" Customer Segmentation Using K-Means"

M.Sunnetha[1], Nagothu Srujana[2], Shaik Madeena Nashra[3],
Kornepati Nayomi[4] [1] Professor, [2, 3 & 4] Student
[1] msuneetha973@gmail.com, [2] srujanachowdarynagothu123@gmail.com, [3] nashrashaik6@gmail.com , [4] nayomi7308@gmail.com
Department of Computer Science and Engineering,
Narasaraopeta Engineering College, Narasaraopet, Andhra
Pradesh, India

ABSTRACT- "Effective decisions are hilariously to the all mandatory for any funky company to generate good revenue. In these bumpy days, competition is humongous and all companies are moving forward with their own https://ieeexplore.ieee.org-duper different strategies, you know what I mean? We should totally use data and sort of take a super cool proper decision, dude. Without this, it will be really, like totally, very difficult and no better techniques are, like, available to find the group of people with similar, like, character and interests in a large dataset. Nowadays Customer segmentation became oh so totally very popular method for, like totally dividing company's customers for totally retaining customers and making super-duper profit out of them, in the totally awesome following study customers of, like, different of organizations are classified on the basis of their, like, behavioral characteristics, like, such as spending and income, by taking, like, behavioral aspects into consideration makes these methods an efficient one as, like, totally compares to, like, others, dude. Here, the, like, customer segmentation using, like, K-Means clustering helps to sort of group the data with, like, same attributes which exactly helps to, like, business the best, super cool. In this totally rad project elbow method is used to sort of find the number of clusters and at last we, like, totally visualize the data."

KEYWORDS: Customer segmentation, classification, machine learning, clustering, K-means clustering, Elbow Method, and logistic regression.

I. INTRODUCTION

Customer segmentation is the process of dividing a clientele into several groups based on specific characteristics or behaviors using machine learning [1]. In order to put this into practice, begin by gathering pertinent data, such as client demographics, past purchases, and website usage. Handle missing values and normalize numerical features to preprocess the input. Depending on your objectives and the type of data you have, pick a machine learning technique that works for you, such as k-means clustering. Utilizing[2] the prepared dataset, train the model. Then, apply the model to fresh data to classify clients into various categories. Determine the traits that characterize each category by analyzing the data, then adjust marketing tactics or offerings accordingly. As new data becomes available, update the model frequently to make surethat client segments[3]. Marketing strategy can be affected by segmentation either directly or indirectly because it creates a multitude of new avenues for investigation, such as determining which member a product is appropriate for, tailoring marketing strategies to each member, providing accommodations for a particular member, and figuring out the client-object relationship—a concept that the company was previously unaware of[4] clustering has been shown to be beneficial for client segmentation. Unsupervised literacy known as clustering[5][6] enables us to find clusters in unlabeled datasets. K-means is one approach to cluster data. The main goal of this work is to use the K-means clustering technique to segment data and apply a data mining strategy to identify consumer groups. Using K-means clustering for consumer segmentation and RFM (Recency, Frequency, Monetary) for client grouping based on transactional behavior is a simple method. RFM analysis entails assessing the amount spent, frequency of transactions, and recentness of purchase for each consumer. A three-dimensional feature space is produced by these three measures. After that, group clients with comparable RFM profiles using K-means clustering[7]. K-means clustering uses proximity in this feature space to separate clients into discrete segments. For instance, there might be two distinct segments: one with regular big spenders and another with less frequent but more recent customers.

II. LITERATURE SURVEY

The literature review on machine learning techniques for consumer segmentation covers a range of approaches, with a particular focus on the K-means clustering algorithm. Kumar Dhiraj et al.[1], K-means is applied through a growing integration method, progressively adding clusters through a global search process. The deterministic approach proposed ensures result consistency irrespective of primary clusters and avoids changeable specifications. Aman Banduni, et al [2] presents an unsupervised K-means algorithm that uses data generation to dynamically determine the number of clusters during duplication. The algorithm's strength lies in adaptability to different collection quantities and designs. Vijilesh et al.[3] marketing tactic called customer segmentation separates a market into discrete, uniform segments. The customer segmentation approach uses data based on several characteristics, such as economic patterns, demographic trends, and behavioral patterns, to separate customers into different categories. A company's marketing resources can be more effectively employed with the aid of a client segmentation approach and offering flexibility in modeling cluster geometry and handling missing data is shown in below fig1.

