2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) | 979-8-3503-1590-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACCAI58221.2023.10199324

Retail Business Convenience Segmentation using Clustering and Data Visualization

Thirunavukkarasu.J, Sanjanaa.J, Sivarakshana.M and Yuvashree.R

Computer Science and Engineering Sri Sai Ram Institute of Technology Chennai, India

E-mail : thirunavukkarasu.cse@sairamit.edu.in sit20cs013@sairamtap.edu.in sit20cs037@sairamtap.edu.in sit20cs029@sairamtap.edu.in

Abstract- The conventional approach to launching a business is to research and gather data regarding the past performance of rival businesses unless they were profitable or unsuccessful. Innovation is the ethos of the modern day, as everyone is engaged in a struggle to outperform one another. The objective of our suggested research is to create knowledge that will be helpful to aspiring business owners and small companies that are losing money. Our main aim is to assist small-scale manufacturers in becoming successful marketers. In return for the dataset, which must be provided as input, we will provide them with clear instructions on how to start a profitable business and recover from their loss. In order to analyse data more effectively, our planned work will segment clients based on stock input, weekly updates of stocks sold, and waste products. In this work, two different clustering techniques (k-Means and hierarchical) are used to classify the products into subsets, and their respective results are compared. Data will be segmented using clustering algorithms, allowing for much more focused production of the final result.

Keywords— k-means clustering, hierarchical clustering, dataset, small-scale manufactures, marketers, stocks)

I.INTRODUCTION

The development, maintenance, and cultivation of profitable long-term customer connections has historically depended heavily on the management and maintenance of client relationships. Producer segmentation divides a market into numerous unique groupings of producers who have common traits. Clustering is an iterative process to extract knowledge from large amounts of raw and unstructured data.. Traditional marketing has historically been seen more as a tool for consumer consumption than as a tool for generating revenue for businesses. Client segmentation is currently performed through the processing of the Client Database i.e., Demographics or buying history. The business plan must be consistent with the present circumstances of the modern period of innovation, when there is intense competition to be better than everyone else. Customer segmentation refers to the customer division of the organization. Based on demographic characteristics (age, gender, marital status) and behavior. Considering that customers of similar ages may have vastly diverse interests, behavioral elements provide a more effective method for consumer segmentation than demographic variables because of their emphasis on individuality and the fact that they allow for accurate segmentation. The various methods available for user segmentation through aggregation techniques, and the reason that this is such an essential part of customer relationship management. For text datasets, the conceptual vectors generated by the spherical K-means methods serve as a localized and sparse "Base". The main problem is that when a beginner wants to start a company, they need to analyze various scenarios. They primarily wanted to investigate the location of the store where it was to be located. In order to determine whether or not this is the best place to start, the store data set from kaggle is gathered.

Through these tests, they can start and manage a healthier business or they can benefit from their current business The hierarchical algorithm and the k-means algorithm are used to build the proposed work, respectively. K-means and hierarchical [1] have proven to be effective as it beat another algorithm like SVM. Such hierarchical representations can be either agglomerative or divisive, depending on whether they are constructed from the top down or the bottom up. These techniques take a bottom-up strategy, building up from individual data points by merging clusters in a hierarchical fashion. Machine learning, classification, and pattern recognition are just a few of the many areas that put clustering to use. The segmentation process starts with analyzing customer behavior and progresses to discovering the demographic and psychographic traits of these clients. The dataset that would be gathered from each end user was segmented using a hierarchical clustering algorithm.

Our model takes the dataset as input. The collection contains various product information including stock sales and stock inflow. The company's billing system can be used to quickly get the datasets. These forms of input encourage the end-users to have a billing system in their firm, no matter how little. The government gains since the current store pays the GST tax. The supplied dataset in Excel format will be gathered by our model. Although the dataset won't contain any clean data, it will be clustered afterwards to produce useful results. Employing the right method, the visible graph provides the desired outcome, highlighting the gain and loss with the help of efficient algorithms. The user visually evaluates the output. This study uses two algorithms to split small-scale business sellers' data into clusters and present a variety of visualizations for their advantage.

LITERATURE SURVEY

According to Asith Ishanth [2], businesses are willing to spend money on research and development aimed at attracting, retaining, and expanding their clientele. To better understand their customers and develop more effective strategies for reaching out to them, businesses can make great use of business intelligence tools.

Kansal et. al. [3] all authors presented Overloaded information can be overcome by implementation of personalization in eCommerce services such as providing product recommendations, links recommendations, ads or text and graphics that correspond to the users. Their paper classifies customer segmentation based on data processing. According to Sharat et al. [4], the silhouette score is derived by averaging the silhouette coefficient over all instances in the dataset. How closely points through one cluster are to points in neighboring clusters is measured by the silhouette coefficient, which could also vary from -1 to 1.

