Promotional Suggestions for Target Customers Using Machine Learning Algorithm

A Project Report submitted to the

Department of Computer Science and Engineering, Jahangirnagar University

in partial fulfillment of the requirements for the degree of

M.Sc. in Computer Science and Engineering

By

Md Fahim Faisal

Exam Roll: 200488

Registration No: 41600

Session: 2019 - 2020

Supervised by

Md Abul Kalam Azad

Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

JAHANGIRNAGAR UNIVERSITY

November 2022

ABSTRACT

Promotion is a big part of business nowadays. A significant portion of the budget is allotted for this purpose. With the changing and competitive market people are more focused on effective Advertisement. This is where dataset segmentation comes into play. A lot of work is being done on customer segmentation using different machine learning algorithms. The goal of customer segmentation is to develop a relationship with the most lucrative clients by developing the most effective marketing plan. Our project aims to find out potential customers. By using the K-means clustering algorithm We've divided a customer dataset into clusters and separated the special clusters for another test set to predict if a new customer is potential or not. So that the promotional cost reduces.

DECLARATION

The research work entitled "**Promotional Suggestions for Target Customers Using Machine Learning Algorithm**" carried done at Jahangirnagar University's Department of Computer Science and Engineering is unique and complies with all university rules.

I am aware of the University's policy against plagiarism and hereby affirm that no material from this project has been previously published or submitted for credit toward a degree or certificate elsewhere.

_____

Md Fahim Faisal

Exam Roll: 200488

(Candidate)

Counter Signed by

_____

Dr. Md Abul Kalam Azad

Professor,

Department of Computer Science and Engineering

(Supervisor)

ACKNOWLEDGEMENT

I would want to sincerely thank my respected supervisor, **Prof. Md Abul Kalam Azad** for all of his help and support in getting my project finished. Additionally, I want to thank him for giving me the chance to work on a project about promotional suggestions for target customers. Without his assistance and suggestions, the project could not have been finished.

CONTENTS

Abstract ii

Declaration iii

Acknowledgement iv

List of Figures vii

List of Tables viii

List of Symbols ix

List of Algorithms x

1 Introduction ……1

1.1 Background and Motivation ……1

1.2 Objective ……...2

1.3 Related Works ……...3

2 Methodology ……5

2.1 Machine Learning ……5

2.2 Clustering ……..5

2.2.1 K means clustering ……....6

2.2.2 Elbow Method ……...8

2.3 Methodology ……9

2.3.1 Steps ……..9

2.3.2 Flowchart ……11

2.3.3 System Description ……..12

3 System Model ……...13

3.1 Data Description ……13

3.1.1 Exploring Dataset ……....13

3.1.2 Data Visualization ……...15

3.1.3 Data Splitting and Preprocessing ……18

3.2 Model Building ……18

3.2.1 Finding Optimal number of clusters ……..18

3.2.2 Model Building For Train Set ……19

3.2.3 Predicting New Customer's Cluster ……...19

4 Results Analysis ……..21

4.1 Model For Incoming Customers ……....21

4.2 Model For Suggested Customer ……22

5 Conclusion ……23

5.1 Summary ……23

5.2 Limitations & Future Work ……23

References ……25

LIST OF FIGURES

2.2.1 K-means clustering steps ……..7

2.2.2 Elbow Method for selection of optimal K clusters ……...8

2.3.2 Methodology Flowchart ……..11

3.1.2.1 Gender Count ……...15

3.1.2.2 Age Histogram ……15

3.1.2.3 Scatterplot of Age Vs Gender Vs Customer Rating ……...16

3.1.2.4 Jointplot of Age Vs Customer Rating ……..16

3.1.2.5 Scatterplot of Income Vs Customer Rating ……....17

3.1.2.6 Joint plot of Income Vs Customer Rating ……....17

3.2.1 Elbow Method for Optimal 5 Clusters ……....18

3.2.2 Model for Train Set Data ……....19

3.2.3 Predicting New Customer's Cluster ……20

4.1 Clusters of incoming customers ……21

4.2 Customers to Send Ad ……...22

LIST OF TABLES

3.1.1.1: First Five Sample Data ……14

3.1.1.2: Last Five Sample Data ……14

List of Algorithms

2.2.1 K means clustering ……6

CHAPTER 1

1 Introduction

1.1 Background and Motivation

Older businesses have been obliged to use marketing techniques to maintain market competitiveness as survival has grown increasingly challenging as a result of the market's growth. Consumers are becoming more numerous every day, and businesses are finding it challenging to meet each one of their wants. To effectively target each category, customer segmentation is employed, which is the process of grouping clients based on their shared qualities. Through analysis, businesses are able to learn what their actual customers are purchasing, which improves customer happiness. The analysis enables companies to identify their target market and enhance their marketing strategies to increase sales. Knowing your target market and where they are in the purchase process, as well as customer segmentation, are essential for successful marketing. Customer segmentation used to be a laborious and time-consuming process that required hours of hand-reading over various tables and data-querying in the hopes of discovering strategies to keep consumers together. However, using machine learning algorithms that adapt statistical regularity and data to such challenges has made it considerably easier in recent years. Customer data can be processed by machine learning models to find recurrent trends across various features. In many instances, marketing analysts can use machine learning algorithms to locate client segments that would be very challenging to identify through intuition and manual data analysis. Customer segmentation is an ideal illustration of how machine learning and human intuition may be combined to produce something that is more than the sum of its parts.