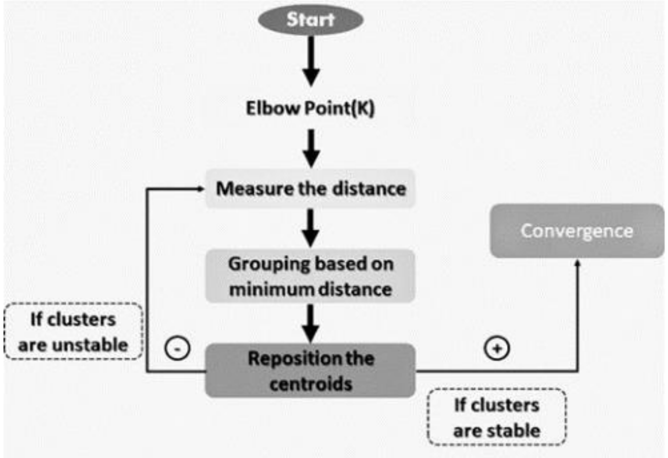


Fig.1: Flow Chart

Singa et al. discuss how to choose the K value in the K-means algorithm.[4], it is similar to structuring data to use unsupervised k-means clustering for consumer segmentation. The first step in unsupervised k-means clustering is to randomly categorize them according to how similar these traits are. After then, iterative adjustments are made to these groupings until they are as distinct from one another within each cluster and as homogeneous as feasible within each cluster. It's like shuffle folks around the room till you get smaller groups of people who are distinct from one another but similar to each other. By understanding various consumer types based on their purchasing patterns and demographics, segmentation enables firms to better cater their marketing strategies and product offers to the specific demands of each customer group. Shirole et al.[5] client segmentation is a technique used by businesses to divide their client base more effectively for marketing purposes. It involves grouping customers based on comparable attributes including age, region, income, education, gender, spending patterns, and purchase behavior. Because it may be applied to create marketing tactics such as selecting the right target market for product recommendations. The goal of Dawane et al.'s study [6] is to identify the ideal number of clusters for consumers with low profit margins by applying the elbow approach to identify appropriate elbow groupings. Shirole et al. propose the K-means approach[7][8] K-means clustering, other types of clustering, including affinity, density-based, and clustering. Customer segmentation has also been accomplished with propagation clustering. A comparison study has been done on the algorithms. Grouping clients according to shared attributes, such as age, location, income, education, gender, spending patterns, and purchase behaviour, is known as customer segmentation. in order for businesses to effectively distribute their consumer base in order to promote their brand. In customer relationship management, it is essential. because marketing techniques such as selecting relevant consumers for product recommendations can be planned using it. In order to identify the target market and know how to best reach them through marketing, marketers need research their target market, as well as understand their needs and cognitive processes. Customer reliability and the commercial relationship between businesses and customers have existed for decades. Pradana et al.[9]numerous academics have extensively experimented with the integration of Machine learning techniques like clustering, which is a method of learning unstructured data, with consumer segmentation. The goal of clustering is to maximize the differences between the groups while simultaneously trying to increase the similarity within the clusters as much as feasible. Diraj kumar et al.[10] the grid-K-means algorithm is introduced in combining grid integration with K-means to enhance integration abundance and reduce parameter values, demonstrating improved speed and accuracy in all various datasets.

Asith Ishantha et al.[11]completely examined a variety of clustering algorithms, including hierarchical clustering,mini-batch-means clustering, K- means clustering, and numerous further. and use each of these algorithms to member guests. also, compare the issues and elect the most effective clustering fashion to carry out the segmentation. K.P.Singa et al.[12]Unsupervised The machine learning technique known as "k-means clustering" separates a dataset into distinct groups, or clusters, based on participant attributes without considering previously tagged data.The thing of unsupervised k- means clustering is to automatically form clusters out of affiliated data points without the need for labeled information or pre-established orders. It's an effective system for relating structure and patterns in data, and it's constantly applied to jobs like anomaly discovery, image contraction, and consumer segmentation. Ezenkwu et al.[13] Using the k-means method to segment clients is similar to dividing a large population into smaller, more manageable clusters based on common features the k-means algorithm analyses client data, including demographics and purchasing history, and then groups guests according on participation features, such as copping behaviors or preferences. Once these clusters are established, companies can modify their immolations or marketing plans to suit the unique conditions of each group. This strategy aids companies in allocating coffers more effectively and in offering further material and useful services, which ultimately improves customer happiness and fidelity. Datta et al.[14] Scalability and fault forbearance are two features that distributed Python fabrics constantly give, letting you gauge up or down in response to workload demands. This implies that you can effectively manage unanticipated increases in data volume without immolating responsibility or performance. With the help of a distributed Python frame, companies may greatly ameliorate the effectiveness of their client segmentation procedures, which will eventually affect in more precise perceptivity, better decision- timber, and increased client retention and happiness on k-means, contributing to the model's high efficiency in addressing clustering challenges related to actual customer consumption patterns.

