

Customer Segmentation using Data Mining Technique: A case study of Retail Industries in Nigeria

Onyianta John Chiedozi^{1*}, Musa Ibrahim Umar² and Agbo Jonathan Chukwunwike³

¹Department of Computer Science, University of Nigeria, Nsukka

Email: jonnyspencer210@gmail.com, Phone: +2348062923028

²Department of Extension and Management, College of Agriculture, Jalingo, Taraba State, Nigeria. Email:

musa.umar.pg76718@unn.edu.ng, Phone: +23437615708

³University of Kairouan, Tunisia and Department of Computer Science, University of Nigeria, Nsukka.

Email: agbo.jonathan@unn.edu.ng, Phone: +2348035287760



Abstract – The importance of retail industries are seen in the exchange of goods and services. Retail industries are the meeting point where different producers and wholesalers have direct contact with the final consumer. Other than distributing goods to the consumers, they are also involved in providing essential services to their clients. Recently, due to the rapid growth in information technology, such as availability of networks, fast internet and mobile devices, the present-day retail industries has also witnessed a significant change. There are challenges involves in the use of IT in retail business such as security risk and operation cost. Despite these risks, in other to stay ahead of other competitors, it is important to adopt the use of information technology to maximise profit and improve customer satisfaction. The aim of this study is to apply K-nearest Neighbour (KNN) classification algorithm on customer driven data such as their shopping pattern, choice of product, amount spent after shopping and average time spent on shopping to properly predict and classify the customer's class of income earning.

Keywords – Customer Segmentation, Data Mining Technique, Retail Industries.

I. INTRODUCTION

Retail industries are vital in the exchange of goods and services. Retail industries are the meeting point where different producers and wholesalers have direct contact with the final consumer. Apart from distributing goods to the consumers, they also provide other essential services. Recently, customers are paying attention to retailers that can provided quality services, and are attracted with well decorated stores equipped with varieties of products and point of sale (POS). It is the concern of every retailer to build plans on how to attract more customers in other to increase patronages. There is no doubt that shopping malls with well arranged service structures will attract more customers. This should include the strategic plan to easily locate and collect items on the shelf and how to reduce queue when customers are making payments. For this reason, one can simply say that services in retail industries are usually tailored towards the customers need. Customers' behaviour varies from customer to customer due to the variation in their shopping time and changes in their shopping needs, desire and discrepancy in their choice of product brands. In order to make decision on how to serve the customer well, it is important to keep records of information relating to customer behaviour.

Recently, due to the rapid growth in information technology, such as availability of networks, fast internet and mobile devices, the present-day retail industries has also witnessed a significant change. Retail industries are now a global community, because the advancement of information technology has affected the production chain, the various means of product sales, distribution and delivery. The introduction of information technology in retail business has considerably changed the traditional way of transacting businesses. There are so many advantages attached to the use of technology in retail industries. By introducing information technology into retail business, it is easier to achieve improved customer experience and services, proper management of inventory and supply chain. For instance, with the help of wireless technologies, customers can use wireless devices such mobile phones to communicate with retailers at different location in the world. There are challenges involves in the use of IT in retail business such as security risk and operation cost. Despite these risks, in other to stay ahead of other competitors, it is important to adopt the use of information technology to maximise profit and improve customer satisfaction.

II. AIM AND OBJECTIVES OF THE STUDY

The aim of this study is to apply Knearest Neighbour (KNN) classification algorithm on customer driven data such as their shopping pattern, choice of product, amount spent after shopping and average time spent on shopping to properly predict and classify the customer's class of income earning. The class of earning is of three main categories: High income earner, average income earner and low income earner.

To achieve the above stated aim, the study must meet the following objectives.

1. To discuss the various information systems used in retail industries.
2. To examine the usefulness of information technology in reducing technological and operational.
3. To visualize the different categories of customer using KNN algorithm.

III. LITERATURE REVIEW

This work applied a qualitative research approach in reviewing literatures from different researchers regarding the role of information technology in understanding customer behaviour, delivering quick and quality services in retail industries. This is to acquire clear understanding of the topic of discuss. It was carried out using various sources such as library resources and online materials to discuss the views of other scholars, followed by creating a literature gap that will accommodate new idea. The review is based on decisive evaluation of the approaches and methodologies as used by other scholars.