Customer segmentation has some benefits. Price optimization is the first benefit of client segmentation. Understanding the consumer and their financial situation will enable the business

to optimize prices appropriately and better allocate resources, which in turn enables businesses to achieve economies of scale. The second benefit is that it makes businesses more competitive. Customer retention affects revenue generation, which increases a company's ability to compete in the market. If a business segments the market, it will know who its clients are and can develop new items or variants in response to changing consumer preferences. The corporation may potentially benefit from the first mover advantage if it is watchful enough. Brand awareness is the third benefit. The corporation can effectively market to its consumer base by segmenting them. Customers who recognize the brand are more likely to buy the goods directly, which improves the company's reputation in the marketplace and its brand value relative to rivals. The fourth benefit is customer acquisition and retention through a personalized connection, which will assist the business win the satisfied customer. Customers that are happy with an organization's service are more inclined to stay with it. Building stronger relationships with potential clients will result from improved customer segmentation. The organization may better understand its clients by using customer segmentation, which enables it to better meet their demands. Increased income and interest returns would be the final benefit. The company's revenue will rise as a result of improving its marketing message because customers are more inclined to buy when they receive what they need.

1.2 Objective

Today, a significant portion of the company is marketing. The budget has set aside a considerable amount of money for this. People are increasingly focused on successful advertising due to the evolving and competitive market. Segmenting the dataset becomes important in this situation. Numerous efforts are being made to segment customers using various machine-learning techniques. By creating the best marketing strategy, customer segmentation aims to establish relationships with the most profitable customers. Identifying potential clients is the goal of our project. We separated a customer dataset into clusters using the K-means clustering algorithm, and we selected the special clusters for a different test set to determine if a new customer is promising or not. so that the expense of promotion is reduced.

1.3 Related Works

Anuj Vyas [1] Mall Customer Segmentation. Using k means clustering, he constructed customer clusters from a dataset of consumers based on their purchasing scores and annual salary.

Amit Kumar [2] Customer Segmentation of Shopping Mall Users Using K-Means Clustering. In this study, customer segmentation was carried out using preliminary data analysis on the data from the supermarket and k-means clustering. Two separate clusters were formed based on the annual salary vs. spending level and the age vs. spending score. Four optimal groups for age and expenditure scores and five optimal clusters for yearly income and spending scores were discovered using the elbow graph approach. First off, when respondents were categorized by age and expenditure score, those in higher age groups had lower spending scores. Second, groups with high annual income scores and very low spending scores on the basis of annual income and spending scores. The retailer can therefore provide these clustered clients with incentives to visit.

Kyoung-jaeKim, HyunchulAhn [3] A recommender system using GA *K*-means clustering in an online shopping market. This study recommends GA K-means as a fresh clustering approach. From the standpoint of intraclass inertia, we observed that GA K-means may produce superior segmentation than other conventional clustering algorithms, including plain K-means and SOM, in a real-world application for market segmentation in internet commerce. The utility of GA K-means as a preprocessing tool for recommendation models was also empirically investigated.

Musthofa Galih Pradana, Hoang Thi Ha [4] Maximizing Strategy Improvement in Mall Customer Segmentation using K-means Clustering. In this research, K-means clustering is utilized as the cornerstone for the segmentation that will be carried out, with other models, of course, being used to back up the study findings. Customers with high-salary levels and high spending ratings were found to be ideal targets for launching marketing campaigns. As a result, they were able to divide consumers into 5 clusters based on the association between their purchase score and yearly salary.

Ioannis Mesionis [5] Mall Customer Segmentation - Unsupervised learning. He used k means clustering to group customers into groups based on their expenditure scores and annual income.

Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury [6] Customer Segmentation using K-means Clustering. In this study, the buyers are divided into segments using 3 different clustering approaches (k-Means, Agglomerative, and Meanshift), and the results of the segments created by the methodologies are then compared. A dataset with two features—200 training samples collected from local retail outlets and a common scaler—has been used to train Python software. Both characteristics are made up of the typical annual store visits and the typical quantity of purchases made by customers. Five cluster segments—Careless, Rigorous, Regular, Targeted, and Reasonable clients been produced through the use of clustering. But after applying mean shift clustering, two more groups—High buyers and frequent visitors and Strong buyers and infrequent visitors—appeared.

Patel Monil, Patel Darshan, Rana Jacky, Chauhan Vimarsh, Prof. B. R. Bhatt [7] Customer Segmentation using Machine Learning. In order to segment the consumer and use various marketing methods appropriately, many clustering approaches have been discussed in this research. It has also been explored whether a hybrid clustering method could do better than a single model.

Şükrü Ozan [8] A Case Study on Customer Segmentation by using Machine Learning Methods. The study uses actual data on payments made by customers to address the company's data segmentation issue. Machine learning techniques are used to find solutions to data management issues since they are effective in doing so. In order to distinguish between premium and standard clients in a company's database, various classification techniques are evaluated. Customers' two-dimensional payment information is used as input data (features), and the separation performances of the various approaches are contrasted.

.

CHAPTER 2

2 Methodology

2.1 Machine Learning

Studying machine learning focuses on understanding and creating "learning" techniques, or techniques that use data to improve performance on a particular set of tasks. It's thought of as a part of artificial intelligence. Algorithms for machine learning build a model from sample data, also known as training data, to make predictions or judgments without being explicitly trained to do so. In many fields where it is difficult or impractical to develop conventional algorithms that can complete the necessary tasks, such as computer vision, speech recognition, email filtering, medicine, and agriculture, machine learning algorithms are used. A portion of machine learning and computational statistics, which focuses on using computers to make predictions, are closely connected. The field of machine learning benefits from the tools, theory, and application domains that come from the study of mathematical optimization. Data mining is a related area of study that focuses on unsupervised learning for exploratory data analysis. Some machine learning applications employ data and neural networks in a way that closely resembles how a biological brain functions. Machine learning is also known as predictive analytics when it comes to solving business problems.

2.2 Clustering

A machine learning approach called clustering or cluster analysis groups the unlabeled dataset. It is characterized as "a technique for arranging the data points into several clusters made up of related data points. The items with potential similarity continue to be in a group that shares little to no similarity with another group."