Shuai et. al. [5] studies have the same overarching goal: to describe Market niches that have only a loose connection to consumers' likely actions in the store. This approach to market segmentation is predicated on the idea that the primary distinction between "fake" and "real" market divisions is the nature of the benefits consumers hope to reap from using a product. Fanlin et al. [6] acknowledge that client form clusters are useful for narrowing in on certain audiences and advertising content that will pique their interest via marketing and social media. Customer segmentation is a crucial part of customer relationship management, and Fanlin discusses why, along with several models for doing it with the help of a clustering algorithm.

According to Maurya et. al. [7], the idea vectors produced by the perfectly circular K-means algorithms provide a robust localised, sparse "Base" for text data sets. Sreedevi et. al. [8] proposed that, the method runs more quickly as the distance between clusters widens and a variety of empirical study on real data set and synthetically generated data from application colour quantisation, data compression, and image segmentation.

According to Abdulnassar et. al. [9], the idea vectors produced by the perfectly circular K-means algorithms provide a robust localised, sparse "Base" for text data sets. Samples drawn from a predetermined group structure are also feasible within the models, as shown by Hou et al. [10].

PROPOSED WORK

It is the work where the seller could maintain a proper dataset of the customer to improve their marketing. The seller feeds the data collected from the stock. Based on the input given the system analyses and provide them with the statistics of their sales for the day through data visualization. These datasets are maintained properly and provide an outlook based on their profit and loss on the sale.

Architecture Diagram

The architecture shows how the data analyst gathers the information required for the analysis database, formats it by removing any NA values, and produces data that is ready for processing. It selects features that refine the model, in our case the features are annual income and expenditure scores for efficient analysis. K indicates that the classifier uses the characteristics at its disposal when performing clustering. With the clusters established, the marketing team can develop several tactics for more effective client targeting.

The dataset is what our model uses as input. The dataset includes different product information on stock inflow and sold stock. We can readily get these datasets by leveraging the organization's billing system. Even if it is a tiny business, these kinds of input encourage the end-users to have a billing system in their organization. The fact that the existing store pays the GST tax benefits the government.

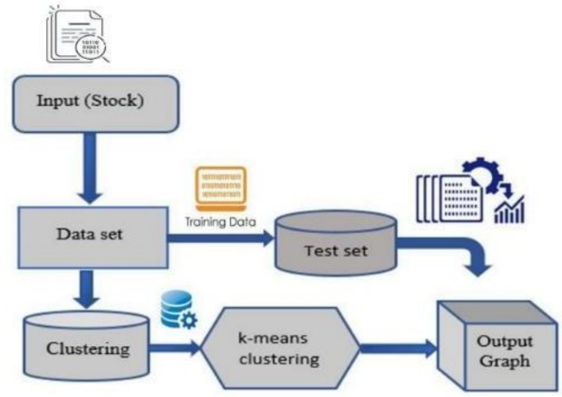


Fig. 1.Architecture Diagram of proposed work.

Our model will only collect the specified dataset in Excel format. The data supplied in the dataset won't be clean, but it will be clustered later to produce effective results.

Clustering algorithms seek to locate natural clusters in data, and there are numerous ways in which data clustering methods can be tweaked and updated. Clustering is predicated on the assumption that components inside the same cluster should be comparable. The information is consolidated such that related pieces are close together. The data will be clustered using 2 different algorithms (k- means, hierarchical clustering).

Using an appropriate technique, the visible graph offers the required result, indicating the gain and loss utilizing effective algorithms. The output is analyzed visually by the user.

Algorithm

Step1: Collection of data from a store's sold stocks and its stock intake, either through the billing system or manually.

Step 2: Data are fed into our model, and using k-means and hierarchical clustering methods, they are clustered.

Step 3: For instance, the characteristics include grocery store fruits and cereal. The k terms are separated using the K-means algorithm based on quality. Different qualities in a collection of data are distinguished from one another by their values and characteristics. It will cluster each fruit and cereal item separately and return the clean data.

K cluster centers are represented by the symbol. The sign stands for K different classes. Here are how the two sets of data differ. The central point can be obtained by (1).

Step 4: Centroids are discovered and grouped using hierarchical clustering. the procedure is carried out till it reaches the clustered final data. Hierarchical clustering is not always supported. Most often, the k-means is employed to group the data.

Step 5: After the data is cleansed, different graphs (such as a line chart and a dendrogram) are generated so that the user may determine whether product sales are increasing or decreasing. The entire data analysis is initially presented visually. Second, the graph for each attribute is defined using the 24-hour market price. Thirdly, a weekly market analysis is offered. Finally, it illustrates how to invest and what products to recommend.

