อุปทานด้านการเกษตรได้

Acknowledgments

First and foremost, I must thank my advisor, Dr. Somchai Limsiroratana, for understanding and supporting me over the years and giving me so much freedom to choose, explore, and pursue this research.

My sincere thanks also go to Dr. Andrew Davison for his support and patience with my academic writing. His classes and manuscript revisions have proved to be one of my learning experiences.

I want to thank the rest of my thesis committees, Asst. Prof. Dr. Nichnan Kittiphattanabawon, Assoc. Prof. Dr. Pichaya Tandayya, Asst. Prof. Dr. Panyayot Chaikan, and Dr. Anant Choksuriwong for their encouragement, insightful comments, and inspired questions.

I gratefully acknowledge the funding received towards my PhD from the Thailand education hub for ASEAN scholarship in doctoral degree funded by Prince of Songkla University, Thailand.

I would also like to give special thanks to the experts at Central Rubber Market in Hat Yai, Songkhla, Thailand, for their helpful comments, suggestions, and data.

Finally, I must express my profound gratitude to my family and friends for providing unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis.

Mallika Kliangkhlao

Contents

Content Page

Chapter 1 Introduction 12
1.1 Motivation 12
1.2 Original Contributions 14
1.3 Research Objectives 14
1.5 Research Scopes 14
1.6 Thesis Structure 14
Chapter 2 Background Knowledge and Literature Review 16
2.1 Agricultural Supply Chain and Big Data 16
2.2 Agricultural Supply Chain Management 18
2.3 Causal Bayesian Networks 20
2.4 Research Challenges 22
Chapter 3 Methodology 23
3.1 Automatic Approach for Agricultural Supply Chains Explanation 23
3.2 Streaming Data Digitization for Agricultural Supply Chains 23
3.3 CBNs Model Design and Development 26
3.3.1 Qualitative CBNs Model 26
3.3.2 Quantitative CBNs Model 29
Chapter 4 Results and Discussion 32
4.1 Experiments 32
4.1.1 Streaming Data Digitization 33
4.1.2 Predictive Performance Measurement 35
4.1.3 Explanation Measurement 36
4.2 Discussion 43
Chapter 5 Conclusion 44
5.1 Research Summary 44
5.2 Claim of Originality 45
5.2.1 The New Framework 45
5.2.2 The Original Model 45
5.3 Future Directions 46

Contents (continued)

Content Page

References......48

Appendix 53

Appendix 1 54

Appendix 2 61

Appendix 3 79

VITAE 94

List of Tables

Table Page

Table 2.1 Previous works related to the agricultural supply chain. 18

Table 2.3 Causal Structure Conclusion.. 21

Table 3.1 Summarization of natural rubber supply chains data.................................. 29

Table 3.2 Summarization of natural rubber supply chains’ contextual variables........ 29

Table 4.1 The different causal relationship between CBNs model and BS model...... 39

List of Figures

Figure Page

Figure 2.1 Conceptual of causation relationship..16

Figure 2.2 CBNs represent a hypothetical crop yield production...............................20

Figure 2.3 CBNs represent a hypothetical crop yield production...............................21

Figure 3.1 The ASCs explanation conceptual framework. 23

Figure 3.2 The data preprocessing approach based on time-series decomposition. 25

Figure 3.3 Causal assumption of natural rubber SCs using CBNs. 28

Figure 4.1 The tuned CBNs model. 38

Figure 4.2 The tuned BS model. 38

Figure 4.3 The results of the first question. 39

Figure 4.4 The results of the second question...40

Figure 4.5 The results of the third question. 41

Chapter 1
Introduction

1.1 Motivation

The agricultural sector plays an essential part in many countries, especially Thailand. There are 12.6 million farmers across the country, the largest occupation, who earn an annual income below the poverty line with a lot of debt [1]. Climate change, education, and lack of modern technology are limitations that cause high risks in the productivity of Thai farming. Even though Thai farmers face problems in their job, the Thai government claims they are “the kitchen of the world”, and their food industries and businesses play the most significant part of the country’s GDP [2]. While farmers—suppliers receive a minor income, the food industries—demand earns an enormous profit. This paradox undeniably reflects that the nature of agricultural supply chain management is changed and dynamic, which is too far needed to reform the policy to strengthen farmers’ productivity, income, and security. It depends upon the decision-making that requires understanding the supply chain situation thoroughly.

Management of the Agricultural Supply Chains (ASCs) then needs to adapt to understand the current situation and the upcoming event for deciding policies to support and stabilize ASCs. Significantly the world-class market is transformed into a modern supply chain [3]. This modern style creates shortcuts in the product flows from farmers to consumers by transferring the trading process to the commodity market that runs on a digital platform. Then, the relationship between suppliers and consumers is more dynamic because there is no market monopoly, which strengthens suppliers' bargaining power. The demand and supply data emerged continuously in various digital platforms called Big Data. The decision-making performance depends upon Big Data analysis to gain intensive knowledge, including the product flows, climate impacts, business strategies, and factors that affect price movement. This work covers data sensing, preprocessing, analyzing, understanding, and decision-making, which is time-consuming and labor-intensive for the respondent.

In the era of automatic systems, many studies share contributions in supply chain management using Machine Learning (ML) and data-driven approaches for dealing with human limitations. Crop yield prediction with deep learning was studied by focusing on supply information from farm sensing data to help increase productivity in ASCs [4]–[6]. Punia *et al.* [7] proposed a deep learning-based approach for demand prediction. They contributed that demand prediction from point-of-sale data helps decision-makers make strategic, tactical, and operational decisions for ASCs. Chen *et al.* [8] focused on automatic agricultural commodity price prediction to help the government detect the market balance and plan ASCs management policy. These studies performed good results in demand and supply prediction. Although prediction, an absolute data-dependency approach, is perfectly fine for estimating a situation in a regular supply chain, the nature of the modern ASCs is too dynamic and beyond the power of prediction. Not only the situation prediction, but the modern ASCs management also requires an explanation that provides details for supporting decision-making.

An ASCs explanation means the details of each ASCs operation in the whole process, covering pre-production to retail [9]. It should answer basic questions likes '*How about the crop yield production and why is it?*' and '*Why does demand drop?*'. These questions depend on an expert's intensive knowledge to monitor information for deep understanding.

This research aims to employ Causal Bayesian Networks (CBNs) to encode human-like knowledge into the ML model towards improved automatic ASCs explanation. The research question is '*How to analyze streaming data incorporating with CBNs model to detect and explain the ASCs situation*?'. The approach for ASCs explanation is then a challenge to deal with the dynamics of ASCs and the data sources.

1.2 Original Contributions

1. An automatic framework for analyzing an ASCs situation from real-time big data.
2. The CBNs model incorporated with prior knowledge for explaining ASCs situations based on expertise manners.

1.3 Research Objectives

1. To explore an intuitive approach for sensing ASCs information from digital platforms for supporting ASCs explanation.
2. To design and develop a CBNs model encoded from human-like intelligence to generate ASCs knowledge from big data.

1.5 Research Scopes

1. The ASCs background knowledge is declared based on Thailand's rubber market.
2. The dataset used in this thesis was collected from multiple open sources between 2017-2019.
3. The ASCs knowledge covers the demand, supply, and market situation required for agricultural market management.

1.6 Thesis Structure

Even though some of these research contributions have already been published, they are some subjects that may include contribute to a clearer understanding of this thesis. The rest of this thesis is concerned with Causal Bayesian Models (CBNs) representation. Using CBNs makes the ability to represent knowledge of Agricultural Supply Chains (ASCs) into a model. This thesis is represented as follows.

In Chapter 2, the background knowledge of the ASCs and the reviews of the current limitation of machine learning models for ASCs management will be examined. Chapter 3 introduces the methodology of design and development for the CBNs model.

In particular, the framework for supporting the CBNs model with big data illustrates how to apply the model in an automatic system. The results and discussion of the CBNs model for explanation of an ASCs situation to prove the research question are detailed in Chapter 4. Furthermore, the conclusions, research trends, and future directions will be recommended in Chapter 5. In appendices contain the publications that previously contributed. The appendices contain a list of previously published materials that are the ground through this thesis. Appendix 1 presents an overview of the agricultural market understanding. Appendix 2 proposes a concept, development, and experiments of big data digitization for supporting machine learning modeling. And Appendix C presents a vital knowledge of the contributed CBNs modeling. It contains an original idea, background, design and development, and experiment of the CBNs model for the agricultural supply chains explanation.

Chapter 2

Background Knowledge and Literature Review

In this chapter, the characteristics of big data are developed based on the agricultural supply chain. Section 2.1 and Section 2.2 show recent studies on market situation detection with big data and machine learning technologies. The perspective of Causal Bayesian Networks is introduced to deal with the complexity of agricultural big data and the market explanation in Section 2.3. In Section 2.4, the research challenges are discussed.

2.1 Agricultural Supply Chain and Big Data

The agricultural supply chain concerns activities that transport crop yield from suppliers to process and dispatch to final consumers that have impacts on each other in the chain manners, as shown in Figure 2.1.

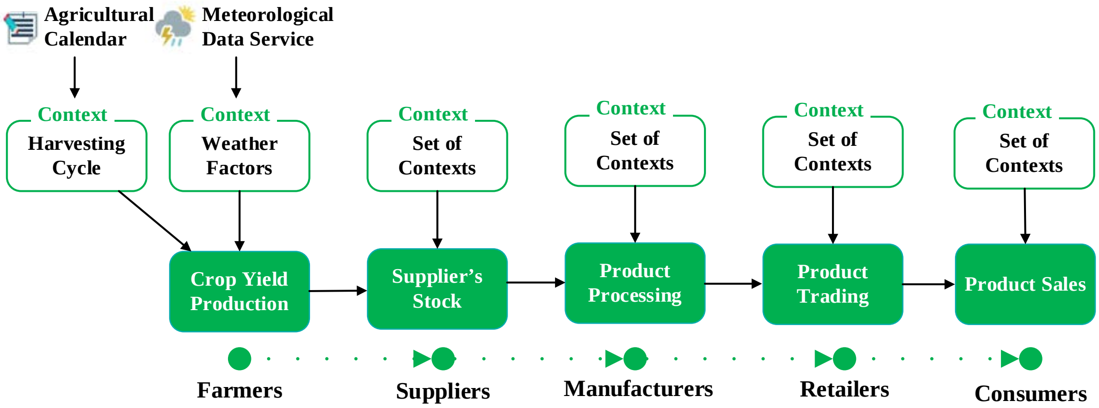


Figure 0.1 Conceptual of causation relationship.

Figure 2.1 shows the diagram of the concept of the supply chain for the agricultural market. This supply chain shows the dependencies between activities that reflect demand and supply in the short-term or long run. The processes in the supply chain start with crop yield production by farmers—a supplier. The product flow runs along the chain to the supplier's stock, product processing, and product trading until ends with the product sales activity which depended upon consumers. Therefore, decision makers need diversely supportive information to understand the supply chain.

Focused on the first activity—crop yield production, the production quantity may be naively considered from the plantation area using on-farm sensors. However, it is not easy as its looks, the crop yield production is vulnerably depended upon numerous factors, such as the harvesting cycle, weather, or even crop infectious diseases. Then, decision makers always manage supply chain situations using competency questions, such as "Will supply quantity drop if it is going to be monsoon season?". These factors are called contexts which are causes for inferencing the result of activities. The context is contributed to discover the contextual information of supply chain activities. It shows that the effect of 'monsoon season' is long-run with hidden time-dependent information.

According to Figure 2.1, crop yield production can be estimated using its contexts; harvesting cycle, which is seasonality according to agricultural calendar, and weather information from meteorological data service. It shows the requirement for contextual information and knowledge to detect and explain the supply chain situation.

In the era of big data, it is related and well-timed data that can be found in multisource sensors whether the smart devices, crowdsourcing, open data, and internal data warehouse. The details of applying big data in the agricultural field have been discussed and published, detailed in Appendix 1.

Although big data is full of contextual information, it inherits the complex characteristics of big data: volume, velocity, variety, veracity, and valorization [10]. Particularly, I focus more on the variety and uncertainty that requires human intelligence to fuse, infer, and transform data into information. However, manual agricultural big data processing is time-consuming and labor-intensive. Machine learning then comes across to perform an automatic concept.

2.2 Agricultural Supply Chain Management

In the agricultural domain, Machine Learning (ML) has been proposed as a solution to deal with gigantic experiences from big data [11], [12]. It is because it is an approach to achieving an automatic model to deal with complex problems. It has three main components; task as a goal of it to perform analysis, experience as a source for learning, and performance measurement to prove the ability of task performing. ML has been adopted in the agricultural domain related to demand and supply exploration using multimedia data. Koirala *et al.* [13], Haghverdi *et al.* [14], Zhang *et al.* [15], and Akbar *et al.* [16] proposed a deep learning model for crop yield production prediction as a supply exploration. Simple regression was proposed to deal with large and multi datasets of statistical data for rubber demand and supply prediction [17]. Furthermore, the hybrid model of deep learning and regression was proposed as a novel approach for supply prediction [18]. These studies have been perfectly proved for demand and/or supply prediction. However, that predictions still lack the contexts required in supply chain explanation management.

Table 0.1 Previous works related to the agricultural supply chain.

Author	Supply		Demand		Price
	Short-term	Long-term	Short-term	Long-term	
Bocca, F. and Rodrigues, L. (2016)	✓	×	×	×	×
Arunwarakorn *et al.* (2017)	✓	×	✓	×	×
Shynkevich *et al.* (2017)	×	×	×	×	✓
Stein and Steinmann (2018)	×	✓	×	×	×
Zhang *et al.* (2018)	×	×	×	×	✓
Chen *et al.* (2018)	×	×	×	×	✓
Zhu *et al.* (2019)	×	×	×	×	✓

The relationship between demand and supply is standard law in economics which encodes a supply chain prior knowledge to reveal market price movement [19], [20]. The interrelationships between activities affect each other as short-long-term impacts according to the production cycle and time lag. Short-term impacts directly affect activities, while long-term impacts are indirect. This thesis reviews related works that focus on contexts of supply chain based on short-term and long-term impacts, as shown in Table 2.1.

Table 2.1 summarizes relevant studies focusing on the impacts of short-and-long-term demand and supply. In addition, the reviewed papers show the lack of short-and-long-term demand and supply recognition that has a goal for supply chain management. Bocca and Rodrigues [21] and Arunwarakorn *et al.* [17] proposed short-term supply prediction using several factors, such as weather data, stock, and crop prices. While Stein, S. and Steinmann, H. [22] proposed long-term supply using annual weather data. Shynkevich *et al.* [23], Chen *et al.* [24], Zhu *et al.* [25], and Zhang *et al.* [26] proposed price behavior detection without mentioned on market demand and supply. Moreover, this ignores discovering the relationship among factors in the supply chain, which is an invaluable opportunity to discover contexts.

As a result, causation is a vision to explore contextual information based on a cause-and-effect relationship. It is a prerequisite adopted to select and explore the valuable data from the multi-data source, which can be used to recognize short-and-long-term of demand and supply.

Bayesian Networks (BNs) transparently model knowledge of supply chain relationships to produce such information, which decision-makers employ to create policies. BNs are probabilistic graphical models that capture the uncertainty and relationships among relevant factors in the supply chain decision-making process. Random variables represent these factors, and their relationships are encoded by conditional probabilities using Bayes' theorem. Sharma and Sharma [27], Chhimwal *et al.* [28], Lawrence *et al.* [29], and Ojha *et al.* [30] proposed for BNs-based risk assessment approach for supply chain management using historical data. They summarized that the approach could help the supply chain managers identify the risk factors early. El Amrani *et al.* [31] studied the sustainability of the supply chain network. These methods were successful because they focused on predicted outcomes and contextual explanations. However, they still did not consider explaining the context of demand and supply. It means that the model cannot answer ASCs management questions such as '*What is the situation of demand and supply*? Furthermore, *why were these outcomes produced?*'. The burden of causal interpretation and rational explanation is left to humans.

2.3 Causal Bayesian Networks

Causal Bayesian Networks (CBNs) are originally contributed by applying causality with Bayesian Networks [32]. CBNs represent directed acyclic graphs that encode the causal assumption among variables, as shown in Figure 2.2.

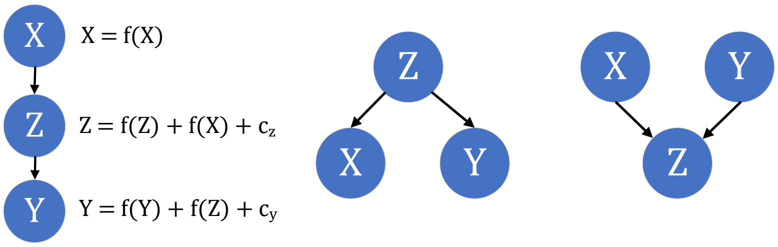


Figure 0.2 CBNs represent a hypothetical crop yield production.

CBNs model consists of a set of variables (nodes) connected by a set of functions (f), where **X**: a set of input (cause), **Y**: a set of output (effect), **Z**: a set of mediators, and c: set of uncertainty. CBNs model determines interdependency among variables using prior knowledge and big data. It represents a relationship between a pair of nodes which is a causal assumption that shows the impact of the parent on the child. The causal assumption, graphical representation, and meaning of the causal structures are concluded in Table 2.3.

The causal structure inherits the conditional dependencies concept to connect the nodes in networks with causal relationships and block the paths between nodes with independencies, called "d-separation" [55]. Causal relationships benefit the ASCs to model knowledge and help decision-makers discover the reasons behind the complex environment.

Table 0.2 Causal Structure Conclusion

Causal Assumption	**Representation**	**Axiom**
Chain	$\boldsymbol{X \rightarrow Z \rightarrow Y}$	***X*** indirectly causes ***Y*** through ***Z***
Fork	$\boldsymbol{X \leftarrow Z \rightarrow Y}$	***X*** and ***Y*** are correlated caused by ***Z***
Collider	$\boldsymbol{X \rightarrow Z \leftarrow Y}$	There is no relationship between ***X*** and ***Y***, but they connect somehow through ***Z***

CBNs help encodes causality in supply chain knowledge and formalizes probabilism in an underlying dataset into mathematic forms using the Bayes theorem.

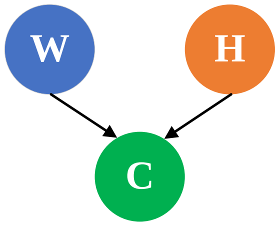


Figure 0.3 CBNs represent a hypothetical crop yield production.

Figure 2.3 shows a simple CBNs that is encoded from a hypothetical crop yield production. This consists of three random variables: weather (**W**), harvesting cycle (**H**), and crop yield production (**C**). They are connected using a causal assumption that **W** and **H** are parents or causes affecting the event of **C**. This graph visuals a qualitative model while a quantitative model is a mathematical form applying with Bayesian Networks shown in (2.1).

$$P(X|Pa(X)) = \prod_{i=1}^{N} P(X^i|Pa(X^i)) \tag{2.1}$$

(2.1) shows a Bayesian Network's chain rule concept which consists of a component of **X**, and a parent of **X** (Pa(**X**)). The relationship between variables has an associated conditional probability distribution (CPD). This approach is applied with the assumption in Figure 2.2, and formalized as follows:

$$P(C|W,H) = \prod P(C|W_{t1}, H_{t2}) \tag{2.2}$$

(2.2) shows a quantitative model of the causal assumption. As mentioned, the parents can affect the child in either the short-term or long-term. Then, an event of W and H can be either the same time slice or the variously different periods (*t1*, *t2*).

The causal structure uses conditional dependencies to connect nodes with causal relationships and block the paths between nodes with independencies; a process known as *d*-separation [33]. Causal discovery algorithms have been studied to structure a CBNs model using statistical properties from the observational data [34]. The algorithms are widely studied, including constraint-based and score-based methods. The constraint-based methods apply conditional independence constraints (e.g., Fast Causal Inference or FCI, and PC), while the score-based methods are based on the posterior probability of the candidate model (e.g. Greedy Equivalence Search or GES, and Greedy FCI). However, the resulted model' performance is hard to be tested without a gold standard [35]. Then, expert-based modeling is the answer for discovering causal relationships in a domain that lacks a baseline.

Therefore, the CBNs model relies on two perspectives: (1) the causal assumptions that show both causalities among the variables, and (2) the impact velocity that shows the time-dependent among the causal assumptions. The effectiveness of CBNs is not just focused on the event prediction, it covers an ability to explain the event in a human sense. The explanation gives both prediction results and its contexts that support decision makers deciding a policy in ASCs management.

2.4 Research Challenges

This thesis is contributed CBNs to encode knowledge of human intelligence by focusing on ASCs management. The challenges are how to construct CBNs and how to apply them in a real-world application.

Chapter 3
Methodology

The goals of this chapter are 1) to propose a framework for supporting ASCs explanation, 2) to present a data processing approach for exploring ASCs data from digital platforms, and 3) to detail a CBNs model design and development methodology.

3.1 Automatic Approach for Agricultural Supply Chains Explanation

The key to ASCs management is to make a thoughtful decision for defining response and review for proactive planning that require humanlike intelligence to explain the situation rationally. It is a real-time and continuous process that is a cause of time-consuming and labor-intensive tasks for decision-makers. This thesis contributed a conceptual automatic framework to sense and respond through the descriptive ASCs management framework, shown in Figure 3.1.

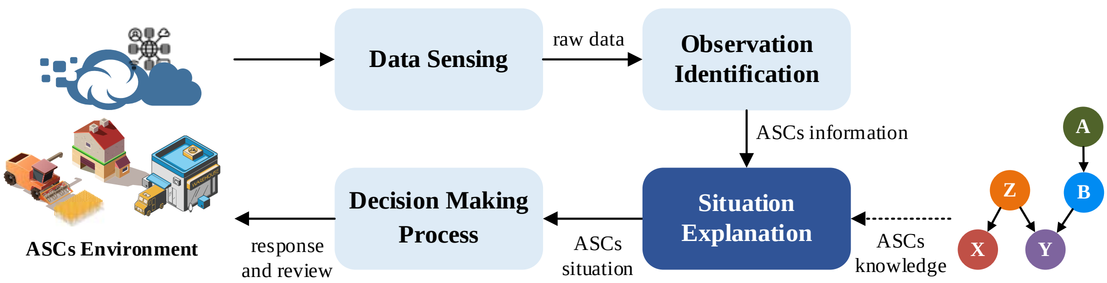


Figure 0.1 The ASCs explanation conceptual framework.

The proposed framework consists of four components: data sensing, observation identification, and decision-making process. Data sensing acts as an agent to gather real-time data from data sources. It can be applied with open data Application Programming Interfaces (APIs), widely available on digital platforms and in-house Internet of Things (IoT). This component's output is raw data sent to observation identification. Observation identification performs the task of identifying and digitizing the raw data into market factors as ASCs information. Then, the information will be sent to the situation explanation component, an advanced component for making the framework descriptive. This component needs ASCs knowledge from the CBNs model for situation detection and explanation. Effective and high-quality knowledge of the CBNs model is the most important because it provides a situation detail for decision-making. Then, the decision depends upon that ASCs knowledge, which provides responses to relieve the ineffective ASCs and reviews for proactive planning. The impacts of the response and review will affect the ASCs environment to produce the new observation for this continuous framework. Then, the CBNs model needs to be designed and developed as a pre-process to encode the ASCs knowledge into machine-interpretable form. The model will be added to the situation explanation component as a machine intelligence in the descriptive framework.

The proposed framework's initial requirements are observation identification and the CBNs model. Then, data processing and CBNs design and development are contributed.

3.2 Streaming Data Digitization for Agricultural Supply Chains

Data preprocessing is the contributed approach for observation identification to automatically digitize the streaming data from multiple sources into ASCs information. The ASCs information is required for ASCs explanation since it is based on the contexts of short-term and long-term impacts of ASCs management. Therefore, the data preprocessing is applied with time-series decomposition to decompose time-related information hidden in the streaming data.

Time-series decomposition technique concerns data change according to time movement by identifying the features based on frequency. Dimensionality reduction transforms high-dimensional data into low-dimensional with essential information [36]. Moreover, it helps the framework decomposing ASCs information that is structured into the time-series components: level, movement, trend, and seasonality

The marketer and decision maker define the interval. The level represents a single point value (e.g. hourly, daily, or monthly), movement is the distance between one level and another, and the trend is a fixed interval made up of a set of movements that represent semantic meaning of the direction of the change. Lastly, seasonality is a long-term scale representing the repeated pattern of trends impacting decision-maker plans for future directions. These time-series components are time-dependent information and can also be applied to short-long term impacts in ASCs information. The data preprocessing approach based on time-series decomposition, shown in Figure 3.2.

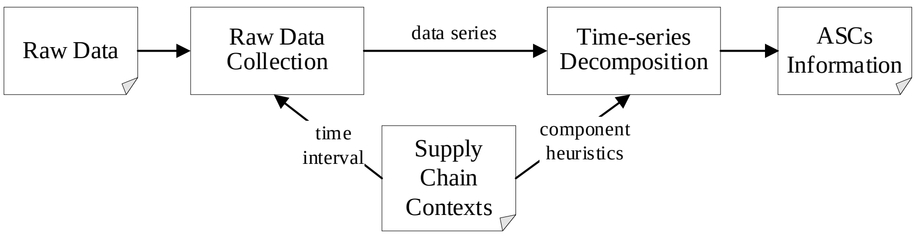


Figure 0.2 The data preprocessing approach based on time-series decomposition.

The data preprocessing approach's main requirement is supply chain contexts, which detailed time intervals and component heuristics for the time-series decomposition component. This contribution has already been published [37]; see Appendix 2 for more detail. The output of this component is ASCs information that detected the event of the activities in ASCs, such as a trend of the rainfall, a movement of market prices, a trend in crop yield production, and processed food price. This bunch of events is just a piece of information that still needs humanlike intelligence to analyze, understand, and infer critical knowledge—demand, supply, and the ASCs situation. Then, the next requirement for the framework is the CBNs model that performs as a humanlike brain to transform ASCs information into knowledge for supporting the decision-making process.

3.3 CBNs Model Design and Development

A fundamental task in CBNs design and development is to find causal relationships in disciplines of knowledge. Causal discovery methods can conduct causal relationships through observational data. These methods generate statistical correlations among observations from well-structured, comprehensive, and complete data covering all possible events. However, the supply chain introduces uncertainty and change into the ASCs environment dependent on expertise to analyze the situation rationally. It means that an absolute data-driven approach cannot produce accurate causal-and-effect explanations [35].

