ISSN: 2710-2599

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Data Science and Big Data Analytics

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper

Open Access

Supply Chain Management Techniques Using Big Data for Agro-Based Food Products in Bangladesh

Md. Miftahul Alam[1] and Akinul Islam Jony[2*]

[1]Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh. E-mail: 18-38839-3@student.aiub.edu

[2]Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh. E-mail: akinul@aiub.edu

Article Info

Volume 3, Issue 2, November 2023

Received : 22 August 2023

Accepted : 19 October 2023

Published : 05 November 2023

doi: 10.51483/IJDSBDA.3.2.2023.19-34

Abstract

Agriculture is the primary source of income for 75% of Bangladesh's rural population. The agro-based food industry faces Supply Chain Management (SCM) issues, including inadequate infrastructure, insufficient logistics, and limited access to accurate information. Big data can help overcome these challenges by providing precise and up-to-date information for decision-making procedures. This study aims to develop industrial application frameworks to harness big data's power to solve agro-food supply chain problems in Bangladesh. A conceptual big data architecture model is proposed to address the problems faced by all stakeholders, including farmers, processors, wholesalers, and retailers. The proposed model will contribute to the knowledge on big data's application to supply chain management and offer recommendations for enhancing Bangladesh's agro-based food sector.

Keywords: *Agriculture, Big data, Agro food, Supply chain management, Data analytics Architecture, Machine learning*

1. Introduction

With 42.7% of the workforce employed and 14.2% of Bangladesh's GDP in 2017, agriculture is the country's largest employer (CIA, 2023). Bangladesh is an agricultural country and so its major workforce directly and indirectly involved with agriculture. It employs around 40.6% of the total rural employment (Agricultural Labour Force Survey 2016-17). So, for all this reason and food being the basic right of human being, agro-food supply chain is crucial sector for Bangladesh present and future wellbeing.

The most common subject of this research study is agro-based food which is a fundamental need of human life. It is essential for human growth, development, and survival, providing nutrients, energy, and

* *Corresponding author: Akinul Islam Jony, Department of Computer Science, American International University-Bangladesh, Dhaka 1229, Bangladesh. E-mail: akinul@aiub.edu*

sustenance. Access to safe, nutritious food is a fundamental human right, ensuring food security and eradicating hunger globally. It also impacts culture, traditions, and social gatherings, promoting good health and well-being (Marvin *et al.*, 2017).

Besides, inefficiencies and knowledge asymmetry in Bangladesh's agriculture supply chain led to high transaction costs and poor product quality. In addition to these, infrastructure issues, especially in rural areas, cause significant post-harvest losses. Middlemen in the supply chain result in high transaction costs and low farmer returns. But the most challenging part is getting information, analyzing the present and predicting the future. If Bangladesh Govt. and food-based conglomerates had the power to collect all data throughout the process and stages to control and understand the market, it would be easier to take cost-effective and long-term initiatives to address the problem. In summary, it can be said that the use of big data analysis can improve supply chain management and demand-wise forecasting zone for agro-based food products in Bangladesh, leading to a more efficient and profitable agro-industry.

To attain this goal, we will address techniques and propose a conceptual model of a big data architecture which will help policymakers to make quick decisions for helping agro-based food supply chain management and minimize waste or loss.

2. Background

2.1. Overview of Agro-Based Food Products in Bangladesh

Demand for high-quality agricultural products has been spurred by a growing middle class (estimated at over 30 million (International Trade Administration, 2022). According to the Bangladesh Bureau of Statistics (*Yearbook of Agricultural Statistics*, 2019), various crops, fisheries, forest products, and livestock are accounted for approximately 12.5% of Bangladesh's total GDP and employed about 40% of the entire population in FY 2020-21 (*Yearbook of Agricultural Statistics*, 2019). The majority of Bangladesh's agricultural output is classified as traditional subsistence farming. Bangladesh is known for producing a substantial amount of agricultural goods, including dairy products, grain food (such as rice, wheat, corn), different types of beans, various fruits and vegetables, meat (chicken, beef, mutton), fish, and other seafood.

Bangladesh's main crops include rice, wheat, mango, and jute, with rice being the main crop due to its soil and water supply (*Yearbook of Agricultural Statistics*, 2019). But the primary dietary component of Bangladeshis is rice. Other domestic crops include potatoes, sweet potatoes, oil seeds, and fruits. Sugarcane is grown and refined for raw sugar. However, pesticide use poses environmental and financial risks. Wheat is usually imported with assistance from the US and EU.

Table 1 presents the toal market size for agricultural products from 2016 to 2021. The amount of each year is presented for toal local production, total exports, total imports in $ mn unit. The toal market size is calculated by using the following formula.

Total Market Size = [(Total Local Production + Imports) – Exports]

On the other hand, Table 2 shows the total food grain production from 2014-15 to 2018-19 for Aus, Aman, Boro, total rice, wheat, and Maize.

Table 1: Total Market Size for Agricultural Products: 2016-2021($)

	2016	2017	2018	2019	2020	2021
Total Local Production	30,424	33,374	35,882	38,095	37,568	38,801
Total Exports	1488	1625	1418	1600	1343	1586
Total Imports	5991	9470	5952	8000	8442	10334
Total Market Size	34,927	41,219	40,417	44,495	44,667	47,549

Source: International Trade Administration (2022)

Table 2: Total Food Grain Production: 2014-2018 (Metric Tons)

Food Grains	2014-15	2015-16	2016-17	2017-18	2018-19
Aus	23.28	22.89	21.34	27.09	27.02
Aman	131.90	134.83	136.56	139.94	141.34
Boro	191.92	189.93	180.16	195.76	196.23
Total Rice	347.10	347.18	338.06	362.79	364.59
Wheat	13.48	13.48	13.12	11.53	12.87
Maize	23.61	388.17	35.78	38.93	38.28

Source: Yearbook of Agricultural Statistics (2019)

2.2. *Big Data and its Importance in Supply Chain for Agro-Based Food Products of Bangladesh*

According to Subudhi *et al.* (2019), big data is defined as: *"Conglomeration of the booming volume of heterogeneous data sets, which is so huge and intricate that processing it becomes difficult, using the existing database management tools."*

Big data refers to the huge amount of structured and unstructured data generated by businesses, individuals, and other organizations on a daily basis. It is complex and diverse, requiring processing at a high speed (Jony, 2016). The importance of big data lies in its ability to help organizations make better decisions by providing insights that would be difficult to obtain otherwise (Morshed and Jony, 2016). Big data is crucial for supply chain management due to its ability to improve decision-making (Tanvir *et al.*, 2023), forecasting (Lisun-Ul-Islam *et al.*, 2023), efficiency, productivity (Shanto *et al.*, 2023), and cost savings (Shovon *et al.*, 2022). Agriculture stakeholders can use big data to identify patterns and trends in enormous amounts of data, enabling them to make informed decisions. This data-driven approach helps them make more accurate predictions about future (Sultana *et al.*, 2022) demand and adjust their operations accordingly. Additionally, big data helps identify bottlenecks and inefficiencies, enabling them to make necessary changes to improve efficiency and productivity. By analyzing data on various aspects of SCM, big data can help organizations streamline operations and reduce waste and leading to improved overall efficiency and productivity. The majority of the farmers in Bangladesh are smallholders, with farms ranging from 0.1 to 1 hectare (*Yearbook of Agricultural Statistics*, 2019).