One can also review historical store data and examine how the most recent sale performed to determine whether to open a specific store or not.

K-means is a commonly used technique for calculating clusters because of its efficiency. As a result, it has high status and impact and is widely employed in both theoretical study and practical manufacturing. This paper acquaint with a symbol 'D'.

$$D = \{y_i \in R^n, i = 1, 2, ..., n\} \quad (1)$$

The symbol represents K cluster centers. K distinct classes are represented by the symbol. The gap between the two data is as follows.

$$dis(y_1, y_n) = \lim_{\lambda \to 0} \sqrt{\sum (y_i(p), y_{i+1} + (p))^2}, \lambda \quad (2)$$

The center point can be defined using the following eq. 3.

$$k_j = \lim_{n \to \infty} \frac{\sum y_i(p)}{n_j} \quad (3)$$

where n_j refers to the number of the same class. Iterative approaches include k-means. We choose k for the cluster. The clustering center is then updated often. Unfortunately, the procedure still has a lot of issues.

- The results are not consistent every time because of the impact of the initial value and outliers.

- It is simple to reach the regionally ideal answer.
- The number of clusters must be predetermined.
- Centre for clustering you may not always be a part of the data set.
- Because to the usage of the L2 distance function,
- k-means is easily impacted by noise.

We have enhanced k-means in an effort to address these issues.

An algorithm called hierarchical clustering, commonly referred to as hierarchical cluster analysis, divides comparable objects into clusters. The result is a collection of clusters, each of which differs from the others while having things that are generally similar to one another.

The centroids are discovered and categorized according to their characteristics on either side, after which they are linked, and as a result, several distinct qualities are generated. By computing its Euclidean distance, the dendrogram shows two dissimilar materials. The consumers are on the x axis, while the y axis shows the Euclidean distance.

Min approach: Distance here is the minimum of the pairwise distances, like in above example, d ([A, B], C) = min (d (A, C), d (B, C)). This is also called single linkage.

Max approach: Max of the pairwise distances taken.

Average: Average of the pairwise distances taken.

Like d ([A, B], C) = (d (A, C) + d (B, C)) / 2

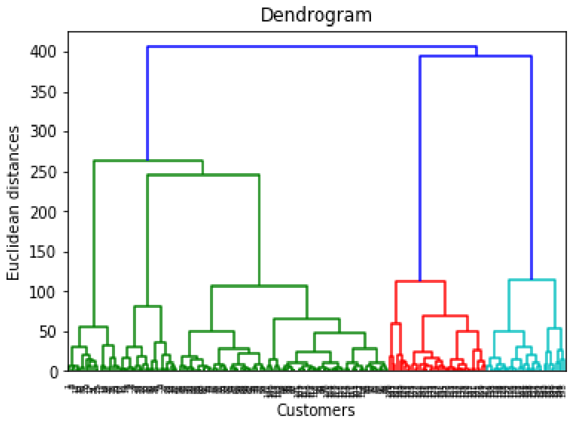


Fig. 2.Visualization of dataset using dendrogram.

The Support Vector Machine (SVM) is a common supervised learning technique that is used to handle classification and regression problems, forecast unknown data, and is excellent for training data. The choice of an appropriate core function (for processing non-linear data) is complex. What happens is that when you use a large kernel function, you run the risk of creating too many support vectors, which reduces the drive speed. The SVM requires a lot of memory to store all the media vectors in the memory. This number continues to grow as training data sets grow in size and requires extensive training on large data sets. A method to justify SVM refers to the loss function.

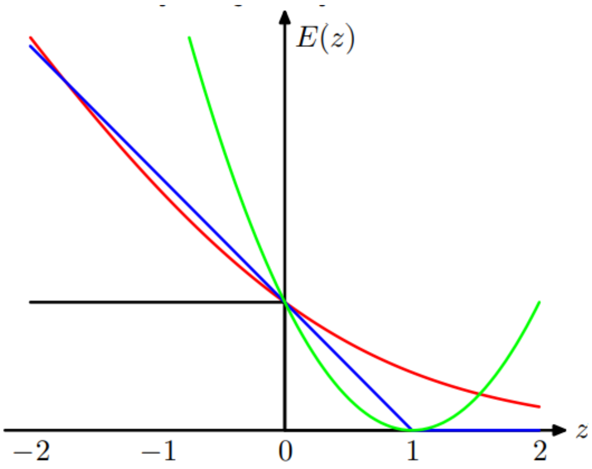


Fig. 3.Loss function.