It accomplishes this by identifying comparable patterns in the unlabeled dataset, such as shape, size, color, behavior, etc., then classifying the data according to the presence or absence of these patterns.
It uses an unsupervised learning approach, which means the algorithm receives no supervision, and it works with an unlabeled dataset.

Each cluster or group is given a cluster ID after using this clustering technique, which ML systems can employ to streamline the processing of big and complicated datasets. The clustering method is frequently employed for analyzing statistical data.

2.2.1 K means clustering

The goal of k-means clustering, a vector quantization technique that originated in signal processing, is to divide n observations into k clusters, where each observation belongs to the cluster that has the closest mean (also known as the cluster centroid or cluster center), which serves as a prototype for the cluster. As a result, the data space is divided into Voronoi cells. The geometric median is the only one that minimizes Euclidean distances; k-means clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more challenging Weber problem. For instance, k-medians and k-medoids can be used to find better Euclidean solutions. Although the problem is computationally challenging (NP-hard), effective heuristic algorithms quickly reach a local optimum. These often follow an iterative refining strategy used by both k-means and Gaussian mixture modeling, which is similar to the expectation-maximization procedure for mixtures of Gaussian distributions. They both employ cluster centers to represent the data, but the Gaussian mixture model allows for different-shaped clusters whereas k-means clustering tries to discover clusters of equivalent spatial dimensions.

The stages listed below can help us understand how the K-Means clustering technique operates:

- Step 1: We must first specify how many clusters, K, must be produced by this process.

- Step 2: Next, choose K data points at random and group each one into a cluster. Simply said, categorize the data according to the number of data points.
- Step 3: At this point, the cluster centroids will be computed.
- Step 4 After that, repeat the following iteratively until we find the optimal centroid, which is the assignment of data points to clusters that do not change anymore.

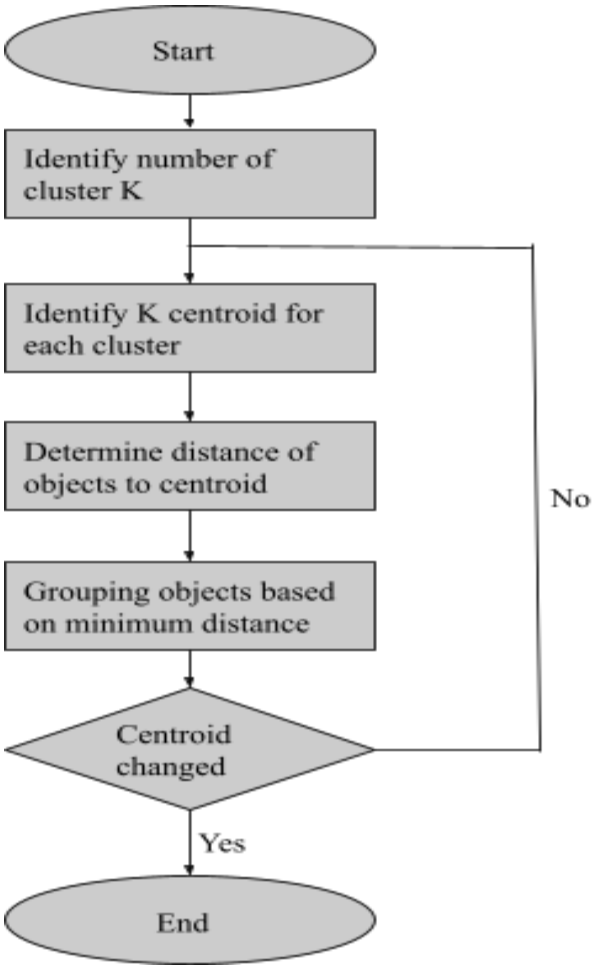

Fig. 2.2.1 K-means clustering steps

2.2.2 Elbow Method

Finding the ideal number of clusters to divide the data into is an important point in any unsupervised technique. One of the most prominent techniques for figuring out this ideal value of k is the elbow method. By using various choices of k, the elbow approach plots the value of the cost function. When you are aware, as k rises, average distortion decreases, the number of constituent examples in each cluster decreases, and the instances move closer to their respective centroids. However, as k rises, the gains in average distortion will diminish. We should stop further clustering the data at the elbow, or the value of k at which improvement in distortion drops the most.