Expert-based modeling is an answer for initialing a CBNs model as a gold standard. The gold standard is a concept to model prior knowledge from everyday situations that covers regular events, rare events, and theoretical events that may never have happened practically. This idea can deal with rapid adaptation in the supply chain. However, the CBNs model is based on probability theory that quite impossible for humans to measure concrete statistics for tuning the model's parameter. This research then applied it with the data-driven approach for parameter learning.

CBNs model is constructed with two foundations: qualitative and quantitative models. The qualitative model represents an assumption of knowledge that experts use for reasoning in the domain. The quantitative model represents a mathematical form encoded using the qualitative model.

3.3.1 Qualitative CBNs Model

The qualitative CBNs model is encoded from prior knowledge into a graphical model with random variables as a node of interesting and causal assumptions as an edge among them.

In terms of Thailand's natural rubber supply chain, the futures market controls the demand for the natural rubber products consumed by the automotive and tire industries [17], [38]. The products in that market are rubber sheets locally produced which depend on climatic conditions [39]. Indeed, climatic problems are the leading cause of decreased source production, while the future market influences consumers' preferences and impacts the market demand.

This research employs this information to model the causal assumptions between the random variables. Causal assumptions are constructed from familiar questions in ASCs management, such as:

- *Will crop yield be undersupplied if prolonged rainfall in the monsoon season?*
- *Will demanded quantity in bidding activity drop if there is a downtrend in the future market?*
- *If the market price rises with a low crop yield production, will consumer preference increase?*
- *What are the factors that cause market equilibrium?*

The answer is discovered using an ASCs prior knowledge that depends upon the experts. It consisted of: (1) interviews with three experts and two practitioners from the Central Rubber Market (CRM) in Hat Yai, Songkhla, Thailand; (2) reviews of a CRM database of 5 years provided by the Thai government.

Figure 3.3 shows the full mode of graphical causal assumptions between the random variables, with cause(s) pointing directly to effect(s) (cause(s) → effect(s)). After consulting with experts, some causal assumptions (grey nodes with dash lines) were cut off because lack of data sources, out-of-date information, and inaccurate data problems.

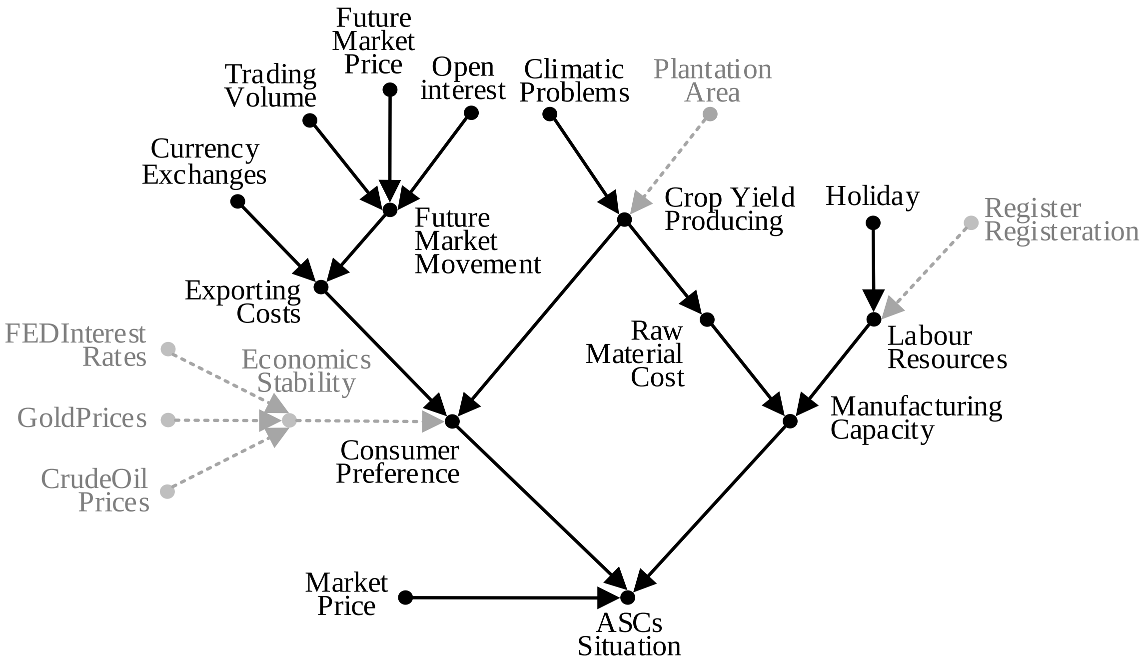


Figure 0.3 Causal assumption of natural rubber SCs using CBNs.

Then, the assumptions are causally structured for explaining the **ASCs Situation** in terms of **Manufacturing Capacity**, **Consumer Preference**, and **Market Price**, and most of them are encoded as collides. For example, **Trading Volume**, **Open Interest**, and **Future Market Price** explain the liquidity and activity of **Future Market Movement**. **Trading Volume** reflects the short-term demanded quantity throughout the trading day, while **Open Interest** shows the number of futures contracts that are still open. **Trading Volume** and **Open Interest** are independent unless **Future Market Movement** is questioned, and then they become causally dependent. In other words, the causes are causally independent of each other, but conditioning on **Crop Yield Production** makes them dependent. Moreover, **Crop Yield Production** affects the behavior of **Raw Material Cost**, which passes its information to **Manufacturing Capacity**. This graphical model can be interpreted into quantitative CBNs model.

3.3.2 Quantitative CBNs Model

The quantitative CBNs model uses a dataset as training data for tuning model parameters. Natural rubber supply chain data for tuning prior and likelihood functions were collected between 2015 and 2019 from the CRM in Hat Yai, Songkhla, Thailand. The data obtained from the training data are summarized in Table 3.1.

Table 0.1 Summarization of natural rubber supply chains data.

Data Sources	Random Variables	States
Climatological Center [40]	**Climatic Problem**	*normal* (48%), *drought* (8%), *monsoon* (13%), *flood* (31%)
Agricultural Production Data [41]	**Labor Resources**	*down* (10%), *stable* (81%), *up* (9%)
	Raw Material Cost	*downtrend* (36%), *sideway* (7%), *uptrend* (47%), *fluctuation* (9%)
Thailand Daily Rubber Price [42]	**Market Price**	*down* (19%), *stable* (69%), *up* (19%)
Bank of Thailand [43]	**Currency Exchanges**	*strengthening* (53%), *stable* (4%), *weakening* (43%)
Markets Insider [44]	**Exporting Costs**	*down* (7%), *stable* (75%), *up* (19%)
Tokyo Commodity Exchange (TOCOM) [45]	**Trading Volume**	*downtrend* (47%), *sideway* (6%), *uptrend* (47%), *fluctuation* (0%)
	Future Market Price	*downtrend* (24%), *sideway* (11%), *uptrend* (34%), *fluctuation* (31%)
	Open Interest	*downtrend* (47%), *sideway* (1%), *uptrend* (51%), *fluctuation* (12%)

Table 3.1 summarizes data in the form of observed random variables. At the same time, contextual variables were retrieved from the CRM database. They are labeled using experts, as shown in Table 3.2.

Table 0.2 Summarization of natural rubber supply chains' contextual variables.

Random Variables	States
Crop Yield Producing	*down* (39%), *stable* (14%), *up* (48%)
Manufacturing Capacity	*low* (17%), *normal* (31%), *high* (52%)
Consumer Preference	*low* (9%), *normal* (52%), *high* (39%)
Future Market Movement	*down* (33%), *stable* (51%), *up* (15%)
ASCs Situation	*equilibrium* (7%), *abnormal-equilibrium* (29%), *shortage* (13%), *abnormal-shortage* (8%), *surplus* (24%), *abnormal-surplus* (20%)

The significant proportion of **Crop Yield Producing** is *up* (48%), which causes **Manufacturing Capacity** to be *high* (52%), which accounts for over half of the dataset. The summarization shows an imbalanced market that reflects the inefficient supply chain.

These data become the priors of the random variables. For example, let *cp* be a set of *m*-possible outcomes of **Climatic Problem** (**CP**), and *P*(**CP**) be the prior for **Climatic Problem**, defined as:

$$P(\mathbf{CP}) = \prod_{i}^{m} P(\mathbf{CP} = cp_i) \tag{3.1}$$

The cp_m is the set of the *m*-possible outcomes of **Climatic Problem.** The probabilities distribution of **CP** is $P(cp_{normal}) = 0.48$, $P(cp_{drought}) = 0.08$, $P(cp_{monsoon}) = 0.13$, and $P(cp_{flood}) = 0.31$.

This research also uses these data to tune the likelihood parameters by using Maximum Likelihood Estimation [46] that functioned using Conditional Probability Distribution (CPD). For example, the causal assumption shows that **Crop Yield Producing** (**CYP**) is affected by **Climatic Problem** (**CP**), defined as:

$$P(\mathbf{CYP}, \mathbf{CP}) = \prod_{i,j}^{m,n} P(\mathbf{CYP} = cyp_{i:m} | \mathbf{CP} = cp_{j:n}) \tag{3.2}$$

The cyp_m is the set of the *m*-possible outcomes of **Crop Yield Producing**, and cp_n is the set of *n*-possible outcomes of **Climatic Problem** that passes their information to **Crop Yield Producing**.

The causal structure represents the scientific assumption that a prior-based process integrates with a data-driven process to produce a gold standard of the CBNs model.

This research employs 10-fold cross-validation to estimate model performance. The validation shows that the proposed possesses good model performance and can be applied to this case study. The design, development, and validation of the proposed CBNs model have been contributed, as explained in [47], according to Appendix 3.

The validation shows that the overall performance is high of 94%. The accuracies of *equilibrium* and *abnormal-shortage* are lower than the others because they are rare events, occurring at around 7% and 8% in the sample proportion, respectively. The *equilibrium* market is ideal and rarely occurs because the market context changes dynamically. Similarly, *abnormal-shortage* means a shortage of supply with decreasing price, which is an extraordinary situation that contradicts the laws of demand and supply. It is also a rare event with a small sample for training the model. In conclusion, our proposed possesses good model performance and can be applied to this case study.

Although k-fold cross-validation shows model performance, it does not yet convince a satisfying performance in explanation ability. A significant advantage of the proposed CBNs model is that it can detect market events using dynamic streaming data and explain the market situation correctly and reasonably to support agricultural market management.

Chapter 4
Results and Discussion

This section aims to show the effectiveness of the research contributions by proving the research question, '*How to analyze streaming data incorporating with CBNs model to detect and explain the ASCs situation?*'. This research divides effectiveness into three perspectives: correctness of real-time data processing, the correctness of the CBNs model in ASCs situation detection, and reasonableness of the CBNs model in ASCs situation explanation.

4.1 Experiments

This research arranged three experiments to answer the research question. Firstly, the streaming data digitization experiment is set to prove an intuitive approach for sensing the streaming ASCs information from digital platforms. Secondly, the predictive performance measurement is proposed to test the CBNs model correctness in ASCs situation detection. Lastly, the sensitivity analysis is adopted to analyze that the CBNs model can encode reasonableness from human-like intelligence for supporting ASCs explanation.

First, data are collected from rubber auction events between 2015 and 2019 from the Central Rubber Market (CRM) in Hat Yai, Songkla, Thailand. This data collection was employed for experiments. It consisted of 111,250 transactions from seven primary sources: (1) the Thai calendar is a holiday and shedding season data, (2) Thai rainfall data [48], (3) fresh latex and reserved auction prices from the Rubber Authority of Thailand [49], (4) supplied the commodity prices, open interests, and trading volumes from the Tokyo Commodity Exchange (TOCOM) [45], (5) currency exchange rates for the Thai Baht/US Dollar (THB) and Japan Yen/US Dollar (JPY) from Bank of Thailand

[43], and (6) crude oil prices, included West Texas Intermediate (WTI) and Brent, from Markets Insider [44], and CRM database of 5 years provided by the Thai government. The data was preprocessed and divided into two subsets. This experiment randomly split the dataset into two subsets. The first subset is for model training and validation, and the second is for testing. The data splitting method was performed using the scikit-learn Python library [50].

4.1.1 Streaming Data Digitization

Experiment Objective

The objective of this experiment was to examine the performance of the streaming data digitization based on time-series decomposition. This method's performance is proven by data significance testing and predictive analytics.

Experiment Setting

The most common form of statistical significance is the correlation coefficient which measures the relationships between independent and dependent variables. This experiment set supply (**Manufacturing Capacity**) and demand (**Consumer Preference**) as dependent variables because these two variables conclude an ASCs situation. At the same time, the rest variables are considered independent variables.

The hypothesis (H1) and the null hypothesis (H0) were defined as follows: H1: the digitized data had a significant correlation for recognizing demand and supply in the market. Furthermore, H0: there is no significant correlation between the digitized data for recognizing demand and supply in the market.

The H1 is proved if (1) the digitized data shows a significant correlation, and (2) it provides good classification results. Four well-known classification algorithms were utilized: Decision Trees (DT), Neural Networks (NN), Support Vector Machines (SVM), and Naïve Bayes (NB) [12]. These algorithms employ correlations and relative odds for the dependent variable outcomes given independent variables. These models' classification results can reflect the digitized data's significance.

Metrics

For significant testing, the multinomial logit model [51], [52]was employed to evaluate the dependent variables against the reference group. This method measures the correlation coefficient in the dataset based on p-values with a significant-alpha and odds ratio (OR) to determine whether the digitized data is significantly different from the null hypothesis. The significance-alpha level for all the statistical tests was set to 0.05, which gives a 5% chance of error rates. If the p-value was less than or equal to this alpha, this experiment rejected the H0. Furthermore, if the OR is 1, then the association between independent variables is deemed insignificant. Otherwise, the association of the independent variables significantly influences the dependent variables. The calculations used the Statsmodels Python library [53] for multinomial logit modeling.

For predictive analytics, The F-measure (F_1) was used to evaluate classification in terms of overall outcome accuracy. The metric is $F_1 = \frac{tp}{tp+\frac{1}{2}(fp+fn)}$. The true positive (*tp*) is a correct outcome from the prediction results, the false positive (*fp*) is an incorrect result, and false negative *(fn*) is an unclassified outcome. The scikit-learn Python library [50] was employed to tune the model's hyperparameters, with training data.

Results

The complete result of this experiment has been attached in Appendix 2, page 57. The overall p-values for supply are mostly highly significant (computed the average of 0.037), and the OR result is acceptable and highly significant (computed the average <0.001). While overall p-values for demand are insignificant (computed at the average of 0.504), OR result shows acceptable significance (computed at the average of 0.223).

For predictive analytics, the highest average scores are with the DT algorithm (0.98 and 0.95 for supply and demand). It appears to be the most compatible with the proposed approach because it uses a logical model that can very successfully handle category-based data. In contrast, NB, a probabilistic model based on Bayes' theorem to compute posterior probabilities, gives the lowest average scores. It may cause by the training data imbalance that generates many false positives. At the same time, NN and SVM algorithms are geometric models employing optimization methods that need to tune the best hyperparameters. The overall accuracy of NN and SVM with default parameter settings are acceptable for this research, but there might be some benefit in adjusting NN's complex layers' parameters for deep learning or SVM's margin to improve accuracy.

The best contribution from this finding is to provide an automatic approach for sensing and digitizing ASCs information from digital platforms for supporting ASCs explanation.

4.1.2 Predictive Performance Measurement

Experiment Objective

The objective of this experiment was to examine the predictive performance of the proposed CBNs model to detect the ASCs situation.

Experiment Setting

The target class is the states of the **ASCs Situation** random variable since it helps to provide an overview of the market and help decision makers manage the supply chain.

This experiment used standard classification algorithms, including Decision Trees (DT), Neural Networks (NN), Support Vector Machines (SVM), and Naïve Bayes (NB) [12] for predictive performance comparison. It was arranged to highlight the CBNs model's predictive ability.

Metrics

The states of **ASCs Situation** were measured based on Precision, Recall, and F-Measure. Precision (PS) is a proportion of the correction of the positive prediction, which is computed as $PS = \frac{TP}{TP+FP}$. Recall (RC) is a proportion of the correction of the prediction, which is computed as $RC = \frac{TP}{TP+FN}$. Lastly, F-Measure (FM) is a balance between Precision and Recall, which is computed as $FM = \frac{2\times Precision\times Recall}{Precision+Recall}$. TP is a true positive prediction, FP is a false positive prediction, and FN is a false negative prediction.

The scikit-learn Python library [50] was employed to tune the model's hyperparameters, with training data.

Results

The complete result of this experiment has been attached in Appendix 3, on page 84. The average results are acceptable for prediction by reaching over 80%. The lowest is 84% from the NB model since the NB has a "naïve" assumption that its features are independent and only dependent on the outcomes that are the paradox assumption to the supply chain since independent features rely on each other. The rest are 93%, 94%, 94%, 93%, and 95% for NN, SVM, DT, BS, and CBNs, respectively. These results are high performances since all models have been trained and validated using well-prepared data. In this way, the models are ready to apply to decision support systems to help decision-makers understand the **ASCs Situation**.

Even though the CBNs model's results are acceptable, it does not show the explanation ability to support decision-making. This research conducted the sensitivity analysis to show how the model will explain the rationale. It is set to highlight this work's best knowledge.

4.1.3 Explanation Measurement

Experiment Objective

The objective of this experiment was to examine whether the causal relationships from the CBNs model express a rational explanation.

Experiment Setting

This experiment selected scenario-based testing for the explanation. It was selected from the ASCs management questions:

1) *Will crop yield be undersupplied if prolonged rainfall in the monsoon season?*
2) *Will demanded quantity in bidding activity drop if there is a downtrend in the future market?*

3) *If the market price rises with a low crop yield production, will consumer preference increase?*
4) *What are the factors that cause market equilibrium?*

It is because these questions asking for the final event that provides high impact on decision-making. Decision-makers need to understand the factors that support their decisions and wish to keep the market stable as long as possible. As a base case, this work used the most sensitive scenario, "**ASCs Situation** is *equilibrium*".

According to the hypothesis, the posterior probabilities of the **questioned variable** should be affected by its **cause(s)** according to the causal assumptions from qualitative model and corresponding with the **evidence(s)**. Then, the assumption is that the base case is sensitive to variations of the states from its relevant cause random variable(s).

The Bayesian Search-based model (BS) was compared with the CBNs model because of its use of conditional dependency of a Bayesian Network [54], which produces relationships based on a score-based structural learning.

Metrics

Sensitivity analysis calculates the posterior probability distributions over the **questioned variable** parameters (each causes random variables' states). It can be calculated as $Sensitivity(\boldsymbol{x}) = \frac{\partial p(x_t|e)}{\partial \boldsymbol{x}}$, $\boldsymbol{x}$ is a target variable, with interest in $\boldsymbol{x} = \boldsymbol{x}_t$ as a base case, and $p(\boldsymbol{x}_t|e)$ posterior distribution of the base case given evidence. The average sensitivity conditioned from all parameters is between *zero* and *one*. The *zero* means that the changes of the **questioned variable**'s causes reduce the chance of a base case, while the *one* influences the **questioned variable** to occur. Sensitivity analysis can measure a miner change of cause parameters sensitive to **questioned variable** 's posteriors (e.g., cause of non-equilibrium). This analysis computes sensitivity between cause and effect in the manner of human-like intelligence based on the uncertainty of the **questioned variable**.

Results

Firstly, the CBNs model and BS model were constructed and tuned parameters with the training data, as shown in Figure 4.1.

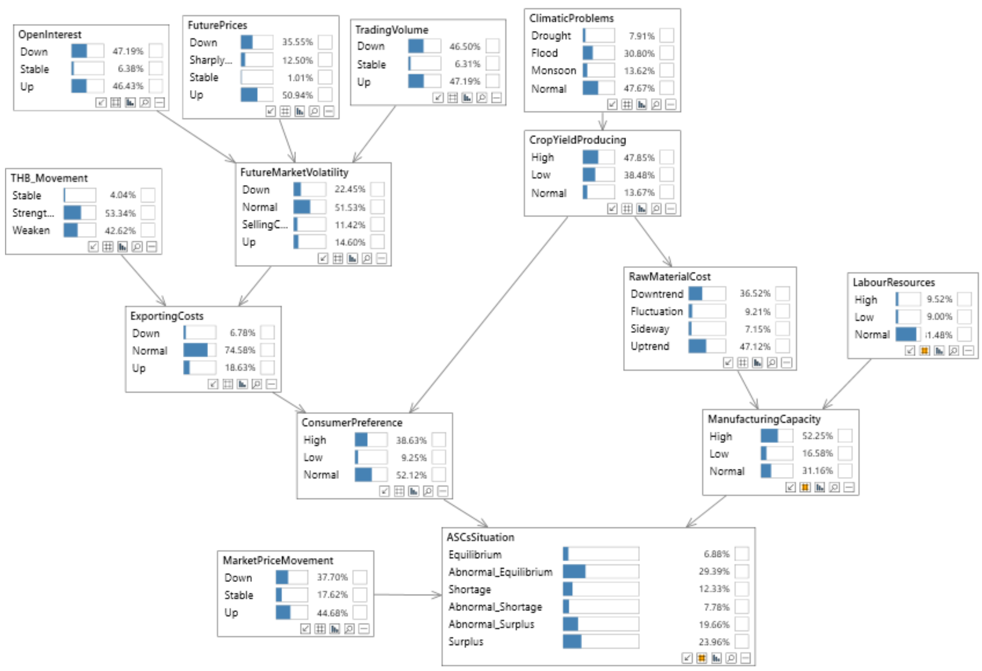


Figure 4.1 The tuned CBNs model.

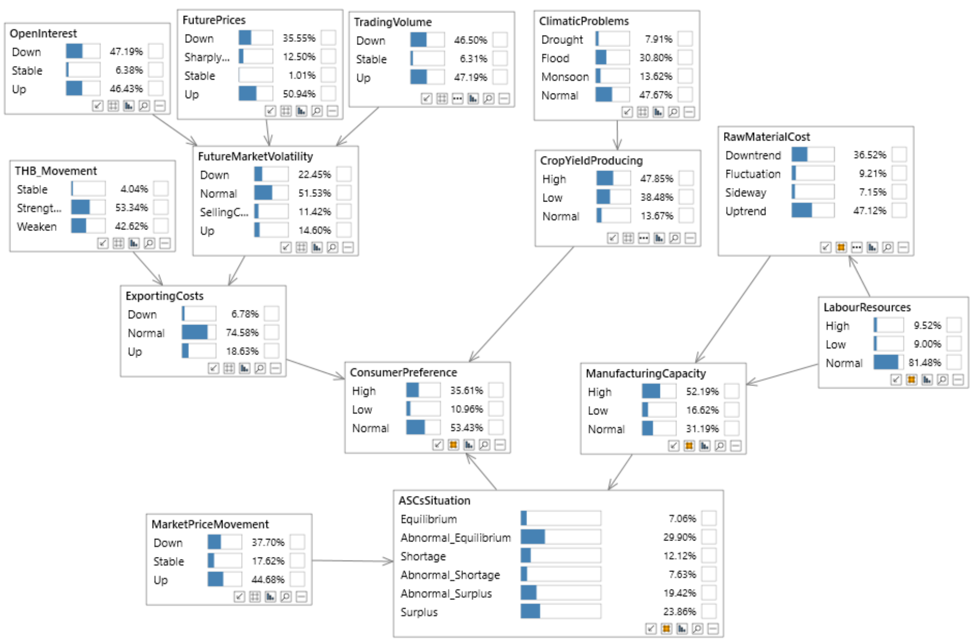


Figure 4.2 The tuned BS model.

Figure 4.1 and 4.2 showed some similarity between the tuned CBNs model and BS model. This is because even the BS's structure learning is score-based method that discovered causal relationship among the data, the data fed for model learning were well collected and prepared. Their differences are showed in Table 4.1.

Table 4.1 The different causal relationship between CBNs model and BS model.

No	Proposed CBNs Model	BS Model
1	Market Price → ASCs Situation Consumer Preference → ASCs Situation Manufacturing Capacity → ASCs Situation	Market Price → ASCs Situation ASCs Situation → Consumer Preference Manufacturing Capacity → ASCs Situation
2	RawMaterial Cost → Manufacturing Capacity Labour Resources → Manufacturing Capacity	Labour Resources → RawMaterial Cost Labour Resources → Manufacturing Capacity RawMaterial Cost → Manufacturing Capacity

Each of the questions was extracted into (1) the final event—questioned random variable, and (2) the evidence —observed random variable(s). These random variables were inferenced by our proposed CBNs model and the BS model. The outputs consist of the posterior distribution of the final event and its causal as the explanation. The comparison of the models' explanation are as follows:

Q1: Will crop yield be undersupplied if prolonged rainfall in the monsoon season?

Final event: **CropYieldProducing** = *low*

Evidence: **ClimaticProblems** = *monsoon*

(a) Proposed CBNs Model (b) BS Model

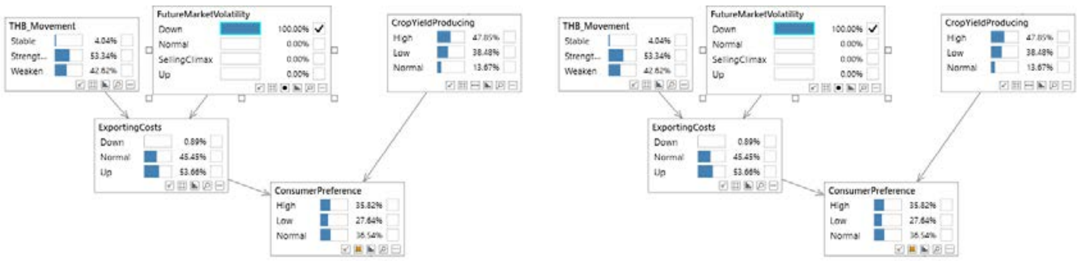
Figure 4.3 The results of the first question.

Figure 4.3 showed that both models answer that there is a 0.45% chance of *low* **CropYieldProducing** given *monsoon* **ClimaticProblems**. This is because they share the same causal assumption that **ClimaticProblems** is an only cause of **CropYieldProducing**.

Q2: Will demanded quantity in bidding activity drop if there is a downtrend in the future market?

