Technoarete Transactions on Advances in Data Science and Analytics
Vol-1, Issue-2, August 2022
e-ISSN: 2583-2956

Use of Big Data in the Process of Customer Segmentation in the Retail Sector

Mudunuri Ananya Varma[1], Anaparthi Santosh Akhil[2], Kokku Sai Anoushka[3], Prithvi Sriman Maddukuri [4]

[1] Research Assistant, Department of Computer Science, Gokaraju Rangaraju Institute of Engineering and Technology, India.
[2] Documentation Assistant, Department of Computer Science, Gokaraju Rangaraju Institute of Engineering and Technology, India.
[3] Teaching Assistant, Department of Computer Science, Gokaraju Rangaraju Institute of Engineering and Technology, India.
[4] Testing Assistant, Department of Computer Science, Gokaraju Rangaraju Institute of Engineering and Technology, India.
*Corresponding Author Email: [1]mudunuri.ananya@gmail.com

Abstract

The study has clearly demonstrated the role of big data analytics on customer segmentation in the retail sector. The customer segmentation can be defined as separating consumers depending on previous purchase behaviour. In this study retail markets took the assistance of big data analytics to gather lots of information about their targeted consumers.

In the literature part of the study, it has been clearly demonstrated that the segmentation of consumers is necessary to enhance the sales value of the retail market. The retail marketers also took the assistance of automation, cloud software, and AI to collect data of consumers for business purposes. The literature part also helps to understand that a lack of expertise in management systems can lead to gathering wrong data of consumers and it may impact negatively the retail market.

The methodology part of the study is an essential factor to obtain an appropriate result. In this context, a descriptive research design has been chosen to conduct the study. In addition, the methodology part also took care of the authenticity of the research. A deductive approach has been selected for this study. The data has been gathered with the help of a secondary data collection method.

The discussion part has evaluated that the retail market has grown a lot depending on digitalisation. Retail marketers have been using big data to understand market trends and also trying to understand consumer purchase behaviour.

Keywords

AI, big data, Customer segmentation, consumer purchase behaviour.

INTRODUCTION

The concept of customer segmentation became essential for the emerging digitised market and also the retail sector. The term *"customer segmentation"* can be defined as the method of separating customers depending on their behaviour and demographics. Therefore, customer segmentation can be denoted as one of the essential business tools for retail marketers and those retail marketers can easily target their consumers for enhancing the brand's value. The study will shed light on the role of *big data* in segmenting consumers and its impact on the overall business. The term big data can be defined as, gathering a huge amount of data within a short time. In this process, the volume of collecting data becomes enhanced a lot and no traditional tool can manage this kind of large data. In addition, it can be observed that the retail sector is one of the important parts of the country that contributes lots of capital to strengthen the economy of any country. The *retail industry* always incorporates the sales of products to their consumers directly without any other individual's interference. The study will further discuss multiple issues faced by the retail industry while implementing big data and analytics within their business strategies.

LITERATURE REVIEW

"Concept of Big data in retail industry"

The term big data can be explained as a huge amount of data set that has been unitised by different business sectors to know about their consumers in a proper way. The process of collecting big data is a complicated framework and it can be observed that traditional software cannot manage the huge data set. One of the fundamental objectives of the retail industry in the use *of Big Data* is to detect complicated problems arising in business and also address it with the help of a large dataset.

