**CUSTOMER SEGMENTATION FOR E-COMMERCE PLATFORM**

| Describe the project in 2-3 sentences | The project targets to analyze consumer data from the e-commerce platform Autofurnish.com by using K-means clustering algorithms to segment customers and identify patterns in tehri behavior for 2018 and 2019. This analysis is performed by using the RapidMiner Tool and the findings are compared to provide recommendations for improving consumer influx and sales performance. | |
| --- | --- | --- |
| What is the contribution of this project? (technical, application etc.) | The selected project focuses to the application of data mining techniques, specially consumer segmentation by clustering to gain data into consumer behavior in the e-commerce domain, enabling data driven decision making strategies | |
| Describe the dataset in 2-3 sentences | The dataset consists of consumer data from autofurnish.com for the years 2018 and 2019. Thai includes attributes like organic searchers, age, city, users with over 600 brown of data based on demographic, behavioral and psychographic areas. | |

| Novelty  How novel is this project? | The novelty lies in applying established data mining techniques like K-means clustering to the e-commerce domain for customer segmentation and behavior analysis. | |
| --- | --- | --- |
| Strengths  What are the strengths of this project? | An extensive dataset spanning two years, a systematic approach using a better clustering algorithm into consumer demographic and behavior | |
| Weaknesses  What are the weaknesses of this project? | A limited scope focused solely on consumer segmentation, a lack of evaluation of different segmentation methods. | |
| Suggested changes & Recommendations | Some of the recommendations can be included like incorporating additional data sources, evaluating alternative segmentation methods, enhancing interpretability, validating findings by pilot tests, analyzing predictive modeling and developing a scalable and auto automated solution for continuous monitoring and adaptation. | |

# Introduction

Whether it is e-commerce or any other type of fast-paced business environment, the knowledge and strategy for segmentation of the customers is fundamental in keeping the business afloat as well as to cater to the different needs of the clients. Target marketing is the process of classifying clients into specific categories giving similar attributes such as behavior, interest or preferences. The online platforms thus get the ability to personalize their marketing strategies, product assortments and customer experiences to cater to the diverse needs of each segment.

# Research question

* What are the segmentation methods and evaluation metric for identifying and targeting different consumer segments in e-commerce?
* How can temporary trends in consumer behavior effectively capture into the consumer segmentation process for e-commerce platforms?
* To What extent can integrate multiple data sources like web analytics, purchase history and demographic data develop the accuracy and relevance of consumer segmentation models for e-commerce businesses?

# Dataset

The researchers have utilized the rapidMiner tool that is an open source predictive analytics software for analyzing the consumer data. They have employed the L-means clustering algorithm that is a mostly used method for grouping similar data points together (Kh Khaled Kalam, 2020). This analysis targets to segment consumers based on their behavior, identify patterns and makes comparisons among the two years for reflecting some recommendations for improving consumer influx and sales performance.

Researchers have made several analyses including calculating the DBI (Davies Bouldin Index) to determine the optimal number of clusters (k) for the k-means algorithms (Kim *et al.* 2022). They have also implemented aggregate functions to extract information like the most active and least active cities, age groups that have the highest and lowest engagement and the number of organic searches and users.

| **Metric** | **Value** |
| --- | --- |
| Least Active City | Agartala |
| Most Active City | Maximum |
| Minimum Age Group | 65+ |
| Maximum Age Group | 25-34 |
| Total Organic Searches | 598,367 |
| Minimum Users | 56 |
| Maximum Users | 74,395 |
| Total Users | 101,038 |

**Table 1: Analysis of the year 2018**

(Source: Choosed literature)

| **Metric** | **Value** |
| --- | --- |
| Minimum Active City | Ajman |
| Maximum Active City | Agartala |
| Minimum Age Group | 65+ |
| Maximum Age Group | 25-34 |
| Total Organic Searches | 345,175 |
| Minimum Users | 65 |
| Maximum Users | 60,734 |
| Total Users | 889,816 |

**Table 2: Analysis of the Year 2019**

(Source: Choosed literature)

# Methodology

In this journal a primary analysis has been conducted based on customer segments in e-commerce by using a data mining method. RapidMiner toolkits were employed by the researcher to achieve pre-processing and data mining using the K-means algorithm for client segmentation. Consumer attributes from an e-commerce hardware organization Autofurnish (Liu *et al.* 2022). However, the com mailbox has been conserved on a regular basis since December 2018 and to date, January 2019. The dataset includes descriptors such as age, city, clean/without make-up, and amount of clients. The data was subsequently imported into the RapidMiner device where pretreatment tasks such assessing values and data normalization were performed (Shirole *et al.* 2021). The k-means clustering calculation was performed on the figure process to obtain the result. The number of clusters (k) was moved from 3 to 10 and for each k-value, the DBI computations were performed. The number of clusters was selected based on the way each card's k with the least comparison of DBI differed from each other since it was a precise value for the yearly information. For 2018, the supposed value of K was 4, while for 2019 it was 8. The analysis, therefore, revolves around clusters of customers which are formed and categorized by means of total abilities such as whole, variable, maximum, minimum etc. to understand customers' real behavior based on age, city, gender and number of users. The clustering, then, took place in the course of 2018 and 2019 to pick out the patterns related to the behavior of the clients during a specified period and for carrying out the experiences to focus on the separate client categories.

