



Data Analytics Approach for Retail Business Prospects Using Machine Learning Algorithms

Raksha Raju¹, A.M Prasad²

Department of CSE, Dayananda Sagar College of Engineering, VTU, Bengaluru, Karnataka, India^{1,2}

ABSTRACT: Regarding business, distribution, and customer support Big Data has been influenced by many companies. A new approach called data analytics has become more important for any organization in today's business environment. To increase company profits in industries, structuring business strategies like customer segmentation is very useful. Clustering algorithm is said to extract valuable details from the records by looking at which individuals of input variables drop into which group when the algorithm is applied. Throughout this work, a comparison between the K-Means clustering algorithm with the Fuzzy C-Means is made. The main focus of this study is to identify the finest potential cluster centers for validity using K-Means and Fuzzy C-Means to look for a better calculation value. That is, the consumer behavior for that specific datasets are obtained by measuring the accuracy. The company is B2C, a source of data collected from Optical Stores in Bangalore located in three different areas. K-Means has a better measure of precision compared to the Fuzzy C-Means, between two different resampling methods. This project has also authorized the use of data analytics to provide clear knowledge of consumer value and deliver a beneficial Customer Relationship Management (CRM) process.

KEYWORDS: Big Data, Clustering Algorithm, Customer Relationship Management, Customer Segmentation, Data Analytics.

I. INTRODUCTION

Small-scale and medium-scale enterprises now plays a significant role in our daily routine. Usually, the term "entrepreneurs" applies when a supplier meets the small orders of several consumers most of whom are end-users, rather than just massive orders of a limited number of the wholesale. India has 1.3 billion people and is among the largest rising retail markets in the world. The company finds it hard to gain new clients as well as to retain old clients in an established industry [1]. An enterprise should know the pros and cons of its buyers in those tough crowded markets.

Small companies can lead to ineffective decision making, unsatisfactory retailing, or a level of social interest in the pricing of the retail store. A retail company fails because of "lack of resource, unsafe salesperson, improper business location, consumer expectations," etc. One approach to group the buyers is that suppliers may make decisions to be targeted to a particular group. With the company and client, the buyer pays a significant part throughout each stage of the buying process experience through every single point of interaction. Popular areas in visualization. Recent days, a lot of business firms gather, store, and analyze lots of data. These data are generally called "Big Data" due to its high scale, the pace at which it arrives, and the range of types it takes. The results provide a new generation of assistance for problem-solving in business intelligence.

The use of analytics is the main factor in deriving value from massive data [2]. Documents need to be studied and the result used by decisionmakers and team owners. Pinpointing the needs of the consumers, the demands, requirements, preferences, etc. through a 'customize' process called CRM. Under the growth of modern computer technology, more retailers are gradually prioritizing CRM [3]. To retain the consumer interactions correctly CRM procedures and software are helping the firm. CRM is about to run by software in which entrepreneurs have consumer data as follows: name, contact details, age, address, location where buyer tends to buy. The firm has a tool that sets all the contacts. Specific methods are hired to make company needs [4]. Methods are data mining, visualization of data, data analytics, etc.

This study has the goal of comparing centroid-based grouping method with the Fuzzy C-Means classifier to pick the ideal values for categorizing. This methodology is thus willing to choose a detailed impact on consumer choices and review the impact of huge details by CRM. The cluster center k is assessed by the reliability factors of the groups. When the unit range of user information is bunched with classifiers for grouping and Fuzzy C-Means algorithms [5]. Evaluation checks to figure out which method achieves the good are assessed [6].



Comparison of the process with maximum testing steps with the reliability of Fuzzy set. When the groupings have been established, the user samples are reviewed to look deeply into each consumer sector’s behavioral tendency.

II. RESEARCH METHODOLOGY

Analysis technique in this study is tried to carry out from detailed repository to stage of data interpretation to produce awaited evidence.

A. Customer Segmentation

The firm should consider the attributes like desires, interests, outlooks, and viewpoints of its consumers in honor of goods offered. B2C sells straight to users [7]. So can use relational data such as brands purchased, money paid, etc. to establish successful extraction. Throughout the retail market, these kinds of relational data have been collected and registered at the checkout each consumer relational history information that enables service to carry out further analyzes of their client’s tendencies.

B. Clustering Technique

K-Means clustering is correlated throughout these studies with Fuzzy C-Means for consumer insights.