Technoarete Transactions on Advances in Data Science and Analytics
Vol-1, Issue-2, August 2022
e-ISSN: 2583-1127

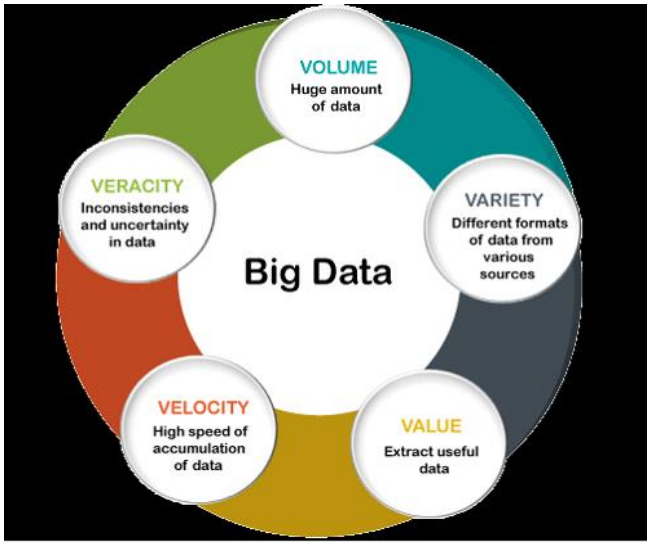

Figure 1: Characteristics of big data
Source: [1]

On the other hand, the study helps to know about the working process of big data and it includes three different steps, *"Integrate", "manage"* and *"analyse"* [1]. In this concept, it can be observed that *"Big Data"* can accumulate lots of information of targeted audiences with the help of new technical methods such as "*petabyte"* and "*terabyte"* scales. In the *"integration"* stage, the accumulated data need to be processed in a systematic way to accomplish the further analysing method. In the second phase, retailers need to store data after implementing advanced technologies. The upgradation of *"cloud-based software"* within the business sector leads to storing the processed information of audiences in a proper way [2]. In the last phase, *retail marketers* are generally analysing that information depending on the visual representation. In this process, retailers also use *"machine learning"*, *"artificial intelligence"* and *analytical models* to clarify those datasets for the customer segmentation process.

"Role of big data in customer segmentation in the retail sector"

From the last few years, the retail sector has grown a lot in different parts of the world. Due to the enhancing population, it can be observed that the retail market has grown near about *27%* by the year 2022 [3]. In this extreme competitiveness, the retail industry can cope up with the situation, only if they reach to their targeted consumers properly. With the help of Big Data, retailers are able to use multiple complex algorithms to predict their consumers' behaviours. In that case, *the machine learning* process helps those retailers to segregate their consumers according to *demographics, age, sex* and *gender*. Utilisation of Big data helps retail markets to understand the issues emerging in the competitive market. In this context, it can be observed that the retail market always works a chain wise, hence gathering information with the help of big data leads them to obtain *real time information* regarding business and market trends. Furthermore, real-time information provides opportunities to retail marketers to set *"optimum pricing"* of any product.

On the other hand, maximum retail markets target their consumers depending on "cluster methods". Sometimes, they focus on the income of consumers and collect huge amounts of information with the help of big data and analytics. In this process, retail marketers *separate* their audiences into small groups and also monitor their previous purchase history. All information they have collected with the help of *analytics* and big data. In contrast, those marketers also get data from online and also offline markets, after that they combine it to predict the *purchase trends* of targeted audiences [4].The analytics technology help to understand those retail marketers about their existing consumers and also the alternation of choices over time. The graph demonstrates that by the year 2022, big data has collected revenue of *70 billion US Dollars* and by the year 2027, it can be predicted that the value will enhance up to*103 billion US dollars* across the world.

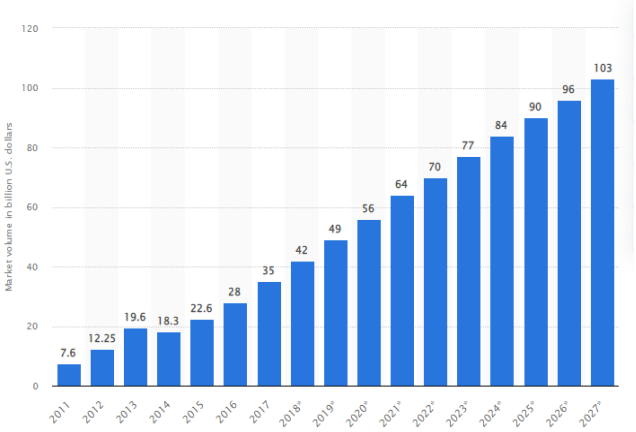

Figure 2: Revenue collected by big data Market
Source: [4]

"Issues while implementing Big data"

The previously discussed information help to know that Big Data" help retail industries to gather real time data of their audiences. Despite all these positive aspects, the implementation of big data can bring some challenges to those retail marketers.