Final event: **ConsumerPreference** = *low*

Evidence: **FutureMarketVolatility** = *down*



(a) Proposed CBNs Model (b) BS Model

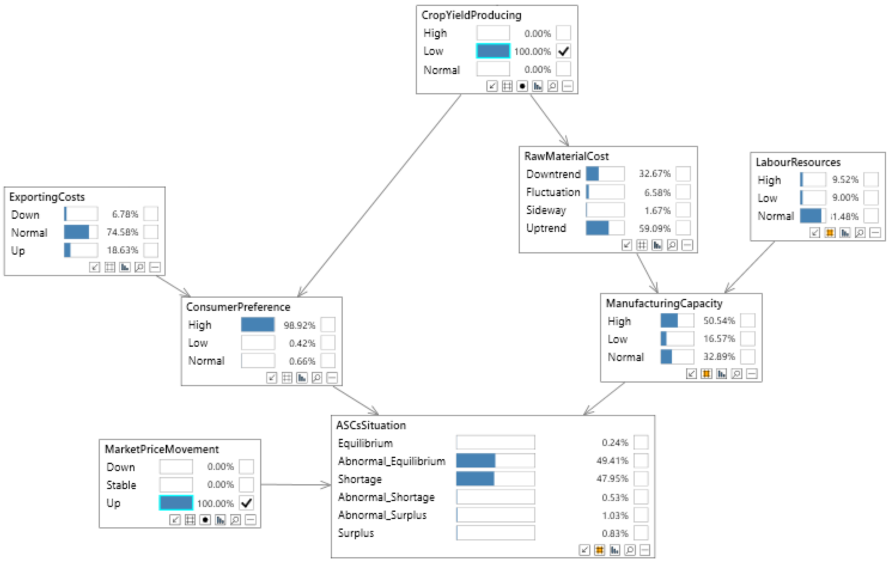
Figure 4.4 The results of the second question.

Figure 4.4 showed that both models also answer that there is a 25.71% chance of *low* **ConsumerPreference** given *down* **FutureMarketVolatility**. This is because they also share the same causal assumption that *low* **FutureMarketVolatility** causes to a 53.66% chance of *up* **ExportingCosts** that passes an effect to *low* **ConsumerPreference**.

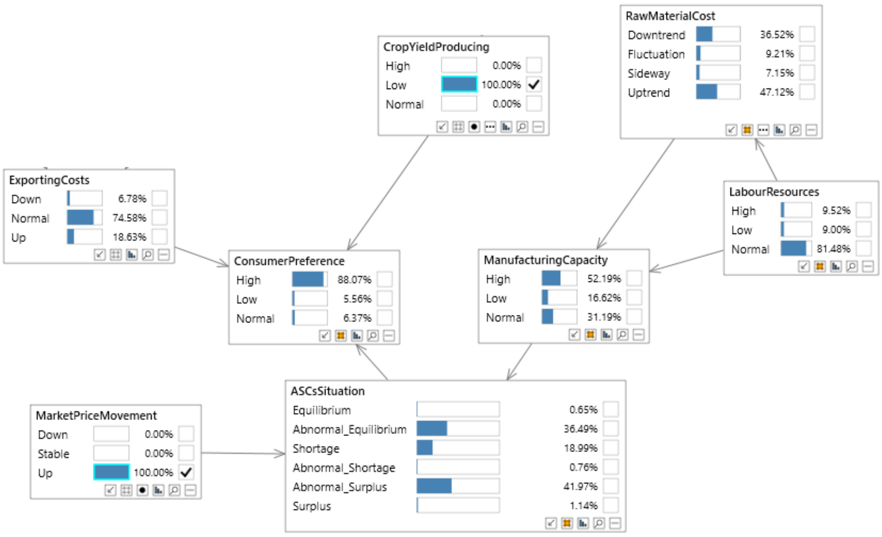
Q3: If the market price rises with a low in crop yield production, will consumer preference increase?

Final event: **ConsumerPreference** = *high*

Evidence: **MarketPrice** = *up*, **CropYieldProducing** = *low*



(a) Proposed CBNs Model



(b) BS Model

Figure 4.5 The results of the third question.

Figure 4.5(a) showed the CBNs model answer that there is a 98.92% chance of *high* **ConsumerPreference** given *up* **MarketPrice** and *low* **CropYieldProducing**. It also explained the *low* **CropYieldProducing** gives a 59.09% chance of **RawMaterialCost** that affects a 50.54% chance of **ManufacturingCapacity**. These information leads to either a 49.41% chance of *abnormal_equilibrium* **ASCsSituation** or a 47.95% chance of *shortage* **ASCsSituation**.

Figure 4.5(b) showed the BS model answer that there is an 88.07% chance of *high* **ConsumerPreference** given *up* **MarketPrice** and *low* **CropYieldProducing**. Despite the CBNs model, BS model has no causal assumption between **CropYieldProducing** and **RawMaterialCost**. Then, these evidence leads to either a 41.97% chance of *surplus* **ASCsSituation** or a 36.49% chance of *abnormal_equilibrium* **ASCsSituation** or an 18.99% chance of *shortage* **ASCsSituation**.

These results show that the models can both explained the situation, but the fact that BS model did not consider causal assumption between **CropYieldProducing** and **RawMaterialCost** makes it performed less reasonably than the CBNs model.

Q4: What are the factors that cause market equilibrium?

Final event: **ASCsSituation** = *equilibrium*

Evidence: -

Despite the previous questions, the fourth question has no evidence. Then, the sensitivity analysis was employed to find the parameter that has influence on this base case. The result of this test has been attached in Appendix 3, page 85. The sensitive degrees for CBNs and BS are 0.069 and 0.071, respectively. One difference between CBNs and BS is the number of random variables affecting the sensitivity of the base case. CBNs and CBNs share the top three sensitive factors, which are **Market Price**, **Manufacturing Capacity**, and **Consumer Preference**.

The first three parameters from the models show that *equilibrium* has converged to *zero*. It means that changes to **Manufacturing Capacity**, **Consumer Preference**, and **Market Price** cause **ASCs Situation** to become unbalanced (¬*equilibrium*, *shortage*, or *surplus*). The posterior distributions of **ASCs Situation** for both CBNs and BS are highly sensitive to **Market Price**. Experts understand that consumer and supplier behaviors are principal factors affecting **ASCs Situation**, so BS and CBNs can help people interpret events using something close to expert reasoning.

While BS has more sensitive factors, **Trading Volume**, and **ASCs Situation**, because the training data may provide high correlations, this difference shows that the number of variables reflects upon resources and processing time. Moreover, this relationship is considered irrational because **Trading Volume** is never used to explain **ASCs Situation** directly. Experts understand that **Trading Volume** is the root cause of **ASCs Situation** that transfers its effect through **Future Market**, **Exporting Costs**, and **Consumer Preference**.

The sensitivity represents how domain experts view environment changes and what they should consider adjusting.

4.2 Discussion

The experiments show that CBNs provide predicted outcomes and relevant parameters to help decision-makers understand the ASCs situation.

The first experiment proved that the proposed streaming data pre-processing could transform sequential data into discretized data. Whether reducing an overloaded data dimension in processing cost, the discretized data also explore important information required for machine learning-based algorithms. Primarily, the proposed time-series decomposition method is based on prior knowledge of the supply chain. It brings vital ASCs information for CBNs modeling since the data representation works consistently with human interpretation.

The second experiment confirms that the CBNs model performs satisfactorily for market situation detection. CBNs can reach an accuracy of around 95%, which works well within traditional supply chain management, where many companies employ experts to examine the probabilities of shortage or surplus. However, small companies lack this expertise, which makes their analysis much more labor-intensive and time-consuming.

The third experiment shows that CBNs offer a new dimension of decision support for the supply chain management. It provides market interpretable explanations based on cause-and-effect, which companies need.

This research can conclude that the CBNs model incorporates prior knowledge for analyzing ASCs situations based on expertise.

Chapter 5
Conclusion

5.1 Research Summary

An imbalance between demand and supply causes an abnormal situation in an agricultural market, especially for the modern market, which goes far beyond changes from traditional ones. Agricultural supply chains (ASCs) management is a foundation for detecting a market situation and supporting policy planning. This duty depends upon the decision-making that requires understanding the supply chain situation thoroughly and rationally. It still depends upon expert people to explore new data and analyze new information for making decisions, which is time-consuming and labor-intensive. This research proposed a machine learning-based ASCs explanation model to discover the supply chain process details. The proposed model is designed and developed using the cause-and-effect assumption represented using Causal Bayesian Networks (CBNs). The CBNs model automatically encodes human-like knowledge to detect the market situation with the ASCs explanation.

Finally, this study arranged experiments to prove the research performance from three perspectives: the streaming data processing correctness, the model predictive correctness, and the model's sensitivity. The results showed that the proposed streaming data processing performs acceptable results that can digitize sequential data into meaningful categorical. The CBNs model has an excellent predictive ability to detect the market situation with a rational and supportive context as an explanation. It shows that the CBNs model offers a new dimension of decision support for the supply chain management. These results proved that this study could answer the research question.

5.2 Claim of Originality

This study answers the research question, '*How to analyze streaming data incorporating with CBNs model to detect and explain the ASCs situation?*'. Therefore, the originality is the approach for ASCs explanation which proposed using the conceptual framework for utilizing big streaming data and applying with the original CBNs model in an automatic system.

5.2.1 The Conceptual Framework

This study proposes the original concept of the ASCs explanation framework for an automatic system which is a tool for guiding fellow researchers in solving their system design. This conceptual framework covers the vision to sense ASCs-related streaming data from various sources for detecting and explaining the market situation. The conceptual framework can be applied to various agricultural markets by adjusting the factors affecting suppliers, consumers, and product lines according to the market characteristic. The adjusted factors reflect the data source and market situation required from this framework. For example, the data sensing component can be applied with data sources that provide ASCs contexts, e.g., sensing technologies such as sensors, open access satellite data, point-of-sale data warehouse, and web API. While the decision-making process can be applied with a decision support system. This component enables farmers, agricultural experts, research workers, or market analyzers to drive real-time operational decisions and reinvent for modern business models. This study contributes the framework for an automatic ASCs explanation, which aims to deal with the time-consuming and labor-intensive work problem.

5.2.2 The Original Model

This study proposes the original CBNs model that was constructed using cause-and-effect vision. It encodes prior knowledge of supply chains into the qualitative model using the causal assumption concept. The qualitative model is a graphical model constructed using random variables as a node of interesting and causal assumptions as an edge among them. Furthermore, this study transforms it into mathematical form as a quantitative model based on the Bayes theorem. It uses a data-driven approach to tune prior and likelihood distribution functions according to probabilistic modeling. The functions support the CBNs model adaptively calculate a posterior for ASCs management question. The CBNs model is the golden standard model ready to be applied to the ASCs management system.

5.3 Future Directions

This thesis has been mainly focused on how to encode human knowledge to construct a gold standard model to answer the research question, which was tested and discussed methodologically. However, the more focused and deeper consideration in the research leaves some ideas outside the scope of the thesis, but they are research gaps and challenges to improve and drive the future ASCs market management. The following ideas could be performed:

The ongoing agricultural market and supply chains glow continuously, modernly, globally, and digitally. The farmers likely develop a marketing strategy to a digital platform that generates enormous data with various new data sources, such as an online market that generates product quantity trends, supplier competition promotions, consumer preferences, and demand trends. It helps data sensing expand opportunities for examining new and exciting market factors to discover new causal assumptions for ASCs management.

Digital marketing creates a significant change in supply chains. It cuts off some activity and makes some shortcuts. For example, it directly connects farmers to consumers, which generates short-term impacts between crop yield processing and consumer preference, while it eliminates impacts on the middle market and traders. These phenomena may create a flaw in the streaming data digitization in Chapter 3.2. This is because the current time-series decomposition approach considers transforming nonstationary time-series data into categorical data according to prior knowledge. Even the categorical distribution represents states of random variables in human-friendly form, ultimately binding to prior knowledge that may abandon some new meaningful information. The alternative approach is a challenge in dealing with nonstationary events.

The causal discovery algorithms are challenging to determine causal relationships in the CBNs model. Besides it reduces labor-intensive and time-consuming tasks in CBNs model learning, it might produce some hidden knowledge beyond the golden standard that might help generate modern, rich, and exciting events for better ASCs management. Furthermore, it could be interesting to consider an automatic causal discovery to learn more inclusive and new knowledge from the enormous data from digital marketing.

References

[1] M. of D. E. and S. National Statistical Office of Thailand, “Labour Force Survey Whole Kingdom Quarter 3: July - September 2021,” 2022.

[2] Thailand Board of Investment, “Thailand: The Kitchen of The World.”

[3] L. Chen, Y. Lu, and R. Zhao, “Analysis and application of modern supply chain system in China,” *Modern Supply Chain Research and Applications*, vol. 1, no. 2, pp. 106–119, May 2019, doi: 10.1108/MSCRA-01-2019-0004.

[4] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin, and N. Khan, “A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction,” *IEEE Access*, vol. 9, pp. 63406–63439, 2021, doi: 10.1109/ACCESS.2021.3075159.

[5] T. Khan, H. H. R. Sherazi, M. Ali, S. Letchmunan, and U. M. Butt, “Deep learning-based growth prediction system: A use case of china agriculture,” *Agronomy*, vol. 11, no. 8, pp. 1–18, 2021, doi: 10.3390/agronomy11081551.

[6] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, “Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt,” *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021, doi: 10.1038/s41598-020-80820-1.

[7] S. Punia, S. P. Singh, and J. K. Madaan, “A cross-temporal hierarchical framework and deep learning for supply chain forecasting,” *Computers and Industrial Engineering*, vol. 149, p. 106796, Nov. 2020, doi: 10.1016/j.cie.2020.106796.

[8] Z. Chen, H. S. Goh, K. L. Sin, K. Lim, N. K. H. Chung, and X. Y. Liew, “Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques,” *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 4, pp. 376–384, 2021, doi: 10.25046/aj060442.

[9] 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 and Operations Research*, vol. 119, p. 104926, Jul. 2020, doi: 10.1016/j.cor.2020.104926.

[10] M. Lezoche, H. Panetto, J. Kacprzyk, J. E. Hernandez, and M. M. E. Alemany Díaz, “Agri-food 4.0: A survey of the Supply Chains and Technologies for the Future Agriculture,” *Computers in Industry*, vol. 117. Elsevier B.V., p. 103187, May 01, 2020. doi: 10.1016/j.compind.2020.103187.

[11] A. Chlingaryan, S. Sukkarieh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” *Computers and Electronics in Agriculture*, vol. 151. Elsevier, pp. 61–69, Aug. 01, 2018. doi: 10.1016/j.compag.2018.05.012.

[12] R. H. L. Ip, L. M. Ang, K. P. Seng, J. C. Broster, and J. E. Pratley, “Big data and machine learning for crop protection,” *Computers and Electronics in Agriculture*, vol. 151, pp. 376–383, Aug. 2018, doi: 10.1016/j.compag.2018.06.008.

[13] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, “Deep learning – Method overview and review of use for fruit detection and yield estimation,” *Computers and Electronics in Agriculture*, vol. 162. Elsevier, pp. 219–234, Jul. 01, 2019. doi: 10.1016/j.compag.2019.04.017.

[14] A. Haghverdi, R. A. Washington-Allen, and B. G. Leib, “Prediction of cotton lint yield from phenology of crop indices using artificial neural networks,” *Computers and Electronics in Agriculture*, vol. 152, pp. 186–197, Sep. 2018, doi: 10.1016/j.compag.2018.07.021.

[15] C. Zhang, L. Di, L. Lin, and L. Guo, “Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps,” *Computers and Electronics in Agriculture*, vol. 166, pp. 1–15, Nov. 2019, doi: 10.1016/j.compag.2019.104989.

[16] A. Akbar, A. Kuanar, J. Patnaik, A. Mishra, and S. Nayak, “Application of Artificial Neural Network modeling for optimization and prediction of essential oil yield in turmeric (Curcuma longa L.),” *Computers and Electronics in Agriculture*, vol. 148, pp. 160–178, May 2018, doi: 10.1016/j.compag.2018.03.002.

[17] S. Arunwarakorn, K. Suthiwartnarueput, and P. Pornchaiwiseskul, “Forecasting equilibrium quantity and price on the world natural rubber market,” *Kasetsart Journal of Social Sciences*, vol. 40, pp. 8–16, Sep. 2017, doi: 10.1016/j.kjss.2017.07.013.

[18] P. S. Maya Gopal and R. Bhargavi, “A novel approach for efficient crop yield prediction,” *Computers and Electronics in Agriculture*, vol. 165, p. 104968, Oct. 2019, doi: 10.1016/j.compag.2019.104968.

[19] G. Yuan, J. Han, L. Zhou, H. Liang, and Y. Zhang, “Supply and demand law under variable information,” *Physica A: Statistical Mechanics and its Applications*, vol. 536, p. 121004, May 2019, doi: 10.1016/j.physa.2019.04.240.

[20] A. Golmohammadi and E. Hassini, “Capacity, pricing and production under supply and demand uncertainties with an application in agriculture,” *European Journal of Operational Research*, vol. 275, no. 3, pp. 1037–1049, Jun. 2019, doi: 10.1016/j.ejor.2018.12.027.

[21] F. F. Bocca and L. H. A. Rodrigues, “The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling,” *Computers and Electronics in Agriculture*, vol. 128, pp. 67–76, Oct. 2016.

[22] S. Stein and H. H. Steinmann, “Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe,” *European Journal of Agronomy*, vol. 92, pp. 30–40, Jan. 2018, doi: 10.1016/j.eja.2017.09.010.

[23] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, “Forecasting price movements using technical indicators: Investigating the impact of varying input window length,” *Neurocomputing*, vol. 264, pp. 71–88, Nov. 2017, doi: 10.1016/j.neucom.2016.11.095.

[24] C. H. Cheng and J. H. Yang, "Fuzzy time-series model based on rough set rule induction for forecasting stock price," *Neurocomputing*, vol. 302, pp. 33–45, Aug. 2018, doi: 10.1016/j.neucom.2018.04.014.

[25] Q. Zhu, F. Zhang, S. Liu, Y. Wu, and L. Wang, "A hybrid VMD–BiGRU model for rubber futures time series forecasting," *Applied Soft Computing Journal*, vol. 84, p. 105739, Sep. 2019, doi: 10.1016/j.asoc.2019.105739.

[26] J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system," *Expert Systems with Applications*, vol. 97, pp. 60–69, May 2018, doi: 10.1016/j.eswa.2017.12.026.

[27] S. Kumar Sharma and S. Sharma, "Developing a Bayesian Network Model for Supply Chain Risk Assessment," *Supply Chain Forum*, vol. 16, no. 4, pp. 50–72, Jan. 2015, doi: 10.1080/16258312.2015.11728693.

[28] M. Chhimwal, S. Agrawal, and G. Kumar, "Measuring circular supply chain risk: A bayesian network methodology," *Sustainability (Switzerland)*, vol. 13, no. 15, pp. 1–22, 2021, doi: 10.3390/su13158448.

[29] J. M. Lawrence, N. U. Ibne Hossain, R. Jaradat, and M. Hamilton, "Leveraging a Bayesian network approach to model and analyze supplier vulnerability to severe weather risk: A case study of the U.S. pharmaceutical supply chain following Hurricane Maria," *International Journal of Disaster Risk Reduction*, vol. 49, no. April, p. 101607, 2020, doi: 10.1016/j.ijdrr.2020.101607.

[30] R. Ojha, A. Ghadge, M. K. Tiwari, and U. S. Bititci, "Bayesian network modelling for supply chain risk propagation," *International Journal of Production Research*, vol. 56, no. 17, pp. 5795–5819, 2018, doi: 10.1080/00207543.2018.1467059.

[31] S. El Amrani *et al.*, "Modelling and assessing sustainability of a supply chain network leveraging multi Echelon Bayesian Network," *Journal of Cleaner Production*, vol. 302, 2021, doi: 10.1016/j.jclepro.2021.126855.

[32] J. Pearl, M. M. Glymour, and N. P. Jewell, *Causal inference in statistics: a primer*, 1st ed., vol. 53, no. 9. John Wiley & Sons, Ltd, 2016. Accessed: Jan. 28, 2021. [Online]. Available: https://books.google.co.th/books?hl=en&lr=&id=L3G-CgAAQBAJ&oi=fnd&pg=PR9&ots=CsXjFSsLgt&sig=43Gy8yBsvYr7G6M6kq3quNW_q5A&redir_esc=y#v=onepage&q&f=false

[33] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Commun ACM*, vol. 62, no. 3, pp. 54–60, 2019, doi: 10.1145/3241036.

[34] D. Malinsky and D. Danks, "Causal discovery algorithms: A practical guide," *Philosophy Compass*, vol. 13, no. 1, p. e12470, Jan. 2018, doi: 10.1111/phc3.12470.

[35] X. Shen *et al.*, "Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-59669-x.

[36] K. P. Murphy, *Machine learning: A Probabilistic Perspective*, no. 9783319056029. The MIT Press, 2012.

[37] M. Kliangkhlao and S. Limsiroratana, "Harnessing the power of big data digitization for market factors awareness in supply chain management," *Multimedia Tools and Applications*, 2022, doi: https://doi.org/10.1007/s11042-022-13309-w.

[38] J. Chanchaichujit and J. F. Saavedra-Rosas, "The Elements of the Natural Rubber Industry Supply Chain," in *Using Simulation Tools to Model Renewable Resources*, Cham: Springer International Publishing, 2018, pp. 19–40. doi: 10.1007/978-3-319-55816-5_2.

[39] R. Dejchanchaiwong, A. Kumar, and P. Tekasakul, "Performance and economic analysis of natural convection based rubber smoking room for rubber cooperatives in Thailand," *Renewable Energy*, vol. 132, pp. 233–242, Mar. 2019, doi: 10.1016/J.RENENE.2018.07.145.

[40] Thai Meteorological Department, "Climatological Center," 2021. http://climate.tmd.go.th/gge/index.html (accessed Jul. 28, 2021).

[41] Office of Agricultural Economics, "Agricultural Production Data," 2021. https://www.oae.go.th/ (accessed Jul. 28, 2021).

[42] Office Of The Rubber Replanting Aid Fund, "Thailand Daily Rubber Price," 2021. https://www.raot.co.th/ewtadmin/ewt/rubber_eng/main.php?filename=index (accessed Jul. 28, 2021).

[43] Bank of Thailand, "Daily Foreign Exchange Rate," 2021. https://www.bot.or.th/english/_layouts/application/exchangerate/exchangerate.aspx (accessed Jul. 28, 2021).

[44] Markets Insider, "Crude Oil Price Today | BRENT OIL PRICE CHART," 2021. https://markets.businessinsider.com/commodities/oil-price (accessed Jul. 28, 2021).

[45] Tokyo Commodity Exchange Inc., "TOCOM : Quotes for RSS3 Rubber," 2021. https://www.tocom.or.jp/market/kobetu/rubber.html (accessed Jul. 28, 2021).

[46] D. Barber, *Bayesian Reasoning and Machine Learning*, 1 st. United Kingdom: Cambridge University Press, 2011. doi: 10.1017/cbo9780511804779.

[47] M. Kliangkhlao, S. Limsiroratana, and B. Sahoh, "The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains," *IEEE Access*, pp. 1–1, 2022, doi: 10.1109/ACCESS.2022.3199353.

[48] Thai Meteorological Department, "Thailand Meteorological Department API : TMDAPI." https://data.tmd.go.th/api/index1.php (accessed May 31, 2022).

[49] Rubber Authority of Thailand, "Rubber Price - Rubber Authority of Thailand," 2019. http://www.raot.co.th/rubber2012/menu5.php (accessed Mar. 16, 2020).

[50] Scikit-learn, “scikit-learn: Machine Learning in Python,” 2022. https://scikit-learn.org/stable/ (accessed May 28, 2022).

[51] S. Sperandei, “Understanding logistic regression analysis,” *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014, doi: 10.11613/BM.2014.003.

[52] D. Reynard and M. Shirgaokar, “Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?,” *Transportation Research Part D: Transport and Environment*, vol. 77, pp. 449–463, Dec. 2019, doi: 10.1016/j.trd.2019.03.002.

[53] J. Perktold, S. Seabold, and J. Taylor, “Statsmodels.” https://www.statsmodels.org/stable/index.html (accessed Aug. 20, 2020).

[54] B. Sahoh and A. Choksuriwong, “A proof-of-concept and feasibility analysis of using social sensors in the context of causal machine learning-based emergency management,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, p. 3, May 2021, doi: 10.1007/s12652-021-03317-3.

[55] J. Pearl, “An introduction to causal inference,” *International Journal of Biostatistics*, vol. 6, no. 2. Berkeley Electronic Press, 2010. doi: 10.2202/1557-4679.1203.

Appendix

Appendix 1

M. Kliangkhlao and S. Limsiroratana, "Towards the idea of agricultural market understanding for automatic event detection.” In *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pp. 81-86. 2019.

Towards the Idea of Agricultural Market Understanding for Automatic Event Detection

Mallika Kliangkhlao
Department of Computer Engineering
Faculty of Engineering,
Prince of Songkla University, Thailand
kliangkhlao.m@gmail.com

Somchai Limsiroratana
Department of Computer Engineering
Faculty of Engineering,
Prince of Songkla University, Thailand
somchai.l@psu.ac.th

ABSTRACT
Agricultural Market (AM) understanding has the main goal to discover knowledge of market situation for decision making in agricultural management. Agricultural Big Data (agri-big data) is the valuable data for that process. With the uncertainty of agri-big data, it needs expert knowledge to understand AM-related effect from the observation to infer the most complete situation. This manual process causes the cost of time-consuming. It is important to consume the well-timed data and generate knowledge for supporting decision maker to make policy in agribusiness. Therefore, the concept of automatic agri-big data processing using Machine Learning for AM understanding is more focused. This paper shows the application of that idea with agricultural market event detection in case study of Natural Rubber (NR) Market in Thailand. The automatic AM understanding and its challenge are discussed.

CCS Concepts
• **Mathematics of computing~Probabilistic inference problems** • Information systems~Information integration

Keywords
Agri-Big Data; Agribusiness; Automatic Agricultural Management System; Agricultural Market Event Detection; Machine Learning.

1. INTRODUCTION

Agribusiness is concerned to understand the effects of agricultural operation and movement in the Agricultural Market (AM) [1]. AM understanding is critical assessment for improving potential of policies and operation in agriculture management [2]. It can provide knowledge of real-time market phenomena and support decision makers. Many studies are focused on demand recognition [3] and supply recognition [4]. However, they were manually conducted from the history data which may outdated and not support the dynamic of the real-world situation. It hardly informs knowledge to support decision making in real-time which is the significant point in agricultural marketing. Hence, this approach requires collection of well-timed data and model to reveal the behavior of agriculture situation.