Bangladesh faces challenges in its agro-based food supply chain, including poor infrastructure, inadequate storage facilities, limited credit access, and technological advancement. Natural disasters and climate change also disrupt the supply chain. The government has implemented initiatives to improve infrastructure, provide financing for small farmers, and promote technology adoption. Big Data can help understand demand and supply trends, reduce waste, and control product prices.

3. Literature Review

3.1. *Overview of Supply Chain Management and Big Data*

As per the traditional definition proposed by Christopher (1992), *"A supply chain is the network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services delivered to the ultimate consumer."*

In the context of the current business environment, SCM has emerged as a vital component in gaining a competitive advantage. As customer expectations, complexity, and technological advancements increase, SCM practices have evolved from cost efficiency to responsiveness and agility. Member collaboration is vital for success in an unpredictable business climate. An example of supply chain management procedure for products is depicted in Figure 1.

Chen *et al.* (2012) identified three stages in the evolution of business intelligence and analytics: Level 1 relies on structured data and DBMS systems, while Level 2 uses web and unstructured data, and Level 3 uses mobile and sensor-based content. Big Data analytics is crucial for levels 2 and 3, focusing on predictive analytics. Big Data analytics architecture is mainly influenced by factors like data volume, sources of data, quality of data, latency, throughput, data security and privacy, and cost issues. Creating a cloud-based integrated framework for tracking and enhancing firm performance in complex environments is essential.