- In this context, it can be observed that a lack of operational capacity of marketers can cause some issues during collecting "real time accurate data" of their consumers. Therefore, it is necessary to understand the types of consumers and also utilise the "optimal method" to gather data. Hence, gathering poor quality or wrong data can lead a sector to obtain massive losses. The inappropriate process of data collection also hinders the sales rate of the retail sector.
- The incapability of implement of rules and regulations within the business may cause data breaching. A huge risk appears when data is stolen by the outsider and it may lead to the formation of an issue regarding the authentication of the business [11]. This kind of incident occurs when retail marketers do not know about the rules of gathering

e-ISSN: 2583-1127

data with the help of analytics.

- The failure of retail marketers to obtain the trust of their consumers can hinder the capturing of the market with the help of big data. Consumers sometimes do not feel safe sharing information with retailers; As a result, even after the usage of big data can not bring positive result for those retailers.

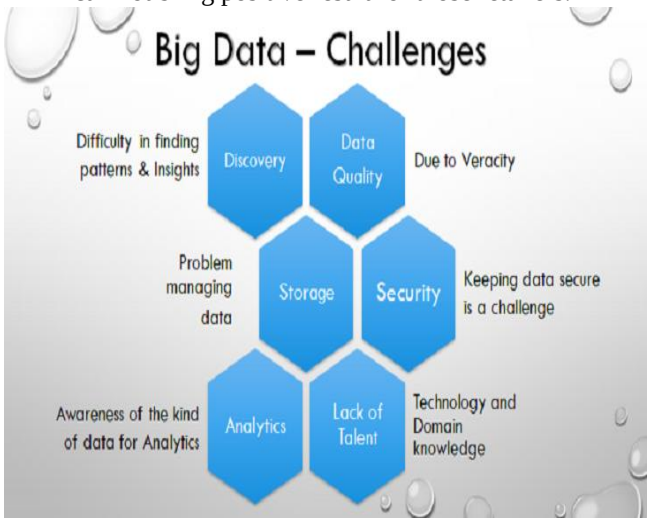

Figure 3: Big Data challenges

Source: [11]

- In this context, it can be observed that the volume of "Big Data" increased at least double after two years. Therefore, every organisation needs to double its storage system of the collected data [5]. Lots of industries are struggling to store data in a proper way.
- Every year "Big Data" technologies have been altered a lot, hence "a lack of understanding" of analytics, "big data", "AI" and "cloud based" software can be considered some issues while implementing big data in the retail market.

"Benefits of implementing big data"

In order to run a business effectively, retail marketers need to implement Big Data within their business strategy. Gathering huge amounts of information from consumers helps to understand the behaviour of audiences properly. As a result, retail marketers can segment their audiences to develop their business plans according to the current market trends. *"Consumer acquisition and retention"* is one of the major benefits of implementing big data in the retail business to accomplish consumer segmentation [6]. It can be observed that, most of the consumers share their preferences, likes and needs in digital platforms. Hence, with the help of "Big Data," retail marketers can predict their consumer's purchase behaviour. It also helps to "enhance consumer engagement and brand loyalty" after analysing the needs of audiences [21]. In this way, retailers can satisfy their audiences with the help of product quality and also depending on the services. In order to run the business for a long time, the retail industry is required to optimise the cost according to the trending market. In that case, "Big Data" and analytics always promote "cost reduction" for the benefit of the business.