# Experiments and results

When focusing on the findings of the journal, then it can be said that it provides necessary information about customer segmentation in the e-commerce platform. In this journal, the focus was on a particular journal. In the 2018 compilations, an optimal number of clusters, i.e., 4, was determined using Davies-Bouldin Index metric. The main pleasant customers' port was from the bus of Agra, and the range of 25-34 yrs was the foremost pleasant. As for the 2019 dataset, the range clustering should have been equal to 8. It was Agartala, with the most clients' AIDS needle exchange program and the dominant age group once again was the age group 25-34. Even though there were more clients in 2018 (1,010,308) than in 2019 (889,816), there was also a lower number of natural looks counted in 2018 (598,367) as opposed to in 2019 (345, 175.) . The test results showed that Agartala, which was the static city in the year 2018, became the one with the least number of customers in the year 2019, while Ajman showed a similar trend. In two years the rise of seniors 65+ was the lowest among this period of time which gives an indication that there are gaps that need to be performed to tend to this group. Based upon the identification of underlying issues, the analysts have recommended Autofurnish to (Yuping *et al.* 2020). The target group to be aged between 25 to 34 years, should be more concentrated upon the people of cities namely Agra (2018) and Agartala (2019), for a robust organizational engagement and maximum return on investment.

**Areas of improvement**

There are many areas that can be observed by which the research can be developed in the future for customer segmentation in e-commerce platforms. Investigate other k-means clustering algorithms and compare their running time to identify the most efficient solution. Use human behavior data streams like purchase history, online browsing knowledge, etc. to develop a complete persona (Sarkar *et al.* 2023). Clustering is defined as the grouping through the usage of independent measurement values or tests of experts. Analyze the connection union of client portions with exchange measurement for instance income, client lifetime value, etc. therefore, it can identify the targeted customers and promote the aim through sing out the recognized customer portions.

# Conclusion and future recommendations

Through studying the literature it can be concluded that the researchers have applied the K-means clustering algorithm to segment e-commerce customers and identify patterns in their behavior. The analysis has found some of the limitations like having a single ecommerce platform and dataset. Future work needs to analyze clustering techniques, incorporate additional data sources like purchase history and web analytics, thereafter develop predictive models to forecast consumer behavior. Validating all the findings by pilot tests and developing scalable, automated solutions for continuous monitoring would enhance the practical applicability of consumer segmentation for improving e-commerce marketing tactics.

# Reference List

**Journals**

Kh Khaled Kalam (2020), Market Segmentation, Targeting and Positioning Strategy Adaptation for the Global Business of Vodafone Telecommunication Company, International Journal of Research and Innovation in Social Science (IJRISS) |Volume IV, Issue VI, June 2020|ISSN 2454-6186

Kim, M., Huh, J., Sa, A., Jun, A. and Lee, H., 2022. A Study on Market Segmentation Based on E-Commerce User Reviews Using Clustering Algorithm. Journal of Society for e-Business Studies, 27(2).

Liu, C.J., Huang, T.S., Ho, P.T., Huang, J.C. and Hsieh, C.T., 2020. Machine learning-based e-commerce platform repurchase customer prediction model. Plos one, 15(12), p.e0243105.

Shirole, R., Salokhe, L. and Jadhav, S., 2021. Customer segmentation using rfm model and k-means clustering. Int. J. Sci. Res. Sci. Technol, 8, pp.591-597.

Yuping, Z., Jílková, P., Guanyu, C. and Weisl, D., 2020, March. New methods of customer segmentation and individual credit evaluation based on machine learning. In “New Silk Road: Business Cooperation and Prospective of Economic Development”(NSRBCPED 2019) (pp. 925-931). Atlantis Press.

Sarkar, M., Ayon, E.H., Mia, M.T., Ray, R.K., Chowdhury, M.S., Ghosh, B.P., Al-Imran, M., Islam, M.T., Tayaba, M. and Puja, A.R., 2023. Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases. Journal of Computer Science and Technology Studies, 5(4), pp.186-193.