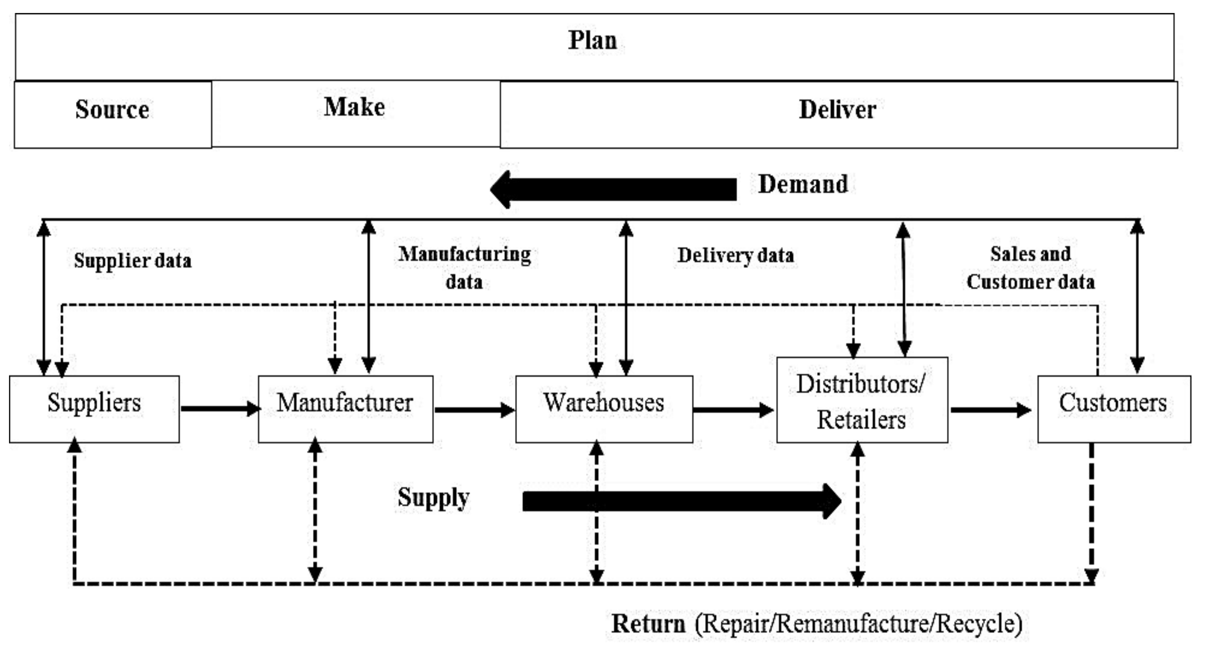

Figure 1: Supply Chain Management Procedure for Products

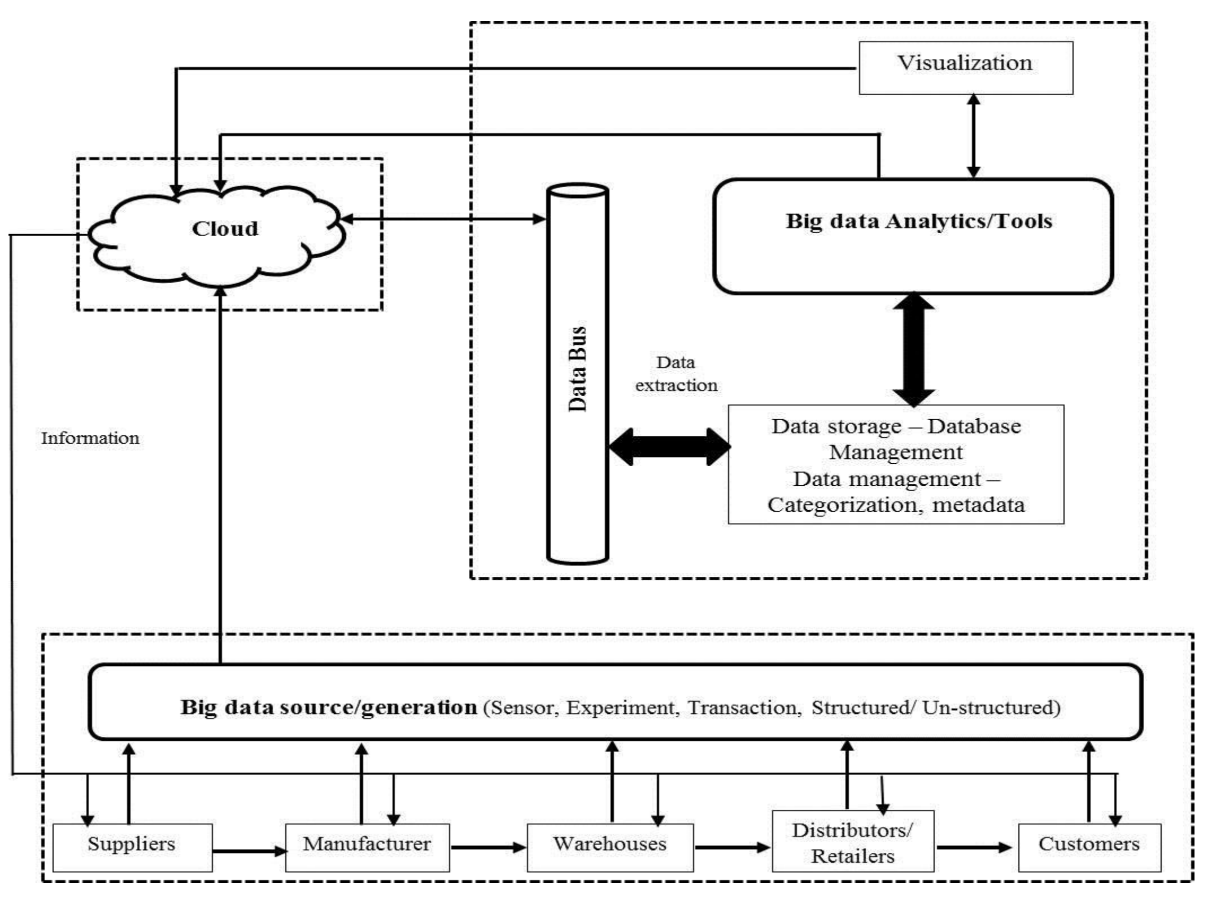

Figure 2: Architecture of Six Components (Proposed by Chan (2013) and Reviewed by Biswas (2016)

Figure 2 depicts an architecture consists of six components: (i) data collection devices (such as sensor); (ii) cloud system and infrastructure; (iii) data bus; (iv) data storage and database management system; (v) data analytics engine or tools; and (vi) data visualization and rendering system.

3.2. Theoretical Frameworks for Big Data Architecture in Agro-Based Food Products Supply Chain

Businesses may get insights into customer behavior, supply chain efficiency, and market trends by combining this data (Ray *et al.*, 2017). Big data-driven supply chain management techniques can revolutionize the agro-based food product industry by improving efficiency, inventory management, and reducing costs. One of the most commonly used SCM techniques that leverage big data for agro-based food products is predictive analytics (Valencia *et al.*, 2022). By integrating data from various sources, predictive analytics is crucial for informed decisions, anticipating future events, and enhancing supply chain planning (Cravero *et al.*, 2022). Businesses may get insights into customer behavior, supply chain efficiency, and market trends by combining this data (Ray *et al.*, 2017; Belaud *et al.*, 2019). Inventory management optimizes levels by reviewing sales trends, reducing excess inventory and costs. Procurement optimization through mobile apps helps identify the best suppliers, and zone-wise consumption vs. production estimation helps optimize production and distribution methods for timely delivery in strong regions. For enterprises in the agro-based food product market, social media may be a valuable source of data (Cécile *et al.*, 2022).

The big data architecture framework for agro-based food products supply chain includes data sources, collection, processing, visualization, and security. Data sources include IoT devices [25], weather stations, mobile apps, social media, and ERP systems. Data collection uses ETL, ingestion, and cleansing technologies, while processing utilizes Hadoop, Spark, and NoSQL. Another SCM technique that leverages big data is blockchain technology (Tarun *et al.*, 2021).

Data visualization helps identify trends, monitor performance, and make data-driven decisions. Data security ensures data integrity throughout the supply chain. Collaboration between stakeholders, such as farmers, suppliers, wholesalers, retailers, and end-users, is facilitated through mobile apps and social media tools.

4. Methodology

4.1. Research Design and Approach

The step-by-step research design and approach is described below as well as depicted in Figure 3.

Step 1: Problem Identification

The first step in research is identifying the problem to effectively manage the agro-based food industry's supply chain using big data architecture. This involves addressing challenges like data collection, processing, analysis, and security, while identifying key stakeholders like farmers, wholesalers, retailers, and processors and understanding their unique needs and challenges.

The problems identified are (i) inadequate information flow recording; (ii) lack of technology infrastructure for real-time analysis; (iii) lack of tools for supply-demand forecasting, and price control.

The proposed conceptual Model will cater to all the above needs and also help farmers, processors, wholesalers, retailers, FMCG Companies and GOVT. Regulators to take data driven decisions. Benefits of the conceptual model are:

1. Increased Knowledge and real time Insights of the market
2. Power to improved decision-making
3. Reliable Forecasting
4. Waste Minimization

Step 2: Primary Information and Analysis

Gathered primary information through surveys, interviews, and observation to evaluate supply chain management practices in Bangladesh's agro-based food industry and identified areas for big data architecture designing.

Step 3: Secondary Information and Analysis

Gathered secondary information through literature reviews, case studies, and other sources to understand big data architecture in supply chain management.

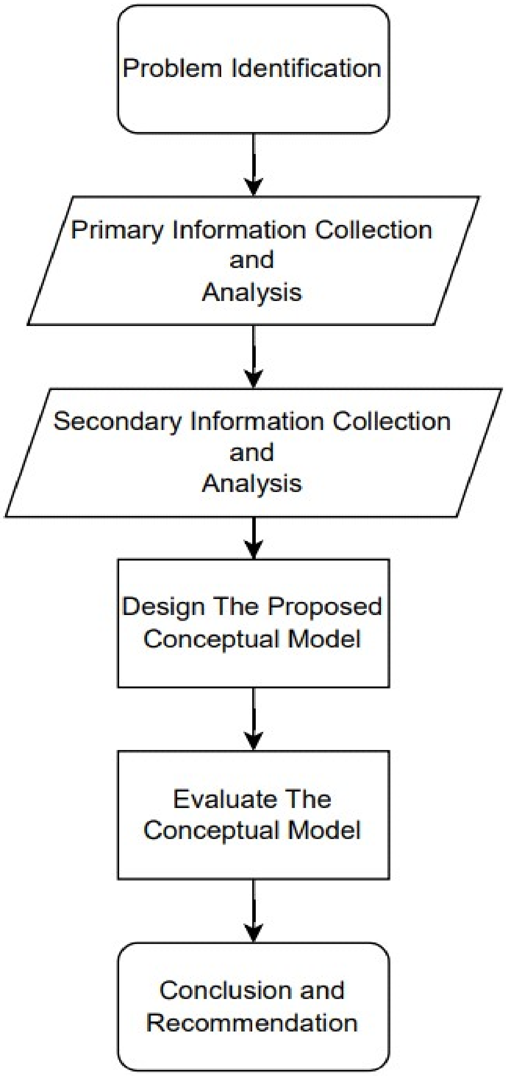

Figure 3: Research Methodology Process

We did thorough analysis of Bangladesh Bureau of Statistics Reports on Agriculture, Ministry of Agriculture reports, Business reports by different FMCG companies. We also studied different research papers from around the globe.