To determine the ideal minimum clusters for the K-means cluster, the existing groundwork uses the Elbow approach. Nevertheless, in some cases, the elbow approach is not effective. For example, refer to the scatter diagram below. Humans may be able to identify the five different groups where the data comes from, but humans have difficulty understanding large-scale data. As two clusters are thus close to each other, the Elbow Method leads us to this conclusion. Indeed, placing a centroid in the middle of the two clusters reduces the distance between the data points. Therefore, finding the right number of clusters for our clustering task requires a more accurate, rigorous and reliable method using K-means clustering.

Feature Extraction

Selects the attributes that make the model more accurate; in our case, the attributes are the annual income and expense score for effective analysis. One characteristic is nothing else than the meaningful representation of an image that can be used for further segmentation and classification. Selecting the feature is the first step to obtaining the images retrieved from the feature. This method is very useful for repeated functionality and selected functionality that does not have data. It will not choose without data, and will not be helpful for the upcoming processing. The selected feature has been extracted to simplify the number of resources required to accurately describe an extensive dataset. Feature Extraction is a general term that describes extracting only valuable information from specified raw data. The main aim is to represent the raw image in its reduced form and also to reduce the original data set by measuring some properties to facilitate the decision-making process for filing.

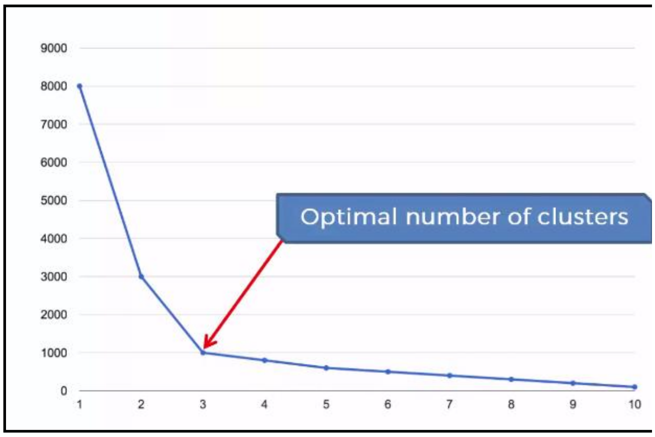


Fig. 4.Elbow approach representation..

Dataset Description

The dataset is obtained from the stock the salesperson holds. The dataset for the classification of products based on sales is taken from the billing records. The data set from the Kaggle website was utilized for our proposed work. The said dataset is a segmentation of mall patrons.

TABLE 1 PARAMETER AND DESCRIPTION OF DATASET

SI.No	Parameters	Description
1	Stock Intake	Data set of Stocks per Month
2	Stock Sold	Product Sold Per Month
3	Billing System	Customer Details and Cart Items

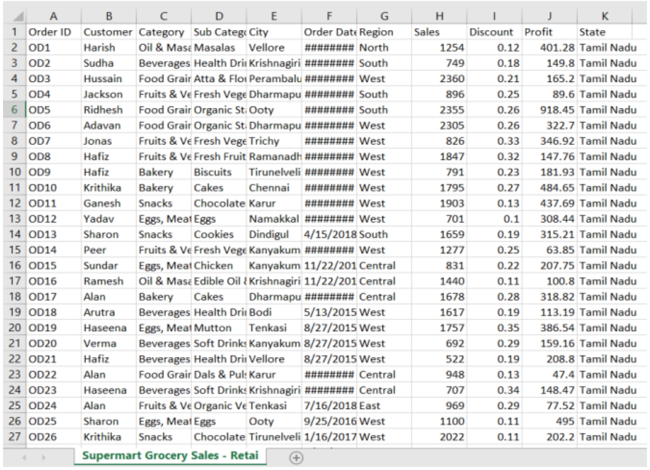
The submitted dataset must be in excel format. The characteristics are information on the products obtained through stock intake and stock sales, which can be gathered via the billing system of that specific store.

RESULT ANALYSIS

Figure shows the hierarchical clustering of the provided dataset. Each attribute is organized and has a graph. In hierarchical clustering the data follows 5 steps

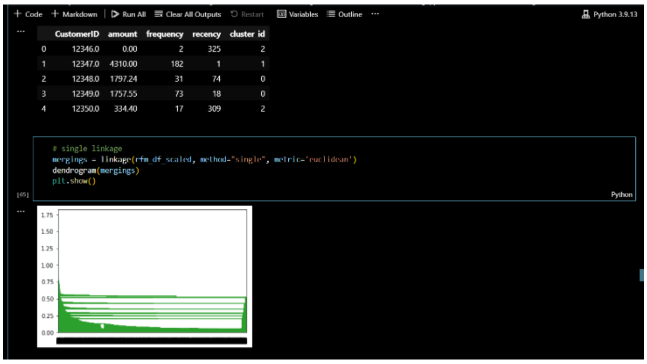
- Read and understand the data
- Clean the data
- Prepare the data
- Data Representation
- Final report of analysis

As a result, a number of clusters are formed, each of which differs from the others but nevertheless sharing many characteristics.