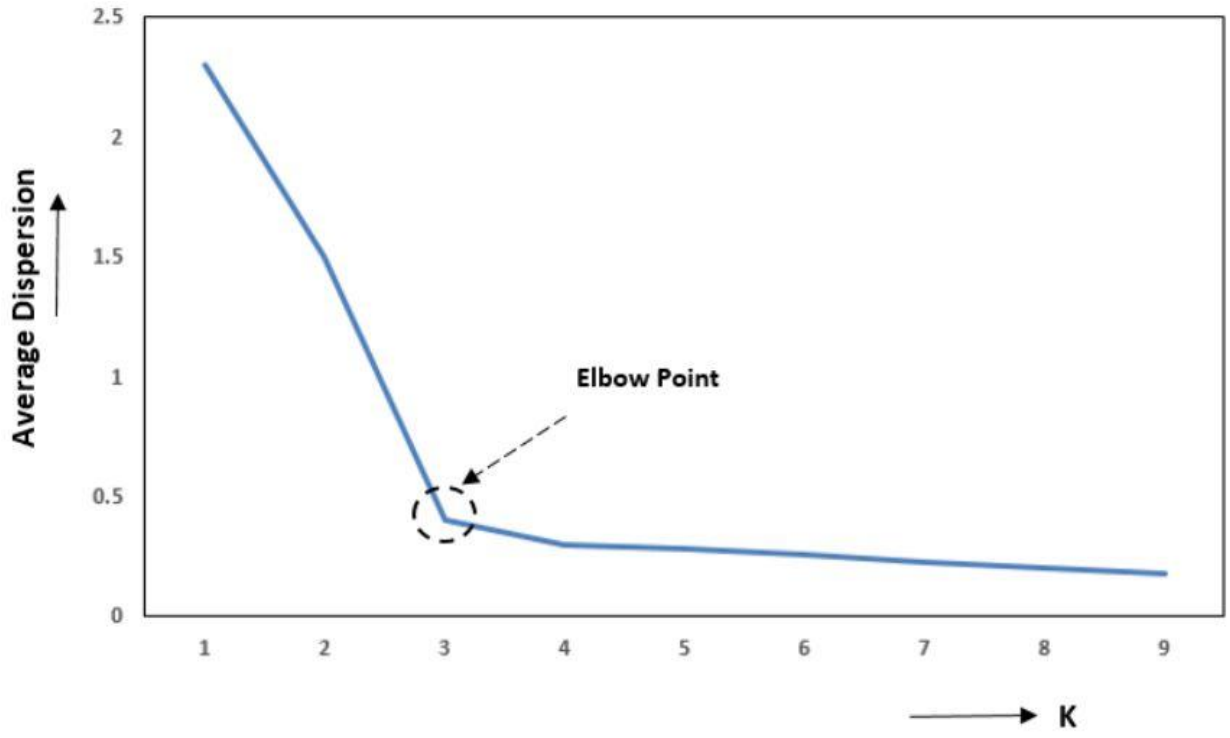

Fig 2.2.2 Elbow Method for selection of optimal K clusters

2.3 Methodology

We used the K-means clustering algorithm to divide a customer dataset into clusters in order to discover potential customers. We split our dataset into train and test sets and built two models from each. By measuring the distance between a new point and a cluster point, we chose the special clusters for the test set and showed them in a separate model to assess whether a new client belongs to a promising cluster or not.

2.3.1 Steps

Using the libraries, customer dataset, and k-means clustering algorithm we created two models for a train set and test set. From the train set model, we get the clusters and from the testing set model, we can get the upcoming customer's cluster finally we built a model for potential customers to invest in. The steps of the project are given below:

- ☐ **Import Libraries And Dataset:** The Google Colab platform is used to develop the project. And to create the model, we'll use Python programming language. First, we must import the necessary libraries, such as Pandas, Sklearn, Numphy, and others. Following the import of To import the dataset from Google Drive, we need certain libraries. The data must then be preprocessed.

- ☐ **Exploring Dataset:** It is crucial to the process of constructing models. To prepare the data for building a flawless model, we must clean the data, transform the data, reduce the data, and do a number of other operations.

- **Data Visualization:** Here dataset was plotted as histograms and scatterplots by their attributes like age, gender, and customer ratings. Annual Income Vs Customer Rating was also plotted using Scatterplot and Jointplot.

- **Data Splitting and Preprocessing:** Here dataset was split into train sets and test sets in a 5:1 ratio.

- **Finding the optimal number of clusters:** Using the Elbow Method we found 5 clusters are optimal for the dataset.

- **Feature Scaling:** Since KMeans uses distance (Euclidean, Manhattan, etc.), feature scaling is done since scaling the values speeds up the model's performance.

- **Model Building For Train Set:** Using K-means clustering we clustered the train data sets here and built a model for the clusters.

- **Predicting New Customer's Cluster:** In this segment, We calculated the distance between a new customer point and the closest cluster point and thus could predict the new customer's cluster.

- **Model Building For Incoming Customers:** We developed a model for this sector to plot the new consumers to clusters to which they belong based on the measured distance.

- **Model For Suggested Customer:** We created a model for this section that only displays clusters with potential clients.