Step 4: Design Proposed Conceptual Model

Based on the primary and secondary information gathered, the next step is to design a proposed conceptual model for implementing big data architecture in SCM for the agro-based food industry in Bangladesh. This involves identifying the key data sources, data processing techniques, data visualization tools, and data security measures required to effectively manage the supply chain.

Step 5: Evaluate the Conceptual Model

After designing the proposed conceptual model, it is important to evaluate its effectiveness in addressing the identified problem.

Step 6: Conclusion and Recommendation

Finally, the research concludes with a conclusion and recommendation. Lastly summarizing the findings, identifying limitations, and offering recommendations for future research and big data architecture implementation in Bangladesh's agri food supply chain.

4.2. Primary Information Collection and Analysis

The study collected data through surveys, interviews, and public opinions in Bangladesh. The study in Bangladesh found that weekly shoppers prioritize food quality over price, facing challenges in buying fresh fruits and vegetables. Challenges include price hikes, middleman involvement, lack of collaboration, technology company involvement, and infrastructure issues. Proper detail-oriented analysis and visualization are needed for effective action.

4.3. Secondary Information Collection and Analysis

Secondary data is information collected from government agencies, research institutions, and private companies. It includes government reports, industry reports, academic publications, news articles, and online databases. Government reports and statistics, industry reports, consulting firms, trade associations, and research organizations provide data and insights on the agro-based food industry in Bangladesh. Academic publications, news articles, and online databases provide valuable insights into the industry's trends, challenges, and developments. These sources help researchers understand the broader context of the agro-based food industry in Bangladesh. Figure 4 shows the country wise number of publications relating to agro-based food supply chain.

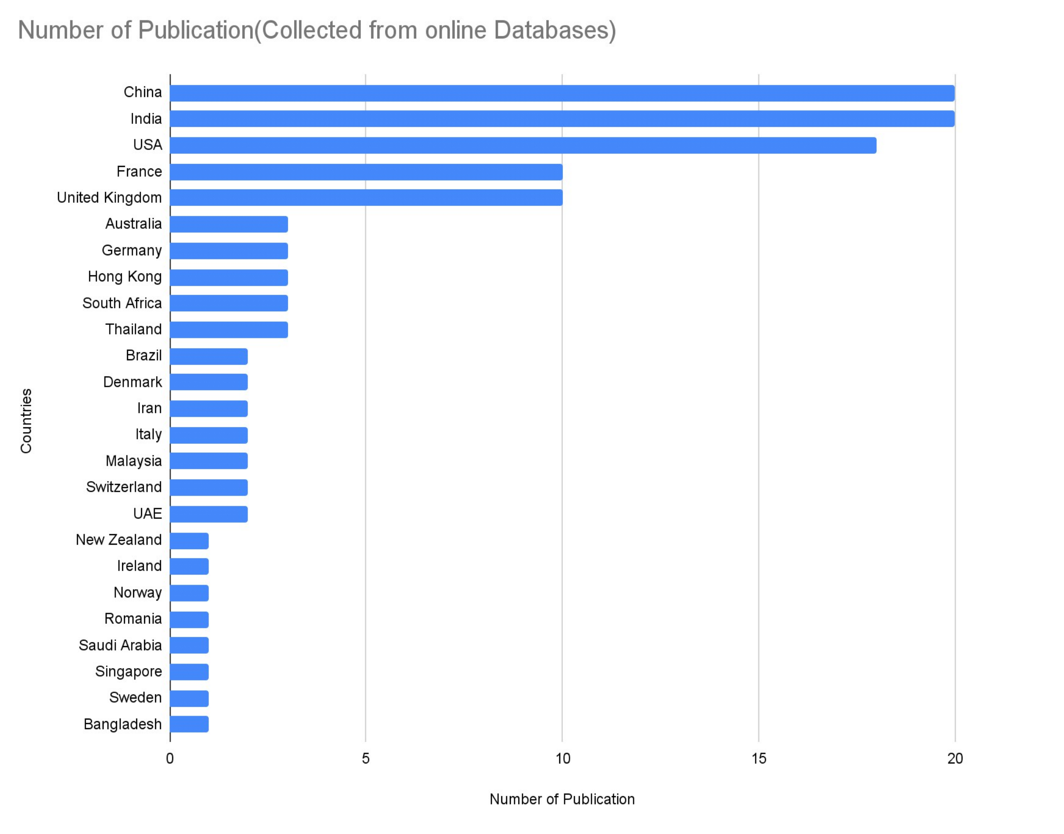

Figure 4: Number of Publications Reviewed Relating to Agro-Based Food Supply Chain by Country

5. Proposed Conceptual Model of the Big Data Architecture

5.1. Introduction to the Proposed Conceptual Model

According to IEEE Standard 1471-2000, "Architecture is the fundamental organization of a system embodied in its components, their relationships to each other and to the environment and the principles guiding its design and evolution" (Maier *et al.,* 2001).

The final part of the whole research work is to propose our model. Our Proposed Big data architecture conceptual model has nine layers which will perform the entire process with a segmented approach to ensure optimum velocity, manage volume and produce quality insights.

The followings are the key stakeholders:

1. Farmer
2. Processor/yard
3. Wholesaler
4. Retailers
5. B2B Clients
6. Govt regulators

The farmer produces crops and raw food and then sells to processors/yard. Then comes wholesalers who buy processed crops, vegetables, and other products from processors/yard. Then they store them and sell to Retailers.

B2B clients are those parties including transportation companies, Conglomerate Companies, Independent Farm Owners, Food, and grocery Ecommerce and many advanced brands. At last but not least comes Govt. Regulators, who keeps tracks of each and every transaction and exchange of products for better understanding and regulations.

The whole product information flow between stakeholders is presented in Figure 5.