	A	B	C	D	E	F	G	H	I	J	K
1	Order ID	Customer	Category	Sub Categ	City	Order Date	Region	Sales	Discount	Profit	State
2	OD1	Harish	Oil & Masa	Masalas	Vellore	########	North	1254	0.12	401.28	Tamil Nadu
3	OD2	Sudha	Beverages	Health Dri	Krishnagiri	########	South	749	0.18	149.8	Tamil Nadu
4	OD3	Hussain	Food Grair	Atta & Flo	Perambalu	########	West	2360	0.21	165.2	Tamil Nadu
5	OD4	Jackson	Fruits & Ve	Fresh Vege	Dharmapu	########	South	896	0.25	89.6	Tamil Nadu
6	OD5	Ridhesh	Food Grair	Organic St	Ooty	########	South	2355	0.26	918.45	Tamil Nadu
7	OD6	Adavan	Food Grair	Organic St	Dharmapu	########	West	2305	0.26	322.7	Tamil Nadu
8	OD7	Jonas	Fruits & Ve	Fresh Vege	Trichy	########	West	826	0.33	346.92	Tamil Nadu
9	OD8	Hafiz	Fruits & Ve	Fresh Fruit	Ramanadh	########	West	1847	0.32	147.76	Tamil Nadu
10	OD9	Hafiz	Bakery	Biscuits	Tirunelveli	########	West	791	0.23	181.93	Tamil Nadu
11	OD10	Krithika	Bakery	Cakes	Chennai	########	West	1795	0.27	484.65	Tamil Nadu
12	OD11	Ganesh	Snacks	Chocolate	Karur	########	West	1903	0.13	437.69	Tamil Nadu
13	OD12	Yadav	Eggs, Meat	Eggs	Namakkal	########	West	701	0.1	308.44	Tamil Nadu
14	OD13	Sharon	Snacks	Cookies	Dindigul	4/15/2018	South	1659	0.19	315.21	Tamil Nadu
15	OD14	Peer	Fruits & Ve	Fresh Vege	Kanyakum	########	West	1277	0.25	63.85	Tamil Nadu
16	OD15	Sundar	Eggs, Meat	Chicken	Kanyakum	11/22/201	Central	831	0.22	207.75	Tamil Nadu
17	OD16	Ramesh	Oil & Masa	Edible Oil &	Krishnagiri	11/22/201	Central	1440	0.11	100.8	Tamil Nadu
18	OD17	Alan	Bakery	Cakes	Dharmapu	########	Central	1678	0.28	318.82	Tamil Nadu
19	OD18	Arutra	Beverages	Health Dri	Bodi	5/13/2015	West	1617	0.19	113.19	Tamil Nadu
20	OD19	Haseena	Eggs, Meat	Mutton	Tenkasi	8/27/2015	West	1757	0.35	386.54	Tamil Nadu
21	OD20	Verma	Beverages	Soft Drinks	Kanyakum	8/27/2015	West	692	0.29	159.16	Tamil Nadu
22	OD21	Hafiz	Beverages	Health Dri	Vellore	8/27/2015	West	522	0.19	208.8	Tamil Nadu
23	OD22	Alan	Food Grair	Dals & Puls	Karur	########	Central	948	0.13	47.4	Tamil Nadu
24	OD23	Haseena	Beverages	Soft Drinks	Krishnagiri	########	Central	707	0.34	148.47	Tamil Nadu
25	OD24	Alan	Fruits & Ve	Organic Ve	Tenkasi	7/16/2018	East	969	0.29	77.52	Tamil Nadu
26	OD25	Sharon	Eggs, Meat	Eggs	Ooty	9/25/2016	West	1100	0.11	495	Tamil Nadu
27	OD26	Krithika	Snacks	Chocolate	Tirunelveli	1/16/2017	West	2022	0.11	202.2	Tamil Nadu

Fig. 5.Sample Dataset.



```
   CustomerID   amount  frequency  recency  cluster id
0     12346.0     0.00          2      325           2
1     12347.0  4310.00        182        1           1
2     12348.0  1797.24         31       74           0
3     12349.0  1757.55         73       18           0
4     12350.0   334.40         17      309           2
```

```
# single linkage
mergings = linkage(rfm_df_scaled, method="single", metric='euclidean')
dendrogram(mergings)
plt.show()
```

Fig. 6.Segmention of hclustering

The centroids are found and categorized based on their attributes on either side, they are linked, and as a result, a variety of unique features are produced. The dendrogram displays two different materials by calculating their Euclidean distance. By grouping points into clusters, Euclidean clustering creates chains of points between any two points in a cluster where the projected distance between the next points in the chain is less than a certain limit. The data is grouped using Euclidean distance.