K-Means

The method K-Means would be a way of figuring patterns or sections in one’s outcomes [8].The grouping application is a composition of steps which recursively try to find the units and spot the data. Target at slicing n occurrences into k region with its nearby mean relate to the closest estimate cluster. Step by step procedure of K-Means algorithm shown below in Fig.1 [9].

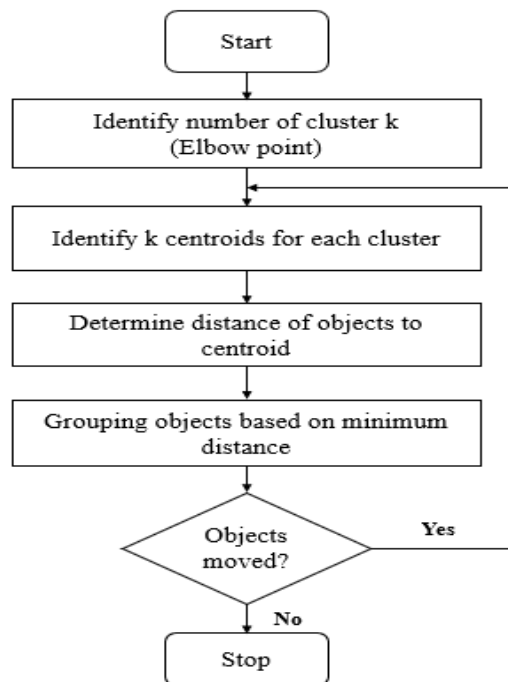


Fig.1. Flow Chart of K-Means

Fuzzy C-Means

Fuzzy logic is a type of classifying by which an attribute may adhere to several cluster. In FCM number of clusters are selected randomly. FCM is K-Means enhanced feature [10]. Grouping or cluster process includes allocating variable to units so that elements from the same set are comparable, although assets across various nodes are as unique as necessary. Step by step procedure of Fuzzy C-Means algorithm is shown below in Fig.2.

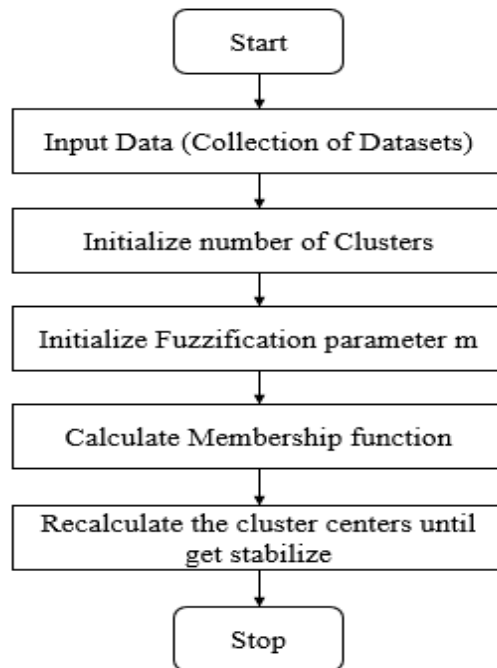


Fig.2. Flow Chart of Fuzzy C-Means

III. EXPERIMENTAL RESULTS

The raw data involved throughout this study is collected from Optical Stores in Bengaluru based in different areas. These Optical Shops operates under configuration B2C. Whose products are sold to end consumers right away. The Point of Sale (POS) program gathers buyer's records from each dealer, and all user information is registered. The information includes the name of the consumers, the amount paid, the exact description of the branch issued, brand names borrowed, etc. All actions of evaluating the data are incorporated in SQL Server and Anaconda (Python Jupyter Notebook).

A. Data Processing

Information is gathered from the database through this study where it has entered all the client relevant data through January 2017- December 2019 of three divisions based in Bengaluru. The steps performed are the replacement of data with null data and anomalies. The next step is to obtain columns needed for the further phase. The phase listed finishes in late December 2019.

B. Number of Clusters Selection

The subset of features will be established first until the classifier is finished. Clusters head numbers are an integral feature of grouping. First will be formed the subset of features before the classifier is done. Cluster head numbers become an important aspect of grouping. Parameters were determined by employing K-Means for $k=3$ till $k=5$. The appropriate training set would tend for being three from test conformation.

C. Comparison between K-Means and Fuzzy C-Means

The results revealed that between two clustering algorithms the best possible results for authentication measures are indeed achieved via K-Means. The findings of the analysis are shown in Fig.3. It displays the validity test samples for both K-Means and Fuzzy C-Means the finest optimized cluster centers secure where $k=3$. While the appropriate set of clusters derived by the K-Means and FCM algorithms are similar, perhaps both algorithms have been proposed using accuracy. From the analysis, we can infer that the K-Means algorithm is more accurate to measure the data contrasted to the FCM algorithm relying on the outcome gained, it can be acknowledged easily that K-Means classification achieves better precision especially in comparison to FCM. Even more research on the information of consumers for each group is then carried forward utilizing the K-Means algorithm.