Because of the strong competition that exists among retailers and the contest between traditional and technological enhanced retail industries. Reinhard Schütte in his work titled "Information Systems for Retail Companies Challenges in the Era of Digitization" noted that globalization brought about a huge growth in both local and international markets [1]. Therefore, retail industries are has witnessed severe competition on concentrated markets, especially; a market in which control over the supply of a commodity is in the hands of a small number of producers and each one can influence prices and affect competitors.

Information technology has evolved so much in recent times, and has offered so many opportunities to be applied in retail industries. Information technology has brought a new way of solving problems in every sphere of life including the business organisation. The innovative solutions target to enhance customer service and maximize the usefulness of retail industries. According to [2], information technology can be used almost in all retail operations which may include logistics, supplies, procurement, accounting and customer services. In his work, he examined some information technology solutions that a mostly used by retailers by carrying out a study based on their usefulness among retailers and customers of a selected retail industries. He also noted that the benefits of information technology don't end with enhancing business processes but also the way in which the enterprise relates with the customer. Improved retailer-customer relationship can be accomplished by improving the way the retail industries offers services to their customer and minimizing some of the challenges that customers come across during shopping. Some of these shopping challenges encountered by customers are (i) inability to locate product (ii) difficulty in finding product information (iii) change in price product during customer checkout.

Development in information technology and the availability of internet connectivity has affected work processes globally. In [3], titled "The Effect of ICTs on Retail Space – a South Africa Perspective", is a study on how the incorporation of ICT and internet technology affected the management of retail space in retail industries. The study was carried out using a survey research technique. Commercial centre managers and investors were interviewed to ascertain the effect of information technology on retail space. The

result of the study shows that both centre managers and investors are getting ready to harness the opportunities offered by the development of online shopping. The retail shops can be expanded by introducing other services such as internet shops and gaming devices.

Information and technology growth is enormous globally. It has become a marketing means used by retailers to reach out to more potential customers. Use of information technology in retail industries is perhaps the fastest and cheapest of selling goods and services to potential clients. For instance, E-commerce is an example of technological innovations used in retailing. In a research paper by Muhammad Najib Razali et al, the effect of information technology implementation on retail space among retailers in major cities of Malaysia [4]. The research was conducted using both qualitative and quantitative research approach. Hence, questionnaires and interview was used a technique for data collection to look into the views of retailers who has applied information technology in their business processes. They used 233 key retailers in the cities of Malaysia as the study sample size. The study results noted that a good number of retail industries in Malaysia use various IT solutions in their daily retail services. The study also revealed that a negligible number of retailers in Malaysia encountered loss while using IT in their business routine. The research paper exposed the usefulness of information technology in retailing, adding that information technology is capable of changing the method of making demand and supply in Malaysian retail industry.

Over the past few years, innovation in technology has undergone extraordinary growth and has not left retailing industries behind. Because of this technological advancement, retailing in India has changed, moving from urban market to a semi-developed market and local markets. This has equally changed customer behaviour due to increasing retailing opportunities. Ponduri, S. B. and V. Sailaja conducted a qualitative research that tries to examine the key roles played by information technology in rendering quick and quality service delivery [5]. The research data was gotten by observing and interviewing clients of different retail store; research data was also gotten by interviewing retail centre managers. The research explained the need to improve customer services through the use of evolving information technology solutions. The authors found that the quality of services provided by retail centres can be realized through correct execution of information technology solutions that a manned by skilled and qualified staff.

According to [6] [7], some of the major area of retail industry that needed information technology implementation includes (i) information about a product, this includes their availability, shelf life, promotion, demand and supply (ii) information about the customer; this has to do with customers profile, their behaviour, and their choice of product (iii) operations information; this includes product allocation, inventory, procurement and shelf space. In [7], some of the advance technology used in retail industries was discussed alongside with the associated benefits. These technology solutions includes, Point of Sale (POS), Radio Frequency Identification (RFID), Smart Operating System, Warehouse Management System (WMS) and ERP – Enterprise Resource Planning with their benefits

E-commerce as an example of information technology solution has contributed to an improved way of combining manual solutions with digital solution such as clicking on item to add them to shopping carts. Clients are influenced by retailers with real-time item availability characteristics. Therefore, failure to meet customers need through implementation of evolving information technology solutions that will enhance product demand and distribution chain are identified as some of the major risks facing retail industries [8]. The conventional retailers in a physical building are no longer the major channel of purchasing goods. It has shifted from the traditional stores to a digital point of connection between the investors, retailers and customers. With the help of information technology, retailers have witnessed lots of change in the traditional business process. According to [8], apart from procurement and supply chain management, there are various means by which retailers apply ICT in their daily activities. These include (i) location sensing technologies for new customers (ii) social media (iii) customization of product and services (iv) E-commerce.