In the era of Big Data, it is a related and well-timed data which can be found in multisource sensors whether the smart devices, crowdsourcing, open data, and internal data warehouse. This paper focuses on Agricultural Big Data (agri-big data), which relates to geospatial technology, machinery data, production information, weather, and marketplace. Decision maker can use such information for agricultural management. Therefore, it is the beneficial data to provide knowledge for estimating the real-time AM. However, agri-big data inherits the complex characteristics of Big Data which are high in volume, velocity, variety, veracity, and valorization [5]. Particularly, this paper focuses on the uncertainty of data emerging. It needs the expertise of human to infer by fusing and transforming agri-big data into knowledge. However, the manual of agri-big data processing is time-consuming. Therefore, the automatic concept for agri-big data processing is the first prerequisite.

Machine learning is the idea that improves machines' learning ability and turns them to be learners. In this case, machine learning can possibly support the automatic AM understanding goal by learning the behavior of agri-big data. There are studies apply Machine Learning for agricultural field [4], [6]. However, related studies considered analysis model only for recognizing output as a deterministic information. This recognized output still needs experts to infer with their knowledge for understanding the broader AM.

To deal with that limitation, the objective of this study is to propose idea of the agri-big data processing in

ICSCA '19, February 19–21, 2019, Penang, Malaysia

ACM ISBN 978-1-4503-6573-4/19/02...$15.00

agricultural management. This has main benefit to design the overview architecture of the automatic agri-big data processing. This paper is proposed to extend the concept of agri-big data fusion by combining an ability of Machine Learning with expertise knowledge for supporting AM understanding. This is for supporting the decision makers in agribusiness to understand the broader picture of real-world AM situation to support their real-time decision and policy making.

In the remainder, this paper explains the background of AM understanding and agribusiness in Section 2. In Section 3, we show the literature review of agri-big data processing in agricultural management using Machine Learning. The overview architecture of AM understanding with the automatic agri-big data analysis is presented in Section 4. The case study of AM understanding for NR Market Event Detection is detailed in Section 5. Finally, the conclusion of the proposed idea and the direction of future work are detailed in Section 6.

2. BACKGROUND OF AGRICULTURAL MARKET UNDERSTANDING

2.1 Force Analysis with Agri-Big Data

AM understanding is the concept to reveal the conclusion from real-world AM environment using observations from agri-big data. This concept is important for supporting decision-making in agribusiness management. In this paper, it is more focused on the broader AM situation, especially the market stability during the phenomena of anomaly market situation.

In agribusiness, five-force analysis is the agribusiness framework to understand the market factors that shape and drive the AM [7], [8]. This paper adopts this framework to fuse the agri-big data and discover knowledge to summarize each of five factors, including:

(F1) The situation of the agricultural products from sale transaction, and environmental data. This can be the cause of the supply situation in the agribusiness.

(F2) The power of buyers from the buyer behavior, warehouse stocks, and economic status. This shows the demand behavior which can be increasing, decreasing or stable.

(F3) The threat from the price situation of substitute products. This force can infer the shifting in demand (F2) and pricing in both short-term and long-term effects.

(F4) The product distribution cost, such as currency exchange, logistic cost, effect of government policy and regulations. This affects to shift supply (F1) which leads to the dynamic pricing in the AM.

(F5) The external drivers, such as future market price, and holiday. This has the relationship with (F2) by altering the buyer's decision-making. Moreover, it has effect on (F1) for supplier's decision in production rate.

In AM understanding, the AM situation depends on the relationship between those five factors. When decision makers observed the AM-related data, they will use them to recognize the five forces information for support AM understanding. Examples of the five-force recognition is detailed in Table 1.

Table 1. Five-force Recognition from Agri-big Data Example

Five Forces	Agri-Big Data Sources	Observed Data	Recognized Information
(F1)	Environmental Data Service	Flooding in crop production area	Negative effect to the crop production
(F2)	Product Trading Data Warehouse	Increasing of product consuming	Demand is increasing
(F3)	Commodity Exchange Market	Substitute products price is rising	Positive effect to demand in main production
(F4)	Crude Oil Market, and Petroleum Price Service in Thailand	Petroleum price in Thailand keep rising for a month	Negative to the logistic in product distribution cost
(F5)	Crop Exporting Trading Prices	Rising of crop prices	Negative effect to demand

From Table 1, the agri-big data is observed in form of raw observation according to the data generation from multiple sources. It will be recognized and transformed into the recognized five forces information. Furthermore, the interrelationship between the force information are the significance that needed to be understood using expertise knowledge to reveal the AM situation. In this case, decision makers can infer the anomaly AM event which is the surplus of demand. This is caused from supply decreasing from (F1) and consumption rising in (F2). Although, (F4) and (F5) shows the causes in price rising that can shift demand, (F3) shows the limit choice of buyers. Then, they can decide the AM management policy which may increase the supply power to balance the AM again, such as expanding the central price ceiling.

2.2 Agri-Big Data Process for Agricultural Management

From the concept of five-force recognition, it is the requirement to understand AM for the decision-making in agricultural management. This study separates the agri-big data process into four stages, including agri-big data discovering, agri-big data fusion, AM understanding, and agricultural decision-making, as shown in Figure 1.

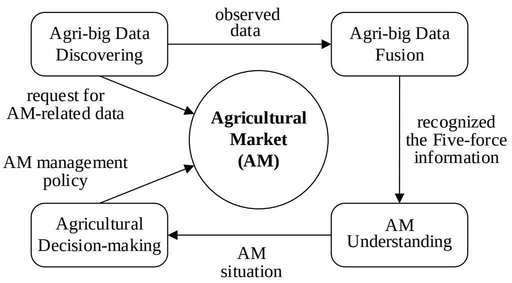


Figure 1. Agri-big Data Process for Agricultural Management

From Figure 1, agri-big data processing is for AM monitoring in agricultural management. Agri-big data discovering is used to request and consume the AM-related data which is generated in that time from multi-sources. Secondly, data fusion is the process that decision-makers need to transform data, remove irrelevant data, recognize the five forces using the relationship between observed data. Third, AM understanding is the process to discover knowledge about AM situation by inferring the relationship between recognized five forces using expertise knowledge. The output of the AM understanding can be used to support agricultural decision-making in the last process.

In addition, decision makers need to keep in touch with the real-world situation to monitor and understand AM situation for agribusiness management effectively. However, it is time-consuming and hard-working behind the agri-big data processing for human. In addition, it needs the concept of automatic approach to reach the requirement about automatic agricultural management system. Especially, an automatic approach to fuse the agri-big data and recognize it into factors for supporting AM understanding process is needed.

3. AGRI-BIG DATA PROCESSING WITH MACHINE LEARNING

3.1 Machine Learning in Agricultural Management

Innovation technology is required to deal with the complex and variety of agri-big data analytics in agricultural management. Machine Learning is the well-established method for learning the behavior of data which has main ability for pattern recognition [9]. In the point of agri-big data processing, the related studies of agri-big data using machine learning is showed in Table 2.

Table 2. Agricultural-related Data Analysis Studies based on Machine Learning

Five Forces	Objective	Agri-big Data Characteristic	Data Sources	Techniques Analysis
(F1)	Planting Analysis	Static and Multivariate	Multiple Sources	U (ARIMA and SVM)
(F2)	Product Positioning, Consumer Satisfaction Analysis	Static and Multivariate	Single Source	U (K-means) and S (What-if Analysis, SVM)
(F3)	Equilibrium Quantity and Price Forecasting	Time-series, and Multivariate	Multiple Sources	S (Simultaneous Equations)
(F4)	Tactical Supply Planning	Static and Multivariate	Single Source	S (MINLP)
(F5)	Dynamic Dependence Analysis	Time-series, and Multivariate	Multiple Sources	S (Wavelet and Copula)

* 'U' means unsupervised learning, 'S' means supervised learning.

From Table 2, the related studies are detailed in the dimensions of five forces consideration, objective, data characteristic, data sources, and techniques. Wen et al. proposed the (F1) analysis using the combination of ARIMA and SVM to describe the linear relationship between time-series history data which fused from environmental sensors [10]. For (F2) consideration, clustering and what-if analysis is proposed to recognize the pattern of consumer's demand from multiple sources [11]. On the other hand, consumer satisfaction analysis was proposed using social data for demand factor recognition that also related to (F2) [6], [12]. Moreover, there are studies proposed to recognize (F3) and (F4) for supporting supply planning using supervised learning [13], [14]. For (F5), there is study focused on the dynamic interdependence among the financial and agricultural markets over time [15]. Therefore, Machine Learning is the important approach to apply with agri-big data processing for supporting in agricultural management field.

3.2 Interdependence between Five Forces for Data Fusion

From the related papers in previous section, those studies showed the agri-big data utilization for supporting agribusiness using Machine Learning. However, they focused only one factor in their works which not support to understand broader AM situation. From this paper's goal, the AM understanding is depended on the interdependence between five-force factors. This is because the knowledge of interdependence between factors can be used to infer the AM situation when any factor is lost according to the uncertainty of agi-big data emerging. Tai et al. proposed the idea of big data fusion and inference using Probability Distribution to understand the multi-data [16]. However, they represented the interdependence in association rules which is static and not support the dynamic in real-world situation. Causal Inference is a significant fundamental in Machine Learning to identify the causalities of multivariate data [17]. In addition, this paper can extend that ideas to understand the causal and dependent relationship among that five-force factors. This idea can be supported by the casual inference model which calculated from Probability Distribution among the factors from agri-big data. Therefore, this paper needs to focus on the agri-big data analyzing from multiple data sources for supporting AM understanding.

4. AM UNDERSTANDING FOR AM EVENT DETECTION ARCHITECTURE

In this section, this paper presents the overview architecture of AM understanding for AM event detection. This overview architecture is based on the concept of manual agri-big data processing for agricultural management which shown in Figure 1. This idea is focused on the automatic process for handling with agri-big data that originated from multiple sources simultaneously. Especially, Data Fusion and AM Understanding are focused because these two processes can be applied with the expertise knowledge of human as an initial model for supporting the whole model learning and adapting to the uncertainty of agri-big data. From that idea, the proposed overview is separated into two main parts which are environment part and agent part, as shown in Figure 2.

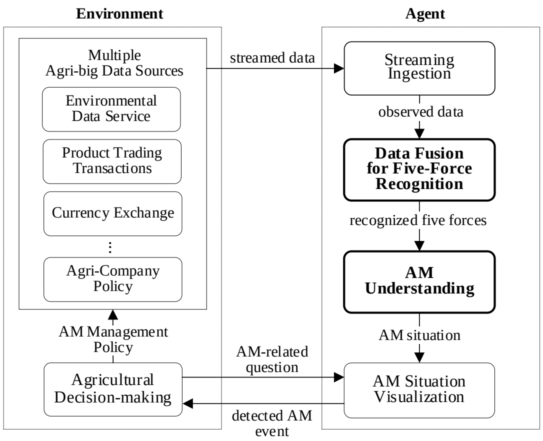


Figure 2. The Overview of AM understanding for AM Event Detection Architecture

From Figure 2, agent is the actor of the automatic system that observe the agri-big data from multiple data sources from environment to process and act reasonably. The multiple agri-big data sources were chosen based on their generated data that important for five-force recognition, such as the environmental data service, and product trading transaction from data warehouse. Therefore, agent has the first job to perceive the streamed agri-big data from multi-sources, which called streaming ingestion. The observed data will be sent to the data fusion for five-force recognition. In addition, this process needs the approach to fuse the observed data and transform it into the significant five forces which is the requirement for AM Understanding. AM Understanding will act like an expert to discover knowledge from fused data to rationally summarize the current AM situation. Therefore, this process will use that knowledge to predict the possibly forward event according to AM-related question from decision makers in agricultural decision-making process. Then, agent will give the feedback via AM situation visualization application. The decision makers, which is the part of environment, can adopt the detected AM event to support their decision and create policy to manage their company policy. This overview architecture can be applied to the AM event detection system for AM management.

From literature review in section 3.1, Machine Learning can take the role in Data Fusion using supervised and/or unsupervised techniques to recognize the pattern of data from past observation. The specific technique will be discovered by the pre-modeling or training process according to the characteristics of each data. Then, AM event detection system will understand AM and detect the AM situation perfectly by recognizing complete five forces from the comprehensive observation. However, this case is rarely happened according to the uncertain characteristic of agri-big data. Each data source is independent from each other. It generates data in different time, format and frequency. Therefore, when decision-maker questions to the AM event detection system, the observation may be incomplete and cannot be recognized to all five forces. For example according to the observation from Table 1, if the observations are consisting of (F1) and (F2), then decision makers can recognize only two factors out of five. They still can use their expertise knowledge to reveal the AM situation.

AM Understanding needs the ability of Machine Learning to teach it the background knowledge of inference. Causal Inference and Probability Distribution is proposed to learn the dependency between the five forces from past observation and represented into causal model. That probability dependency can be used to design the causal model to select the relevant data and eliminate the irrelevant one. This can represent the background knowledge to make AM understanding process deal with the lack of five forces information. Therefore, it is the base of data fusion to integrate required agri-big data to answer the question as an AM understanding.

5. AM UNDERSTANDING IN NATURAL RUBBER MARKET EVENT DETECTION

In this section, we apply the idea of AM Understanding to detect the market situation of Natural Rubber (NR) Market in Thailand.

Event in NR market is the occurrences of rubber situation that have abnormal behaviors and possibly affect the future situation, such as the suddenly falling of rubber trading prices in Tokyo market that has high probability effect to the local trading prices, or the event of monsoon season with heavy raining in the south of Thailand that may lead to the event of shortage supply. That event mainly reflects the uncertainty of demand and supply which experts use to timely estimate the next state of the situation for planning the strategy. Consequently, the NR market event detection is the needed to monitor and analyze the related factors for predicting the next state of the rubber event.

For AM Understanding application, we firstly model the causal knowledge to represent the causation between variables according to five-force analysis for NR market. This model is the background knowledge for Data Fusion to give the relation to the observed data from agri-big data. In this section, we example the simplest case of NR market event understanding. The process of data analysis is based on the overview of AM event detection architecture which showed in Figure 2. In addition, the AM understanding process is divided into two parts which are recognition and prediction processes. Therefore, we briefly show the AM Understanding example, which detailed in Figure 3.

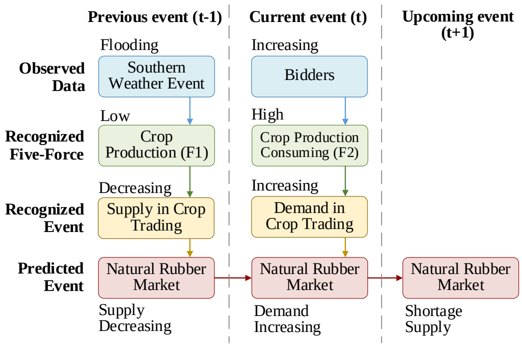


Figure 3. AM Event Understanding in case of NR Market in Thailand

According to Figure 3, the detail diagram of AM event understanding for NR market is separated into four rows and three columns. In recognition process, rows have three layers, which are ‘Observed data’ that is the raw data from data ingestions, ‘Recognized Five-Force’ is the output of Data Fusion for five-force recognition. ‘Recognized Event’ is the current AM event that recognized from observations. The prediction process is represented in row ‘Predicted Event’ which is the predicted output by inferencing from previous event and recognized event in current time. Columns are the considered times, which are ‘previous event (t-1)’ is the previous event detection, ‘current event (t)’ is the processing in current time, and ‘upcoming event (t+1)’ is the event that possibly emerge from the sequence event.

In example, we observed ‘Flooding in the southern of Thailand’ as an observation in the time (t-1) from Thailand Meteorological Service. In Data Fusion, model can recognize that (F1)-related information to ‘Low rate of Crop Production’. This is because from the background of causal knowledge the flooding in southern of Thailand, which is the main NR producing, can decrease the ability of rubber harvesting. This model will recognize event of ‘Decreasing of Supply in Crop Trading’ which used to predict ‘Supply Decreasing event in NR Market’. After that, we observe the ‘Increasing of Bidders in the Central Rubber Market Bidding’ from market data warehouse in the current time (t). Model can recognize that it will affect to (F2)-related information as 'High rate of Crop Production Consuming'. With the causal knowledge, that recognized event has relation to the ‘Increasing of Demand in Crop Trading’ which used to predict ‘Demand Increasing event in NR Market’.

In AM understanding, the current event also has effect from previous event. The model will find that previous predicted event and current predicted event are leaded to the sequence of events. The event understanding model will analyze the chain of ‘Supply Decreasing event’ and ‘Demand Increasing event’ to the ‘Shortage Supply event’ as a summarization. From that output, rubber market manager can use that summarization for making decision or planning the strategy to deal with that event.

From that idea, we can see that the chain of event is varying and dynamic to the changing behaviors of data-especially, the random variables that observed from streaming and multiple data sources. This is totally challenging in present study to make model understand AM and discover the forward event that can answer question dynamically and rationally. In this paper, the challenge is the approach on the data fusion to support AM understanding for dealing with the uncertainty of AM.

In consequent, causal inference model can be structured the sequence of events that shows the dynamic behaviors of time-series events which can be used to summarize the predicted event by discover the possible transition among sequentially recognized events. Therefore, it needs the causal model with probability distribution to fuse the incoming data by understanding the relationship between them. This concept is the main challenge in this research to study and apply for AM event understanding to dealing with the uncertainty of event from real-time data.

6. CONCLUSIONS

AM understanding is the basic and required process in agricultural management, especially for the decision-making in industry level. With the data-driven industrial, decision makers can discover the knowledge from agri-big data which up-to-date with real-world situation. This paper presents the idea for AM understanding using agri-big data with machine learning. The literature review shows the related work focused only on the AM factor recognition which not support to summarize the AM situation. Therefore, this paper proposes the idea of automatic AM understanding using agri-big data for agricultural management.

Although, machine learning is proposed to fulfill the goal of automation, the exact technique for data fusion process is out of our scope. This is because from the literature review, technique is chosen depending on the character and behavior of data. In addition, we more focus on how to infer the relationship between five-force which have uncertainty behavior according to AM environment. Therefore, the challenge from this idea is AM understanding which needs the intuitive experience from the expertise knowledge including the ability of machine learning to calculate the dependent and independent relationship between knowledge. From that idea, it needs the inferencing model to learn the dependency between the observation. The causal model can be construct by the concept of Probability Distribution to prove the dependence and independence between the data nodes. In addition, the entire structure of the causal knowledge model may be needed to estimate, which is still the big challenge in this work.

The limitation of causal model is it is the knowledge-based model that need the expertise knowledge and training set for measuring the conditional dependency in the initial model. Therefore, it still needs the vision that lead model for adapting and learning the knowledge according to the dynamic and changeable behavior of agri-big data in model environment. In addition, causal is the initial stage for supporting the idea of extensible and reusable knowledge. This is because causal model is represented in graph-based structure which can be extended by connecting to another graph. Hence, causal knowledge model will be the next generation of knowledge modeling which support knowledge sharing including the dynamic knowledge according to the new discovery. Finally, the challenge from this paper is how to define and model the AM understanding which can be used to summarize the AM situations. In future work, we will more focus on the dynamic learning model to deal with the uncertainly of agri-big data.

7. REFERENCES

[1] K. L. P. Ho, C. N. Nguyen, R. Adhikari, M. P. Miles, and L. Bonney, “Exploring market orientation, innovation, and financial performance in agricultural value chains in emerging economies,” Journal of Innovation & Knowledge, vol. 3, no. 3, pp. 154–163, Sep. 2018.

[2] E. Castellari, D. Moro, S. Platoni, and P. Sckokai, “Retailers’ strategies and food price dynamics: Evidence from dairy scanner data,” Food Policy, vol. 74, pp. 212–224, 2018.

[3] Q. Shen, J. Zhang, Y. Hou, J. Yu, and J. Hu, “Quality control of the agricultural products supply chain based on ‘Internet +,’” Information Processing in Agriculture, vol. 5, no. 3, pp. 394–400, Sep. 2018.

[4] J. Oh and B. Jeong, “Tactical supply planning in smart manufacturing supply chain,” Robotics and Computer-Integrated Manufacturing, Apr. 2018.

[5] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, “A review on the practice of big data analysis in agriculture,” Computers and Electronics in Agriculture, vol. 143, pp. 23–37, Dec. 2017.

[6] A. Singh, N. Shukla, and N. Mishra, “Social media data analytics to improve supply chain management in food industries,” Transportation Research Part E: Logistics and Transportation Review, vol. 114, pp. 398–415, Jun. 2018.

[7] M. A. Gunderson, M. D. Boehlje, M. F. Neves, and S. T. Sonka, “Agribusiness Organization and Management,” Encyclopedia of Agriculture and Food Systems, pp. 51–70, Jan. 2014.

[8] G. Behzadi, M. J. O’Sullivan, T. L. Olsen, and A. Zhang, “Agribusiness supply chain risk management: A review of quantitative decision models,” Omega, vol. 79, pp. 21–42, Sep. 2018.

[9] M. K. Saggi and S. Jain, “A survey towards an integration of big data analytics to big insights for value-creation,” Information Processing & Management, vol. 54, no. 5, pp. 758–790, Sep. 2018.

[10] Q. Wen, Y. Wang, H. Zhang, and Z. Li, “Application of ARIMA and SVM mixed model in agricultural management under the background of intellectual agriculture,” Cluster Computing, pp. 1–10, Mar. 2018.

[11] N. Lei and S. K. Moon, “A Decision Support System for market-driven product positioning and design,” Decision Support Systems, vol. 69, pp. 82–91, Jan. 2015.

[12] L. Nkurunziza et al., “Understanding effects of multiple farm management practices on barley performance,” European Journal of Agronomy, vol. 90, pp. 43–52, Oct. 2017.

[13] S. Arunwarakorn, K. Suthiwartnarueput, and P. Pornchaiwiseskul, “Forecasting equilibrium quantity and price on the world natural rubber market,” Kasetsart Journal of Social Sciences, Sep. 2017.

[14] F. T. Ramos et al., “Defining Management Zones Based On Soil Attributes And Soybean Productivity,” Revista Caatinga, vol. 30, no. 2, pp. 427–436, Jun. 2017.

[15] Y. Jiang, J. Lao, B. Mo, and H. Nie, “Dynamic linkages among global oil market, agricultural raw material markets and metal markets: An application of wavelet and copula approaches,” Physica A: Statistical Mechanics and its Applications, vol. 508, pp. 265–279, Oct. 2018.

[16] C. H. Tai, C. T. Chang, and Y. S. Chang, “Hybrid knowledge fusion and inference on cloud environment,” Future Generation Computer Systems, vol. 87, pp. 568–579, Oct. 2018.

[17] S. Du, G. Song, and H. Hong, “Collective causal inference with lag estimation,” Neurocomputing, vol. 323, pp. 299–310, Jan. 2019.

Appendix 2

M. Kliangkhlao and S. Limsiroratana, “Harnessing the power of big data digitization for market factors awareness in supply chain management,” *Multimedia Tools and Applications*, 2022, doi: https://doi.org/10.1007/s11042-022-13309-w.

Harnessing the power of big data digitization for market factors awareness in supply chain management

Mallika Kliangkhlao and Somchai Limsiroratana

Received: 23 February 2021 / Revised: 19 May 2021 / Accepted: 30 May 2022

Abstract
An increasing complication due to the rise of dynamic trades and global industry causes a burden in decision-making. There is a need for multi-level perspective factors in supply chain management, such as short-long terms of demand and supply, and their impact on agricultural market dynamics. In this study, Big data is proposed as supply chain open data sensors for data digitization to deal with the problem. Although Big data supports comprehensive, real-time sources, and provides information about market functions, traditional machine learning technologies have proved insufficient for dealing with Big data characteristics. We then propose a time-series decomposition approach for extracting contexts about short-long term impacts to provide insights into Big data for determining market demand and supply. Our agri-big data digitization reveals the significant information about Big data with the better predictive ability and can support agri-big data analysis using any kind of machine learning model.

Keywords Time-series Decomposition . Demand and supply . Open Data Sensors . Machine learning . Statistical significance . Big Data Analysis

1 Introduction

New technologies in the agri-food supply chain, such as the Internet of Things, farming robots, and the agricultural commodity markets, have modified its behavior, by emphasizing dynamic trades and global marketing [41, 44]. These phenomena generate agricultural big data (agri-big data) which offer key information for manufacturing the supply chain's balance in agri-market management as it expands to cover worldwide agricultural industries [21, 48].

The agri-food supply chain can become uncertain because the collaboration along the supply chain involves complex factors such as an uncontrolled environment [47]. This can unbalance demand and supply, and when it becomes critical (e.g., by causing under or oversupply and demand), then the damage can be severe. Agri-market management is needed for planning effective strategies for supply chain performance [12, 26].

The decision-making process requires comprehensive and real-time information based on supply chain integrated knowledge synthesized from multiple and heterogeneous sources, to determine the demand and supply for good agri-market management while handling data analytics [20, 24]. The utilization of agri-big data for market management depends upon experts to integrate the data and extract contextual information to offer insights into the agri-market [4, 12], which is both time-consuming and labor-intensive. Our research challenge is the extraction of contextual information from agri-big data to support a machine learning model.

The contextual computing approach for extracting contextual information (e.g., the acquisition, transformation, visualization, and representation of data) from agri-big data is an essential ingredient in decision-making system. Golmohammadi and Hassini captured contex- tual information about farm preferences in the almond industry by using climatic sensors and internal data [14]. Their approach dealt with uncertainty by interpreting the contextual information which enhanced the production quantity predictions. Arunwarakorn et al. [2] and Chen and Wang [8] applied contextual analysis to open data sensors which focused on imbalances in the world natural rubber market for auction investment. The resulting contextual information learnt to understand demand and supply gradually.

Although previous studies have shown the utility of demand and supply recognition based on context activity in the supply chain, decision-makers must still combine that information with the supply chain flow. This introduces time-dependent information about the short-long term impacts in the market which has not been previously considered. Time-dependent information is fundamental to dynamic market business decisions in agri-market management [6, 12], and requires a novel digitization approach for the contextual information based on short-long term impacts from supply chain integration.

Our research utilizes information extraction based on contextual computing, which recognizes the short-long term impacts of demand and supply. The digitization employs time-series decomposition aligned with supply chain flow, with the domestic natural rubber market as a case study. The objectives of our work can be summarized as follow: first, to propose a new framework for agri-market decision making by integrating agri-big data using time-series decomposition for demand and supply. The second is to compare our information extraction with outcomes from a traditional approach using a Natural Rubber market management case study.