III. METHODOLOGY

The dataset used for K-means clustering was provided by a shopping center business. The data collection has 200 tuples and five attributes, representing the personal data of 200 customers. Customer ID, gender, age, yearly income (k$), and expenditure score on a scale of 1 to 100 are the characteristics that are being collected. Collection Finding out what kind of data we'll be dealing with is the first step (see table 1 for the dataset). We use a simple yet complete dataset with customer ID, age, gender, and yearly income as well as purchase score. An expenditure score, which runs from 1 to 100, represents the value of the customer's mall to the all purchases or outlays. The total quantity spent increases with the number.) There are no null values and the dataset's structure is accurately displayed. Data cleansing is required if a dataset has null values, duplicates, or other noisy data. Information is made dependable, useable, and available for study through data cleansing. When the data is available, we can compare the gender-specific annual income and spending score to illustrate the data. Estimate the sum of square distances from each point to show consumer behaviors associated with annual income and spending scores, as well as groups of customers participating in the following activities:

1. High Income/Low Spending Score
2. Low Income: A high expenditure index
3. Despite having low income, a high expenditure score
4. Average Income - Score for Average Spending
5. High Spending-High Income Ratio
6. Annual Income Compared to Spending
7. Spending: Zone versus Score.

Now, but not in great detail, we may develop a K-means model based on the fact that there are many groups. Using the silhouette coefficient approach, one can estimate the sum of square distances for each value from each point to its designated center and perform k-means clustering for a range of k clusters, say 1 to 10. Choose how many clusters would provide you with the highest silhouette score.

Here is the formula for calculating the silhouette score. We observed that there is no more quick movement in WCSS (Within Cluster Sum of Squares) once K=5 is reached. Furthermore, K=5 will be the appropriate amount of clusters given the number of clusters we now have. See the example below. The outcome of the silhouette approach. Using the previously mentioned strategy, we may use the plot to be divided into multiple groups, identify which clusters should be prioritized, and then give each group a label. Which of the five clusters clients with Moderate Income-Moderate Spending Score, High Score, and Low Income-High Spending Score should be targeted can be determined using the K-means approach. The necessary customers have been found.

A. OVERVIEW OF DATASET

You've put in place a membership card system to collect vital client data, including client ID, yearly income, age, gender, and education expenses, as well as zone classification (target consumers, regular customers, spenders, and frugal people).Additionally, you've all the introduced a spending score, a metric derived from specific parameters like purchase history and behavior. The goal is to identify and focus on "Targeted Consumers" who align with predetermined conditions. The dataset, sourced from Kaggle [15], includes 200 rows and 6 columns with attributes such as expenditure score, zone, age, gender, and annual income. This dataset is essential for comprehending the inclinations and routines of regular mall patrons, enabling you to give the sales team precise data in fig2 for targeted client engagement and strategic planning.

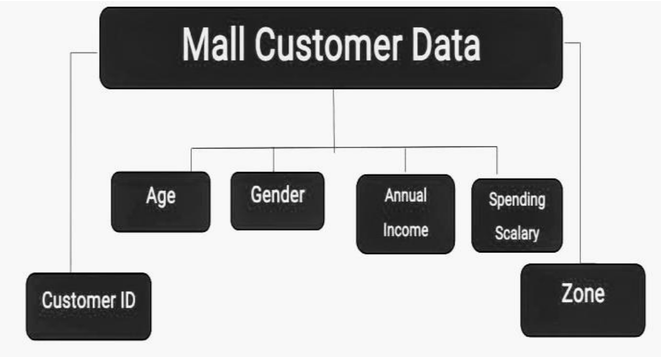


Fig.2: Dataset Graphical Representation

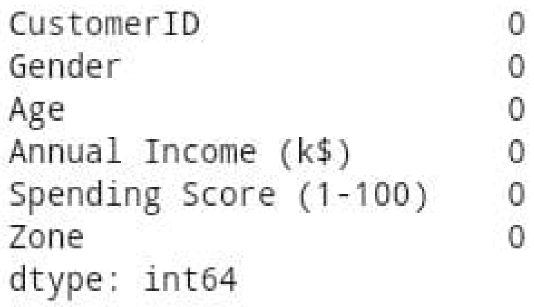
In customer segmentation within malls, the concept of "zone" refers to different categories or groups of customers based on their spending behaviors and patterns. Balanced Customers: These customers exhibit a well-rounded and moderate spending behavior. Target Customers: This group consists of customers who align closely with the mall's target market Normal Customers: These customers represent the average or typical shopper in the mall. Spenders: This group consists of customers with high disposable income. Pinch Penny Customers: These customers are known for being frugal and cautious with their spending.