2.3.2 Flowcharts

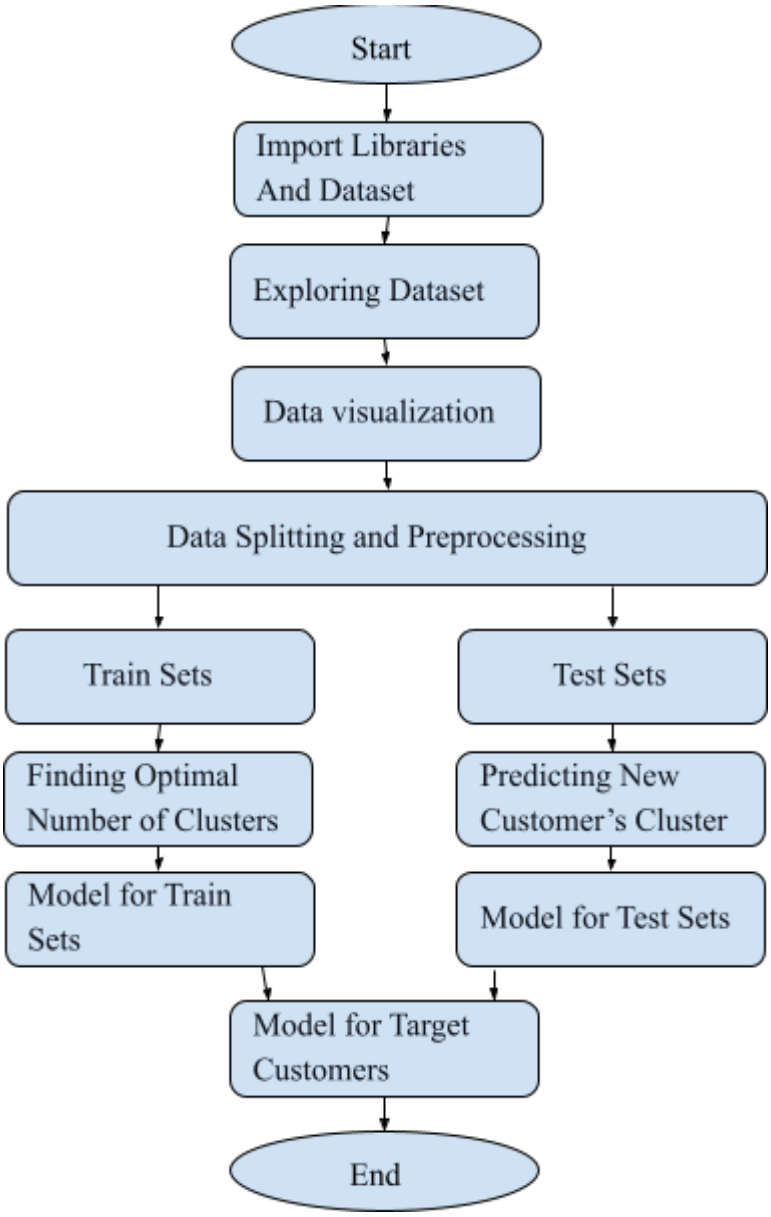

Fig 2.3.2 Methodology Flowchart

2.3.3 System Description

Hardware Description:

- ➢ Asus Vivobook Pro
- ➢ 8GB System Memory
- ➢ 128KB L1 cache memory and 512KiB L2 cache
- ➢ Intel Processor Core i5-8G
- ➢ 256KB BIOS

Software Synopsis:

- ➢ NumPy
- ➢ Python
- ➢ Google Colab
- ➢ Matplotlib

We set up a framework utilizing Google Colab to import functions for the implementation. The following are a few functions:

Pandas: Read data from a CSV file using Pandas.
Numpy: The data will be converted using Numpy into a format that is suited for our classification model and is simple to use.
Sklearn: Import the logistic regression into our model using Sklearn.
Matplotlib: As an alternative to Matlab graph plotting, Matplotlib provides an open-source framework for viewing and plotting graphs in Python.
Seaborn: Similar to Matplotlib, Seaborn can be used to plot and view graphs.

CHAPTER 3

3 System Model

3.1 Data Description

For our dataset, we have an Excel data sheet. We developed a new dataset for my project by compiling data from numerous sources, including Google, GitHub, Kaggle, and many others. And a CSV file with our dataset is saved. The document is called Book1.csv. We employ pandas to read our dataset. The data frame is known as df.

3.1.1 Exploring Dataset

Information on 270 customers is included in our dataset. Every person has a unique set of qualities or characteristics to take into account. The characteristics are listed below:

- CustomerID
- Gender
- Age
- Annual Income in K BDT
- CustomerRating (Out of 100)

Below is a sample screenshot of the dataset table. (The table includes the first five sample's data together with their fractures.)

	CustomerID	Gender	Age	Annual Income in K BDT	CustomerRating (Out of 100)
0	1	Female	50	67	57
1	2	Male	27	67	56
2	3	Female	38	67	40
3	4	Female	40	69	58
4	5	Male	39	69	91

Table 3.1.1.1: First Five Sample Data

An example screenshot of the dataset table is provided below. (The data and fractures from the last five samples are included in the table.)

	CustomerID	Gender	Age	Annual Income in K BDT	CustomerRating (Out of 100)
265	266	Female	24	20	77
266	267	Male	37	20	13
267	268	Male	22	20	79
268	269	Female	35	21	35
269	270	Male	20	21	66

Table 3.1.1.2: Last Five Sample Data

3.1.2 Data Visualization

All the information columns(Age, Gender, Annual Income, Customer Rating) are visualized with a histogram, scatterplot, and joint plot.

Gender Count: Using Countplot we display the column labeled "Gender"

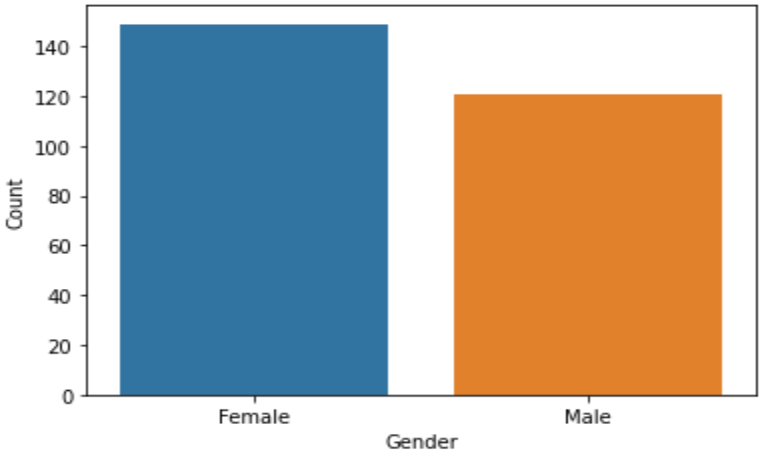

Fig 3.1.2.1: Gender Count

Age Histogram: Using a histogram we display the columns of "Age"

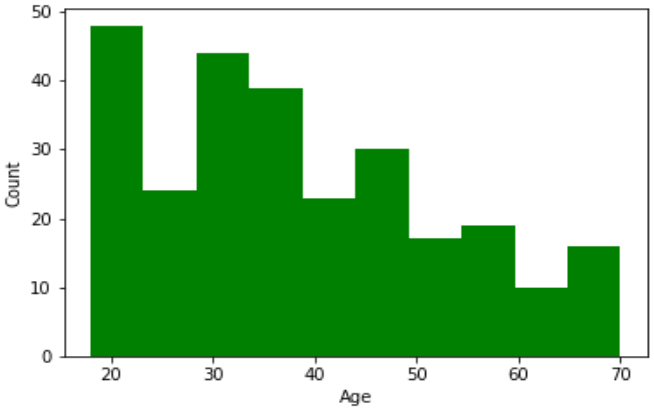

Fig 3.1.2.2 Age Histogram

Age Vs Gender Vs Customer Rating: Age and Customer Rating (Out of 100) visualizations using Scatterplot and Jointplot.