The proposed conceptual model consists of 9 layers as given below as well as depicted in Figure 6.

1. Data Source Layer
2. Data Collection layer
3. Data Storage Layer (Central Repository)
4. Data Processing Layer
5. Real Time Database Layer
6. Data Analysis and mining Layer
7. Machine Learning Layer (Predictive analytics)
8. Data Visualization Layer
9. Data Sharing Layer

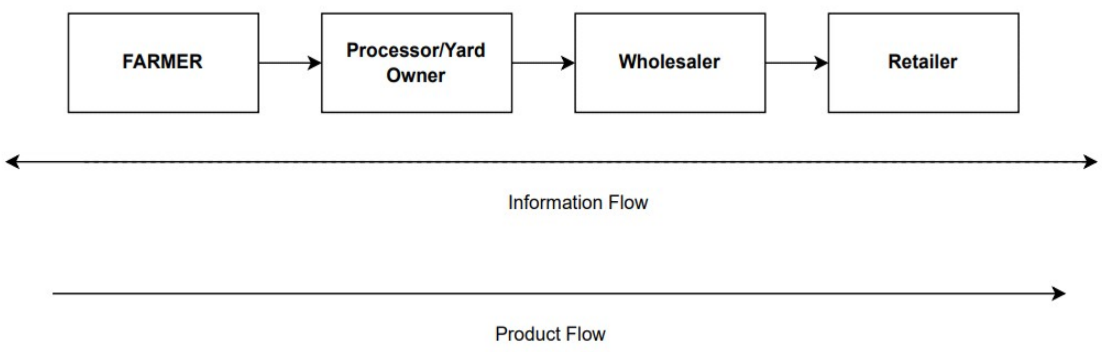

Figure 5: Product Information Flow Between Stakeholders

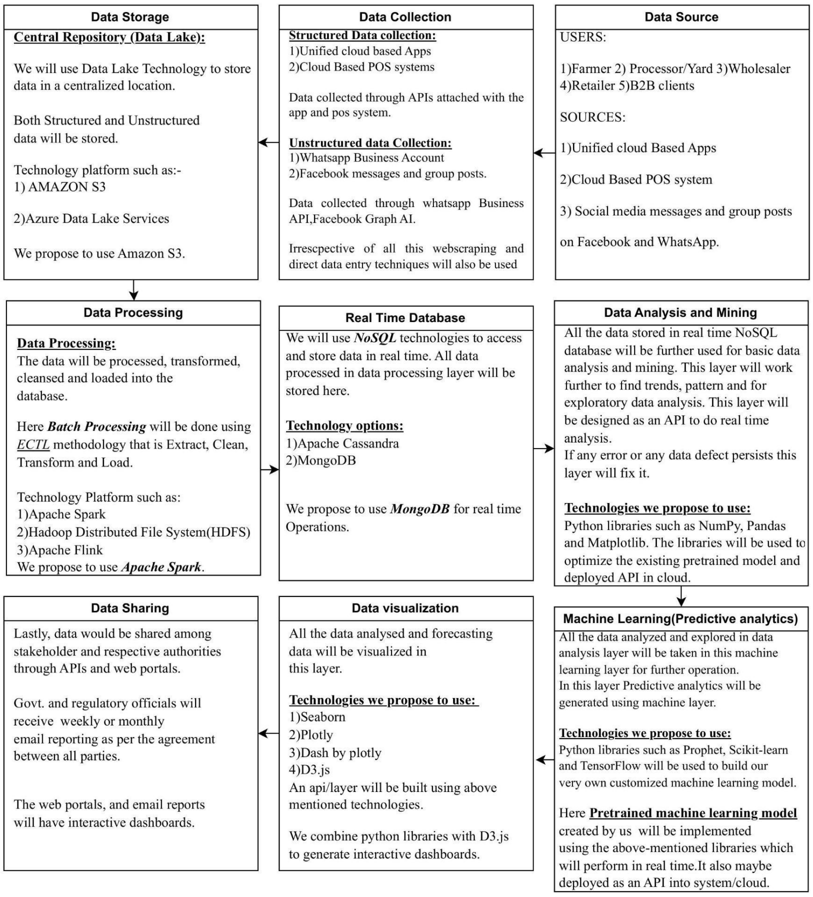

Figure 6: Process Diagram of the Conceptual Model

5.2. Explanation of the Proposed Model

5.2.1. Data Source Layer

The data source layer is the first layer of the model.

Users of the system and the data provided are given below:

1. Farmer: Data Provided: Crop/produce type, harvest date, quantity, location, sales, transaction etc.
2. Processor/Yard: Data provided: Receiving date, processing date, quantity, location, transaction etc.
3. Wholesaler: Data provided: Quantity, location, price, transaction, inventory, and sales etc.
4. Retailer: Data provided: Quantity, location, inventory, transaction, sales, price etc.

Data sources for the system include three categories:

1. Unified Cloud-Based Apps: Unified cloud-based apps are lightweight, lightweight software applications that run on the cloud, requiring minimal processing power. These apps enable stakeholders to upload data and gain insights, enabling structured information for farmers, processors, wholesalers, retailers, and B2B clients.
2. Cloud Based POS (Point-of-Sale) System: Cloud-based POS software processes payments for food products and crops, collecting data from POS systems used by retailers, wholesalers, and B2B clients. The system collects structured information on sales, inventory, and customer data.
3. Social Media Messages and Group Posts: Since it is a developing nation and a lot of people are not comfortable using apps or POS system, they can simply send messages through WhatsApp or Facebook messages. Besides, data would also be collected from Facebook posts through zone or district wise groups.

5.2.2. Data Collection Layer

The data collection layer captures and stores structured and unstructured data from various sources for processing and analysis. Structured data is easily searchable and stored in databases, while unstructured data follows predefined models.

5.2.3. Sources of Structured Data

I. Unified cloud-based Apps

II. Cloud Based POS Systems

Data will be collected through APIs attached to the app and POS system.

5.2.4. Sources of Unstructured Data

I. WhatsApp Business Account

II. Facebook Messages and group posts.

Data will be collected from social media through WhatsApp Business API and Facebook Graph API. Irrespective of all this, web scraping and direct data entry techniques will also be used.

5.2.5. Data Storage Layer

Data Lake Technology is a cost-effective and scalable solution for storing structured and unstructured data in a centralized location. It is built using cloud-based platforms like Amazon S3 or Microsoft Azure, allowing easy access from anywhere with an internet connection. The technology provides data governance, security features, and access controls to ensure proper management and protection. Available technology platforms include Amazon S3, Apache Hadoop HDFS, Microsoft Azure Data Lake, and IBM Cloud Object Storage. We propose to use Amazon S3.