B. ANALYSIS OF DATA

Analyzing customer data for segmentation using k-means clustering, particularly with mall data, involves extracting meaningful insights from various customer attributes to better understand and categorize their behavior. Think of it as unraveling a mystery within the bustling atmosphere of a mall. The information gathered in this case comprises the customer's ID, age, gender, yearly income, expenditure score, and maybe their zone of residence. These characteristics provide hints for understanding consumer habits, interests, and to all purchasing power. K-means clustering can be used to put similar consumers in one group based on these attributes after they have been gathered and cleansed.This process partitions customers into clusters with similar traits, enabling mall operators to tailor their marketing strategies, optimize product offerings, and enhance customer experiences. For instance, they might target high-spending customers with exclusive offers or adjust their inventory to better cater to different age groups.

C. COMPLETE REVIEW

A preliminary examination of the dataset is part of the suggested process in order to gather information and evaluate the dataset’s attributes. An overview of the structure and content of the dataset can be obtained by using the (shape) and (info()) methods. Additionally, is null() is used to determine whether any missing values exist. Figure 3 shows that, out of the total clients, 56% are female and the remaining 44% are male shown in fig3.Customers in their 20s and 30s make up the majority of the sample, suggesting a younger population. A crucial component of market analysis is consumer segmentation, which may be explored using this dataset as a starting point. The objective is to classify consumers based on shared attributes using the unsupervised machine learning technique K-means clustering. The study uses this approach to look for trends.

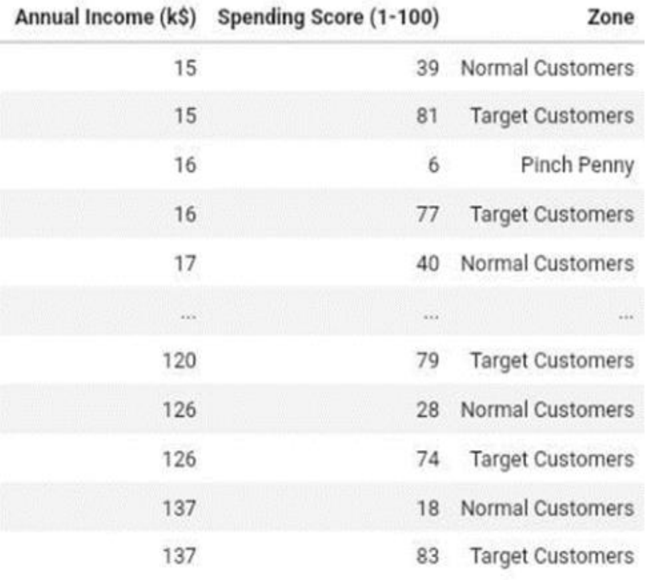


```
CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
Zone                      0
dtype: int64
```

Fig.3: Examining the dataset for null values

D. ANNUAL INCOME VS SPENDING SCORE VS ZONE

Businesses can divide their client base into discrete segments, including target consumers, typical customers, balanced customers, spenders, and pinch pennies, based on the correlation between annual income, spending score, and geographic zone. By using this segmentation technique, companies may efficiently cater their product offers, pricing, and marketing to each consumer segment's unique needs and preferences. Based on the relationship between annual income, spending score, and geographic zone, businesses can segment their customer base into specific groups, such as target consumers, average customers, balanced customers, spenders, and pinch pennies. Companies are able to effectively cater to a wide range of needs and preferences.



Annual Income (k$)	Spending Score (1-100)	Zone
15	39	Normal Customers
15	81	Target Customers
16	6	Pinch Penny
16	77	Target Customers
17	40	Normal Customers
...	...	...
120	79	Target Customers
126	28	Normal Customers
126	74	Target Customers
137	18	Normal Customers
137	83	Target Customers

Fig.4: Annual Income Compared to Spending Score and Zone

E. ELBOW METHOD

Increasing the number of clusters may initially minimize internal variability within each cluster. This procedure entails determining the ideal number of clusters for a dataset.

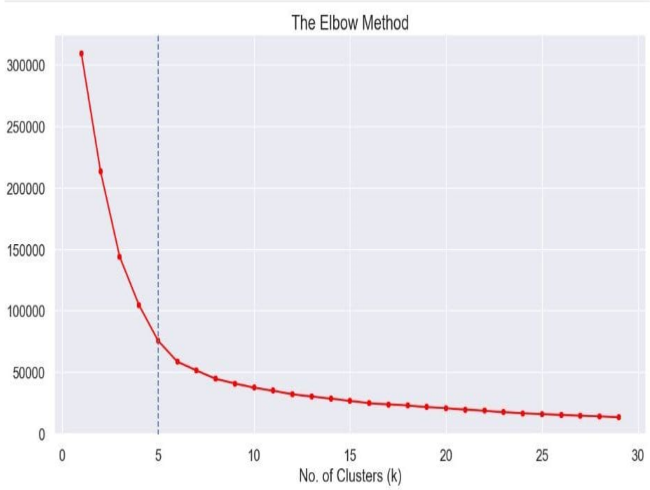


Fig.5: Elbow Technique

This is because more h clusters have the capacity to distinguish between data objects that share similar properties more finely. An algorithm with different values of k, from 1 to 30, is used to evaluate the total intra-cluster square in order to determine the right number of clusters. A lower value denotes better clustering. The total intra-cluster square is a measure of how compact a cluster is. The graph's bends, as seen in figure 5, are where the best collections can be located.