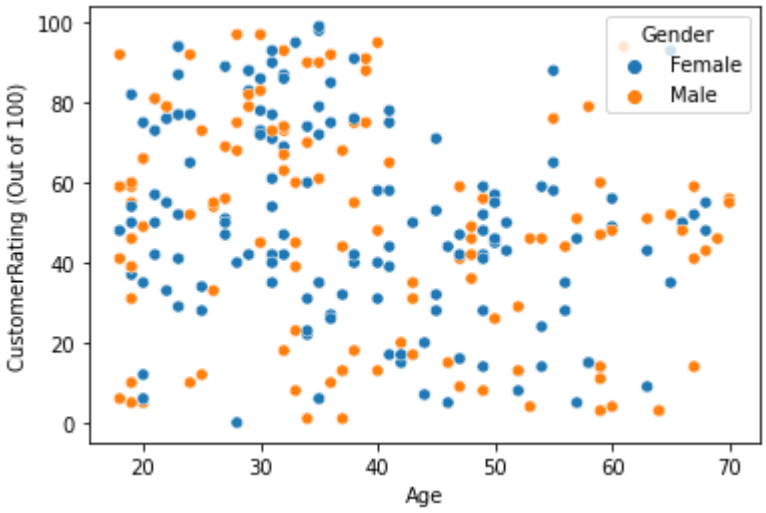

Fig 3.1.2.3 Scatterplot of Age Vs Gender Vs Customer Rating

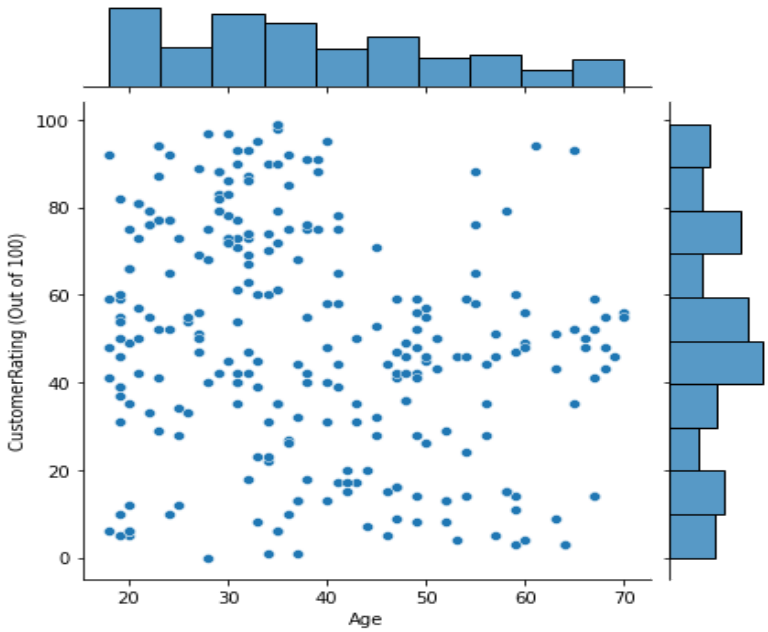

Fig 3.1.2.4 Jointplot of Age Vs Customer Rating

Annual Income Vs Rating: Utilizing a scatterplot and joint plot to visualize the columns "Annual Income in K BDT" and "CustomerRating (Out of 100)".