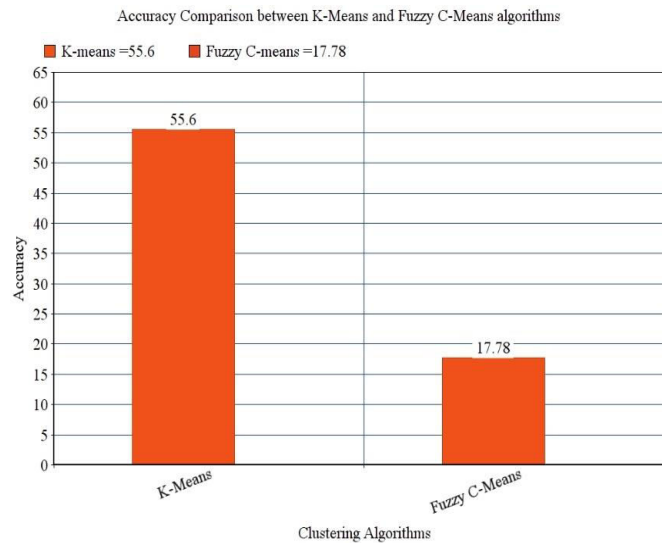


Fig.3. Accuracy Comparison

IV. CONCLUSION

The firm needs to assess the qualities of the interests, desires, beliefs, and viewpoints of its consumers regarding the current quality of the product in offers. A sample size retrieved from Bengaluru's Optical Stores operating in various areas. This work has the goal of evaluating K-Means implementations with the Fuzzy programming interface to determine the right classification outcome. From the conclusions, the maximum value of k obtained by determining the internal node quality tests corresponds to three. Evaluation measures to determine which technique works efficiently. In that K-Means secures the best classification measurement result, it is used to correlate that to the Fuzzy C-Means method for the next phase of the cycle. The algorithm with optimum validation measure will be compared to Fuzzy C-Means based on accuracy.

When the samples are established, the target audience is assessed to dive deeper into each buyer section's person's perceptions. More specifics from every cluster's client information encourage businesses to build cost leadership for a particular subset of clients. This paper aims to recognize the lowest and the most valuable consumers and therefore helps the organization to focus on those who are most willing to purchase products. Tends to create a lifelong client relationship by cultivating and providing them the goods and facilities they desire.

REFERENCES

- [1] C. W. Lamb, J. F. Hair, and C. McDaniel, "Essentials of Marketing", (Lachina Publishing Services, 2011)
- [2] Watson, Hugh J. "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," AIS Journals, Vol. 34, Article 65, 2014.
- [3] Pratyosh Rai and Dr Ritu Bhargava, "Analysis of various Customer Relationship Management", International Journal of Engineering Development and Research, Volume 6, Issue 1, 2018.
- [4] K. Tsipis and A. Chorianopoulos, "Data mining techniques in CRM: Inside customer segmentation", (Hoboken, NJ: Wiley, 2010).
- [5] Nadhira Riska Maulina, Isti Surjandari and Annisa Marlin Masbar Rus, "Data Mining Approach for Customer Segmentation in B2B Settings using Centroid-Based Clustering", 16th International Conference on Service Systems and Service Management (ICSSSM), 2019.
- [6] J. C. Bezdek†, "Cluster Validity with Fuzzy Sets." Journal of Cybernetics 3(3), pp. 58-73, 1973.
- [7] A. K. Mann and N. Kaur, "Survey Paper on Clustering Techniques", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, 2013.
- [8] K. Singh, D. Malik, N. Sharma, "Evolving limitations in K-means algorithm in data mining and their removal", IJCEM International Journal of Computational Engineering & Management, 2011.



- [9] Shruti Kapil, Meenu Chawla, Mohd Dishad Ansari, "On K-means Data Clustering Algorithm with Genetic Algorithm", Fourth International Conference on Parallel, Distributed and Grid Computing, January 2016
- [10] Fahmida Afrin, Md. Al-Amin, Mehnaz Tabassum, "Comparative Performance of Using PCA with K-Means and Fuzzy C Means Clustering for Customer Segmentation", International Journal of Scientific & Technology Research, Volume 4, Issue 10, October 2015