The past few years has witnessed a remarkable change in understanding customer behaviour due to a speedy growth of the information and communication technology environment that has transformed the retail industries. Hence, retailers are deeply involved in competition in order to attract new customers. Therefore, in a competitive environment, the survival of any retailer is dependent on the number of customers they were able to attract. For their business to stand the test of time, the must be devoted to services that are centred on the customer satisfaction. Nowadays, client does not go to retail stores for the sole reason to shop; rather, they are attracted by retails that can offer the most favourable purchasing experience when placing order and receiving the purchased product at any given time and location [9]. The author in [9] noted that those retailers focusing on customer satisfaction, improved operational experience and better IT solution implementation will survive in the newly evolved retail environment. He also pointed out some of the changes that affected the retail industry's environments. They are: (i) digital shift, (ii) service shift and (iii) power

shift from manufacturers to the consumer. The digital shift is concerned with the various ways consumers make purchases using their mobile phones and other smart devices irrespective of time and their location. With the growth in internet technology, consumers may decide to either purchase the product online through e-commerce or from a retail shop. He explained that service shift is a type of change in retail environment, concerned with the variety of goods and intangible services offered by the retailers, while power shift is change of power from the manufacturer side to the client (consumer) side.

Almost all retailers are mapping out strategies to improve their quality service in order to meet up the competitive demand of the retail industries. Success rate of any retail industries depends on the ability to minimize operation cost and increase their customer service simultaneously [10]. This can be achieved upon the retailer ability to implementing technology solution to satisfy their customer in all the sections of the retail industry. Therefore, retailers must have second thought to understand their business processes and IT operations to ascertain the relevant area they can reduce cost and increase operations.

IV. METHODOLOGY

4.1. Customer segmentation using data mining approach

In order to improve customer satisfaction in retail industries, we have decided to show concern on customer behaviour pattern and appropriately categorize them using a clustering algorithm. The segmentation is based on customer's demographic information such as their age, location and gender, Sales information about Day & Time, Customer account information about bills and payments and so on. Customer demographic profile will enable retailers to appropriately modify their business processes to satisfy their customers well.

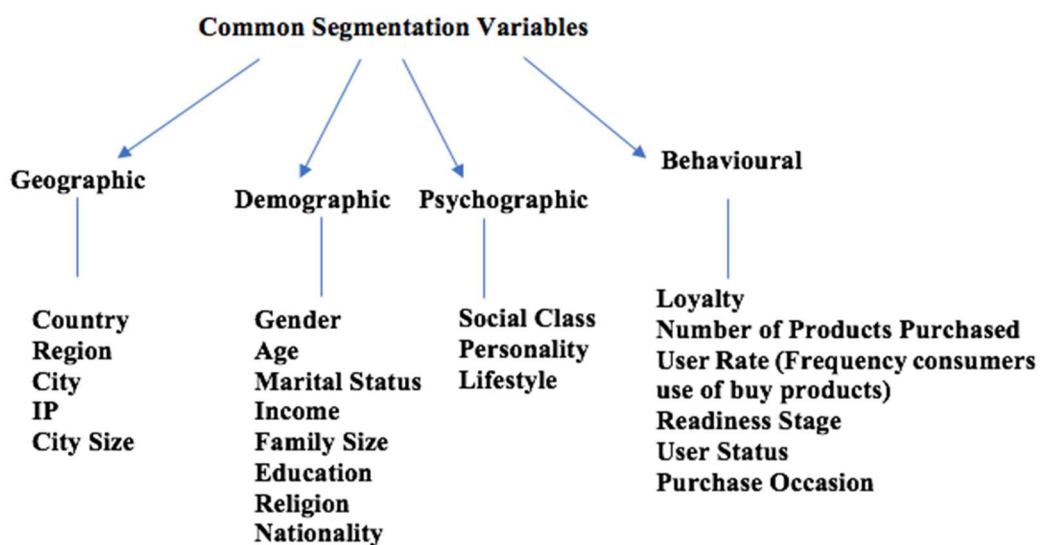
Segmenting customers into various clusters is another form of market segmentation. Customer segmentation involves making a sub-group of customers of the related demographic profile within a universal group of customers with different demographic profile. After the segmentation process, it is easier to the identify the group or cluster that are valuable to the retailers, hence allowing the retail industries to concentrate on these important group of customers in order to run the risk of losing them to their competitors.