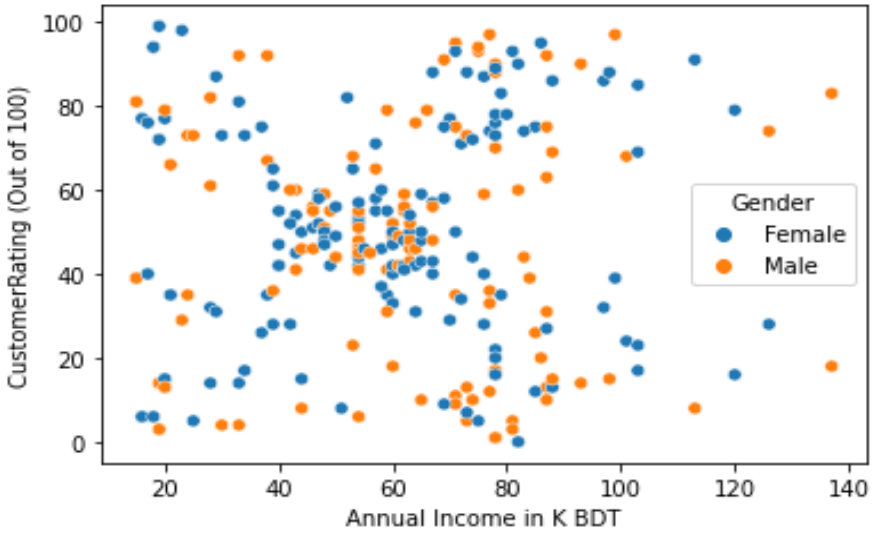

Fig 3.1.2.5 Scatterplot of Income Vs Customer Rating

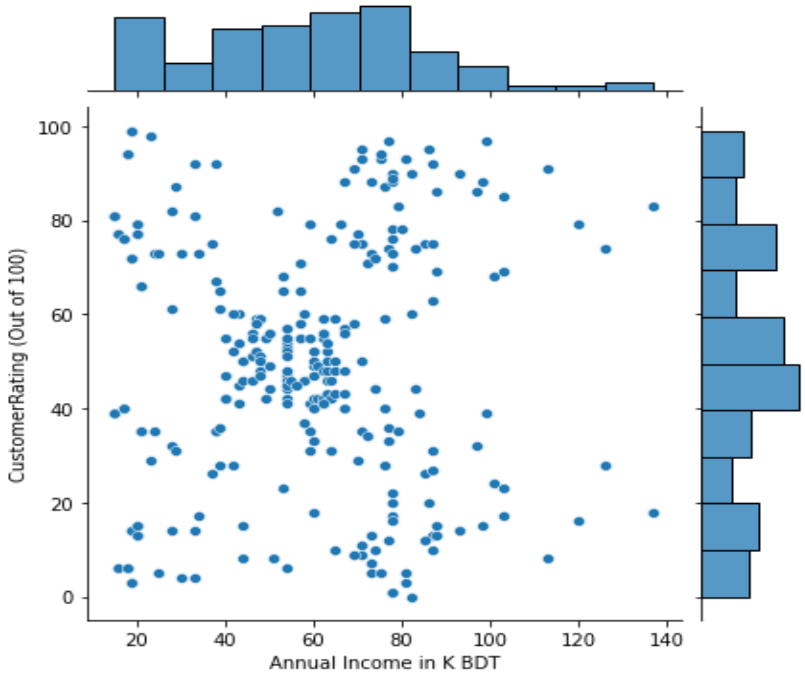

Fig 3.1.2.6 Joint plot of Income Vs Customer Rating

3.1.3 Data Splitting and Preprocessing

It is the most important element. Data in the physical universe could be inconsistent, loud, or lacking. The data needs to be processed earlier, thus. then we'll begin the modeling process. Finding a better or more accurate result would be challenging if the data quality was poor. Also here we split the dataset into Train and Testing Datasets in a 5:1 ratio.

3.2 Model Building

After Processing the data We need to find an optimal number of clusters and build models accordingly.

3.2.1 Finding an Optimal number of clusters

Using the Elbow Method thus Calculating WCSS values for 1 to 10 clusters We found that the optimal number of clusters is 5. By plotting the WCSS values we got:

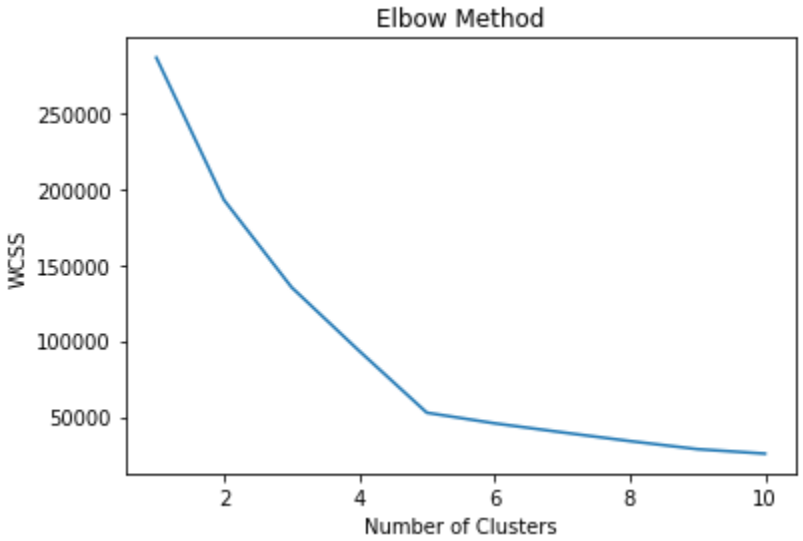

Fig 3.2.1 Elbow Method for Optimal 5 Clusters

3.2.2 Model Building For Train Set

In this section, we used the k-means technique to create a model using the data from our train set. The model is given below:

Fig 3.2.2 Model for Train Set Data

3.2.3 Predicting New Customer's Cluster

We were able to anticipate the new customer's cluster in this segment by calculating the distance between a new customer point and the closest cluster point. We used our test dataset for new customer points. The results are given below:

```
New Customer  1  belongs to cluster  3  with distance of  1.171600175179626
New Customer  2  belongs to cluster  3  with distance of  1.1980884894539727
New Customer  3  belongs to cluster  5  with distance of  1.154688314239751
New Customer  4  belongs to cluster  3  with distance of  1.0051444317443785
New Customer  5  belongs to cluster  3  with distance of  0.25031739424249855
New Customer  6  belongs to cluster  5  with distance of  0.5807824392253169
New Customer  7  belongs to cluster  3  with distance of  0.39611134239219
New Customer  8  belongs to cluster  5  with distance of  0.7233290931049225
New Customer  9  belongs to cluster  3  with distance of  0.19608058581060442
New Customer  10  belongs to cluster  5  with distance of  0.04527939760378433
New Customer  11  belongs to cluster  3  with distance of  0.4642297885365338
New Customer  12  belongs to cluster  5  with distance of  0.08835815089072346
New Customer  13  belongs to cluster  3  with distance of  0.4642297885365338
New Customer  14  belongs to cluster  5  with distance of  0.7044257684943966
New Customer  15  belongs to cluster  3  with distance of  0.6415635532670644
New Customer  16  belongs to cluster  5  with distance of  0.3715025494454491
New Customer  17  belongs to cluster  3  with distance of  0.46596673314792264
New Customer  18  belongs to cluster  5  with distance of  0.3422322267638826
New Customer  19  belongs to cluster  3  with distance of  0.6910983992113491
New Customer  20  belongs to cluster  5  with distance of  0.49245772903176294
New Customer  21  belongs to cluster  2  with distance of  0.8286756589616445
New Customer  22  belongs to cluster  5  with distance of  0.6972259770453567
New Customer  23  belongs to cluster  3  with distance of  0.8244902768945316
New Customer  24  belongs to cluster  2  with distance of  0.35412926937288536
New Customer  25  belongs to cluster  3  with distance of  0.9990713966218168
New Customer  26  belongs to cluster  5  with distance of  1.0229072262998182
New Customer  27  belongs to cluster  3  with distance of  1.194597068606786
New Customer  28  belongs to cluster  2  with distance of  0.44741401956576177
New Customer  29  belongs to cluster  2  with distance of  0.7232349010520527
New Customer  30  belongs to cluster  2  with distance of  0.901653561493962
```