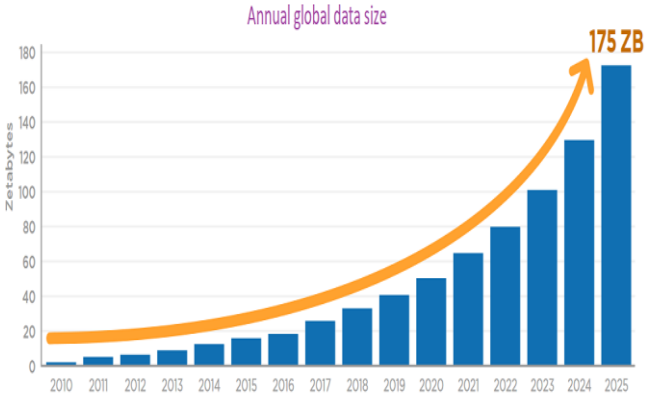

Figure 4: Annual Global size data

Source:[6]

METHODOLOGY

The term *research design* can be explained as a process or framework to accomplish the study properly. The particular method helps to sharpen the study and also help to gather accurate information for the study. Therefore, following an exact research design for the methodology part can reduce the chances of external influence. As a result, the study can be more authentic and reliable for other individuals. In this context, *"a descriptive research design"* has been selected to conduct the study properly [7]. The descriptive research design includes gathering data with the help of "a theory-based design". In addition, the "descriptive design" helps individuals to understand the study in a better way.

The "research approach" can be considered as one of the important parts in research methodology. The approach can help to assume the outcome of the study and plays an important role in collecting data for this study. In order to conduct the study properly, "a deductive research approach" has been chosen for the study. The *"deductive research approach"* is always done in the case of theory-based studies [8].Different hypotheses have been used in this method to find out the gaps present in the study. The "deductive research approach" helps to avoid the bias behaviour and also help to eliminate risks.

The term data collection can be explained as the method of gathering information from various sources and then analysing those previously collected data for research purposes. In this context, "a secondary qualitative data collection method" has been selected to obtain an appropriate result [9]. In the 'secondary data collection process," information is collected from already existing sources. In that case, previously available information from the internet and other sources have accumulated and utilised in this research to obtain effective results. Due to the "cost-effectiveness" of the secondary data, most of industrialists have been preferring to use this type of data. The sources of "secondary data" are, "websites", "journals", "newspapers" and already published articles [10]. In the different government and non-government websites, "secondary data" can be easily available and downloaded after just one click. The

e-ISSN: 2583-1127

authentication of the data should be maintained after collecting information from various authentic websites of multiple “companies”. In this way, consumers' authentic information can be collected through a “secondary data collection method". On the other hand, "other commercial sources" always contain information of secondary data, and those sources are "magazines", "journals" and " newspapers". Some advantages of the secondary data collection method are

- Anyone can gather a huge volume of information within a short time
- Well-structured information
- “Time saving”
- Cleanliness of information
- Availability of authentic data.
- “Easily accessible”

After gathering data properly, it is required to be analyse those data for business use. Therefore, the data need to be analysed with efficiency to enhance the value of the study. In addition, the study should not entertain any external influence to maintain its authenticity of the study. Depending on the proper analysis process, the outcome of the study becomes more accurate. In this context, “a qualitative analysis” method has been selected to analyse the previously collected data [13]. The “qualitative data analysis” process is generally a non-numeric method. This kind of process required a detailed observation of the topic. The "qualitative data analysis" method mainly includes a "proper insight" or "reasoning" of the topic, so that individuals can dig in deep for the research.

DISCUSSION

From the above-mentioned information, it can be said that the retail sector has developed a lot after the digitization process. The retail industry has contributed to the economic sector of any country by enhancing their productivity. There are two major characteristics that can be observed in the retail sector, include "sale of products" and "services". In this market, stores or offline facilities are dominating and the online retail market is also gaining popularity too.By the year 2021, the sales value of the "global retail market" was 26 trillion US dollar and it can be predicted that the sales value will enhance up to 30 trillion US dollars by the year 2024 [12]. In the case of many countries, the growth, well-being and maintenance of a healthy economy can be possible with the help of the growth of the “retail industry”. The retail industries have experienced rapid small transactions, hence the cash flow within the business has continued.