1.1 Visualizing a common customer segment method

It is recommended to make the segmentation as simple as possible. Depending on the size of the retail shop, it is important to group the customers into two or more segments according to their individual needs.

A simple customer segment method includes:

- Geographic
- Demographic
- Psychographic
- Behavioural



4.2 Data mining and its approaches

Data mining is a machine learning technique that can be used to obtain useful information from a big data set that may not be significant for a specific reason. In this context, data mining is a technique used in extracting useful information out of a big data set that is unprocessed. The extracted data are transformed into useful information that may be used for quicker decision making.

KNearest Neighbour (KNN) and K-Means clustering

KNearest Neighbour (KNN) algorithm and K-Means clustering are used to segment the customers into clusters based on their demographic information.

Customer Segmentation with KNN

Given a data set of customers known as the training set, a selected customer can be classified to a segment using KNN algorithm based on the nearest neighbour in the training set. Once you query the algorithm requesting a unknown customer (test data) to be classified into a segment, the algorithm will search through the entire training set, then calculates the distances between the unknown customer and all the known customer in the training set. The corresponding segment (class) in the training set nearest to the unknown customer is the distance with the smallest value. Hence, the unknown customer is classified using the class of its nearest neighbour.

KNN algorithm

```
for all the unknown samples UnSample(i)
for all the known samples Sample(j)
    compute the distance between UnSamples(i) and Sample(j)
end for
    find the k smallest distances
    locate the corresponding samples Sample(j1),...,Sample(jk)
    assign UnSample(i) to the class which appears more frequently
end for
```

Euclidean distance formula

Given a distance x,y d^2 to denote the distance between vectors x and y .

$$x,y \ d^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2$$

that is, the distance itself is the square root

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

This can be expanded to

$$d_{x,y} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

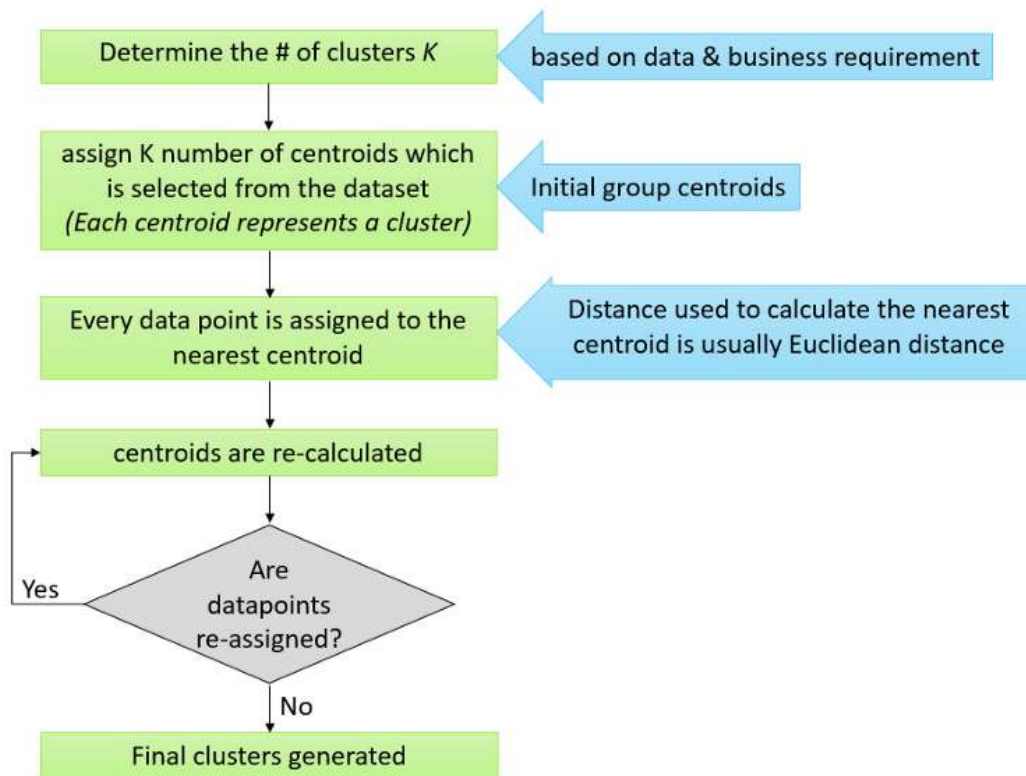
And can be summarized as

$$d_{x,y} = \sqrt{\sum_{j=1}^j (x_j - y_j)^2}$$

Customer segmentation with K-means clustering

K-means clustering is a method that is used to segment customers into clusters using python programming language and their libraries. Some of the known python libraries used are numpy and pandas. K-means uses iterative method to determine the number of clusters K and the data point is assigned to one of the clusters K using the similarity feature.