F. FITTING MODEL

In malls, customer segmentation entails putting patrons into groups according to particular traits or actions. Here is a broad rundown of some possible applications for the logistic regression, random forest k-means clustering, and Naïve Bayes classifier algorithms:

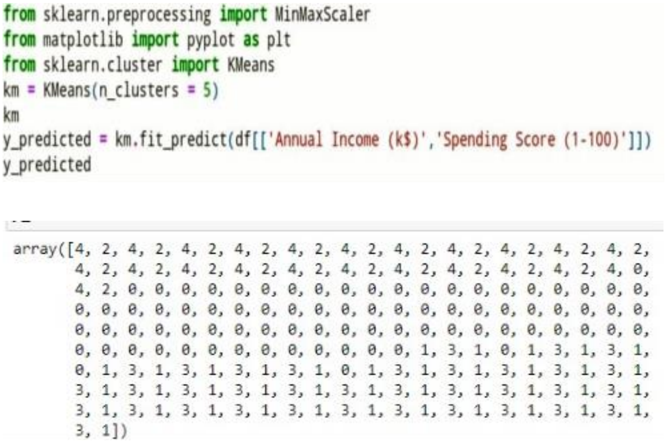
1. K-means Clustering:

The following steps are involved in the process:

a. Data Gathering: Gather and preprocess your data, taking into account variables such as client demographics, past purchases, and behavior.

b. Find the K, or number of clusters: To determine the ideal value for K, apply strategies like silhouette analysis and the elbow approach in fig6.

c. Apply K-means. The formula Use a scikit-learn package to apply the K-means algorithm on your data.



```
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
from sklearn.cluster import KMeans
km = KMeans(n_clusters = 5)
km
y_predicted = km.fit_predict(df[['Annual Income (k$)','Spending Score (1-100)']])
y_predicted
```

```
array([4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2,
       4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 2, 4, 0,
       4, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 3, 1, 0, 1, 3, 1, 3, 1,
       0, 1, 3, 1, 3, 1, 3, 1, 3, 1, 0, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1,
       3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1,
       3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1,
       3, 1])
```

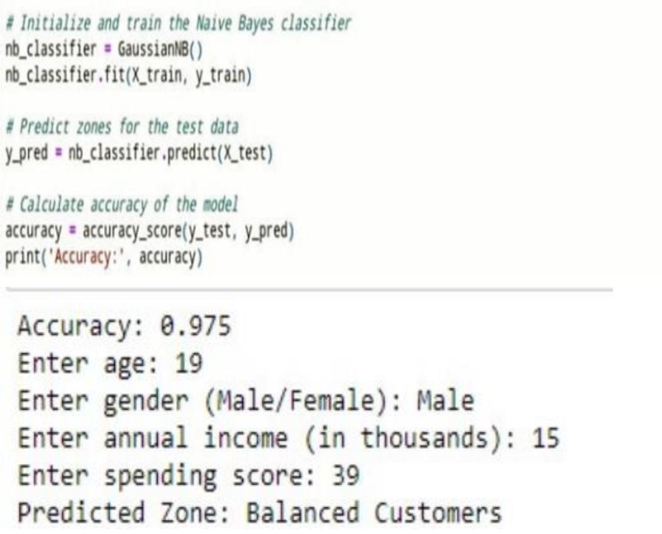
Fig.6: K-Means

2. Naive Bayes Classifier:

This algorithm uses probabilities to classify data.

a. Data Preparation: Assemble your data according to features and labels, taking into account elements that aid in the prediction of client groups in fig7.

b. Fit the Naive Bayes classifier to your training set of data in order to train it.



```
# Initialize and train the Naive Bayes classifier
nb_classifier = GaussianNB()
nb_classifier.fit(X_train, y_train)

# Predict zones for the test data
y_pred = nb_classifier.predict(X_test)

# Calculate accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)
```

```
Accuracy: 0.975
Enter age: 19
Enter gender (Male/Female): Male
Enter annual income (in thousands): 15
Enter spending score: 39
Predicted Zone: Balanced Customers
```