Our experiments show the significance of agri-big data decomposition, including significant test and more accurate predictive analytics compared to the traditional approach.

The rest of this paper is organized as follows. In Section 2, we present background on the agri supply chain, big data exploration, and time-series data digitization. Section 3 gives details on our agri-big data digitization based on the supply chain. In Section 4, we describe a natural rubber (NR) market management case study, which digitizes its agri-big data into NR-market contextual information. The experiments presented in Section 5 measure the significance and prediction ability of extracted information determined using well-known machine learning algorithms. Section 6 concludes, along with a discussion of some future directions.

2 Background

Agri-big data can provide valuable insights into how contextual information affects supply chain factors [19, 39]. The supply chain is a series of sequential activities. Each of the activities has a factors context that must be understood by the processing method in its present state. For example, crop yield production can be determined from plantation and harvesting information as direct farming factors. This means that agri-market management requires information from each activity, and knowledge of the supply chain, to discover the consequential effects in both the short and long terms.

Fortunately, a collaboration of big data, open data, and sensor networks provide the necessary sensors to support information discovery in the agricultural domain [41], such as meteorological sensors and the agricultural calendar for farming factors [13, 14, 22, 25], trading commodity data service for crop price, warehousing, and logistics costs [2, 15]. The contextual information must utilize current activity states using observations from data sensors in order to ascertain demand and supply.

Although contextual information for the supply chain has been studied previously [12], it lacks integration between the activities' contexts. This makes the information incomplete, and marketers are left to determine demand and supply. We hope to contribute a short-long term concept that considers time-dependent effects on demand and supply. Time is a key factor affecting the ripeness and deterioration of an agricultural product.

2.1 Short-long term impacts

The relationship between demand and supply is standard law in economics which encodes a supply chain prior knowledge to reveal market price movement [14, 45]. The interrelationships between activities affect each other as short-long-term impacts according to the production cycle and time-lag. Short-term impacts directly affect activities, while long-term impacts are indirect. For example, procurement factors illustrate short-term impacts that are explicit in the auction process. However, hidden contextual information depends upon long-term impacts from farming factors that affect supply in the long-term.

This knowledge can be only discovered only the contextual variables are extracted, and short-long term impacts are the key to revealing such time-dependent information since they cause the demand and supply in the activity, such as in material price movement. In addition, indirect causes are long-term impacts, such as when the harvesting season brings products to market. Time-series in big data produce time-dependent information for demand and supply determination.

2.2 Time-series decomposition

Big data streams coming from multiple sources are hard for software agents to consume [41, 44]. It needs dimensionality reduction to transform high-dimensional data into low-dimensional with essential information [27]. Briefly, dimensionality reduction can be categorized into three techniques: subspace, feature selection, and time-series decomposition. Table 1 details the summary of feature between subspace, feature selection, and time-series decomposition.

Table 1 summarizes the method, objective, and target of subspace, feature selection and time-series decomposition technique. Subspace techniques are suitable for data represented by complex dimensions. This technique has main contribution to lower the multidimensional feature space of input data, such as PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis). Feature selection techniques are statistical filter and wrapper methods to order and select the best features that are tradeoff between accuracy and complexity. It uses accuracy-based method to evaluate the features that can provide highest correlation according to predicted outcomes. Time-series decomposition techniques concern with a change of data according to time movement by identifying the features based on frequency.

Table 1 Summary of the dimensionality reduction techniques

Technique	Method	Objective	Target	Refs
Subspace	Matrix factorization	Redundant variable removal	Multidimensional data	[16, 23]
Feature selection	Filter and wrapper	Irrelevant variable removal	Structured data	[25, 46]
Time-series decomposition	Frequency based interpolation function	Sampling interval discretization	Time-series data	[29, 35, 43]

Data characteristic in supply chain management often represent in time-series and decision- makers employ time-dependent information to understand short-long term impacts.

Typically, the decomposed information is structured into the time-series components: level, movement, trend, and seasonality, as shown in Fig. 1.

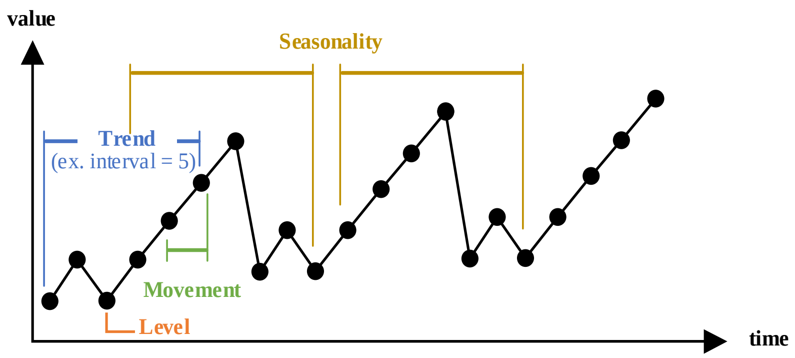


Fig. 1 Time-series components

Level represents a single point value (e.g. hourly, daily, or monthly), movement is the distance between one level and another, and trend is a fixed interval made up of a set of movements that represent semantic meaning and the time interval. The interval is defined by the marketer and decision-maker who control the market movement direction. Lastly, seasonality a long-term scale that represents the repeated pattern of trends that impact decision-maker plans for future directions. These time-series components are essentially time-dependent information for decision-makers, and can also be applied to our short-long term impacts to discover knowledge.

In the last decade, many contributors have used time-series components to extract knowledge from big data. Bocca and Rodrigues [5] and Arunwarakorn et al. [2] propose short-term supply-based approaches using weather, stock, and crop price data. Stein and Steinmann [37] employ annual weather data to analyze long-term supply, but interpret the contexts differently depending on the stakeholders problems. They utilized a simple moving average (SMA) to compute the most likely value, but do not consider agri-market trends required by marketers. Zhang et al. [46] propose a trend-based approach using the random forests technique to recognize the short-term impact on price behavior with the features of up, down, and stable. Zhu et al. [49] integrate variational mode decomposition (VMD) for long-medium-short term price extraction.

Although these studies reveal time-dependent information for decision making, none of them consider contexts obtained from supply chain integration. Instead, they focus on a single activity, overlooking the indirect information from other activities, which will cause the decision-makers to have insufficient information about the supply chain.

3 Overview of agri-big data digitization

Supply chain context information is currently implicit and is hided in agri-big data. It needs prior knowledge to convert that data into the short-long term impacts. Knowledge engineering can be employed to make implicit knowledge explicit [28, 42], and is applied in this study with time-series decomposition to transform agri-big data into time-dependent information. The following sections describe the conceptual framework for supply chain-based agri-big data with the extraction of time-dependent information.

3.1 Supply chain-based agri-big data framework

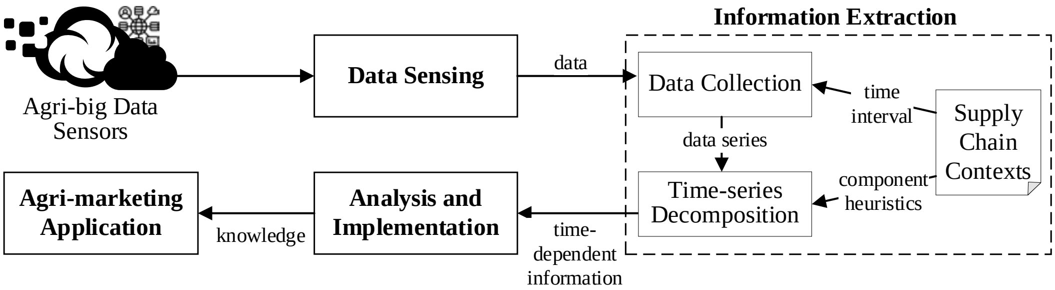


Fig. 2 Supply chain-based agri-big data framework

Figure 2 shows how the framework is made up of four main components: (1) data sensing, (2) information extraction, (3) analysis and implementation, and (4) agri-food marketing application.

Data sensing gets raw data from sensors that is transformed into supply chain information using data collection and time-series decomposition. The aim is to extract time-dependent information based on supply chain contexts, short-long term contexts, by referring to time- series component heuristics and time intervals. The time interval encodes the number of data points used to construct the series during data collection, while the heuristics convert the series into information via time-series decomposition. In this way, the time-dependent information becomes a fundamental ingredient for the machine learning model for supporting agri-food marketing applications such as market monitoring, market analysis, and decision support systems.

Gardas et al. [12] and Schniederjans [33] employed knowledge management with the agricultural supply chain. Their processes used both data and expert knowledge to encode the short-long term knowledge used by supply chain contexts. We use information extraction to transforming tacit knowledge into explicit knowledge as detailed in the next section.

3.2 Time-series decomposition for short-long term impact extraction

Agri-big time-series data is defined as:

Definition 1 Time-series data, D

$$D = \{d_1, d_2, \dots, d_n\}$$

Time-series data in a time interval (*n*) consists of *n* sequential-data point (*d*). For example, time-series components in Fig 1 are representative of this kind of time-dependent information. Our information extraction approach uses time-series decompositions to extract information according to the principal time-series components.

The level component is defined as follows:

Definition 2 The level, L

$$\boldsymbol{L}_i = f(d_{i-n}, \dots, d_{i-1}, d_i)$$

L represents time-series data during a time interval (n). The time interval will vary depending on the agri-food market's supply chain fields, such as hourly climatic data for harvesting [1], or monthly product consumption [2].

The level detection function typically employs the simple moving average technique [5, 35, 37], and is used to support short-term impact decisions (e.g., current raining forces harvesting activities to stop). This affects routine operations in agri-market management and is simple enough that administrators might already be using it.

However, when data levels are detected continuously, they pose a more complex problem, and impact tactical planning. This requires more contextual information to observe changes to activities in the supply chain. This is defined as follows:

Definition 3 The movement, M

$$\boldsymbol{M} \in \{m_1, m_2, \ldots, m_i\}$$

Movement detection explores the meaning of pairs of levels, representing them using states (m_i). This is indirect information, obtained from the level sequence. A heuristic to discover the movement is defined as:

Heuristic 1 Movement is the difference between data levels in the time-series data.

Movement (**M**) differentiates between the current level ($\mathbf{L}_t$) and its previous value ($\mathbf{L}_{t-1}$), and is discretized by considering short-term behavior (e.g. increasing, decreasing, and stable). For example, the current level might compare the second day of rain to the previous day, and be used by decision-makers considering the shortage of agricultural products since infrequent rain may lead to a shortage.

Situations do not recover immediately which requires long-term tactical planning that can explore the behavior of the data over the long term. It is defined as:

Definition 4 The trend, T

$$\boldsymbol{T} \in \{tr_1, tr_2, \ldots, tr_i\}$$

T determines the activity direction uses states (tr_i) that can be linear or non-linear. Linear trend (a straight-line pattern) appears for up-trend, down-trend, and side-ways. Non-linearity represents unstable market phenomena. A heuristic to discover the trend is defined as:

Heuristic 2 Trend is the most frequent movement during a specific period.

For example, the activity movement for a daily rainfall series might look like Fig. 3.

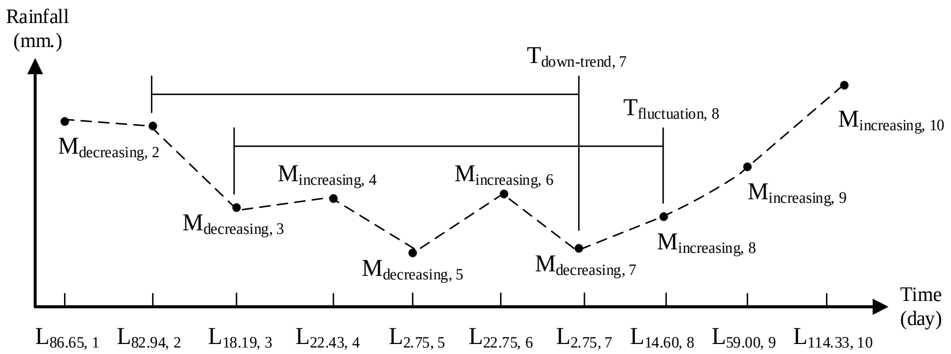


Fig. 3 Trend detection in a rainfall sequence

Figure 3 shows the sequence of daily rainfall ($\mathbf{L}_{state,\ i}$), movement ($\mathbf{M}_{state,\ i}$), and trend ($\mathbf{T}_{state,\ i}$), where *state* is a subset of value, and *i* is a timestamp. The set of levels ($\mathbf{L}_{86.65,\ 1}$ and $\mathbf{L}_{114.33,\ 10}$) are decomposed into the set of movements ($\mathbf{M}_{decreasing,\ 2}$ and $\mathbf{M}_{increasing,\ 10}$), while the rainfall trend during a week can be detected using movements.

In Fig. 3, $\mathbf{M}_{decreasing,\ 2}$ and $\mathbf{M}_{decreasing,\ 7}$ show a decreasing rainfall most often so the trend at $\mathbf{T}_{down\text{-}trend,\ 7}$ is labeled as "down-trend". However, $\mathbf{M}_{decreasing,\ 3}$ and $\mathbf{M}_{increasing,\ 8}$ shows a decreasing rainfall and an increasing rainfall are equally likely, then the trend at $\mathbf{T}_{fluctuation,\ 8}$ can be understood as a fluctuation.

A linear trend helps long-term tactical decision-making, while a non-linear trend detects an unstable situation.

Although a trend provides long-term information for agri-market management, top-level supply chain policy decisions requires different conditions. To this end, we propose a seasonal factor (S), which is defined as:

Definition 5 The seasonality, S

$$\boldsymbol{S} \in \{ss_1, ss_2, \dots, ss_i\}$$

S is the period affecting the administrative decision-makers, and employs trends to determine seasonality states (ss_i) such as shedding or the monsoon season. We propose the heuristic:

Heuristic 3 *A seasonality state is determined by a trend observed during the given period.*

For example, if a set of trends contains repeated down-trends for rainfall and up-trends for temperature, then **S** is implied to be a shedding season, which will help decision-makers deal with upcoming supply issues.

In the next section, natural rubber market management is employed as a case study to show how time-series decomposition is used for supply chain-based agri-big data digitization.

4 Case study: Natural rubber market management

Natural rubber market (NR-market) systems combine activities in the domestic and commodity markets, and the resulting supply chain contexts are utilized by the Thailand Rubber Authority [32] during auctions.

4.1 The supply chain of Thailand's domestic NR-market

Balance between demand and supply reflects an ideal supply chain. We can interpret the balance through product releasing and bidding favor during the Auction activity [7]. However, such activity depends upon another context in supply chain behavior that fills with the short- long term impacts affecting the demand and supply. Then, the contextual factors are the prior knowledge of supply chain as first prerequisite to aware the auction mechanism [11, 34]. The contextual factors in supply chain activities can be presented in sequence-order process, and each one employs both direct and indirect factors that have short-long term impacts. We depict the Thailand's domestic NR-market supply chain activities in Fig. 4.

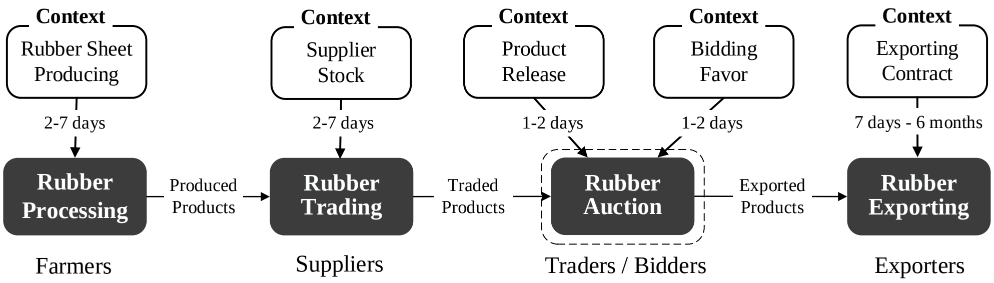


Fig. 4 Thai NR-market supply chain contexts

Each activity in Fig. 4 has associated contexts and time lags. For example, Rubber Processing is defined by two to seven days of *Rubber Sheet Processing*, and its output feeds into Rubber Trading. on the short-term factors for a Rubber Auction are product release, bidding favor, and the traded product, and the long-term factors: rubber sheet production, supplier stock, and the finished product. These direct and indirect factors influence the design of the contexts for the time-series decomposition.

Prior knowledge in supply chain is prerequisite for the proposed framework. It helps decision makers understand market dynamics and respond with situation property. In other words, the Auction activity represents demand and supply that can be employed to model dependent variables while the prior knowledge represents its contextual factors that can be employed to model independent variable.

4.2 Time-series decomposition for NR-market

Supply chain activities utilize a time-lag to decompose time-series data for information extraction. First, we collected rubber auction events data between 2015 and 2019 from the Central Rubber Market (CRM) in Hat Yai, Songkla, Thailand. It consisted of 111,250- transactions from six main sources: (1) the Thai calendar is holiday and shedding season data, (2) Thai rainfall data [9], (3) fresh latex and reserved auction prices from the Rubber Authority of Thailand [32], (4) supplied the commodity prices, open interests, and trading volumes from the Tokyo Commodity Exchange (TOCOM) [40] (5) currency exchange rates for the Thai Baht/US Dollar (THB) and Japan Yen/US Dollar (JPY) from Bank of Thailand [3], and (6) crude oil prices, included West Texas Intermediate (WTI) and Brent, from Markets Insider [17]. We did not consider plantation area factors since they were almost stable for a long-live plant such as the natural rubber tree. The data was decomposed according to short-long term impacts using the time-series components, according to Table 2.

Table 2 Thai NR-market agri-big data decomposition

Supply Chain Activity's Factor	level	movement	trend	seasonality
Rainfall			✓	
Shedding Season				✓
Holiday Boundary				✓
Fresh Latex Price		✓	✓	
Reserved Price	✓	✓		
Commodity Price	✓	✓	✓	
THB Currency Exchange		✓		
JPY Currency Exchange		✓		
Commodity Open Interests			✓	
Commodity Trading Volume			✓	
WTI Crude Oil Price			✓	
Brent Crude Oil Price			✓	

The time-lags employed in Fig. 4 were used to determine suitable time intervals for decomposing the time-series components. For example, Rainfall is considered to have a long-term impact on rubber auction spending of around 7–14 days. Also, the Shedding Season factor is seasonal variables lasting around 7 days and 6 months.

Movement and trend detection, as outlined in Section 3.2, employs a time-series decom- position based on heuristics implemented in Algorithm 1 and Algorithm 2.

Algorithm 1 Movement Detection

Input: the level series $L = \{L_1, L_2, \ldots, L_n\}$
Output: a movement label (mLabel), where mLabel ∈ {up, stable, down}

```
set mLabel = [ ];
for L_i in L do
    set m = 0;
    m = L_i − L_{i−1};
    if (m > 0) then add "up" to mLabel;
    else if (m < 0) then add "down" to mLabel;
    else add "stable" to mLabel;
    end if
end for
return mLabel;
```

Algorithm 2 Trend Detection

Input: the movement series $M = \{M_1, M_2, \ldots, M_n\}$, and the time interval (t)
Output: a trend label (tLabel), where Label ∈ {up-trend, side-ways, down-trend, fluctuation}

```
set tLabel = [ ], t = 7, temp = [ ];
for M_i in M do
    add M_i to temp;
    if (i mod t == 0) then set m = argmax(temp);
        case m of
            "up": add "up − trend" to tLabel;
            "down": add "down − trend" to tLabel;
            "stable": add "side − ways" to tLabel;
        otherwise add "fluctuation" to tLabel;
        set temp = [ ];
        end case
    end if
end for
return tLabel;
```

Algorithm 1 and Algorithm 2 exemplify how to encode the detection concepts, and the other processes are modeled similarly. The outputs are time-dependent information which highlight supply chain contexts, which are fundamental requirements for demand and supply. The collected data was then divided into demand and supply data, in order to understand the data collection behavior from the multiple sources.

Demand and Supply were classified into four-levels: '*low*', '*moderate*', '*high*', and '*very high*', by an expert who is the chief of a CRM market analyzer. Sampled test sets of 25% of the data were utilized by the expert which had been selected by a normality test.

The independent variables were structured according to the time-series decompositions for the contextual information. They were divided into numeric variables (e.g. Level of Reserved Price and Commodity Price), and categorical variable (e.g. Trend of Rainfall and Shedding Season).

4.3 Summary of variables based on demand and supply

A summary of the variables affecting Supply is given in Table 3 and details of the Demand variable are given in Table 4.

Table 3 Summary of variables based on Supply

Variable	Range	%value	Mean	S.D.
Dependent Variable				
Supply	*low*	7.54		
	moderate	24.12		
	high	57.28		
	very high	11.06		
Independence Variables				
7-day Trend of Rainfall (mm./day)	*light*	22.48	1.94	2.67
	moderate	18.48	21.13	7.24
	heavy	16.97	58.37	16.71
	very heavy	42.07	241.74	153.87
14-day Trend of Rainfall (mm./day)	*light*	11.94	3.07	3.04
	moderate	20.22	21.61	6.97
	heavy	19.52	63.18	16.70
	very heavy	48.31	206.79	97.72
Shedding Season	*true*	79.87		
	false	20.13		
Holiday Boundary	*true*	19.00		
	false	81.00		
Fresh Latex Price Movement (THB/k.g.)	*down*	35.45	-1.99	1.79
	stable	17.61	0.00	0.00
	up	46.95	1.53	1.50
7-day Trend of Fresh Latex Price	*down-trend*	33.88		
	side-ways	8.00		
	up-trend	45.65		
	fluctuation	12.47		
Reserved Price Level (THB/k.g.)	40.81–77.25		55.49	6.88
Reserved Price Movement (THB/k.g.)	*down*	21.11	-1.05	0.59
	stable	52.26	0.04	0.28
	up	26.53	1.28	0.80
7-day Trend of Reserved Price	*down-trend*	8.20		
	side-ways	50.77		
	up-trend	20.00		
	fluctuation	21.03		

Table 3 shows that a '*high*' Supply level is the most common which suggests an oversupply situation for the NR-market. Also, Supply is affected by the weather and reserved price. For instance, the 7-day Trend and 14-day Trend of Rainfall variables are mostly '*very heavy*' (42.07% and 48.31%, respectively) which could impact harvesting, and cause a shortage in the long term. The Reserved Price Level average of 55.49 with a wide range (40.81–77.25) and standard deviation (6.88) can be interpreted as meaning that the NR-market price is normally uncertain. These might have a negative impact supply but are inconsistent with the Supply information. In contrast, the decomposed Reserved Price Movement and Trend show that '*stable*' movement, and a '*side-ways*' trend occur more than 50% of the time. This means that they can be employs as patterns to identify a price to support the suppliers' decision to release their stock and therefore trigger an oversupply.

Table 4 Summary of Variables based on Demand

Variable	Range	%value	Mean	S.D.
Dependent Variable				
Demand	*low*	52.26		
	moderate	13.07		
	high	18.59		
	very high	16.08		
Independence Variables				
Shedding Season	*true*	79.87		
	false	20.13		
Holiday Boundary	*true*	19.00		
	false	81.00		
Reserved Price Level (THB/k.g.)	40.81–77.25		55.49	6.88
Reserved Price Movement (THB/k.g.)	*down*	21.11	-1.05	0.59
	stable	52.26	0.04	0.28
	up	26.53	1.28	0.80
7-day Trend of Reserved Price	*down-trend*	8.20		
	side-ways	50.77		
	up-trend	20.00		
	fluctuation	21.03		
Commodity Price Level (JPY/k.g.)	148.58–283.22		179.79	30.44
Commodity Price Movement (JPY/k.g.)	*down*	33.17	-3.79	2.34
	stable	23.12	-0.05	0.56
	up	43.72	4.07	3.01
7-day Trend of Commodity Price	*down-trend*	16.41		
	side-ways	9.23		
	up-trend	37.95		
	fluctuation	36.41		
THB Currency Exchange Rate Movement (THB/US Dollar)	*weaken*	47.74	0.08	0.07
	stable	1.51	0.00	0.00
	strengthen	50.75	-0.08	0.05
JPY Currency Exchange Rate Movement (JPY/ US Dollar)	*weaken*	50.25	0.65	0.53
	stable	1.01	0.00	0.00
	strengthen	48.74	-0.63	0.72
7-day Trend of Commodity Open Interests	*down-trend*	10.77		
	side-ways	9.23		
	up-trend	12.30		
	fluctuation	52.30		
7-day Trend of Commodity Trading Volume	*down-trend*	22.05		
	side-ways	8.72		
	up-trend	27.69		
	fluctuation	41.54		
7-day Trend of WTI Crude Oil Price	*down-trend*	18.46		
	side-ways	26.15		
	up-trend	25.64		
	fluctuation	29.74		
7-day Trend of Brent Crude Oil Price	*down-trend*	17.44		
	side-ways	25.13		
	up-trend	26.67		
	fluctuation	25.64		

Table 4 shows that a '*low*' Demand level is reached 52.26% of the time, which directly affected by external factors such as the unstable Commodity Price Level (range = 148.58– 283.22, mean = 179.79, and S.D. = 30.44). The instability is confirmed how many of the independent variables are usually set to '*fluctuation*' (e.g. 7-day Trend of Commodity Open Interests and Commodity Trading Volume). This situation shows how the '*low*' level of Demand is affected by an unstable commodity market that suppresses purchasing power.

In summary, the demand and supply behaviors in the NR-market are out of balance, with a '*high*' level of Supply and a '*low*' level of Demand; this has long been seen as a major problem for NR-market management. This situation has also been observed in NR-market data obtained from traditional databases [2], but our approach uses real-time multiple heterogeneous sources.

5 Experiment setup

Our objective was to examine the correlations between independent and dependent variables derived from multiple sources in terms of their short-long term impacts on demand and supply. This was done using agri-big data significance testing and predictive analytics.

5.1 Significance testing

The most common form of statistical significance is the correlation coefficient which measures the relationships between independent and dependent variables. The test data from Section 4.2 was divided into two groups: (1) data which had the Supply variable as its dependent variable, and (2) data which had the Demand variable as its dependent variable.

5.1.1 Experiment setting

The hypothesis (H1) and null hypothesis (H0) were defined as follows:

H1 the extracted information had a significant correlation for recognizing demand and supply in the NR-market.

H0 there is no significant correlation between the extracted information for recognizing demand and supply in the NR-market.

The multinomial logit model [31, 36] was employed to evaluate the dependent variables against the reference group. This method measures the correlation coefficient in the dataset as an association between independent and dependent variables based on p-values with a significant-alpha to determine whether the observed data is significantly different from the null hypothesis. The significance-alpha level for all the statistical tests was set to 0.05, which gives a 5% chance of error rates. If the p-value was less than or equal to this alpha, then we rejected the null hypothesis that the result was statistically significant.