Flow chart of a K-means algorithm [11]



Formula for K-means clustering [11]

K-means cluster formula is written as follows:

$$\text{Minimize } \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - c_j)^2$$

where

K = number of clusters

N = number of data points

C = centroid of cluster j

$(x_{ij} - c_j)$ – Distance between data point and centroid to which it is assigned

Data set – the dataset is made up of yearly income of approximately 300 customers and their yearly expenditure on online retail shop. The customer yearly income and expenditure are in (₦000) respectively as shown in the table below.

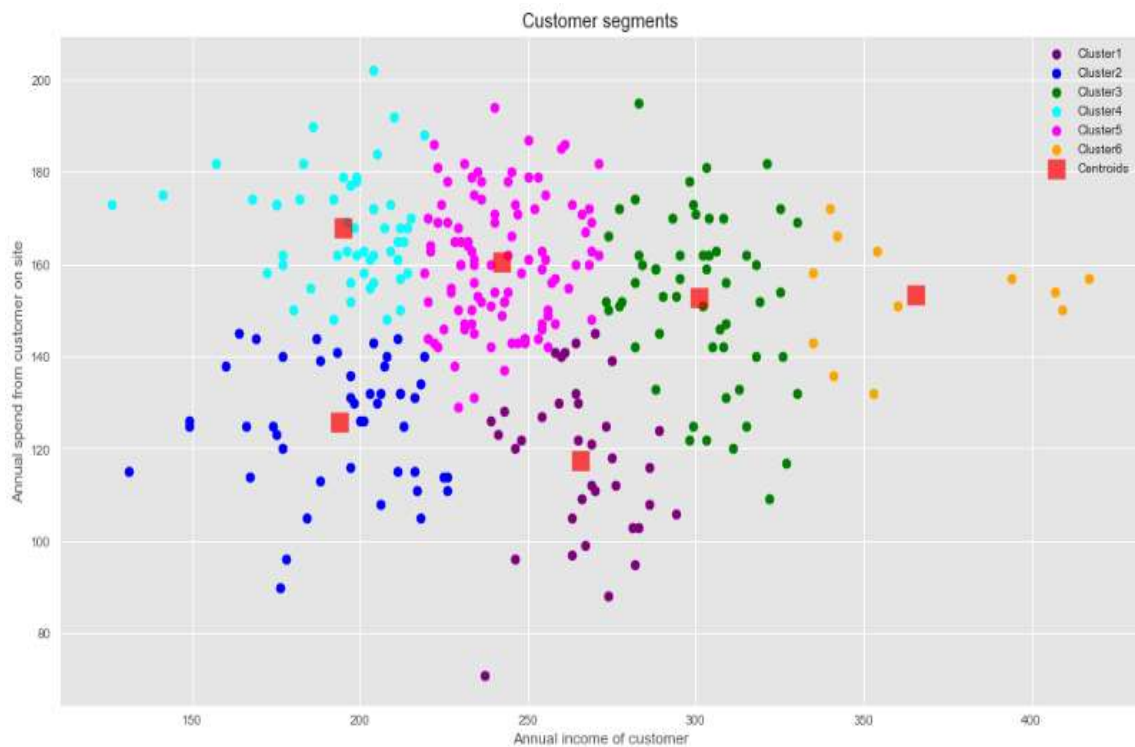
	INCOME	SPEND
0	233	150
1	250	187
2	204	172
3	236	178
4	354	163

Visualizing the Segments

Segments of customer earning and spending were created using a k-means clustering and implemented in python programming language using various python libraries. Each cluster has different colour and a centroid. Therefore, members of a cluster are unique and are different from other cluster

The cluster plot below shows the different segmentation of customer income and their respective yearly spending.

The clusters below were created by fitting the k-means of the data with the value of $k = 6$.