5.2.6. Data Processing Layer

The data processing layer processes collected data from sources and stores it in a data lake. It uses the ECTL methodology to transform structured and unstructured data into a more useful clean format. Batch processing is used in this layer to process large amounts of data simultaneously. The ECTL methodology extracts data, cleans and transforms it into a usable format, and then loads and uploads data to the next layer. Data processing technologies available include Apache Spark, Apache Kafka, Apache Hadoop, and Apache Hive. The proposed conceptual model uses Apache Spark.

5.2.7. Real Time Database Layer

NoSQL technology is utilized for real-time data storage and access, offering high performance, scalability, and flexibility. The processing layer stores processed data in a Real Time Database Layer, which is fast, accessible, and updated for operations like data mining analysis. NoSQL databases like MongoDB, Apache Cassandra, and Amazon DynamoDB are available for high performance and real-time operation. We propose to use a document-oriented database that is MongoDB for the real time database layer of the model.

5.2.8. Data Analysis Layer

Data analysis and mining is crucial for extracting knowledge from data, analyzing real-time databases, and generating valuable insights for business processes and informed decisions. This layer identifies trends, patterns, and exploratory data analysis. It will function as an API for real-time analysis and fix errors or data defects. Python libraries like NumPy, Pandas, and Matplotlib will be used to optimize pretrained models and deploy them as an API in the cloud. Fine-tuning will be done before deployment.

5.2.9. Machine Learning Layer

Machine learning based predictive model uses historical data to predict future outcomes which can be created using Python libraries like Prophet, Scikit-learn, and TensorFlow. The agri-food supply chain's forecasting data is created using ensemble modeling to increase predictive power and reduce overfitting. Supervised learning algorithms like Random Forest predict crop production and demand zone-wise, while unsupervised learning algorithms like K-means clustering identify anomalies. A model is built using libraries/packages, pre-trained models are built as an API using flask, and the API is deployed in the cloud. Regular fine-tuning and updating of the API with new incoming data are performed to improve accuracy. Regular fine-tuning and updating of the API will ensure the proposed conceptual model's effectiveness.

5.2.10. Data Visualization Layer

The Data Visualization Layer is essential for presenting processed data using visual tools like graphs, charts, and dashboards. Utilizing tools like Seaborn, Plotly, Dash by Plotly, and D3.js, an API or layer will be developed, combining Python modules with D3.js to create interactive dashboards for quick and easy data visualization.

5.2.11. Data Sharing

The data will be shared through APIs and web portals, with regular email reports sent to government and regulatory officials. Interactive dashboards will be provided for easy access and visualization of relevant data, enabling informed decisions regarding the agri-food supply chain based on insights and predictions.

6. Evaluation of the Conceptual Model

6.1. Criteria for Evaluating the Proposed Model

The conceptual model's scalability, high performance computing, modularity, consistency, security, real-time operations, and inter-operability are crucial criteria for its success (Chan, 2013). These criteria include handling large amounts of data, automatic adaptation to increasing data volumes, high performance computing, efficient processing and analysis, security, real-time operations, and seamless integration with other systems and technologies. Al-Jaroodi and Mohamed's 2016 survey highlights the importance of these criteria in ensuring the model's ability to handle large amounts of data and adapt to changing data volumes (Gorton and Klein, 2015).

The proposed conceptual model meets the criteria of scalability, high performance computing, modularity, consistency, security, real-time operations, inter-operability, availability, and fault-tolerance. It uses a distributed architecture to handle big data sets and split the workload into smaller jobs for parallel processing. The model uses advanced computing methods, such as distributed computing and parallel processing, resulting in fast and effective data processing. Modularity allows for easy updates and maintenance, while consistency depends on data input quality and system stability. Security measures are implemented to protect against unauthorized access, data breaches, and other security issues. Real-time operations are possible for time-sensitive applications like fraud detection or predictive maintenance, but performance depends on hardware and software settings. Inter-operability is achieved through standardized interfaces and protocols, making the model easier to integrate with various technologies and systems. Availability is achieved through redundant systems, load balancing, and other strategies, ensuring users can access the model whenever they need it.

6.2. Comparison with Other Models

6.2.1. Model 1

Chan (2013) proposes a Big Data analytics architecture called Model 1, which focuses on client-server protocol and a NoSQL database architecture. Hadoop, a popular distributed file system, is used for processing massive data sets across clusters of computers. MapReduce, a distributed processing system, processes files in parallel. The Hadoop architecture consists of parallel computing systems that manage enormous volumes of data processing at a rapid rate. The data warehouse can be used with the business intelligence analytics like OLAP to improve corporate operations and decision processes. The HDFS cluster receives unstructured data, which is then fed into the MapReduce system, which distributes data processing among Hadoop clusters. MapReduce can also perform batch analytics on massive amounts of unstructured data, producing actionable information for operational and analytics applications.

6.2.2. Model 2

Model 2 (Biswas, 2016) is a Big Data analytics architecture and platform for SCM applications, based on

Chan's model (Bangladesh Bureau of Statistics, 2019). The architecture includes structured and unstructured input data from various entities, such as warehouses, manufacturers, suppliers, distributors/retailers, and customers. Structured data is extracted and loaded into a data warehouse using ETL processes, while unstructured data is managed using Hadoop Distributed File System (HDFS) and MapReduce technologies. An Operational Data Store (ODS) is used to combine data from multiple sources, executing operations like preprocessing, redundancy resolution, integrity checks, and business rule compliance. A Real-Time Intelligence (RTI) software accesses the data warehouse as a central repository, enabling fast decision-making through data federation, data virtualization, enterprise application integration, enterprise information integration, and Service-Oriented Architecture (SOA). The output can be sent to a Dimensional Data Store (DDS) for non-real-time analytics. The control system orchestrates and maintains the ETL system using the metadata specified in the data warehouse.