In this competitive market retail industries have been facing huge losses due to improper market segmentation. Therefore, the study has clearly demonstrated the importance of market segmentation for enhancing sales value in this market. The term "customer segmentation" can be explained as the process of separating consumers according to the behaviour of audiences [14]. Multiple large and small brands have decided to build relationships with retailers for the promotion of brands. In this context, it can be observed that, most of the time buyers want to purchase products from retail stores in both offline and online processes. Therefore, retail industries need to segment their audiences depending on age, demographics, income, family and also "job types" [15]. Therefore, it is required to dig deep in "industry-based data" and it can be observed that high-population areas are more suitable to start the business. On the other hand, it can be observed that some products have a relation with demographics. The study depicts that understanding consumers in a better way is the primary step to conduct customer segmentation. The operations in the retail market always occur in chain wise, therefore suppliers, investors, other markets and also consumers are a part of this retail industry. In order to gather proper information about the trending market and also consumers, it is required to separate audiences in small groups. The metrics of segmentation are based on different aspects, such as “post hoc”, “value based” and “need based”.

The study also illustrated the role of “Big data”, and “machine learning” can help in separating consumers on the basis of their previous purchasing behaviour [16]. After the rapid digitalisation process, consumers are generally preferring to buy products and also give their valuable feedback on the websites of the organisation. In this way, "Big data" helps to collect information from Facebook, Twitter and other social sites. In this way, retailers can understand which kinds of products are generally preferred by their audiences. In contrast, the changes of purchasing behaviour of their targeted consumers can be understood after the implementation of "Big data" within the business. On the other hand, the retail industry has started to use cloud-based software to store the huge amount of data in a systematic way [17]. With the help of "Hadoop", the cloud software can secure big data, in this way consumer information can be placed in a safe hand. Retail marketers are required to store data properly because the volume of big data increases at least twice after two years. Hence, cloud software cans smoothly store and transfer big data from one department to another for analysis purposes.

e-ISSN: 2583-1127

Table 1: Raw data

ID	Item Name	Parameter 1	Parameter 2	2013					2012	
				Param 4	Param 5	Param 6	Param 7	Param 8	Param 4	Param5
1	Row 1	4	10	TRUE	487	642	TRUE	10	5	9
2	Row 2	8	7	TRUE	274	837	TRUE	3	7	5
3	Row 3	8	7	FALSE	377	599	FALSE	5	3	6
4	Row 4	7	9	FALSE	280	947	FALSE	9	3	5
5	Row 5	3	3	FALSE	451	129	FALSE	6	5	4
6	Row 6	4	1	FALSE	767	70	FALSE	4	7	5
7	Row 7	1	1	FALSE	374	120	FALSE	9	8	1
8	Row ...	8	7	FALSE	905	241	FALSE	8	8	7
9	Row ...	9	7	FALSE	24	530	FALSE	1	3	7
10	Row ...	5	7	FALSE	813	149	FALSE	8	4	4
11	Row ...	3	6	FALSE	229	426	FALSE	8	3	4
12	Row ...	10	2	FALSE	970	737	FALSE	3	7	10
13	Row ...	1	8	FALSE	954	266	FALSE	7	1	9
14	Row ...	6	8	FALSE	564	667	FALSE	9	2	2
15	Row ...	6	4	FALSE	151	755	FALSE	10	3	10
16	Row ...	7	2	FALSE	582	764	FALSE	6	9	7
17	Row ...	3	2	FALSE	772	287	FALSE	1	10	6
18	Row ...	1	1	FALSE	71	127	FALSE	3	4	2
19	Row ...	6	5	FALSE	751	1000	FALSE	1	3	3
20	Row ...	7	5	FALSE	698	484	FALSE	7	8	2
21	Row ...	5	9	FALSE	259	493	FALSE	9	4	4
22	Row ...	6	4	FALSE	211	96	FALSE	1	10	4
23	Row ...	5	2	FALSE	158	789	FALSE	5	10	8
24	Row ...	4	5	FALSE	53	116	FALSE	7	10	10
25	Row ...	8	8	FALSE	747	220	FALSE	6	7	8
26	Row ...	1	4	FALSE	637	326	FALSE	7	3	5
27	Row ...	10	5	FALSE	393	205	FALSE	1	9	10
28	Row ...	8	3	FALSE	822	719	FALSE	7	4	4
29	Row ...	1	7	FALSE	116	119	FALSE	4	4	5
30	Row ...	5	9	FALSE	878	156	FALSE	3	1	1