V. FINDINGS

5.1 Cluster interpretation and marketing strategies formulation

- Cluster 1 represents customer with **Medium income** and **low yearly spending**
- Cluster 2 represents customers with **Low income** and **low yearly spending**
- Cluster 3 represents customers with **High income** and **high yearly spend**
- Cluster 4 represents customers with **Low income** and **high yearly spending**

- Cluster 5 represents customers with **Medium income** and **low yearly spending**
- Cluster 6 represents customers with **Very high income** and **high yearly spending**

With the data visualization, we are convinced clustering algorithms or segmenting customer based on their demographic profile can provide useful information. Therefore, we can make important decision and recommendation based on the clusters.

Marketing strategies formulation

1. It is recommended that retail industries should put more efforts on product promotion for customers in clusters 6 and 3.
2. Customers in cluster 4 are considered as loyal customers because of their spending strength. Therefore, in order to retain this type of customers, it is recommended that retailers should offer them discount.
3. Customers in cluster 2 are characterized as low income earners with low year spending. Thus, to increase the spending power of this group of customers, it is recommended that the retailers should start up responsive pricing strategies.
4. Clusters 1 and 5 are made up of good income earning customers with very low spending power. It is recommended that retail industries should improve on satisfying the customers better and improve product visibility.

VI. CONCLUSION

From our study, we have seen KNN and K-mean clustering algorithms to be a useful technique used by retailers to clearly segment their customers using their demographical information. This will enable them to understand their customers to be important players in the success of their business. The idea of customer segmentation based on their demographic profile is also a key factor in proper decision making. If well implemented will enable retail industries in formulating the best marketing strategies that will help improve customer satisfaction and product visibility.

REFERENCES

- [1] Reinhard Schütte “Information Systems for Retail Companies Challenges in the Era of Digitization”, Institute for Computer Science and Business Information Systems, University of Duisburg-Essen, Germany reinhard.schuette@icb.uni-due.de
- [2] Barbara Kucharska “INFORMATION TECHNOLOGY IN RETAIL TRADE” University of Economics in Katowice, Poland
- [3] K.S. Cattell, K.A. Michell & R. McGaffin “The Effect of ICTs on Retail Space – A South Africa Perspective” Department of Construction Economics and Management, University of Cape Town, South Africa
- [4] Muhammad Najib Razali, Yasmin Mohd. Adnan & Rohana Abdul Rahman “The Impact of Information Technology on Retail Property in Malaysia” International Real Estate Research and Symposium, 24-25 April 2012, Kuala Lumpur, Malaysia.
- [5] Ponduri, S. B. & V. Sailaja “Information technology in Indian retailing, issues and challenges (A case study on the rise and fall of Subhiksha Retail Chains)” African Journal of Marketing Management Vol. 6(3), pp. 33-38, July, DOI: 10.5897/AJMM2013.0372, (2014), <http://www.academicjournals.org/AJMM>
- [6] Saroja S “**Information Technology – Key Success Factor in Retail**” GIAN JYOTI E-JOURNAL, Volume 1, Issue 2 (Jan – Mar 2012) **ISSN 2250-348X**. Available at: www.gjimt.com/GianJyotiE-Journal.htm
- [7] Aruna R. Shet , “Information Technology in Retail Sector”, International Journal of Scientific Engineering and Research (IJSER) www.ijser.in ISSN (Online): 2347-3878, Volume 4 Issue 5, May 2016
- [8] Dr. C. Mulligan, & Dr. Z. Gurguc “ICT & the future of Retail” Industry Transformation – Horizon scan, JLL, (2014), Long live retail, Retail Research.
- [9] NOTOMI Narumitsu, TSUKAMOTO Michiko, KIMURA Masaaki, & YAMAMOTO Shota “ICT and the Future of the Retail Industry -Consumer-Centric Retailing” NEC Technical Journal/Vol.10 No.1/Special Issue on Enterprise Solutions to Support a Safe, Secure and Comfortable Life.

- [10] Tim Blansett, Namit Kapoor, Denton Newham & Michael Horvath, “Winning in Retail IT-IT *Efficiency and Effectiveness Strategies for the Retail CIO*” Booz & Company (2010).
- [11] Sowmya Vivek “Clustering algorithms for customer segmentation”, August 13, 2018. Available at <https://towardsdatascience.com/clustering-algorithms-for-customer-segmentation-af637c6830ac>