Fig 3.2.3 Predicting New Customer's Cluster

CHAPTER 4

4 Result Analysis

4.1 Model For Incoming Customers

For this area, we created a model to plot new customers to clusters to which they belong based on the distances measured. Following is the model:

Fig 4.1 Clusters of incoming customers

4.2 Model For Suggested Customers

For this area, we developed a model that only shows clusters with potential customers. Listed below is the model:

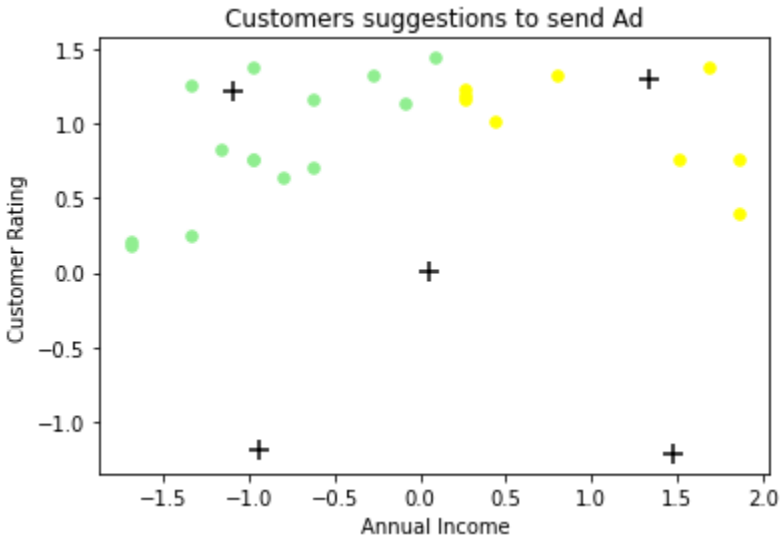

Fig 4.2 Customers to Send Ad

CHAPTER 5

5 Conclusion

5.1 Summary

This dataset's properties, such as age, gender, and customer reviews, were shown as histograms and scatterplots. A scatterplot and joint plot were also used to plot annual income vs customer rating. In a 5:1 ratio, the dataset was divided into train sets and test sets. We discovered that 5 clusters are ideal for the dataset using the Elbow Method. We clustered the train data sets using K-means clustering and created a model for the clusters. By calculating the range between a new customer point and the closest cluster point, we were able to anticipate the cluster for the new client. Based on the measured distance, we created a model to plot the new customers to the clusters to which they belong. And finally, we developed a model that only shows clusters with potential customers.

5.2 Limitations & Future Work

We encountered some unwelcome issues during the project, like null values, missing values, run time limits, etc. In this project, a structured dataset has been chosen. The following are some restrictions:

- ➢ Lack of a larger dataset, more data are required in the dataset.
- ➢ Some data contain Null values, which negatively affect the prediction.
- ➢ The project has chosen the Structured Dataset.

The project can yet be enhanced. We'll include a few extra features. Trying to add more data for training and testing. We will take into account unstructured data to get around the restrictions.

Additionally, We wish to develop a web-based or mobile application for Promotional Suggestion For Target Customers using these categorization and prediction approaches.

REFERENCES

[1] Kumar, A. (2022) "Customer segmentation of shopping mall users using K-means clustering," *Advancing SMEs Toward E-Commerce Policies for Sustainability*, pp. 248–270. Available at: https://doi.org/10.4018/978-1-6684-5727-6.ch013.

[2] Pradana, M. (2021) "Maximizing Strategy Improvement in mall customer segmentation using K-means clustering," *Journal of Applied Data Sciences*, 2(1). Available at: https://doi.org/10.47738/jads.v2i1.18.

[3] Shen, B. (2021) "E-commerce customer segmentation via unsupervised machine learning," *The 2nd International Conference on Computing and Data Science* [Preprint]. Available at: https://doi.org/10.1145/3448734.3450775.

[4] Tabianan, K., Velu, S. and Ravi, V. (2022) "K-means clustering approach for intelligent customer segmentation using Customer Purchase Behavior Data," *Sustainability*, 14(12), p. 7243. Available at: https://doi.org/10.3390/su14127243.

[5] Narayana, V.L. *et al.* (2022) "Mall Customer Segmentation using machine learning," *2022 International Conference on Electronics and Renewable Systems (ICEARS)* [Preprint]. Available at: https://doi.org/10.1109/icears53579.2022.9752447.

[6] Monil, P. (2020) "Customer segmentation using machine learnin," *International Journal for Research in Applied Science and Engineering Technology*, 8(6), pp. 2104–2108. Available at: https://doi.org/10.22214/ijraset.2020.6344.

[7] Kansal, T. *et al.* (2018) "Customer segmentation using K-means clustering," *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* [Preprint]. Available at: https://doi.org/10.1109/ctems.2018.8769171.

[8] KIM, K. and AHN, H. (2008) "A recommender system using ga k-means clustering in an online shopping market," *Expert Systems with Applications*, 34(2), pp. 1200–1209. Available at: https://doi.org/10.1016/j.eswa.2006.12.025.

[9] Ozan, S. (2018) "A case study on customer segmentation by using Machine Learning Methods," *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* [Preprint]. Available at: https://doi.org/10.1109/idap.2018.8620892.