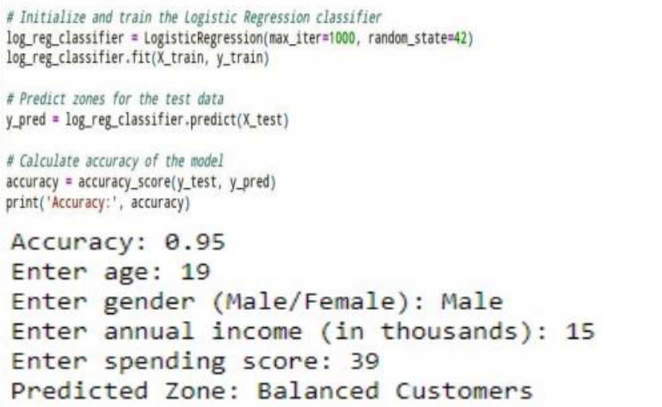
Fig.7: Navie Bayes Classifier

3. Logistic Regression:

This linear model is appropriate for binary categorization in fig8.

a. Preparing the Data: As with Naïve Bayes, group the data according to characteristics and labels.

b. Fit the logistic regression model to your training set of data in order to train it.



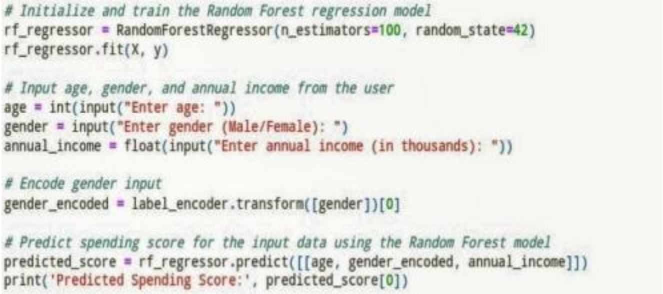
```
# Initialize and train the Logistic Regression classifier
log_reg_classifier = LogisticRegression(max_iter=1000, random_state=42)
log_reg_classifier.fit(X_train, y_train)

# Predict zones for the test data
y_pred = log_reg_classifier.predict(X_test)

# Calculate accuracy of the model
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)
```

```
Accuracy: 0.95
Enter age: 19
Enter gender (Male/Female): Male
Enter annual income (in thousands): 15
Enter spending score: 39
Predicted Zone: Balanced Customers
```

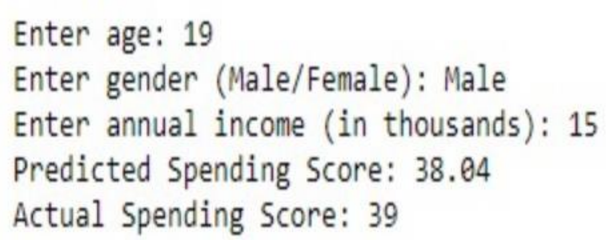
Fig.8: Relative Logistic

4. Random Forest:

Combining several decision trees, Random Forest is an ensemble learning technique a data preparation as with earlier algorithms, get your data ready. Fit the random forest model to your training set of data to train the random forest in fig9.



```
# Initialize and train the Random Forest regression model
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)
rf_regressor.fit(X, y)

# Input age, gender, and annual income from the user
age = int(input("Enter age: "))
gender = input("Enter gender (Male/Female): ")
annual_income = float(input("Enter annual income (in thousands): "))

# Encode gender input
gender_encoded = label_encoder.transform([gender])[0]

# Predict spending score for the input data using the Random Forest model
predicted_score = rf_regressor.predict([[age, gender_encoded, annual_income]])
print('Predicted Spending Score:', predicted_score[0])
```

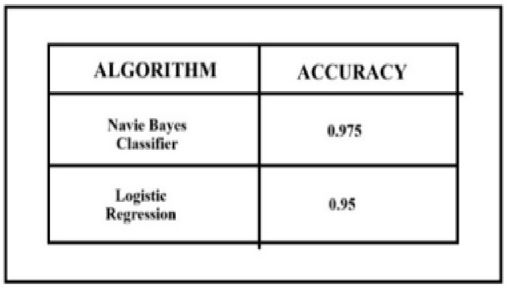


```
Enter age: 19
Enter gender (Male/Female): Male
Enter annual income (in thousands): 15
Predicted Spending Score: 38.04
Actual Spending Score: 39
```

Fig.9: Random Forest

5. Evaluation:

It's critical to assess these models' performance using metrics relevant to your task (e.g., accuracy for classifiers, silhouette score for K-means) once they have been fitted. To evaluate the generalization of your model, don't forget to divide your data into training and testing sets in fig 10. To boost performance, adjust hyperparameters as well. The actual code could change depending on the needs and dataset you have in mind.