An Odds Ratio (OR) [38] was also employed to describe the association between the independent variable levels that affected the dependent variables. If the OR is 1, then the association between independent variables is deemed insignificant, otherwise the association of the independent variables is a significant influence on the dependent variables.

The calculations used the Python Statsmodels library [30] for multinomial logit modeling.

5.1.2 Results and discussions

The classification correlation results for Supply are shown in Table 5, and Demand values appear in Table 6.

Table 5 presents supply as described by the multinomial logit model with Supply's '*low*' level set as the reference group. Many levels in the independent variables have p-values over 0.05 (e.g., 14-day Trend of Rainfall for the '*moderate*' Supply) which labels them as insignificant to supply management. However, the overall p-values for '*moderate*' and '*very high*' Supply are mostly highly significant (i.e. 0.033 and 0.016). This suggests that a single independent variable cannot capture significance to the dependent variable.

Table 5 Summary of the Association between Agri-big Data Variable and Supply

low Supply (Reference Group)	*moderate* Supply			*high* Supply			*very high* Supply		
	Coef.	P	OR	Coef.	P	OR	Coef.	P	OR
7-day Trend of Rainfall	0.893	0.151	2.441	0.766	0.224	2.150	0.750	0.249	2.118
14-day Trend of Rainfall	0.240	0.750	1.271	0.006	0.994	1.006	-0.598	0.442	0.550
Shedding Season	1.282	0.413	3.604	0.461	0.780	1.586	-0.559	0.734	0.572
Holiday Boundary	-6.128	0.004	0.002	-30.11	0.999	0.000	-6.357	0.004	0.002
Fresh Latex Price Movement	1.994	0.034	7.342	2.510	0.009	12.301	2.426	0.011	11.317
7-day Trend of Fresh Latex Price	-0.355	0.273	0.701	-0.334	0.329	0.716	-0.669	0.078	0.512
Reserved Price Level	0.230	0.044	1.258	0.229	0.053	1.257	0.300	0.012	1.350
Reserved Price Movement	2.333	0.062	10.304	2.796	0.026	16.381	1.952	0.126	7.041
7-day Trend of Reserved Price	-0.05	0.887	0.951	-0.114	0.755	0.893	-0.139	0.711	0.870
constant	-12.34	**0.033**	**<0.001**	-11.21	**0.062**	**<0.001**	-14.56	**0.016**	**<0.001**

The bold entries represent random variables in the manuscript

Table 6 Summary of the Association between Agri-big Data Variable and Demand

low Demand (Reference Group)	*moderate* Demand			*high* Demand			*very high* Demand		
	Coef.	P	OR	Coef.	P	OR	Coef.	P	OR
Shedding Season	-0.213	0.814	0.808	-0.262	0.734	0.770	0.998	0.228	2.713
Holiday Boundary	-0.929	0.202	0.395	-1.303	0.037	0.272	-0.706	0.287	0.494
Reserved Price Movement	-0.128	0.171	0.880	-0.111	0.175	0.895	-0.116	0.225	0.891
7-day Trend of Reserved Price	0.132	0.749	1.141	1.562	0.000	4.770	1.720	0.001	5.586
Commodity Price Level	-0.248	0.284	0.781	0.174	0.342	1.190	0.134	0.542	1.143
Commodity Price Movement	0.032	0.097	1.033	0.028	0.107	1.028	0.024	0.246	1.024
7-day Trend of Commodity Price	0.566	0.067	1.761	-0.159	0.594	0.853	0.545	0.123	1.724
THB Currency Rate Movement	0.072	0.777	1.075	-0.331	0.151	0.718	-0.764	0.005	0.466
JPY Currency Rate Movement	-0.217	0.383	0.805	0.120	0.601	1.127	-0.412	0.126	0.662
7-day Trend of Open Interests	0.031	0.868	1.031	0.041	0.812	1.042	0.159	0.396	1.172
7-day Trend of Trading Volume	-0.051	0.740	0.950	0.309	0.041	1.361	0.112	0.514	1.119
7-day Trend of WTI Price	0.215	0.296	1.239	0.187	0.369	1.206	0.297	0.190	1.346
7-day Trend of Brent Price	0.035	0.868	1.035	-0.376	0.089	0.687	-0.183	0.434	0.833
constant	-0.770	**0.769**	**0.463**	-2.347	**0.328**	**0.096**	-2.206	**0.416**	**0.110**

The bold entries represent random variables in the manuscript

Each OR result according to the reference group is acceptable and highly significant ($\neq$ 1) for supply management. However, some features are insignificant, such as the OR result for the 14-day Trend of Rainfall for the '*high*' Supply, which is 1.006. Nevertheless, the constant OR results for '*moderate*', '*high*' and '*very high*' Supply and '*low*' Supply are less than 0.001. This indicates that these independent variables are unique, and clearly describe the supply behavior. Such correlations show the insights that our time-series decomposition approach can offer to agri-big data.

Table 6 shows results for the demand data with '*low*' Demand set as the reference group. Most of the independent variables have p-values over 0.05, indicating that H0 cannot be rejected. Since the independent variables are only significant to the dependent variables. Even though the p-values of the demand independent variables are insignificant separately, the OR values show a significant correlation among them, compared to the reference group (i.e. 0.463, 0.328, and 0.110 for '*moderate*', '*high*', and '*very high*' Demand). This suggests that '*moderate*', '*high*', and '*very high*' Demand are negatively correlated to '*low*' Demand which represents buying power in the NR-market.

These results show that our approach to agri-big data can provide statistical significance indicators for demand and supply. However, demand's p-value and its OR provide contradict information, which may suggest that evaluating our approach using significance tests might not offer a complete picture. Therefore, we also evaluated our work using predictive analysis.

5.2 Predictive analytics

The objective of our predictive ability measurements is to provide additional classification results based on significant associations between the independent variables.

5.2.1 Experiment setting

Four well-known classification algorithms were utilized: Decision Trees (DT), Neural Networks (NN), Support Vector Machines (SVM), and Naïve Bayes (NB) [18]. All of these algorithms employ correlations and relative odds for the dependent variable outcomes given independent variables.

The scikit-learn Python library for machine learning [10] was employed to tune the model's hyperparameters, with training data from Section 4.2, and a 10-fold cross-validation used during testing to avoid overfitting. The metric is $F_1 = \frac{tp}{tp+\frac{1}{2}(fp+fn)}$. The true positive *(tp)* is a correct outcome from the prediction results, the false positive (*fp*) is an incorrect result, and false negative *(fn)* is an unclassified outcome.

5.2.2 Results and discussions

Table 7 shows the F-measure results for the four Supply and Demand classes according to the dependent variables.

Table 7 F-measure results for Demand and Supply classifications

	Dependent Variables	Supply					Demand				
		low	*moderate*	*high*	*very high*	avg.	*low*	*moderate*	*high*	*very high*	avg.
Algorithms	DT	1.00	0.96	0.98	1.00	**0.98**	0.98	0.92	0.92	0.98	**0.95**
	NN	0.36	0.44	0.77	0.21	**0.45**	0.84	0.78	0.59	0.76	**0.74**
	SVM	0.57	0.66	0.85	0.54	**0.66**	0.83	0.68	0.62	0.74	**0.72**
	NB	0.19	0.33	0.67	0.17	**0.34**	0.70	0.48	0.40	0.59	**0.54**
	Average	**0.53**	**0.60**	**0.82**	**0.48**	<u>**0.61**</u>	**0.84**	**0.72**	**0.63**	**0.77**	<u>**0.74**</u>

Bold entries highlight the average values of each states of demand and supply that we mentioned in text, while bold-underlined entries signify the overall average values

The F-measure for each algorithm in Table 6 varied depending on the compatibility of the data and algorithms. The highest average scores are with the DT algorithm (0.98 and 0.95 for Supply and Demand). The lowest is for the NB algorithm (0.34 and 0.54 for Supply and Demand).

The DT algorithm appears to be the most compatible with the extracted information. Perhaps because it uses a logical model that can be very successfully handle category-based data and our time-dependent approach transforms raw data into a categorical format.

In contrast, NB is a probabilistic model that uses Bayes' theorem to compute posterior probabilities for the dependent variables given independent random variables. As a result, the F-measures for the '*very high*' Supply and the '*high*' Demand are low cause the probabilistic model depends on the prior probabilities obtained from the training data. The NR-market training data imbalance will produce high number of false positive that affect the overall F- measure. However, this suggests that the probabilistic model would be good for balanced data classification.

NN and SVM algorithms are geometric models that compute the distances and weights between key features to classify the outcomes. They employ optimization methods that need to determine the best hyperparameters based on the highest performance with complex patterns. The overall accuracy of NN and SVM with default parameter setting are acceptable for our work but there might be some benefit in adjusting NN's complex layers' parameters for deep learning or SVM's margin to improve accuracy.

The significant improvements using our proposed approach can be seen in Fig. 5 which compares it with the predictive accuracies using the traditional SMA [2, 37] approach.

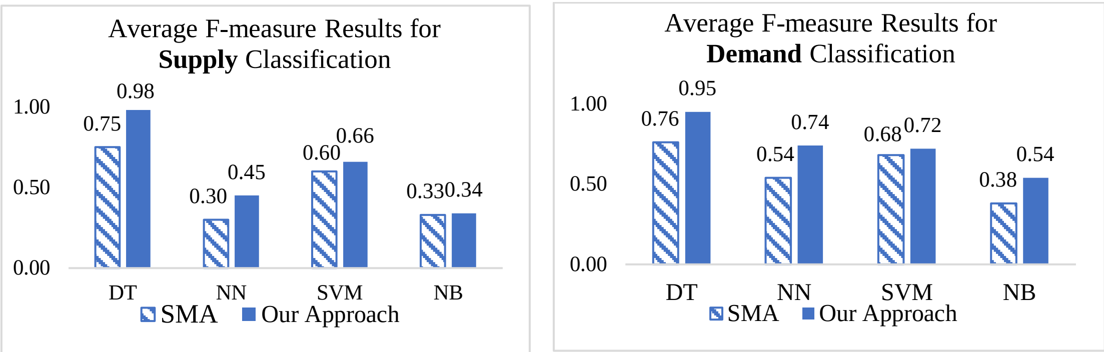


Fig. 5 Comparison of the Predictor's Abilities when applied to Agri-big Data

The overall result of Fig. 5(b) shows F-measures using our approach are higher than with SMA. However, Fig. 5(a) has quite similar results for the NB algorithm using SMA and our approach because the supply data is completely imbalanced, and so limits the performance of the NB.

SMA pre-processes sequential data into a single representation which works well when the decision-makers only need to know a recent situation for the short-term decision without the long-term considered by our proposed. This suggests that SMA is suitable for a short-term report which does not consider data behavior, leaving that task to the decision-makers. However, if the decision-makers need to plan long-term strategies, our approach is more suitable since it deals with time-series data by pre-processing it based on supply chain integration. This situation is much more common in highly uncertain environments.

6 Conclusions

To support decision-making, agri-market management requires contextual information about short-long term impacts based on demand and supply based on agri-big data. However, big data is unstructured, employs heterogeneous data representation, and arrives from multiple sources, all of which suggests the need for agri-big data digitization.

This paper has proposed big data digitization framework for supply chain management with a focus on an information extraction approach based on time-series decomposition. The approach has the main goal to pre-processes sequential data from multiple sources based on supply chain perspectives, with demand and supply acting as contextual information to support decision-making. The case study for our demand and supply awareness was Thailand's domestic natural rubber market. The experiments focused on significance tests using a multinomial logit model and predictive analytics using machine learning. The results show that our proposed approach enhances the agri-big data utilization by extracting its insights for supporting predictive models with higher accuracy than the traditional method.

This proposed framework can be applied in any field that related to market factors awareness in supply chain management, especially with short-long term impacts. The main contribution is the encoding of supply chain knowledge as a prior for model-based analysis. It is based on time-series decomposition that can help decision makers understand market situation using prior knowledge and big data. The future application can directly employ our framework by redefining or determining relevant prior knowledge since it is important scheme in the modern supply chain management in the era of dynamic trades and global marketing.

Our time-series decomposition approach can easily process data from multiple sources, and this will become even more important in areas such as social media for analyzing consumer behavior and satellite imaging of farming production. New sensor types will require new additional solutions to deal with them effectively and accurately.

Acknowledgement This work was supported in part by the Thailand Education Hub for ASEAN Scholarship in Doctoral Degree funded by Prince of Songkla University, Thailand under Grant No. TEH-233/2016.

References

1. Abdollahpour S, Kosari-Moghaddam A, Bannayan M (2020) Prediction of wheat moisture content at harvest time through ANN and SVR modeling techniques. Inf Process Agric 7:500–510. https://doi.org/10.1016/j.inpa.2020.01.003
2. Arunwarakorn S, Suthiwartnarueput K, Pornchaiwiseskul P (2017) Forecasting equilibrium quantity and price on the world natural rubber market. Kasetsart J Soc Sci 40:8–16. https://doi.org/10.1016/j.kjss.2017.07.013
3. Bank of Thailand (2021) Daily Foreign Exchange Rate [Online]. Available: https://www.bot.or.th/english/_layouts/application/exchangerate/exchangerate.aspx. Accessed 28 Jul 2021
4. Belaud J-PP, Prioux N, Vialle C, Sablayrolles C (2019) Big data for agri-food 4.0: Application to sustainability management for by-products supply chain. Comput Ind 111:41–50
5. Bocca FF, Rodrigues LHA (2016) The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Comput Electron Agric 128:67–76
6. Borodin V, Bourtembourg J, Hnaien F, Labadie N (2016) Handling uncertainty in agricultural supply chain management: A state of the art. Eur J Oper Res 254(2):348–359
7. Chanchaichujit J, Saavedra-Rosas JF (2018) The elements of the natural rubber industry supply chain. Using simulation tools to model renewable resources. Springer International Publishing, Cham, pp 19–40
8. Chen JC, Wang XA (2018) Development of fuzzy logic and genetic fuzzy commodity price prediction systems - An industrial case study. In: IEEE International Conference on Industrial Engineering and Engineering Management, vol 2017-Decem, pp 432–435
9. Department TM (2019) Climatological Center. [Online]. Available: http://climate.tmd.go.th/gge/index.html. Accessed 16 Mar 2020
10. Feurer M, Klein A, Eggensperger K, Springenberg JT, Blum M, Hutter F (2019) Auto-sklearn: Efficient and robust automated machine learning. In: Hutter F, Kotthoff L, Vanschoren J (eds) Automated machine learning. Springer, Cham, pp 113–134
11. Fiala P (2016) Supply chain coordination with auctions. J Bus Econ 86:1–2
12. Gardas BB, Raut RD, Cheikhrouhou N, Narkhede BE (2019) A hybrid decision support system for analyzing challenges of the agricultural supply chain. Sustain Prod Consum 18:19–32
13. Gaso DV, Berger AG, Ciganda VS (2019) Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. Comput Electron Agric 159:75–83
14. Golmohammadi A, Hassini E (2019) Capacity, pricing and production under supply and demand uncertainties with an application in agriculture. Eur J Oper Res 275(3):1037–1049
15. Govindan K, Cheng TCE, Mishra N, Shukla N (2018) Big data analytics and application for logistics and supply chain management. Transp Res Part E Logist Transp Rev 114:343–349
16. Hatami-Marbini A, Hekmat S, Agrell PJ (2020) A strategy-based framework for supplier selection: a grey PCA-DEA approach. Oper Res 1–35. https://doi.org/10.1007/s12351-020-00556-8
17. Insider M (2019) [Online]. Available: https://markets.businessinsider.com/commodities/oil-price. Accessed 16 Mar 2020
18. Ip RHL, Ang LM, Seng KP, Broster JC, Pratley JE (2018) Big data and machine learning for crop protection. Comput Electron Agric 151:376–383
19. Kamble SS, Gunasekaran A, Gawankar SA (2020) Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. Int J Prod Econ 219:179–194. Elsevier
20. Kim Y, Lee Y, Chung KY, Lee KD (2015) An investigation on the information systems research in supply chain management: an analysis of research topic and methodology. Multimed Tools Appl 74(20):8849–8860
21. Kliangkhlao M, Limsiroratana S (2019) Towards the idea of agricultural market understanding for auto- matic event detection. In: Proceedings of the 8th International Conference on Software and Computer Applications - ICSCA '19, pp 81–86
22. Koirala A, Walsh KB, Wang Z, McCarthy C (2019) Deep learning – Method overview and review of use for fruit detection and yield estimation. Comput Electron Agric 162:219–234. Elsevier
23. Leng L, Zhang J, Xu J, Khan MK, Alghathbar K (2010) Dynamic weighted discrimination power analysis in DCT domain for face and palmprint recognition. In: 2010 International Conference on Information and Communication Technology Convergence, ICTC 2010, pp 467–471

24. Liu S, Leat M, Moizer J, Megicks P, Kasturiratne D (2013) A decision-focused knowledge management framework to support collaborative decision making for lean supply chain management. Int J Prod Res 51(7):2123–2137
25. Maya Gopal PS, Bhargavi R (2019) A novel approach for efficient crop yield prediction. Comput Electron Agric 165:104968
26. Meng X (2020) Proactive management in the context of construction supply chains. Prod Plan Control 31(7):527–539
27. Murphy KP (2012) Machine learning: A probabilistic perspective. The MIT Press, London
28. Nakamori Y (2020) Fusing systems thinking with knowledge management. J Syst Sci Syst Eng 29(3):291–305
29. Nguyen L, Novák V (2019) Forecasting seasonal time series based on fuzzy techniques. Fuzzy Sets Syst 361:114–129
30. Perktold J, Seabold S, Taylor J. Statsmodels. [Online]. Available: https://www.statsmodels.org/stable/index.html. Accessed 20 Aug 2020
31. Reynard D, Shirgaokar M (2019) Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? Transp Res Part D Transp Environ 77:449–463
32. Rubber Authority of Thailand (2019) Rubber price - Rubber authority of Thailand. [Online]. Available: http://www.raot.co.th/rubber2012/menu5.php. Accessed 16 Mar 2020
33. Schniederjans DG, Curado C, Khalajhedayati M (2020) Supply chain digitisation trends: An integration of knowledge management. Int J Prod Econ 220:1–11. https://doi.org/10.1016/j.ijpe.2019.07.012
34. Shwetha AN, Prabodh CP(2021) Auction system in food supply chain management using blockchain. In: International Conference on Advances in Computer Engineering and Communication Systems, pp 31–40
35. Shynkevich Y, McGinnity TM, Coleman SA, Belatreche A, Li Y (2017) Forecasting price movements using technical indicators: Investigating the impact of varying input window length. Neurocomputing 264: 71–88
36. Sperandei S (2014) Understanding logistic regression analysis. Biochem Med 24(1):12–18
37. Stein S, Steinmann HH (2018) Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. Eur J Agron 92:30–40
38. Szumilas M (2010) Explaining odds ratios. J Can Acad Child Adolesc Psychiatry 19(3):227–229
39. Tiwari S, Wee HMM, Daryanto Y (2018) Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Comput Ind Eng 115:319–330
40. Tokyo Commodity Exchange Inc (2019) [Online]. Available: https://www.tocom.or.jp/market/kobetu/rubber.html. Accessed 16 Mar 2020
41. Verma S, Bhatia A, Chug A, Singh AP (2020) Recent advancements in multimedia big data computing for IoT applications in precision agriculture: Opportunities, issues, and challenges. In: Tanwar S, Tyagi S, Kumar N (eds) Intelligent Systems Reference Library, vol 163. Springer, Singapore, pp 391–416
42. Wang Z, Wu J (2016) Knowledge systems engineering: A complex system view. In: Nakamori Y (ed) Knowledge synthesis: Western and eastern cultural perspectives. Springer, Tokyo, pp 107–149
43. Wang H, Zhang Q, Wu J, Pan S, Chen Y (2019) Time series feature learning with labeled and unlabeled data. Pattern Recognit 89:55–66
44. Wolfert S, Ge L, Verdouw C, Bogaardt M-J (2017) Big data in smart farming – A review. Agric Syst 153: 69–80
45. Yuan G, Han J, Zhou L, Liang H, Zhang Y (2019) Supply and demand law under variable information. Phys A: Stat Mech Appl 536:121004
46. Zhang J, Cui S, Xu Y, Li Q, Li T (2018) A novel data-driven stock price trend prediction system. Expert Syst Appl 97:60–69
47. Zhang C, Di L, Lin L, Guo L (2019) Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps. Comput Electron Agric 166:1–15. https://doi.org/10.1016/j.compag.2019.104989
48. Zhao J, Ji M, Feng B (2020) Smarter supply chain: a literature review and practices. J Data Inf Manag 2(2): 95–110
49. Zhu Q, Zhang F, Liu S, Wu Y, Wang L (2019) A hybrid VMD–BiGRU model for rubber futures time series forecasting. Appl Soft Comput J 84:105739

Appendix 3

M. Kliangkhlao, S. Limsiroratana and B. Sahoh, "The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains," in IEEE Access, vol. 10, pp. 86813-86823, 2022, doi: 10.1109/ACCESS.2022.3199353.

The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains

Mallika Kliangkhlao[1], Somchai Limsiroratana[1], and Bukhoree Sahoh[2,3]
[1]Department of Computer Engineering, Prince of Songkla University, Had Yai, Songkla, 90112, Thailand
[2]School of Informatics, Walailak University, Tha Sala, Nakhonsithammarat, 80160, Thailand
[3]Informatics Innovation Center of Excellence (IICE), Walailak University, Tha Sala, Nakhonsithammarat, 80160, Thailand

Corresponding author: Bukhoree Sahoh (e-mail: bukhoree.sa@wu.ac.th).

This work was supported in part by the Thailand Education Hub for ASEAN Scholarship in Doctoral Degree funded by Prince of Songkla University, Thailand, under Grant No. TEH-233/2016.

ABSTRACT The balancing of demand and supply in the market is complex because of the dynamic supply chain and environment. It causes uncertain situations and is a limitation in decisions making systems that cannot produce reasonable descriptions to help decision makers eliminate uncertainties. This study proposes the design and development of a Causal Bayesian Networks (CBNs) model for market understanding, which encodes a human-like approach to explain demand and supply events for decision makers. A framework for generating reasonable descriptions in Agricultural Supply Chains (ASCs) management is proposed. The qualitative and quantitative design of the CBNs model is developed and proved that the CBNs model can reasonably explain events using predictive performance measurement and sensitivity analysis for producing reasonable descriptions. The results illustrate that the CBNs model is suitable for ASCs situation explanation involving uncertain situations and is ready to apply to real-world applications to support decision-making systems.

INDEX TERMS Explainable Artificial Intelligence; Causal Graph; Machine Learning; Big Data; Demand and Supply Analysis; Supply Chain Management

I. INTRODUCTION

One of the problems in Agricultural Supply Chains (ASCs) is how to deal with the imbalance between demand and supply. For example, agricultural production and consumption may suffer from unexpected seasonal changes, such as an untimely harvest that causes shortages or even surplus market supply [1]. Decision making in this situation is a problem because there is a lack of comprehensive, real-time information.

The expansion of sensors-based-smart farming generates even more extensive data as a source of real-time information. It should allow decision makers to be more aware of demand and supply changes, and apply these variations to the benefit of supply chains. Unfortunately, big data suffer from the problems of enormous volume and complex dimensionality [2], [3]. Machine learning (ML) plays a vital role in a data-driven approach for supporting decision making [4]. It is widely used in agriculture decision-making systems because it can uncover the information needed in ASCs [5]. For example, Punia *et al*. [6] has proposed a retail forecasting approach using extensive point-of-sales data, while Bu and Wang [7] utilized a water consumption approach for crop growth based on IoT sensors. They both employed deep learning to handle identification and classification, but it was less useful when decision makers wanted to ask how and why such outcomes were produced. Deep learning techniques produce black-box models which few people understand, and this lack of reasonable descriptions can cause decision makers to not fully understand the demand-supply situation, which may lead to poor decisions.

Reasonable descriptions utilize technologies that produce contextual information based on supply chain knowledge. The descriptions should be both testable and understandable by both human and agent-based systems by interpreting supply chain knowledge using observational data. Fortunately, Bayesian Networks (BNs) for supply chain knowledge can produce reasonable descriptions since they determine transparent relationships using cause-and-effect as rational contextual information [8]. For example, Qazi *et al*. [1] and Ji *et al*. [9] employed BNs for managing a supply chain, by capturing relationships between supply chain factors from data based on correlation. However, they did not consider the causal assumptions based on rational human knowledge, which allows the model to detect irrational, unexpected ASCs events [10]. Our study addresses this drawback by proposing the use of Causal Bayesian Networks (CBNs) for knowledge in ASCs management.

CBNs determine the consequences and interdependencies among supply chain activities as a context synthesized from prior knowledge and big data. It models expert reasoning to explain demand and supply by producing reasonable descriptions.

The significant contributions of this study are:

- A new framework for knowledge description in ASCs management which addresses unexpected changes in supply chains.
- The development of a CBNs model for suppling descriptions in the natural rubber commodity market.
- Proof that the proposed model converges to expert reasoning by an analysis using predictive performance measurements and sensitivity analysis.

The rest of this paper is organized as follows: background knowledge and related work is presented in section II, and the descriptive supply chain management framework is introduced in section III. Section IV details the design and development of the CBNs model for supply chain management, and section V evaluates the CBN using quantitative experiments. Conclusions and future directions appear in section VI.

II. BACKGROUND KNOWLEDGE AND RELATED WORKS

ASCs management balances demand and supply in the presence of rapidly changing environmental conditions, and so is essential for efficient planning [11]. It utilizes real-time analysis and reaction, which depends upon contextual information in the supply chain [12]. In this section, we give some background on ASCs management, its common tasks, and various analysis approaches.

A. ASCS BACKGROUND

The futures market is an auction-based exchange where buyers and sellers trade contracts for deliveries set for a specified future date based on the quantity, quality, and price of commodities. The futures market help ASCs protect their activities from price fluctuations, which highlighting how ASCs effectiveness depends on adapting to supply network constraints and shifting [13].