RFM (recency, frequency, monetary) analysis can quickly categorize products into homogenous groups using a small number of characteristics. Instead of using the raw, determined FM values for clustering phases, we can achieve better results by applying values. Consequently, RFM scoring should be used to segment, and additional spending behavior analysis should be done on the raw numbers for the targeted cluster to reveal more characteristics and insight. RFM analysis, which primarily relies on buying habits and histories, can be improved by investigating weighted composite values or integrating product demographic and information data. Effective and efficient marketing campaigns can boost profitability at the lowest possible cost with the help of a competent analysis.

K-means provide the efficient clean data for our model. Among the other algorithm k-means works well. To find the best k:

Elbow Method

Finding elbow point "Samples' sum of squared distances to the nearest cluster center". The elbow approach offers a situation in which adding more data samples little affects clusters. Elbow method finds the elbow point.

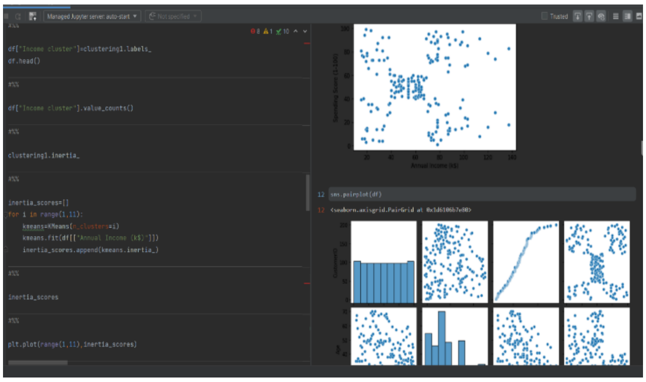
Silhouette Analysis Method

The silhouette score identifies if there are significant gaps between any given sample and every other sample within a cluster or between clusters.

$$Silhouette\,score = \frac{p-q}{\max(p,q)} \quad (4)$$

The data point has an average distance of G within its own cluster, and an average distance of p to the nearest cluster of which it is not a part. The scoring range for silhouettes is between -1 and 1.

The closer a data point's score is near 1, the more similar it is to the other points in its cluster. A value closer to -1 indicates that it cannot be used as a benchmark against the data points in its cluster.



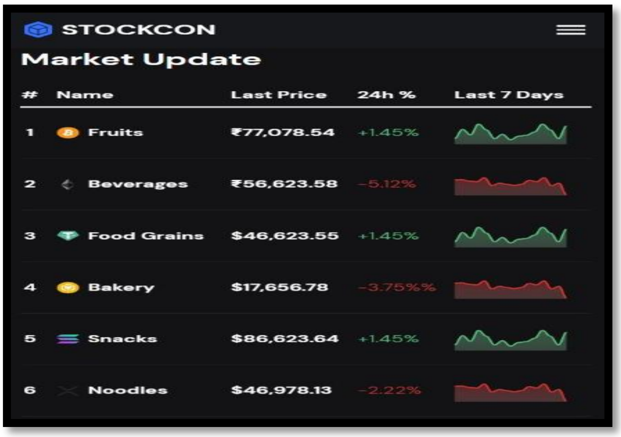
```
#%%

df['Income cluster']=clustering1.labels_
df.head()

#%%

df['Income cluster'].value_counts()

#%%

clustering1.inertia_

#%%

inertia_scores=[]
for i in range(1,11):
    kmeans=KMeans(n_clusters=i)
    kmeans.fit(df[['Annual Income (k$)']])
    inertia_scores.append(kmeans.inertia_)

#%%

inertia_scores

#%%

plt.plot(range(1,11),inertia_scores)
```

```
12 sns.pairplot(df)
12 <seaborn.axisgrid.PairGrid at 0x1d6106b7e08>
```

Fig. 7.Segmentation result of kmeans clustering.

The k-Means clustering of the supplied dataset that is segmented is displayed in fig. 7. Every attribute has a graph and is grouped.

The K-means algorithm constructs the clusters based on the centroid concept. The centroid of the n points, which is merely another point with its own x and y coordinates on an X-Y plane, is a common name for the geometric center of the n points. Following a new calculation, the centroid will now simply be the mean of all the points in each cluster. Then, we will get our new cluster centers.