Source: [17]

In contrast, the study also helps to know that retail marketers have faced lots of issues while implementing big data in consumer segmentation. Due to a lack of understanding about “big data”, “cloud software” and “analytics' ‘, retailers cannot gather real time information about their customers. In addition, some retailers have faced issues during storing the massive amount of data, hence the storage problem can be considered as a major drawback [18]. Maintaining big data, artificial intelligence and cloud software required huge amounts of capital, therefore a shortage of money can be seen within the business of the retail market.

CONCLUSION

Thepreviously discussed information help to understand the role of “Big Data” in case of segmenting consumers for the retail Industry. The term consumer segmentation can be explained as the way of separating audiences according to the consumers' previous purchase behaviour. Therefore, the segmentation can be done depending on a few characteristics such as, "age", "job description", "demographics" and " gender". The study also portrayed the advancement of technology and software that aids the retail sector by enhancing sales value. The big data is the process of collecting huge amounts of information from their consumers with the help of highly advanced technologies. The study helps to know that, there are three steps of big data that can be observed, those are "integrate", "manage" and "analyse". In order to enhance the sales value of the retail market, the experts are trying to focus on their consumers after segmenting them into short groups. It can be observed that, the teenage group has a tendency to purchase products online and they are also concerned about brands. In this way, with the help of Big Data, machine learning and other software, retail markets have successfully segmented their targeted audiences.

RECOMMENDATION:

In order to segment targeted consumers properly with the help of big data, the retail sector needs to implement some new strategies within its business.

- Retail industries are required to work with IT groups in a collaborative way to enhance the accuracy of collected information.
- They need to construct a team, who is capable of monitoring and maintaining the collected data

e-ISSN: 2583-1127

properly [19].

- The team needs to be highly efficient to eliminate data breaching or hacking of information.
- The retail sector needs to train its employees properly, so that those workers can work comfortably with advanced technologies [20].

In order to segment consumers properly, the experts from the retail sector are required to understand the purchase motifs of their consumers, hence improving expertise management can lead to achieved success for the retail sector.

REFERENCES

[1] Dabkiene, V., Balezentis, T. and Streimikiene, D., 2021. Development of agri-environmental footprint indicator using the FADN data: Tracking development of sustainable agricultural development in Eastern Europe. *Sustainable Production and Consumption*, *27*, pp.2121-2133.

[2] Giagnocavo, C., de Cara-García, M., González, M., Juan, M., Marín-Guirao, J.I., Mehrabi, S., Rodríguez, E., Van Der Blom, J. and Crisol-Martínez, E., 2022. Reconnecting farmers with Nature through agroecological transitions: Interacting niches and experimentation and the role of agricultural knowledge and innovation systems. *Agriculture*, *12*(2), p.137.

[3] Moso, J.C., Cormier, S., de Runz, C., Fouchal, H. and Wandeto, J.M., 2021. Anomaly Detection on Data Streams for Smart Agriculture. *Agriculture*, *11*(11), p.1083.