The supply chain manages production, processing, wholesale operations, logistics, and retailer operations that depend upon suppliers, customers, and firms [14]. A supplier is the source for material, while a customer shows interest and demand for the product by considering the supply's trend. A firm's productivity is based on the source, order, and price, determining the supply's trend for the whole chain. In this way, the situation can be elaborated into micro-level and macro-level decisions that let stakeholders monitor these processes and determine the demand and supply at both levels [15]. Relationships between the supplier, customer, and firm explains the market situation that can be in equilibrium or exhibit abnormality as shortages and surplus. The relationship's explanation helps decision makers to decide policies to deal with the supply chain abnormality, which shows how an ASCs explanation so important for management.

B. ASCS MANAGEMENT APPROACH

Big data and ML have been employed for discovering demand and supply situations [16]. ASCs management uses the descriptive ability of ML to adjust operations and troubleshoot situations quickly and efficiently [17].

Kappelman and Sinha [18] proposed an approach based on stochastic optimization methods for dealing with uncertainty in supply chain systems. They claimed that their approach could effectively minimize uncertain problems and optimize time and complexity. Oh and Jeong [19] proposed a tactical supply planning model to overcome the short product life cycle and demand uncertainty. They concluded that their approach could provide the solution based on the optimal trade-off between profit and lead time. Gardas et al. [20] proposed systematic hierarchical structures using cause-and-effect-based relationships supporting decision-making. They discussed that their proposal could help decision-makers improve their understanding. The studies focused on maximizing profits in decision-making but did not consider explaining a situation of unbalancing between demand and supply.

BNs transparently model knowledge of supply chain relationships to produce such information [8], which decision-makers employ to create policies. BNs are probabilistic graphical models that can capture the uncertainty and relationships among relevant factors in the supply chain decision-making process. Random variables represent these factors, and their relationships are encoded by conditional probabilities using Bayes' theorem.

Sharma [21], Chhimwal et al. [22], Lawrence et al. [23], and Ojha et al. [24] proposed for BNs-based risk assessment approach for supply chain management using historical data. They summarized that the approach could help the supply chain managers identify the risk factors early. El Amrani et al. [25] studied the sustainability of the supply chain network. These methods were successful because they focus on predicted outcomes and contextual explanations. However, they still did not consider explaining the context of demand and supply. This means that the model cannot answer ASCs management questions such as 'What is the situation of demand and supply, and why were these outcomes produced?'. The burden of causal interpretation and rational explanation is left to humans.

Causal Bayesian Networks (CBNs) have been proposed to address this problem [26], by modeling both emerging and rare events (e.g. climatic problems) that affect the management of ASCs. This means that CBNs will play an essential role for ASC explanations, even though the learning method for CBNs in ASCs is still far from decided.

C. CAUSAL BAYESIAN NETWORKS

CBNs are a human-like intelligent framework that encodes experience and knowledge based on cause-and-effect assumptions [27], [28]. In this way, CBNs extend traditional BNs by adding an interpretable ability in the manner of human-like understanding to produce explanations [26]. This lets CBNs explain demand and supply behavior to support ASC management.

A cause-and-effect assumption goes beyond correlation because it shows not only a statistical dependency between X and Y, but encodes knowledge that Y happens because of X. A CBNs assumption, $X \rightarrow Y$, is a cause-and-effect relationship that states that "Only X can change Y". For example, Weather → Crop Yield captures the idea that Weather generally influences Crop Yield which means Weather is a cause of Crop Yield. A decision makers may ask "How will crop yield be undersupplied if prolonged rainfall is observed?". This means that a Crop Yield (C) is denoted by undersupplied (us) given Weather (W) by prolonged rainfall (pr), then the query can be written using Bayes' Theorem:

$$P(\boldsymbol{C}= us \mid \boldsymbol{W} = pr) = \frac{P(\boldsymbol{W} = pr\ , \boldsymbol{C}= us) \times P(\boldsymbol{C}= us)}{P(\boldsymbol{W} = pr)} \quad (1)$$

There are four probabilities in (1): the posterior $P(\boldsymbol{C} = us \mid \boldsymbol{W} = pr)$; the likelihood $P(\boldsymbol{W} = pr\ , \boldsymbol{C} = us)$, the prior $P(\boldsymbol{C} = us)$, and the observation $P(\boldsymbol{W} = pr)$. Their definitions are:

P($\boldsymbol{C}$ = *us* | $\boldsymbol{W}$ = *pr)*: the probability that an *under supply* is conditioned on *prolonged rainfall*;

P($\boldsymbol{W}$ = *pr* , $\boldsymbol{C}$ = *us)*: the likelihood that an *under supply* co-occur with *prolonged rainfall*;

P($\boldsymbol{C}$ = *us*): the marginal likelihood of an *under supply* regardless of *prolonged rainfall*;

P($\boldsymbol{W}$ = *pr*): the marginal likelihood of *prolonged rainfall* in the past.

This cause-and-effect assumption helps decision makers deal with market supply when influenced by the weather.

Although such observations are vital ingredients of CBNs for ACSs management, ACSs are a complex domain, which means that many causal assumptions cannot be expressed as direct $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ relations, and may involve hidden factors between the $\boldsymbol{X}$ and $\boldsymbol{Y}$. Pearl *et al.* [29]'s causal model encodes such hidden factors—$\boldsymbol{Z}$ based around three types of causal structures called chains, forks, and colliders. A chain encodes a cause-and-effect relationship in which the factor is involved sequentially. A fork encodes assumptions when a cause-and-effect relationship has a common cause. A collider encodes a cause-and-effect relationship which has a common effect. The structures are summarized in Table 1.

TABLE 1. Causal structures.

Causal Structure	Representation	Axiom
Chain	$\mathbf{X} \rightarrow \mathbf{Z} \rightarrow \mathbf{Y}$	**X** indirectly causes **Y** through **Z**
Fork	$\mathbf{X} \leftarrow \mathbf{Z} \rightarrow \mathbf{Y}$	**X** and **Y** are caused by **Z**
Collider	$\mathbf{X} \rightarrow \mathbf{Z} \leftarrow \mathbf{Y}$	**X** and **Y** are connected through **Z**

The causal structure uses conditional dependencies to connect nodes with causal relationships and block the paths between nodes with independencies; a process known as d-separation [30]. Causal discovery algorithms have been studies to structure a CBNs model from the observational data [31]. The algorithms are widely separated into two types: constraint-based and score-based. The constraint-based algorithms apply conditional independence constraints (e.g., Fast Causal Inference or FCI, and PC), while the score-based algorithms construct model using posterior probability of the candidate model (e.g. Greedy Equivalence Search or GES, and Greedy FCI). However, the resulted model' performance is hard to be tested without a gold standard [32]. Then, expert-based modelling is the answer for discovering causal relationships in domain that lacks a baseline.

Causal relationships help ASCs model knowledge and help decision makers discover the reasons behind complex behaviors. The challenge is determining the semantics of the problem domains, verifying the dependencies among the random variables, and deciding whether they should connect or separate each other. This is the backbone of a descriptive supply chain management framework that can model micro-level and macro-level ASCs situations for decision-making.

III. DESCRIPTIVE SUPPLY CHAIN MANAGEMENT FRAMEWORK

Useful supply chain management must produce proactive planning aligned with evidence, but this is not feasible with traditional technologies. This section proposes a new management framework that can generate explanations based on demand and supply evidence. The framework is summarized in Figure 1.

The framework consists of four components: data sensing, observation identification, situation explanation, and inference for making decisions; it follows the subdivisions employed by Belaud *et al.* [33].

Data sensing retrieves ASCs related data from sources such as global positioning systems (GPS), geographic information systems (GIS), remote sensing technologies, and web- based applications. The raw data is transformed into ASCs observations by the observation identification component. Although these observations detail ASCs information, they do not elaborate the relationships among the ASCs, which need deeper knowledge of the ASCs situation. The situation explanation produces rational explanations based on cause and effect in the manner of human- like reasoning. Lastly, inference supports a decision- making component for proactive planning. It receives a hypothesis from a decision maker, and infers possible outcomes using the current ASCs situation. The resulting response and review help the decision maker to decide upon solutions and plan policy.

The intelligence of this framework depends upon the CBNs model developed as an initial requirement. This is the topic of the next section.

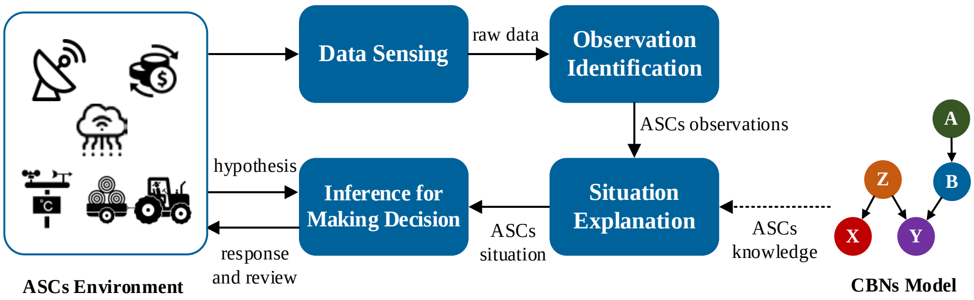


FIGURE 1. The descriptive ASCs management framework.

IV. DESIGN AND DEVELOPMENT OF CBNS MODEL FOR ASCS

CBNs are developed using causal discovery algorithms based on data dependencies that can structure the relations between random variables. Even the automatic algorithms, including Tree Augmented Naïve Bayes (TAN), Bayesian Network augmented Bayesian (BAN), and FCI, are widely studies [34], these algorithms generate statistical correlations among observations derived from well-structured and complete historical data that covers all possible events even it is the rarest. However, the unpredictability of the modern supply chain introduces uncertainty and change into the ASCs environment which generates rare events that do not exist in historical data. This means that a purely data-driven approach cannot produce accurate causal-and-effect explanations [35]. Moreover, the related data is still lacking in the context of the natural rubber ASCs. It lacks in both comprehensive and historical terms that is why the traditional ASCs runs with human.

To deal with that, using expert-based modeling as a gold standard. The prior knowledge depends upon the experts. It consisted of: (1) interviews with three experts, and two practitioners from the Central Rubber Market (CRM) in Hat Yai, Songkhla, Thailand; (2) reviews of a CRM database of 5 years provided by the Thai government. A prior-based process is required to integrate with data-driven process to produce a gold standard of CBNs model, as explained in the rest of this section.

A. RANDOM VARIABLES IN ASCS EXPLANATION

We employ random variables to model possible events, and quantify them based on observational data. Events are fixed as states but can occur randomly, according with natural change. We apply a traditional understanding of demand-supply price based on the structural representation employed in Pearl and Mackenzie [36]. Our case study models the futures market auction system as five ASCs explanation processes: source, supply, demand, market price, and futures market volatility. The random variables' states are designed and built using prior knowledge from ASCs operations [5], and market price considers possible futures market conditions that contribute to our model.

We divide the random variables into three categories: observed, micro-level, and macro-level. Observed random variables model direct environmental observations, micro-level random variables represent supply chain activities, and macro-level random variables model market situations. The three categories of random variables and their states are detailed in Tables 2, 3, and 4.

TABLE 2. Observed random variables.

Context	Random Variables	States
Source	**Climatic Problem**	*normal, drought, monsoon, flood*
	Plantation Area	*downtrend, sideways, uptrend, fluctuation*
Supply	**Raw Material Cost**	*downtrend, sideways, uptrend, fluctuation*
	Labor Resources	*down, stable, up*
Demand	**Exporting Costs**	*down, stable, up*
	Currency Exchanges	*strengthening, stable, weakening*
Future Market Volatility	**Open Interest**	*downtrend, sideways, uptrend, fluctuation*
	Trading Volume	*downtrend, sideways, uptrend, fluctuation*
Market Price	**Future Market Prices**	*downtrend, sideways, uptrend, fluctuation*
	Market Price	*down, stable, up*

1) OBSERVED RANDOM VARIABLES

The observed random variables are based on five ASCs explanations, which are summarized in Table 2.

Climatic Problem affects crop growth and harvesting, and can be obtained from weather station observations or open data services. **Plantation Area** estimates the crop yield quantity, which can be done manually or be automated using sensors. **Climatic Problem** and **Plantation Area** provide information about the source and imply raw materials processing.

Raw Material Cost, such as crop price can be observed from open data services, and shows baseline information that harms secondary production. **Labor Resources** reflects production capacity, obtained through registered labor and official holiday figures. **Raw Material Cost** and **Labor Resources** are essential for estimating the supply context in the supply chain.

For demand, the required information relates to product consumption and logistics. **Exporting Costs** and **Currency Exchanges** movement are the critical factors. **Exporting Costs** information can be obtained from the petroleum prices index and **Currency Exchanges** from web services. In addition, production consumption is varied according to the agricultural product and the nature of the market. Some products may be traded through an agent, while many products are traded by auction, while the commodity product depends upon the futures market. This means that future market volatility is explained using **Open Interest**, **Trading Volume**, and **Futures Market Prices**, which can be observed from business data services. The **Market Price** is the index price for a commodity product reserved by a governmental office or agent, and is directly observable.

Although all the observed random variables in Table 2 are observable through open data, information systems, and services, we need to clarify the state of the variables for the specific market context. For example, in the case of the **Climatic Problem**, we focus on events such as *drought, monsoon,* and *flood* based on the vulnerability of the crop yield. The other variables are categorized based on movements (*down, stable, up*) and trends (*downtrend*, *sideways*, *uptrend, fluctuation*) derived from the ASCs non-stationary characteristics. The criteria for choosing a state is based on how its short long term impact affects the trading process. For instance, **Raw Material Cost** shows the impact on manufacturing, while **Open Interest**, **Trading Volume**, and **Future Market Prices** reflect demand in the futures market. They are indirectly affected by trading processes in the long-term, and so their states are categorized based on trends.

2) MICRO-LEVEL RANDOM VARIABLES

The micro-level random variables are elaborated from prior knowledge of ASCs, which are summarized in Table 3.

TABLE 3. **Micro-level random variables.**

Context	Random Variables	States
Source	**Crop Yield Producing**	*low, normal, high*
Supply	**Manufacturing Capacity**	*low, normal, high*
Demand	**Consumer Preference**	*low, normal, high*
Future Market Volatility	**Future Market Movement**	*down, stable, up*
Source	**Crop Yield Producing**	*low, normal, high*

Crop Yield Producing represents the level of market source. **Manufacturing Capacity** is the intermediate step of market supply production. **Consumer Preference** summarizes the product requirement, which reflects market demand. **Future Market Movement** is the external factor that influences market demand.

These micro-level random variables utilize *low, normal,* and *high* states which reflect their market context. However, **Future Market Movement** is defined using *down*, *stable,* and *up* values since it monitors the futures market situation.

3) MACRO-LEVEL RANDOM VARIABLES

Macro-level random variables summarize supply chain dynamics, which are represented using **ASCs Situation**. It consists of three possible states: *equilibrium* (the quantity demanded and supplied are the same), *shortage* (there is an excess of demand), and *surplus* (there is an excess of supply), but the relationships between demand, supply, and price are complex. For example, if demand is up and supply is down, then the **ASCs Situation** is a *shortage* that increases price according to market theory. In contrast, if the supply and demand relationship trigger a decreasing price, the **ASCs Situation** is still a *shortage* but with abnormal behavior. This latter scenario reflects a dysfunctional market policy, and market managers must implement corrections (i.e., by controlling the reference price or imposing a price ceiling). The states for the macro-level random variable are detailed in Table 4

TABLE 4. **The states of macro-level random variables.**

States	Definition
equilibrium	The market is equilibrium, and the price is stable.
abnormal-equilibrium	The market is equilibrium, but the price is rising or dropping.
shortage	Excess demand and price is stable or rising.
abnormal-shortage	Excess demand, but the price is dropping.
surplus	Excess supply and price is stable or dropping.
abnormal-surplus	Excess supply, but the price is rising.

Table 4 lists the possible states for **ASCs Situation** in the context of demand, supply, and price. They are intended to help managers explain situations involving causal assumptions that interpret market behavior.

B. ASSUMPTIONS IN CBNS MODELING

The futures market controls the demand of the natural rubber productions consumed by the automotive and tire industries [37], [38]. The products in that market are rubber sheets produced locally which depend on climatic conditions [39]. Indeed, climatic problems are the leading cause of decreased source production. We employ this information to model the causal assumptions between the random variables, and the resulting model is shown in Figure 2.

Figure 2 shows the graphical causal assumptions between the random variables, with a random variable for a cause pointing directly to effect random variable(s) (cause(s) → effect(s)). This graphical model can be interpreted into mathematical form using Structural Causal Model [29]. The assumptions are causally structured for explaining the **ASCs Situation** in terms of **Manufacturing Capacity**, **Consumer Preference**, and **Market Price**, and most of them are encoded as collides. For example, **Trading Volume**, **Open Interest**, and **Future Market Price** explain the liquidity and activity of **Future Market Movement**. **Trading Volume** reflects the short-term demanded quantity throughout the trading day, while **Open Interest** shows the number of futures contracts that are still open. **Trading Volume** and **Open Interest** are independent unless **Future Market Movement** is questioned, and then they become causally dependent. **Crop Yield Production** is also a collider, affected by **Plantation Area** and **Climatic Problems**. In other words, the causes are causally independent of each other, but conditioning on **Crop Yield Production** makes them dependent. Moreover, **Crop Yield Production** affects the behavior of **Raw Material Cost**, which passes its information to **Manufacturing Capacity**.

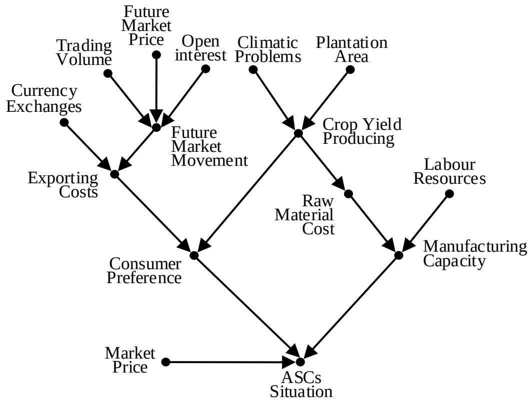


FIGURE 2. **Causal assumption of natural rubber SCs using CBNs.**

The CBNs model is initially constructed by casual assumptions as a rule-based prior knowledge. It is a qualitative knowledge that machine learning-based applications cannot reasonably interpret it. The data-driven approach is then employed to encode the causal assumptions to quantitative knowledge.

C. DATA PREPROCESSING

The character of ASCs-related data is multiple sources that become problematic because they are unstructured, redundant, and streaming. However, the real-world system must analyze and explain the event simultaneously and automatically for decision-making. For example, decision-makers may ask, "What does the trend of future market price look like given current evidence?". The system must transform the input data from multiple sources into information represented with states of random variables to interpret and answer the question.

Section IV.A shows that states of random variables are discrete. For the observed random variable from the data source [40], **Climatic Problem** is categorized as *normal*, *drought*, *monsoon*, and *flood*. The rest are categorized by movement (*down*, *stable*, *up*) and trend (*downtrend*, *sideways*, *uptrend*, and *fluctuation*). Movement is the distance between points, and the trend is the semantic meaning of the direction of movements.

Time-series decomposition approach [41] is applied to produce the states that can extract information from multiple sources. This approach is dimensionality reduction that decomposes using a frequency-based interpolation function. Experts define the heuristics rules according to the short-long term impacts of each random variable on ASCs.

Natural rubber supply chains (SCs) data for tuning prior and likelihood functions were collected between 2015 and 2019 from the CRM in Hat Yai, Songkhla, Thailand. We randomly split the data collection into two subsets. The first subset is utilized for model training and validation, and the second is for model evaluation. The data-splitting method was performed using the scikit-learn Python library [43].

D. MODELING CAUSAL ASSUMPTIONS

CBNs are a qualitative model based on causal assumptions extracted from background knowledge, which are quantified using observational evidence by training their parameters.

The natural rubber SCs obtained from the training data are summarized in Table 5.

TABLE 5. Summarization of natural rubber SCs.

Data Sources	Random Variables	States
Climatological Center [42]	**Climatic Problem**	*normal* (48%), *drought* (8%), *monsoon* (13%), *flood* (31%)
Agricultural Production Data [43]	**Labor Resources**	*down* (10%), *stable* (81%), *up* (9%)
	Plantation Area	*down* (10%), *stable* (81%), *up* (9%)
	Raw Material Cost	*downtrend* (36%), *sideway* (7%), *uptrend* (47%), *fluctuation* (9%)
Thailand Daily Rubber Price [44]	**Market Price**	*down* (19%), *stable* (69%), *up* (19%)
Bank of Thailand [45]	**Currency Exchanges**	*strengthening* (53%), *stable* (4%), *weakening* (43%)
Markets Insider [46]	**Exporting Costs**	*down* (7%), *stable* (75%), *up* (19%)
Tokyo Commodity Exchange (TOCOM) [47]	**Trading Volume**	*downtrend* (47%), *sideway* (6%), *uptrend* (47%), *fluctuation* (0%)
	Future Market Price	*downtrend* (24%), *sideway* (11%), *uptrend* (34%), *fluctuation* (31%)
	Open Interest	*downtrend* (47%), *sideway* (1%), *uptrend* (51%), *fluctuation* (12%)

Table 5 summarizes random variables whose states are distributed and chose to show the movements that affect the market. However, we did not consider **Plantation Area** factors since rubber trees must grow for seven years before their first harvest and live for two decades. This means that they remain stable and less significant during the 5-year data collection period used here.

While, micro and macro-level random variables are contextual variables, retrieved from CRM database. They are labelled using experts, shown in Table 6.

Table 6 shows micro-level and macro-level random variables' prior distribution. The major proportion of **Crop Yield Producing** is *up* (48%), which causes **Manufacturing Capacity** to be *high* (52%), which accounts for over half of the dataset. This suggests that the supply situation for this ASCs market has always been high. In contrast, the **Future Market Movement** value *up* (15%) is the lowest event occurrence, so cannot boost market demand, which results in **Consumer Preference** being *normal* (52%). This shows that the supply and demand situation is unbalanced, causing **ASCs Situation** to have an *equilibrium* value of 7%.

TABLE 6. Summarization of micro-level and macro-level random variables.

Random Variables	States
Crop Yield Producing	*down* (39%), *stable* (14%), *up* (48%)
Manufacturing Capacity	*low* (17%), *normal* (31%), *high* (52%)
Consumer Preference	*low* (9%), *normal* (52%), *high* (39%)
Future Market Movement	*down* (33%), *stable* (51%), *up* (15%)
ASCs Situation	*equilibrium* (7%), *abnormal-equilibrium* (29%), *shortage* (13%), *abnormal-shortage* (8%), *surplus* (24%), *abnormal-surplus* (20%)

The states in Table 5 and Table 6 become the priors of the random variables. For example, let cp be a set of m-possible outcomes of **Climatic Problem** (**CP**), and $P(\mathbf{CP})$ be the prior for **Climatic Problem**, defined as: $P(\mathbf{CP} = normal) = 0.48$, $P(\mathbf{CP} = drought) = 0.08$, $P(\mathbf{CP} = monsoon) = 0.13$, and $P(\mathbf{CP} = flood) = 0.31$, according to the **Climatic Problem** entry in Table 5. This prior informs us that between 2015 and 2019, Thailand suffered from floods and monsoons almost half the time (which fits with typical tropical climate characteristics). Also, the priors of the states are not equally likely because of the nature of the **Climatic Problem**.

This data can be utilized to tune the likelihood parameters, according to Bayes' Theorem. For example, the causal assumption shows that **Crop Yield Producing** (**CYP**) is affected by **Climatic Problem (CP)**. The likelihood can be calculated using joint probability of this causal assumption, represented using a Conditional Probability Distribution (CPD). We use Maximum Likelihood Estimation [48] for tuning the likelihood parameters. For example, the CPD of **Crop Yield Producing** given **Climatic Problem** in the natural rubber supply chain is summarized in Table 7.

TABLE 7. **Conditional probability distribution of Crop Yield Producing given Climatic Problem.**

Crop Yield Producing	Climatic Problem			
	normal	*drought*	*monsoon*	*flood*
high	0.997	0.008	0.005	0.002
low	0.001	0.982	0.005	0.995
normal	0.001	0.008	0.989	0.002

The CPD of **Crop Yield Producing** given **Climatic Problem** shows the *low* production is affected from strange weather (i.e., probabilities of *low* in **Crop Yield Producing** are 0.982 and 0.995 given *drought* and *flood* respectively). CPD show the likelihood between cause and effect random variables that is required for Bayes' Theorem to explore posterior in CBNs model.

Although the CBNs model is encoded from expertise knowledge that makes human sense, it needs model validation to measure its performance for machine understanding.

E. CBNS MODEL VALIDATION

The causal structure represents the scientific assumptions underpinning the ASCs data, and the CBNs-based model exhibits predictive ability with reasonable explanations. The purpose of CBNs model validation is to confirm that our proposed model can predict the ASCs situation.

According to the **ASCs Situation**'s states in Table 6, the target class is distributed over *6*-possible outcomes and is imbalanced. Marcot and Hanea [49] proposed that *10*-fold is the optimal value for *k*-fold cross-validation for a discrete Bayesian-based model. It resamples the data into ten subsets, using nine subsets in each iteration for training, and the rest for testing. Therefore, we have also employed *10*-fold cross-validation to estimate model performance.

The metric for interpreting validation results is accuracy, selected by measuring the model's predictive performance during the learning process; the results are shown in Table 8.

TABLE 8. **ASCs situation prediction accuracy for 10-fold cross-validation.**

ASCs situation *6*-possible outcome	Accuracy (k = 10)
equilibrium	0.86
abnormal-equilibrium	0.97
shortage	0.92
abnormal-shortage	0.88
surplus	0.96
abnormal-surplus	0.95
Average	**0.94**

Table 8 shows that the overall performance is high of 94%. The accuracies of *equilibrium* and *abnormal-shortage* are lower than the others because they are rare events, occurring at around 7% and 8% in the sample proportion, respectively. The *equilibrium* market is ideal and rarely occurs because the market context changes dynamically. Similarly, *abnormal-shortage* means a shortage of supply with decreasing price, which is an extraordinary situation that contradicts the laws of demand and supply. It is also a rare event with a small sample for training the model.

The validation shows that our proposed possesses good model performance and can be applied to this case study. Although *k*-fold cross-validation is fundamental for model testing, it does not provide satisfactory model performance in our explanation requirement. A significant advantage of our proposed CBNs model is that it explains the market situation correctly and reasonably.

V. RESULTS

This section evaluates how well our CBNs model can perform the task correctly and rationally. Consequently, our experiments have two parts: 1) tests of the predictive performance for model correctness, and 2) sensitivity analysis for model reasonableness.