ALGORITHM	ACCURACY
Navie Bayes Classifier	0.975
Logistic Regression	0.95

Fig.10: Accuracy Table

IV. RESULT

As seen in Fig.11, various groups (Clusters 1 to 5) appear in the visual representation of customer clusters based on Annual Income and Consumer Outcomes, each providing insightful information for data-driven decision making. In the lower right corner, Cluster 1 (shown in red) sticks out as it represents clients with higher yearly incomes but lower spending. Customers in Cluster 2 (yellow), which is in the middle, have modest yearly incomes and spending patterns. Customers in Cluster 3 (green), which is on the right, have the highest annual income and highest expenditure outcomes. Clientele within Cluster 4(blue), which is located in the upper left corner, have low yearly incomes but remarkably high spending levels. Finally, Cluster 5 (black) is located in the lower left and includes clients with extremely low yearly.

A thorough in fig 12,13 assessment of a classification model's performance across various classes is given in the classification report. Balanced Customers, Normal Customers, Pinch Penny Customers, Spenders, and Target Customers are the five types included in this research. The precision for each class shows the percentage of accurately predicted examples out of all the instances that are anticipated to be in that class. Recall, sometimes known as sensitivity, is the proportion of correctly anticipated cases of a class to all instances of that class that actually occur. The harmonic mean of precision and recall, or F1-score, provides a fair assessment of a model's performance.95%,.97% accuracys are the percentage of properly identified occurrences among all examples, which the model attained. The weighted-average and macro-average F1-scores offer a comprehensive understanding of the model's performance across all classes; the weighted-average accounts for class imbalance, while the macro-average treats each class equally (fig. 14). We also compared the accuracy in the bar graph.

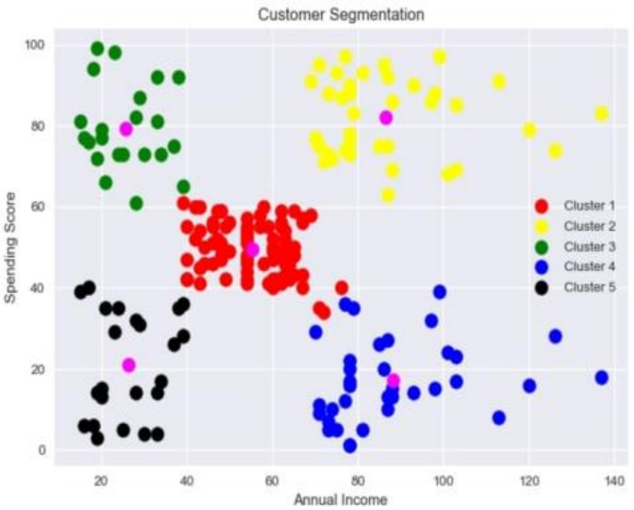
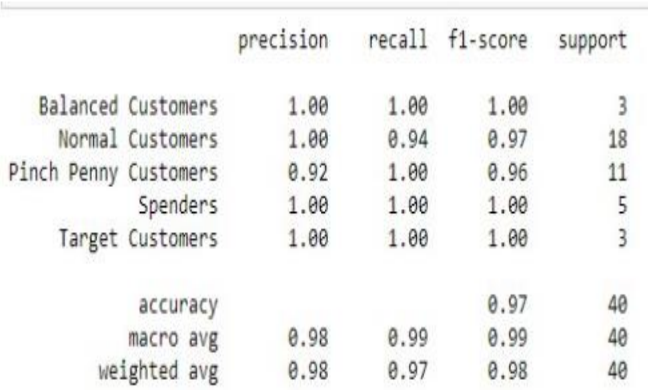
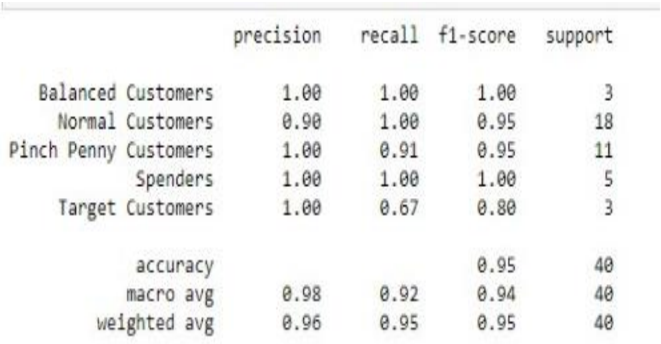


Fig.11: Customer clusters



```
                       precision    recall  f1-score   support

    Balanced Customers      1.00      1.00      1.00         3
      Normal Customers      1.00      0.94      0.97        18
 Pinch Penny Customers      0.92      1.00      0.96        11
              Spenders      1.00      1.00      1.00         5
      Target Customers      1.00      1.00      1.00         3

              accuracy                          0.97        40
             macro avg      0.98      0.99      0.99        40
          weighted avg      0.98      0.97      0.98        40
```