6.2.3. Model 3

Our proposed conceptual model consists of nine layers to improve velocity, handle volume, and create quality insights for stakeholders like farmers, processors/yards, wholesalers, retailers, B2B clients, and

Table 3: Comparison Between Model 1, Model 2 and Model 3

Aspect	**Model 1 (by Chan, 2013)**	**Model 2 (by Biswas, 2016)**	**Model 3 (The Proposed Model)**
Scope and Focus	Supply Chain management	SCM applications	Agri-food supply chain
Data Sources	Suppliers, manufacturers, warehouses, retailers	Suppliers, manufacturers, warehouses, distributors/retailers, customers	Farmers, processors/yard, wholesalers, retailers
Data Types	Structured and unstructured data	Structured and unstructured data	Structured and unstructured data
Data Processing	ETL mechanisms	ETL mechanisms for structured data, HDFS and Map Reduce for unstructured data	ECTL methodology, Apache Spark for batch processing
Data Storage	Data warehouse, operational data store (ODS)	Data warehouse, HDFS, NoSQL database	Data Lake using Amazon S3
Real-Time Intelligence	N/A	Real-time intelligence system	NoSQL database for real-time data storage
Analytics/ML	N/A	RTI module, DDS module, Data Mining module	Data analysis and mining layer, machine learning layer with Python libraries
Data Visualization	N/A	Rich visualization techniques	Seaborn, Plotly, Dash by Plotly, D3.js
Stakeholders	N/A	N/A	Multiple Stakeholders
Data Sharing	N/A	N/A	APIs, web portals, email reports

government regulators. These layers include Data Collection Source, Data Collection Process, Data Storage Management, Data Processing, Real-Time Database Management, Data Analysis and Data Mining, Machine Learning, and Data Visualization. Data Source gathers information from various sources, while Data Collection captures structured and unstructured data using APIs and social media. Data Storage stores data in a centralized repository, while Data Processing processes structured and unstructured data using ECTL approaches. Real-Time Database offers excellent performance, scalability, and adaptability. Machine Learning techniques or algorithms, such as Random Forest and K-means clustering, are used for sophisticated predictive analytics. Data Visualization utilizes visual tools like graphs, charts, and dashboards to present processed data.

A detailed comparison between Model 1, Model 2, and Model 3 is presented in Table 3.

7. Discussion

7.1. Summary of the Research Findings

The summary of the research findings is presented as follows.

- Improved Efficiency: The model employs a segmented method with nine layers, allowing for improved velocity, processing of massive amounts of data, and the generation of quality insights. This can result in increased supply chain efficiency.
- Improved Data Management: The model includes layers for data source, collection, storage, processing, analysis, machine learning, visualization, and sharing, allowing for optimal data management and usage. Better decision-making, forecasting, and planning can arise from this.
- Stakeholder Involvement: Farmers, processors, wholesalers, retailers, B2B clients, and government regulators are among the primary stakeholders involved in the concept. It provides seamless integration, user-friendly interfaces, and data sharing capabilities, boosting stakeholder engagement and transparency.
- Financial Impact: The billing structure, which is dependent on the number of users, can contribute to the model's financial viability.
- Feasibility: Economic viability, operational integration, legal compliance, social effect, and environmental sustainability are all demonstrated by the model. It has the ability to provide economic advantages, increased efficiency, transparency, and market access, while simultaneously addressing environmental issues and encouraging resource management.

7.2. Recommendations for Practitioners and Policymakers

The following are some recommendations for practitioners and policymakers.

- The agri-food supply chain must be a hub of collaboration, data sharing, information interchange, and collaborative decision-making. To achieve this, it is crucial to foster collaboration among farmers, processors, wholesalers, retailers, and policymakers, enabling data sharing, information interchange, and collaborative decision-making.
- Developing robust data infrastructure is essential for better market control and support for stakeholders, practitioners, and policymakers.
- Efficient mechanisms for data collection are crucial in developing nations like Bangladesh, as it is essential to collect quality and real data in real time for better food security and decision-making.
- Data security and privacy policies should be implemented, including data encryption, access restrictions, and compliance with data protection requirements. Data analytics techniques should be promoted, such as machine learning, predictive analytics, and optimization algorithms, to improve decision-making processes.
- Market access for farmers should be increased through the use of the big data paradigm, allowing them to access markets, connect with customers, and facilitate internet trading platforms.

- Creating a supportive policy framework and incentives can encourage the adoption of big data technologies and methodologies in the agri-food supply chain. Financial incentives, tax breaks, and regulatory relief can be employed to encourage stakeholders to adopt data-driven projects.
- Knowledge exchange and learning can also be fostered through a policy framework that encourages the use of big data technologies and methodologies.

7.3. Limitations and Future Work

The limitations of this research work are presented as follows:

- Limited Scope: The research on the Bangladesh agri-food supply chain has limitations due to its focus on a specific component and may not fully represent the entire supply chain. The findings and recommendations may need to be contextualized and adjusted for different regions or countries.
- Data Availability: The data availability and quality of available sources may be limited, and access to more reliable data would strengthen the study.
- Generalizability: The findings are exclusive to Bangladesh and may not be easily transferable to other countries with different agri-food supply chain dynamics, socioeconomic characteristics, and infrastructure.
- External Factors: Additionally, the research does not consider external factors like policy changes, market fluctuations, or natural disasters, which can significantly impact the agri-food supply chain.

The future directions of this research work are described as follows.

- Real World Application: The study outlines the framework for a real-world implementation of the model in Bangladesh's agri-food supply chain, involving collaboration with stakeholders like government agencies, farmer organizations, and industry players.
- Pilot Projects: Pilot projects and case studies can provide insights into the model's practicality and performance, identifying potential problems and refining the paradigm for wider use.
- Stakeholder Involvement and Advanced Technology Integration: Stakeholder involvement is crucial for the model's efficacy, and integrating new technologies like IoT, blockchain, and artificial intelligence can improve efficiency and transparency.
- Continuous Assessment: Continuous assessment and monitoring are necessary to identify areas for improvement, and assessing the model's scalability and replicability in other developing countries can broaden its potential influence. By considering the unique characteristics and concerns of each country, the model can be adjusted to address specific agri-food supply chain demands.

8. Conclusion

The research on supply chain management techniques using big data in Bangladesh demonstrates the potential of leveraging big data analytics to improve the agro-based food supply chain management.

A conceptual model for big data architecture is proposed to address unique challenges, such as distribution inefficiencies, product quality, market access, and supply chain transparency. This can lead to improved operational efficiency, reduced food waste, increased profitability, and sustainable growth.

However, challenges like data privacy, security, infrastructure requirements, and stakeholder capacity building need to be addressed. The adoption of big data analytics in the agro-food supply chain can contribute to the growth and prosperity of Bangladesh's agricultural industry.