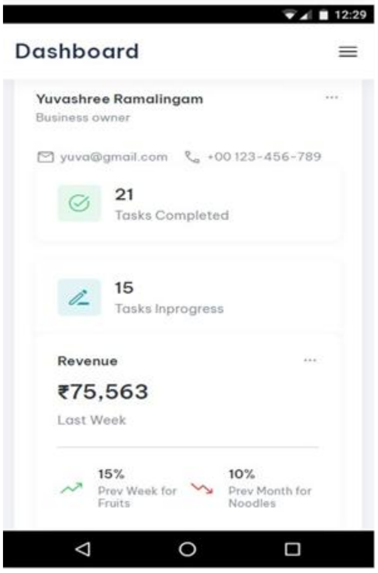


Market Update

#	Name	Last Price	24h %	Last 7 Days
1	Fruits	₹77,078.54	+1.45%	
2	Beverages	₹56,623.58	-5.12%	
3	Food Grains	$46,623.55	+1.45%	
4	Bakery	$17,656.78	-3.75%%	
5	Snacks	$86,623.64	+1.45%	
6	Noodles	$46,978.13	-2.22%	

Fig. 8.Depiction of 24 hr market update.

Fig. 8. information shows the rise and fall in purchases. The customer can analyze what to do to boost the earnings of their business using the offered data on how the market is high or low from stock intake. The provided dataset will be clustered, and the graph will output various results for the specified categories. The line graph visually represents the product's gain and loss. The product's profit or loss on a daily basis is indicated, along with its current and last day market price.



12:29

Dashboard

Yuvashree Ramalingam
Business owner

yuva@gmail.com
+00 123-456-789

21
Tasks Completed

15
Tasks Inprogress

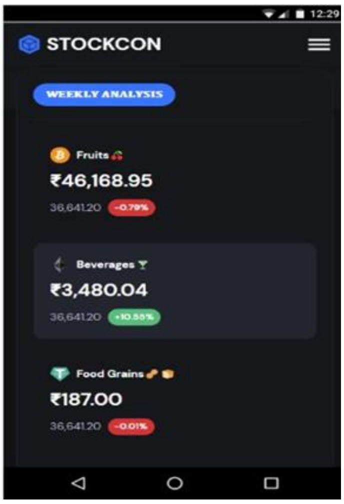
Revenue
₹75,563
Last Week

15%
Prev Week for Fruits

10%
Prev Month for Noodles

Fig. 9.Retailers progress

Fig. 9. represent the proposed work's dashboard, that shows a graph of the previous month's analysis and the current month's analysis. Along with this, the end user profile includes other duties like deciding whether to invest in a product and displaying the total revenue, or the seller's entire budget. Analysis of the market is done on a daily basis for that specific product throughout that season. Later, it will be shown with a price and a profit margin.



12:29

STOCKCON

WEEKLY ANALYSIS

Fruits
₹46,168.95
36,641.20 -0.79%

Beverages
₹3,480.04
36,641.20 +10.55%

Food Grains
₹187.00
36,641.20 -0.01%

Fig. 10.Weekly data analysis

Fig. 10. Shows the dashboard provides the seller a notion of where more money should be spent or concentrated in order to maximize sales and avoid squandering money.

CONCLUSION

The proposed model can be expanded by categorizing the items depending on the amount recommended to be invested in a certain product. The seller's stock investment will be given as input, and the software will generate a list of products and the amount to invest in each product. This study advocates for improved collaboration between the two scientific disciplines of data science and visualization. We conclude that this has become the most convenient means of communication between firms and sellers. This was also demonstrated in our study provided in this paper on the use of visualization approaches in standard data science tools. In contrast, interviews with data scientists reveal a strong interest in using innovative approaches to extract new information from their data sets. Several ways are proposed to better integrate visualization tools into popular data science workflows.

REFERENCES

A. Rosewelt L. and A. Renjit J., "An Intelligent Subtype Fuzzy Cluster based Relevant User Data Retrieval Model for Effective Classification," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 49-54, doi: 10.1109/ICONSTEM.2019.8918796.

G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

T. Kansal, S. Bahuguna, V. Singh and T. Choudhury, "Customer Segmentation using K-means Clustering," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 135-139, doi: 10.1109/CTEMS.2018.8769171.

A. R. L, S. K. P, T. J, A. T. S, P. M and V. K. M, "A Novel Machine Learning Approach to Predict Sales of an Item in E-commerce," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-7, doi: 10.1109/ICSES55317.2022.9914077.

Z. Shuai and L. Wenli, "Game-theoretic analysis of a two-stage dual-channel supply chain coordination in the presence of market segmentation and price discounts," Electronic Commerce Research and Applications, 57, 101222(2023).