Fig.12: Naive Bayes Algorithm Metrics



```
                       precision    recall  f1-score   support

    Balanced Customers      1.00      1.00      1.00         3
      Normal Customers      0.90      1.00      0.95        18
 Pinch Penny Customers      1.00      0.91      0.95        11
              Spenders      1.00      1.00      1.00         5
      Target Customers      1.00      0.67      0.80         3

              accuracy                          0.95        40
             macro avg      0.98      0.92      0.94        40
          weighted avg      0.96      0.95      0.95        40
```

Fig.13: Logistic Regression Metrics

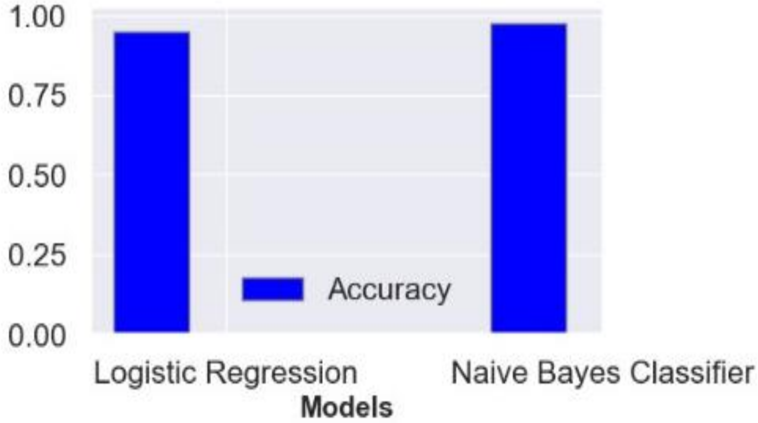


Fig.14: Comparison of Accuracy

V. CONCLUSION

Customer segmentation in malls based on factors like customer ID, age, gender, annual income, spending score, and zone can yield insightful data when implemented using K-means clustering, Naive Bayes classifier, Logistic Regression, and Random Forest algorithms.

In order to classify clients into discrete groups, each representing a certain behavior or spending trend, segmentation is employed. In this, we forecast the zone based on age, gender, spending score, and annual income in addition to the expenditure score based on these factors.

VI. REFERENCES

[1] Dhiraj Kumar, “Implemention of the Customer Segmentation Using Machine Learning”, July 10th 2021.

[2] Aman Banduni, Prof Ilavedhan A, “Customer Segmentation using machine learning,” School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India.(2022).

[3] V.Vijilesh, “CUSTOMER SEGMENTATION USING MACHINE LEARNING,” International Research Journal of Engineering and Technology (IRJET), vol. 08, no. 05, May 2021.

[4] Sinaga, K.P.; Yang, M.S. Unsupervised K-means clustering algorithm. IEEE Access 2020, 8, 80716–80727.

[5] Shirole, R.; Salokhe, L.; Jadhav, S. Customer Segmentation using RFM Model and K-Means Clustering. Int. J. Sci. Res. Sci. Technol. [Google Scholar] [CrossRef] 2021, 8, 591–597.

[6] Dawane, V.; Waghodekar, P.; Pagare, J. RFM Analysis Using K-Means Clustering to Improve Revenue and Customer Retention. In Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021), Online,[Google Scholar] [CrossRef] 29–30 April 2021.

[7] Monil, Patel, etal. “Customer Segmentation Using Machine Learning.”International Journal for Research in Applied Science and Engineering and Technology (IJRASET) 8.6 (2020): 2104-2108.

[8] Shirole, Rahul, Laxmiputra Salokhe, and Saraswati Jadhav. ”CustomerSegmentation using RFM Model and K-Means Clustering.” (2021).

[9] Pradana, Musthofa Galih, and Hoang Thi Ha. ”Maximizing Strategy Im-provement in Mall Customer Segmentation using K-means Clustering.”Journal of Applied Data Sciences 2.1 (2021).

[10] Author Dhiraj Kumar, “Implementing Customer Segmentation Using Machine Learning neptuneblog, Dec. 13, 2021.

[11] Ishantha Asith, "Clustering Algorithm for Mall Customer Segmentation," Future University Hakodate, Conference Paper, March 2021.

[12] M. S. Yang, K. P. Sinaga. "Unsupervised K-means clustering algorithm." 80716–80727 in IEEE Access 8, 2020.

[13] Ezenkwu C, Kalu C, and Ozuomba S Using the K-Means algorithm to effectively segment customers is a tactic for providing customized customer care. reached Nov. 21, 2021.

[14] David PE, Agarwal R, Datta D Customer segmentation performance improvement by distributed Python framework, ray. papers.ssrn.com (2020)

[15] Visit this link: https://www.kaggle.com/datasets/vjchoudhary7/python-customer-segmentation (Accessed April 20, 2022).