References

International Trade Administration (2022). Bangladesh - Agriculture Sectors. https://www.trade.gov/country-commercial-guides/bangladesh-agriculture-sectors

Agricultural Labour Force Survey 2016-17 (BBS). URL:https://mccibd.org/wp-content/uploads/2021/09/Labour-Force-Survey-2016-17.pdf

Bangladesh Bureau of Statistics (2019). Agriculture and Rural Statistics Survey (ARSS) Project 2017. https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/b343a8b4_956b_45ca_872f_4cf9b2f1a6e0/2020-02-02-10-36-84ecf771aa4c2e480f245fb79538ce14.pdf

CIA (2023). Bangladesh - The World Factbook. https://www.cia.gov/the-world-factbook/countries/bangladesh/

Belaud, J., Prioux, N., Vialle, C. and Sablayrolles, C. (2019). Big Data for Agri-food 4.0: Application to Sustainability Management for by-products Supply Chain. *Computers in Industry*, 111, 41–50. https://doi.org/10.1016/j.compind.2019.06.006

Biswas, S. and Sen, J. (2016). A Proposed Framework of Next Generation Supply Chain Management Using Big Data Analytics. Proceedings of National Conference on Emerging Trends in Business and Management: Issues and Challenges, Kolkata, India. SSRN: http://ssrn.com/abstract=2755828

Biswas, Sanjib. (2016). A Proposed Architecture for Big Data Driven Supply Chain Analytics. *SSRN Electronic Journal*. 10.2139/ssrn.2795884.

Cécile Zachlod, Olga Samuel, Andrea Ochsner and Sarah Werthmüller. (2022). Analytics of Social Media Data – State of Characteristics and Application. *Journal of Business Research*, 144, 1064-1076, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2022.02.016.

Chan, J.O. (2013). An Architecture for Big Data Analytics. *Communications of the IIMA*, 13(2), 1-14.

Chen, H., Chiang, R.H.L. and Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188. https://doi.org/10.2307/41703503

Christopher, M. (1992). *Logistics and Supply Chain Management*, Pitmans, London, UK.

Cravero, A., Pardo, S., Galeas, P., López Fenner, J. and Caniupán, M. (2022). Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability*, 14, 16131. https://doi.org/10.3390/su142316131

Gorton, Ian. and Klein, John. (2015). Distribution, Data, Deployment: Convergence in Big Data Systems. Software, *IEEE*, 32. 78-85. 10.1109/MS.2014.51.

Jony, A.I. (2016). Applications of Real-time Big Data Analytics. *International Journal of Computer Applications*, 144(5), 1-5.

Lisun-Ul-Islam, M., Rahat, M.R.H., Esha, S., Faiyaz, A. and Jony, A.I. (2023). Hourly Air Quality Prediction in Dhaka City Using Time Series Forecasting Techniques: Deep Learning Perspectives. Tuijin Jishu/*Journal of Propulsion Technology*, 44(5), 568-579.

Maier, Mark., Emery, David. and Hilliard, Rich. (2001). Software Architecture: Introducing IEEE Standard 1471. *Computer*, 34, 107 - 109. 10.1109/2.917550.

Marvin, H.J.P., Janssen, E.M., Bouzembrak, Y., Hendriksen, P.J.M. and Staats, M. (2017). Big Data in Food Safety: An Overview. *Crit Rev Food Sci Nutr.*, 57, 2286–2295. https://doi.org/10.1080/10408398.2016.1257481

Morshed, M.S.J. and Jony, A.I. (2016). Real Time Social Network Data Analysis for Community Detection. *International Journal of Computer Applications*, 139(6), 1-5.

Ray Zhong., Xun Xu. and Lihui Wang. (2017). Food Supply Chain Management: Systems, Implementations, and Future Research. *Industrial Management & Data Systems*, 117(9), 2085-2114.

Rejeb, A., Keogh, J.G. and Rejeb, K. (2002). Big Data in the Food Supply Chain: A Literature Review. *J. of Data, Inf. and Manag.*, 4, 33–4. https://doi.org/10.1007/s42488-021-00064-0

Shanto, S.S., Ahmed, Z., Hossain, N., Roy, A. and Jony, A.I. (2023). Binary vs. Multiclass Sentiment Classification for Bangla E-commerce Product Reviews: A Comparative Analysis of Machine Learning Models. *International Journal of Information Engineering and Electronic Business*, 15(6), 48-63.

Seyedan, M. and Mafakheri, F. (2020). Predictive Big Data Analytics for Supply Chain Demand Forecasting: Methods, Applications, and Research Opportunities. *J Big Data*, 7, 53. https://doi.org/10.1186/s40537-020-00329-2

Shovon, M.S.H., Islam, M.J., Nabil, M.N.A.K., Molla, M.M., Jony, A.I. and Mridha, M.F. (2022). Strategies for Enhancing the Multi-Stage Classification Performances of HER2 Breast Cancer from Hematoxylin and Eosin Images. *Diagnostics*, 12(11).

Subudhi, B.N., Rout, D.K. and Ghosh, A. (2019). Big Data Analytics for Video Surveillance. *Multimed Tools Appl.*, 78, 26129–26162. https://doi.org/10.1007/s11042-019-07793-w

Sultana, S., Eva, S.R., Moon, N.H., Jony, A. and Nandi, D. (2022). A Comparison of Opinion Mining Algorithms by Using Product Review Data. *International Journal of Information Engineering and Electronic Business*, 14(4), 28-38, DOI:10.5815/ijieeb.2022.04.04.

Tarun Kumar Agrawal., Vijay Kumar., Rudrajeet Pal., Lichuan Wang. and Yan Chen. (2021). Blockchain-based Framework for Supply Chain Traceability: A Case Example of Textile and Clothing Industry. *Computers & Industrial Engineering*, 154(6).

Tanvir, K., Jony, A.I. Haq, M.K., Nazera, F., Dass, M. and Raju, V. (2023). Clinical Insights Through Xception: A Multiclass Classification of Ocular Pathologies. Tuijin Jishu/*Journal of Propulsion Technology*, 44(4), 5876-5885.

Valencia Cárdenas, Marisol., Día-Serna, Francisco. and Restrepo-Morales, Jorge. (2022). Big Data Analytics in the Agribusiness Supply Chain Management. Aibi Revista de Investigación Administración e Ingeniería. 9. 10.15649/ISSN.2346-030X.

Yearbook of Agricultural Statistics (2019). Bangladesh Bureau of Statistics (BBS) Statistics and Informatics Division (SID), Ministry of Planning. URL: https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/1b1eb817_9325_4354_a756_3d18412203e2/2020-05-18-03-38-d3332b38e92c1adb7900d1bb94290177.pdf

Cite this article as: Md. Miftahul Alam and Akinul Islam Jony (2023). Supply Chain Management Techniques Using Big Data for Agro-Based Food Products in Bangladesh. *International Journal of Data Science and Big Data Analytics*, 3(2), 19-34. doi: 10.51483/IJDSBDA.3.2.2023.19-34.