[4] Kittichotsatsawat, Y., Jangkrajarng, V. and Tippayawong, K.Y., 2021. Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability*, *13*(8), p.4593.

[5] Peter, A.P., Koyande, A.K., Chew, K.W., Ho, S.H., Chen, W.H., Chang, J.S., Krishnamoorthy, R., Banat, F. and Show, P.L., 2022. Continuous cultivation of microalgae in photobioreactors as a source of renewable energy: Current status and future challenges. *Renewable and Sustainable Energy Reviews*, *154*, p.111852.

[6] Martos, V., Ahmad, A., Cartujo, P. and Ordoñez, J., 2021. Ensuring agricultural sustainability through remote sensing in the era of agriculture 5.0. *Applied Sciences*, *11*(13), p.5911.

[7] Awan, M.J., Bilal, M.H., Yasin, A., Nobanee, H., Khan, N.S. and Zain, A.M., 2021. Detection of COVID-19 in chest X-ray images: A big data enabled deep learning approach. *International journal of environmental research and public health*, *18*(19), p.10147.

[8] Cravero, A., Pardo, S., Sepúlveda, S. and Muñoz, L., 2022. Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy*, *12*(3), p.748.

[9] Kramer, M.P., Bitsch, L. and Hanf, J., 2021. Blockchain and its impacts on agri-food supply chain network management. *Sustainability*, *13*(4), p.2168.

[10] GuoHua, Z. and Wei, W., 2021. Study of the Game Model of E-Commerce Information Sharing in an Agricultural Product Supply Chain based on fuzzy big data and LSGDM. *Technological Forecasting and Social Change*, *172*, p.121017.

[11] Mathotaarachchi, K.P. and Thilakarathna, K.A.A.N., 2021. Philosophy of Approaches in Social Sciences: A Review of Positivism, Phenomenology and Critical Social Sciences in Qualitative Research. *Technium Soc. Sci. J.*, *20*, p.944.

[12] Haralayya, B., 2021. Ratio Analysis at NSSK, Bidar. *Iconic Research And Engineering Journals*, *4*(12), pp.170-182.

[13] Delgado, J.A., Short Jr, N.M., Roberts, D.P. and Vandenberg, B., 2019. Big data analysis for sustainable agriculture on a geospatial cloud framework. Frontiers in Sustainable Food Systems, 3, p.54.

[14] Kour, V.P. and Arora, S., 2020. Recent developments of the internet of things in agriculture: a survey. Ieee Access, 8, pp.129924-129957.

[15] Goel, R.K., Yadav, C.S., Vishnoi, S. and Rastogi, R., 2021. Smart agriculture–Urgent need of the day in developing countries. Sustainable Computing: Informatics and Systems, 30, p.100512.

[16] Himesh, S., Rao, E.P., Gouda, K.C., Ramesh, K.V., Rakesh, V., Mohapatra, G.N., Rao, B.K., Sahoo, S.K. and Ajilesh, P., 2018. Digital revolution and Big Data: a new revolution in agriculture. CABI Reviews, (2018), pp.1-7.

[17] Tao, D., Yang, P. and Feng, H., 2020. Utilization of text mining as a big data analysis tool for food science and nutrition. Comprehensive reviews in food science and food safety, 19(2), pp.875-894.

[18] Andronie, M., Lăzăroiu, G., Iatagan, M., Hurloiu, I. and Dijmărescu, I., 2021. Sustainable cyber-physical production systems in big data-driven smart urban economy: a systematic literature review. *Sustainability*, *13*(2), p.751.

[19] Daissaoui, A., Boulmakoul, A., Karim, L. and Lbath, A., 2020. IoT and big data analytics for smart buildings: a survey. Procedia computer science, 170, pp.161-168.

[20] Ge, M., Bangui, H. and Buhnova, B., 2018. Big data for internet of things: a survey. Future generation computer systems, 87, pp.601-614.