A. PREDICTIVE PERFORMANCE MEASUREMENT

This experiment measures the CBNs model's predictive performance. The target class are the states of the **ASCs Situation** random variable since it helps to provide final decisions in the supply-chain system.

TABLE 9. **Predictive performance comparison.**

	Equilibrium			*Abnormal-equilibrium*			*Shortage*			*Abnormal-shortage*			*Surplus*			*Abnormal-surplus*			**Avg.**
	PS	RC	FM	PS	RC	FM	PS	RC	FM	PS	RC	FM	PS	RC	FM	PS	RC	FM	
NN	0.87	0.83	0.85	0.96	0.95	0.96	0.93	0.92	0.93	0.91	0.88	0.89	0.93	0.95	0.94	0.93	0.95	0.94	**0.93**
SVM	0.87	0.83	0.85	0.97	0.95	0.96	0.95	0.92	0.94	0.95	0.88	0.91	0.95	0.94	0.92	0.92	0.96	0.94	**0.94**
DT	0.87	0.86	0.87	0.96	0.97	0.96	0.95	0.92	0.93	0.95	0.88	0.91	0.94	0.96	0.95	0.95	0.95	0.95	**0.94**
NB	0.84	0.64	0.72	0.94	0.70	0.81	0.95	0.90	0.93	0.95	0.87	0.90	0.95	0.87	0.90	0.76	0.94	0.84	**0.84**
BS	0.88	0.85	0.87	0.95	0.98	0.96	0.96	0.93	0.95	0.96	0.89	0.92	0.94	0.96	0.95	0.96	0.95	0.96	**0.93**
CBNs	0.86	0.87	0.86	0.96	0.97	0.97	0.95	0.92	0.94	0.96	0.88	0.92	0.94	0.96	0.95	0.95	0.96	0.96	**0.95**

The states of **ASCs Situation** were measured based on Precision, Recall, and F-Measure. Precision (PS) is a proportion of the correction of the positive prediction, which is computed as $PS = \frac{TP}{TP+FP}$. TP is a true positive prediction, and FP a false positive prediction. Recall (RC) is a proportion of the correction of the prediction, which is computed as $RC = \frac{TP}{TP+FN}$, FN is a false negative prediction. F-Measure (FM) is a balance between Precision and Recall, which is computed as $FM = \frac{2\times Precision\times Recall}{Precision+Recall}$.

The measurements employed the dataset described in Section IV.E. Baselines model were evaluated using testing dataset. The average scores of each model are shown in Table 9.

We used standard classification algorithms to compare the predictive performance of our proposed model, including geometric-based models (i.e., Neural Networks (NN) and Support Vector Machines (SVM)), logic-based models (i.e., Decision Trees (DT)), and probabilistic-based models (i.e., Naïve Bayes (NB), Bayesian Search (BS)). As we know that the performance of the classifiers depends upon algorithm's parameters. Then, these models are implemented by scikit-learn [50], a Python library, with default parameter setting. For example, NN was set with *100* hidden layers, *0.001* learning rate, *200* epochs, *ReLU* as the activation function, and *adam* as the optimization. The comparative performance of the predicted results with our CBNs model highlights its predictive ability.

Table 9 shows the Precision, Recall, and F-measure for each model based on the states for **ASCs Situation** states. The average results were well over 80%, which is acceptable for prediction systems. The lowest was 84% from the NB model, since it employs a "naïve" assumption that its features are independent and only depend on the outcomes. This is not true for supply chains where features typically do rely on each other. The other results for NN, SVM, DT, BS, and CBN were 93%, 94%, 94%, 93%, and 95% respectively, which are high since all the models were trained and validated using well-prepared data. This suggests that these models are ready to apply to decision support systems to help understand the **ASCs Situation**.

The FM scores for the *Equilibrium* state are the lowest since it is an infrequent event that is sensitive to the balance of demand and supply, which is affected in various ways. It is also an ideal event, with little chance of occurring, but decision makers need to understand all the factors that support their decisions.

B. SENSITIVITY ANALYSIS FOR CAUSAL ASSUMPTION

Even though the CBN model's results can be acceptably applied to prediction systems, it does not give an explanation for supporting decision making. We addressed this by conducting a sensitivity analysis to show the strength and sensible of connections between random variables. This show how well the CBNs provide guarantees on the query results with rationale explanations. Crucially, this aspect of the CBNs model is missing from the other models. The CBNs model provides contexts for supporting decision making in term of the state parameters of the random variables that impact **ASCs Situation**.

The BS-based model was compared with our model because of its use of conditional dependency of a Bayesian Network [51], which produces relationships based on a DAG of data dependency. We applied scenario-based sensitivity analysis to highlight the rational explanation of both models.

As a base case, we used the most sensitive scenario, "**ASCs Situation** is *equilibrium*". That event has the lowest probability of occurring, but has the highest impact on decision making. According to our hypothesis, the posterior probabilities of **ASCs Situation** may be affected by **Manufacturing Capacity**, **Consumer Preference**, and **Market Price**, and we assume that the base case is sensitive to variations of the states from its random variables.

Sensitivity analysis calculates posterior probability distributions based on the partial derivative over the unknown variables; for example, **ASCs Situation** is questioned given evidence (i.e., the evidence is a state of a cause random variable, *e*). It can be calculated as

$Sensitivity(\boldsymbol{x}) = \frac{\partial p(x_t|e)}{\partial x}$, $\boldsymbol{x}$ is a target variable, with interest in $\boldsymbol{x} = \boldsymbol{x}_t$ as a base case, and $p(\boldsymbol{x}_t|e)$ posterior distribution of the base case given evidence. This computation is based on an algorithm contributed by Kjaerulff and van der Gaag [52].

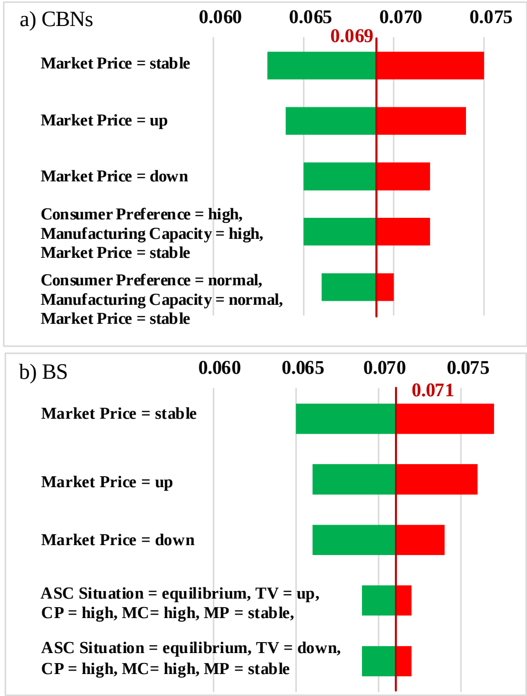
The average sensitivity conditioned from all evidence is between *zero* and *one*. *Zero* means the changes of the **ASCs Situation**'s causes reduce the absolute change in the posterior probability of the base case that shows robustness in posterior distribution calculation, while *one* makes **ASCs Situation** more likely to occur. Sensitivity analysis can measure a minor change in the sensitivity of the **ASCs Situation**'s posteriors (i.e., the causes of non-equilibrium). In essence, this analysis computes the sensitivity between cause and effect in the manner of expert reasoning based on the uncertainty of the **ASCs Situation**.

The degree of sensitivity between causes and effects are represented using tornado diagrams since they are easy to read and interpret [53]. The x-axis in Figure 3 represents the sensitive of "**ASCs Situation** = *equilibrium*" between *zero* and *one*. The y-axis-bar lists the set of parameters as conditions that affect *equilibrium*. The random variables states have 27 possible values, but only the five of the most sensitive parameters appear in Figure 3.

Figure 3 shows the sensitivity levels for the base case from the CBNs and BS models. The sensitive degrees for CBNs and BS are 0.069 and 0.071 respectively.

One difference between CBNs and BS is the number of random variables affecting the sensitivity of the base case. CBNs is highly sensitive to **Market Price**, **Manufacturing Capacity**, and **Consumer Preference**, while BS is highly sensitive to **Market Price**, **Manufacturing Capacity**, **Consumer Preference**, **Trading Volume**, and **ASCs Situation**. The number of variables reflects upon resources and processing time.

The first three parameters from the models show that *equilibrium* has converged to *zero*. It means that changes to **Manufacturing Capacity**, **Consumer Preference**, and **Market Price** cause **ASCs Situation** to become unbalanced (*¬equilibrium*, *shortage*, or *surplus*). The posterior distributions of **ASCs Situation** for both CBNs and BS are highly sensitive to **Market Price**. Experts understand that consumer and supplier behaviors are principal factors affecting **ASCs Situation**, and so BS and CBNs can help people interpret events using something close to expert reasoning.



** **TV**: Trading Volume, **CP**: Consumer Preference, **MC**: Manufacturer Capacity, **MP**: Market Price

FIGURE 3. Tornado diagrams for CBNs and BS for the sensitivity of the base case.

The last two parameters in the CBNs and BS models are different for **ASCs Situation**. The CBNs is highly sensitive to demand (**Consumer Preference**), supply (**Manufacturer Capacity**), and price (**Market Price**), but the BS model is sensitive to **Trading Volume** because the training data may provide high correlations which lets the BS connect it, contradicting human understanding. Indeed, this relationship is considered an irrational explanation because **Trading Volume** is never used to explain **ASCs Situation**. Experts understand that **Trading Volume** is the root cause of **ASCs Situation** that transfers its effect through **Future Market**, **Exporting Costs**, and **Consumer Preference**. This is the situation for the CBNs which show that the *equilibrium* state of the **ASCs Situation** is sensitive to changes in **Manufacturing Capacity**, **Consumer Preference**, and **Market Price**. The sensitivity represents how domain experts view environment changes, and what they should consider adjusting first.

C. DISCUSSION

The experiments show that CBNs provide predicted outcomes and also relevant parameters to help decision makers understand the ASCs situation.

The first experiment confirms that the CBNs model has satisfactory performance in a market situation. CBNs can reach an accuracy of around 95%, which works well within traditional supply chain management, where many companies employ experts to examine the probabilities of *shortage* or *surplus*. However, small companies lack this expertise, which makes their analysis much more labor-intensive and time-consuming.

Recent experiment of model performance employs basic models and the FM of Equilibrium is 0.86, which is quite low. In the future, we plan to use a dynamic CBNs to improve our model performance and may compare with advanced model, such as random forest, gradient boosting, and deep learning.

The second experiment shows that CBNs offer a new dimension of decision support for supply chain management. It provides market interpretable explanations based on cause-and-effect, which is needed by companies.

VI. CONCLUSIONS

This study has proposed a Causal Bayesian Networks (CBNs) model for supporting market understanding. It produces reasonable explanations to aid decision makers dealing with ASC demand and supply uncertainty by interpreting contextual information based on big observational data.

We compared standard machine learning models (Naïve Bayes, Neural Networks, Support Vector Machines, Decision Trees, and Bayesian Search) to our CBNs model. Their performances for predicting unknown events were over 90%, but our model could reach around 95%. Sensitivity analysis confirmed that the CBNs model could produce reasonable descriptions of expert reasoning and that the model was sensitive to contexts utilized by experts. Our model can help decision makers better understand agricultural supply chain situations and successfully adjust supply chain mechanisms.

However, CBNs based on expert knowledge have a subjective quality, which means that markets with different supply chain characteristics will need to adjust the CBNs' causal assumptions, and re-tune parameters with different historical data. In future work, we will examine other market elements, discover additional causal assumptions, and address the issue of exponential numbers of relevant random variables. The ongoing ASCs will grow continuously and modernly and generate many data covering rich and exciting events for better ASCs management. More data create more factors and opportunities to run ASCs management with better performance, which may be beyond the labor work in expert-based modeling. We then hope to perform studying on automatic algorithms to learn more inclusive knowledge from the more modern ASCs and compare it with our proposed gold standard. We plan to employ causal discovery algorithms to determine causal relationships in the CBNs model to reduce labor-intensive and time-consuming tasks.

REFERENCES

[1] A. Qazi, A. Dickson, J. Quigley, and B. Gaudenzi, “Supply chain risk network management: A Bayesian belief network and expected utility based approach for managing supply chain risks,” *International Journal of Production Economics*, vol. 196, pp. 24–42, Feb. 2018, doi: 10.1016/j.ijpe.2017.11.008.

[2] Y. Huang, Z. xin CHEN, T. YU, X. zhi HUANG, and X. fa GU, “Agricultural remote sensing big data: Management and applications,” *Journal of Integrative Agriculture*, vol. 17, no. 9. Elsevier, pp. 1915–1931, Sep. 01, 2018. doi: 10.1016/S2095-3119(17)61859-8.

[3] S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, “Big Data in Smart Farming – A review,” *Agricultural Systems*, vol. 153, pp. 69–80, 2017, doi: 10.1016/j.agsy.2017.01.023.

[4] R. Toorajipour, V. Sohrabpour, A. Nazarpour, P. Oghazi, and M. Fischl, “Artificial intelligence in supply chain management: A systematic literature review,” *Journal of Business Research*, vol. 122, pp. 502–517, Jan. 2021, doi: 10.1016/j.jbusres.2020.09.009.

[5] 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 and Operations Research*, vol. 119, p. 104926, Jul. 2020, doi: 10.1016/j.cor.2020.104926.

[6] S. Punia, S. P. Singh, and J. K. Madaan, “A cross-temporal hierarchical framework and deep learning for supply chain forecasting,” *Computers and Industrial Engineering*, vol. 149, p. 106796, Nov. 2020, doi: 10.1016/j.cie.2020.106796.

[7] F. Bu and X. Wang, “A smart agriculture IoT system based on deep reinforcement learning,” *Future Generation Computer Systems*, vol. 99, pp. 500–507, Oct. 2019, doi: 10.1016/j.future.2019.04.041.

[8] S. Hosseini and D. Ivanov, “Bayesian networks for supply chain risk, resilience and ripple effect analysis: A literature review,” *Expert Systems with Applications*, vol. 161, p. 113649, Dec. 2020, doi: 10.1016/j.eswa.2020.113649.

[9] G. Ji, L. Hu, and K. H. Tan, “A study on decision-making of food supply chain based on big data,” *Journal of Systems Science and Systems Engineering*, vol. 26, no. 2, pp. 183–198, 2017, doi: 10.1007/s11518-016-5320-6.

[10] E. Bareinboim and J. Pearl, “On Pearl’s Hierarchy and the Foundations of Causal Inference,” 2020.

[11] X. Zhao, R. Lovreglio, and D. Nilsson, “Modelling and interpreting pre-evacuation decision-making using machine learning,” *Automation in Construction*, vol. 113, p. 103140, May 2020, doi: 10.1016/j.autcon.2020.103140.

[12] P. Danese, M. Molinaro, and P. Romano, "Investigating fit in supply chain integration: A systematic literature review on context, practices, performance links," *Journal of Purchasing and Supply Management*, vol. 26, no. 5, p. 100634, Dec. 2020, doi: 10.1016/j.pursup.2020.100634.
[13] L. Zhou and J. Huang, "Contagion of future-level sentiment in Chinese Agricultural Futures Markets," *Pacific Basin Finance Journal*, vol. 61, p. 101316, Jun. 2020, doi: 10.1016/j.pacfin.2020.101316.
[14] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 5th ed. Pearson Education, Inc., 2013.
[15] M. J. McGregor, M. F. Rola-Rubzen, and R. Murray-Prior, "Micro and macro-level approaches to modelling decision making," *Agricultural Systems*, vol. 69, no. 1–2, pp. 63–83, Jul. 2001, doi: 10.1016/S0308-521X(01)00018-X.
[16] Y. Zhan and K. H. Tan, "An analytic infrastructure for harvesting big data to enhance supply chain performance," *European Journal of Operational Research*, vol. 281, no. 3, pp. 559–574, Sep. 2020, doi: 10.1016/j.ejor.2018.09.018.
[17] IBM Institute for Business Value, "Smarter supply chains for an unpredictable world," 2020.
[18] A. C. Kappelman and A. K. Sinha, "Optimal control in dynamic food supply chains using big data," *Computers and Operations Research*, vol. 126, p. 105117, Feb. 2021, doi: 10.1016/j.cor.2020.105117.
[19] J. Oh and B. Jeong, "Tactical supply planning in smart manufacturing supply chain," *Robotics and Computer-Integrated Manufacturing*, vol. 55, pp. 217–233, Feb. 2019, doi: 10.1016/j.rcim.2018.04.003.
[20] B. B. Gardas, R. D. Raut, N. Cheikhrouhou, and B. E. Narkhede, "A hybrid decision support system for analyzing challenges of the agricultural supply chain," *Sustainable Production and Consumption*, vol. 18, pp. 19–32, Apr. 2019, doi: 10.1016/j.spc.2018.11.007.
[21] S. Kumar Sharma and S. Sharma, "Developing a Bayesian Network Model for Supply Chain Risk Assessment," *Supply Chain Forum*, vol. 16, no. 4, pp. 50–72, Jan. 2015, doi: 10.1080/16258312.2015.11728693.
[22] M. Chhimwal, S. Agrawal, and G. Kumar, "Measuring circular supply chain risk: A bayesian network methodology," *Sustainability (Switzerland)*, vol. 13, no. 15, pp. 1–22, 2021, doi: 10.3390/su13158448.
[23] J. M. Lawrence, N. U. Ibne Hossain, R. Jaradat, and M. Hamilton, "Leveraging a Bayesian network approach to model and analyze supplier vulnerability to severe weather risk: A case study of the U.S. pharmaceutical supply chain following Hurricane Maria," *International Journal of Disaster Risk Reduction*, vol. 49, no. April, p. 101607, 2020, doi: 10.1016/j.ijdrr.2020.101607.
[24] R. Ojha, A. Ghadge, M. K. Tiwari, and U. S. Bititci, "Bayesian network modelling for supply chain risk propagation," *International Journal of Production Research*, vol. 56, no. 17, pp. 5795–5819, 2018, doi: 10.1080/00207543.2018.1467059.
[25] S. El Amrani *et al.*, "Modelling and assessing sustainability of a supply chain network leveraging multi Echelon Bayesian Network," *Journal of Cleaner Production*, vol. 302, 2021, doi: 10.1016/j.jclepro.2021.126855.
[26] S. Chiappa and W. S. Isaac, "A causal bayesian networks viewpoint on fairness," in *IFIP Advances in Information and Communication Technology*, Jul. 2019, vol. 547, pp. 3–20. doi: 10.1007/978-3-030-16744-8_1.
[27] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Communications of the ACM*, vol. 62, no. 3, pp. 54–60, 2019, doi: 10.1145/3241036.
[28] B. Scholkopf *et al.*, "Toward Causal Representation Learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 612–634, Feb. 2021, doi: 10.1109/JPROC.2021.3058954.
[29] J. Pearl, M. M. Glymour, and N. P. Jewell, *Causal inference in statistics: a primer*, 1st ed., vol. 53, no. 9. John Wiley & Sons, Ltd, 2016.
[30] J. Pearl, "An introduction to causal inference," *International Journal of Biostatistics*, vol. 6, no. 2. Berkeley Electronic Press, 2010. doi: 10.2202/1557-4679.1203.
[31] D. Malinsky and D. Danks, "Causal discovery algorithms: A practical guide," *Philosophy Compass*, vol. 13, no. 1, p. e12470, Jan. 2018, doi: 10.1111/phc3.12470.
[32] X. Shen *et al.*, "Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-59669-x.
[33] J.-P. P. Belaud, N. Prioux, C. Vialle, and C. Sablayrolles, "Big data for agri-food 4.0: Application to sustainability management for by-products supply chain," *Computers in Industry*, vol. 111, pp. 41–50, Oct. 2019, doi: 10.1016/j.compind.2019.06.006.
[34] D. Truong, "Using causal machine learning for predicting the risk of flight delays in air transportation," *Journal of Air Transport Management*, vol. 91, no. May 2020, p. 101993, 2021, doi: 10.1016/j.jairtraman.2020.101993.
[35] M. A. Kłopotek, "Too Fast Causal Inference under Causal Insufficiency," *arXiv*, May 2018.
[36] J. Pearl and D. Mackenzie, "The book of why: the new science of cause and effect," *Basic Books*, p. 402, 2018.
[37] S. Arunwarakorn, K. Suthiwartnarueput, and P. Pornchaiwiseskul, "Forecasting equilibrium quantity and price on the world natural rubber market," *Kasetsart Journal of Social Sciences*, vol. 40, pp. 8–16, Sep. 2017, doi: 10.1016/j.kjss.2017.07.013.
[38] J. Chanchaichujit and J. F. Saavedra-Rosas, "The Elements of the Natural Rubber Industry Supply Chain," in *Using Simulation Tools to Model Renewable Resources*, Cham: Springer International Publishing, 2018, pp. 19–40. doi: 10.1007/978-3-319-55816-5_2.
[39] R. Dejchanchaiwong, A. Kumar, and P. Tekasakul, "Performance and economic analysis of natural convection based rubber smoking room for rubber cooperatives in Thailand," *Renewable Energy*, vol. 132, pp. 233–242, Mar. 2019, doi: 10.1016/J.RENENE.2018.07.145.
[40] Thai Meteorological Department, "Thailand Meteorological Department API : TMDAPI." https://data.tmd.go.th/api/index1.php (accessed May 31, 2022).
[41] M. Kliangkhlao and S. Limsiroratana, "Harnessing the power of big data digitization for market factors awareness in supply chain management," *Multimedia Tools and Applications*, 2022, doi: https://doi.org/10.1007/s11042-022-13309-w.
[42] Thai Meteorological Department, "Climatological Center," 2021. http://climate.tmd.go.th/gge/index.html (accessed Jul. 28, 2021).
[43] Office of Agricultural Economics, "Agricultural Production Data," 2021. https://www.oae.go.th/ (accessed Jul. 28, 2021).
[44] Office Of The Rubber Replanting Aid Fund, "Thailand Daily Rubber Price," 2021.

https://www.raot.co.th/ewtadmin/ewt/rubber_eng/main.php?filename=index (accessed Jul. 28, 2021).
[45] Bank of Thailand, "Daily Foreign Exchange Rate," 2021. https://www.bot.or.th/english/_layouts/application/exchangerate/exchangerate.aspx (accessed Jul. 28, 2021).
[46] Markets Insider, "Crude Oil Price Today | BRENT OIL PRICE CHART," 2021. https://markets.businessinsider.com/commodities/ oil-price(accessed Jul. 28, 2021).
[47] Tokyo Commodity Exchange Inc., "TOCOM : Quotes for RSS3 Rubber," 2021. https://www.tocom.or.jp/market/kobetu/rubber.html (accessed Jul. 28, 2021).
[48] D. Barber, *Bayesian Reasoning and Machine Learning*, 1 st. United Kingdom: Cambridge University Press, 2011. doi: 10.1017/cbo9780511804779.
[49] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Computational Statistics*, vol. 36, no. 3, pp. 2009–2031, Jun. 2021, doi: 10.1007/s00180-020-00999-9.
[50] Scikit-learn, "scikit-learn: Machine Learning in Python," 2022. https://scikit-learn.org/stable/ (accessed May 28, 2022).
[51] B. Sahoh and A. Choksuriwong, "A proof-of-concept and feasibility analysis of using social sensors in the context of causal machine learning-based emergency management," *Journal of Ambient Intelligence and Humanized Computing*, vol. 1, p. 3, May 2021, doi: 10.1007/s12652-021-03317-3.
[52] U. Kjærulff and L. C. van der Gaag, "Making Sensitivity Analysis Computationally Efficient," *UNCERTAINTY IN ARTIFICIAL INTELLIGENCE PROCEEDINGS*, p. 317, Jan. 2013.
[53] D. Rodríguez, E. Brunesi, and R. Nascimbene, "Fragility and sensitivity analysis of steel frames with bolted-angle connections under progressive collapse," *Engineering Structures*, vol. 228, p. 111508, 2021, doi: https://doi.org/10.1016/j.engstruct.2020.111508.

MALLIKA KLIANGKHLAO received her B.S. degree in information and communication technology from Silpakorn University, Thailand, in 2010, and the M.Sc. degree in information technology from Thaksin University, Thailand, in 2014. She is currently pursuing a Ph.D. degree in computer engineering, Prince of Songkla University, Thailand. Her research interests cover various topics, including causal inference, explainable artificial intelligence, and modern supply chain management.

SOMCHAI LIMSIRORATANA received the B.E.E degree in electrical engineering from Prince of Songkla University (PSU), Songkla, Thailand, in 1991. M.A.Sc and Ph.D. in agricultural science from the Division of Environmental Science and Technology, Kyoto University, Japan, in 2000 and 2005, respectively. He is currently a lecturer with the Department of Computer Engineering, PSU, Songkla, Thailand. His research interests are agricultural image processing, medical image processing, digital watermarking and artificial intelligence.

BUKHOREE SAHOH received his B.S. degree in computer science from Prince of Songkla University (PSU), Songkla, Thailand, in 2010, and an M.Sc. degree in information technology from Thaksin University, Phatthalung, Thailand, in 2015. Bukhoree Sahoh earned his Ph.D. in computer engineering at PSU, in 2020. He is currently a lecturer with the School of Informatics, Walailak University (WU), Nakhon Si Thammarat, Thailand, and a member of the Informatics Innovation Center of Excellence (IICE) at WU. His current research emphasizes natural language processing, complex event understanding, explainable artificial intelligence, and causality in machine learning.

VITAE

Name Miss Mallika Kliangkhlao

Student ID 6010130024

Educational Attainment

Degree	Name of Institution	Year of Graduation
Bachelor of Science (Information Technology for Design)	Silpakorn University	2011
Master of Science (Information Technology)	Thaksin University	2014

Scholarship Awards during Enrollment

Thailand Education Hub for ASEAN Scholarship in Doctoral Degree funded by Prince of Songkla University, Thailand under Grant No. TEH-233/2016.

List of Publications and Proceedings

M. Kliangkhlao and S. Limsiroratana, "Towards the idea of agricultural market understanding for automatic event detection." In *Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pp. 81-86. 2019.

M. Kliangkhlao and S. Limsiroratana, “Harnessing the power of big data digitization for market factors awareness in supply chain management,” *Multimedia Tools and Applications*, 2022, doi: 10.1007/s11042-022-13309-w.

M. Kliangkhlao, S. Limsiroratana and B. Sahoh, "The Design and Development of a Causal Bayesian Networks Model for the Explanation of Agricultural Supply Chains," in IEEE Access, vol. 10, pp. 86813-86823, 2022, doi: 10.1109/ACCESS.2022.3199353.