M. Fanlin, M. Qian, L. Zixu, Z. Xiao-Jun, "Multiple dynamic pricing for demand response with adaptive clustering-based customer segmentation in smart grids," Applied Energy, 333, 120626(2023).

S. Maurya and N. K. Verma, "Intelligent Hybrid Scheme for Health Monitoring of Degrading Rotary Machines: An Adaptive Fuzzy c-Means Coupled with 1-D CNN," in IEEE Transactions on Instrumentation and Measurement, doi: 10.1109/TIM.2023.3253887.

A. Rosewelt L, S. B and G. S. C, "An Effective Detection of Version Number Attacks in the IoT using Neural Networks," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-7, doi: 10.1109/ICAECT54875.2022.9807966.

A.A. Abdulnassar and L. R. Nair, "Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model," Measurement: Sensors, 25, 100666(2023).

A. L. Rosewelt, N. D. Raju and S. Ganapathy, "An Effective Spam Message Detection Model using Feature Engineering and Bi-LSTM," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752652.

REFERENCES

[1] Bhanumathi.S., M. Vineeth., Rohit., N (2019). Crop yield Prediction and efficient use of Fertilizers. International Conference on Communication and Single Processing, April 4-6, 2019, India

[2] Chlingaryan, Anna., Sukkarieh, Salah., Whelan, Brett (2018). Machine learning approach for crop yield prediction and nitrogen status estimation in precision agriculture. Computer and Electronics in Agriculture, doi: -https://doi.org/10.1016/j.compag.201 8.05. 012..

[3] Dimitriadis, S., Goumopoulos, Christos., Applying Machine learning to Extract New Knowledge in precision Agriculture application. DOI 10.1109/PCI.2008.3

[4] Ghadge,R., Kulkarni, juilee., More,Pooja,. Nene,Sachee., NL,Priya.,(2018). Predication of crop Yield using machine learning. International Research Journal of engineering and technology (IRJET). Volume:05 Issue: 02| feb-2018, page 22-37.

[5] Goldstein, A., Fink, L., Meitin, A., Bohadana, S., Lutenberg, O., Ravid, G., 2017. Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. Prec. Agric.

[6] Kalimuthu,M., Vaishanvi,P.,Kishore, M.CropPredicition Using machine learning .Proceeding of the Third international Conference on Smart System and Invention Technology(ICSSIT 2020)IEEE Xplore Part Number:CEP20P17-ART;ISBN:978-1-7281-5821-

[7] Kale, Shivani S., and Preeti S. Patil. "A Machine Learning Approach to Predict Crop Yield and Success Rate." 2019 IEEE Pune Section International Conference (PuneCon). IEEE, 2019.

[8] Klompernburg, van Thomas., Kassahum, A., Catal, Cagatay., Crop yield Prediciton using machine learning. Computers and Electronics in agriculture https://doi.org/10.1016/j.compag.2020.105709.

[9] Kumar, Y.Jeevan Nagendra, et al."Supervisd machine learning Approach for Crop yield Predication in Agriculture Sector". 2020 5th International Conference on Communication and Electronic System(ICCSE).IEEE, 2020.

[10] Malik, P., Sengupta, S.,jadon,J.,(2021). Comparative Analysis of SoilProperties to Predict fertility and Cropyield using Machine leaning Algorthium.2021 11th International Conference on cloud Computing DataScience& engineering.

[11] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop Yield Prediction using Machine Learning Techniques." 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE, 2019.

[12] Palmview, K., Surianarayanan, C (2019). An Approach Predicition of Crop yield using Machine learning a leaning and big data technique. International Journal of Computer Engineering and Technology (IJCET) Volume 10, Issue 03, May-June 2019, pp. 110-118, Article ID: IJCET_10_03_013

[13] Priya, Rashmi,Ramesh,D.,(2020). MLbasedsuatainable precision agriculture: A future generation Perspective Computing nfomation and system

[14] Shriya Sahu et al," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach, International Conference on Computing, Communication and Automation, 2017.

[15] Sharma, A., jainA., Gupta,P., Chawdary,V.,(2020).Machine leaning Application for precision Agriculture:Acompreshion Review.Volume9,2021

[16] Y. Zheng, W. Wang, B. Chen, L. Zhang, S. Phangthavong, Z. Su, L. Zhang, G. Xiao, Determining the number of instars in potato tuber moth Phthorimaeaoperculella (zeller) using density-based dbscan clustering, J. Appl. Entomol. 143 (10) (2019) 1080–1088.

[17] Y. Dash, S.K. Mishra, B.K. Panigrahi, Rainfall prediction for the Kerala state of India using artificial intelligence approaches, Comput. Electr. Eng. 70 (2018) 66–73