Marts And Mall Customer Based Segmentation Using Regression and Clustering Techniques.

Vamshi S[1, a)],Priyadarshini[1, b)] Vudagandla VinayKumar[1, c)], Thavva Mohith Ruthviz[1, d)], Vivek Vardhan Reddy[1, e)]
[1]*Department of Computer Science and Engineering, REVA University ,Bangalore – 560 064, Karnataka, India*

***Abstract*— In this analysis, we'll explore a data set on mall visitors to notice similarities or patterns. This analysis is useful in understanding the demographic and psychographic sub-populations in the business field or market field. This helps in the methodology of user persons: This is beneficial in collecting data and gather more data to build a data set that has more features. Because of the more transaction rate, maintaining a balance between the customer demand and inventory is the main part for the business .Therefore , making the right prediction and analysis of the given data in-order to increase the profits of the business. Once this is better understood, you could understand what factors will increase spending score, thus leading to greater profits. Many of the existing analysis, predictions of the business only depends upon extrapolating of the statistical trends .This project helps the marketing team and marts and malls to increase their profits in various strategies .As of more columns or more data there is better understanding of the predicting score**

Keywords—target customers, k-means clustering, supervise and unsupervised learning ,clusters, market analysis

.

I. INTRODUCTION

As of now, Every marts and malls are aware of how to handle data analysis techniques in business strategies. Where they don't know the required knowledge and requirements to analysis. In this paper, we present the use of data analysis techniques in sales-consumers business intelligence for marts and malls business .This prediction serves mainly to assist marketers to know more about their customers so that they can build their marketing strategies more successfully. Based on the machine learning regression models and statistical models, consumers of the marts and malls are divided into different meaningful clusters, where these clusters are divided using the k-means clustering algorithm. The important characteristics of the customers in every cluster it is clearly identified. A set of guides are given to the technician in sales-consumers marketing. Management and maintenance of customer relationships have always played a vital role in providing Organizations to create, manage, and develop valuable long-term customer relationships. The importance of handling customers as an organization's purpose is increasing in value in the day to . Organizations have an interest in investing in the development of customer acquisition, maintenance, and development strategies. Business intelligence has a Technical knowledge can be used to improve customer service. knowledge and programmes for outreach. Customers can benefit from clustering techniques such as k-means. Those with same means are grouped together. This assist the business strategy team to analyzing various customer segments. Purchasing strategies helps in finding out the customers who differ in In terms of preferences, expectations, desires .The primary goal of this Customer segmentation is to group people who have similar interests so that the marketing team can be combined into an effective marketing plan.

II. KEY FEATURES /OBJECTIVES

1. The objectives of this work is two fold, i.e.,

A. The main goal of this project is to create an Marts and mall customer based Segmentation analysis using open-source technologies.

B. Making predictions and analysis on the customer data, where marketing teams makes use of it to raise the profits and branding name of the marts and malls and also increase their customers. Which helps to increase their revenue by making use of the predictions and analysis.

C. From analysis and predictions, not only marts and malls can benefit but also indirectly consumers or customers can also be benefited.

III. Literature Review

Organizations have evolved over time as a result of intense competition in the economic world. They want to increase their earnings and business by meeting client requests and attracting new ones. Consumers based on their requirements Customers must be identified and their needs must be met. Each consumer is a challenging task. This is due to the fact that clients can differ based on their preferences. Their needs, desires, and preferences are not the same as their needs, wants, and preferences. Customers are segmented into groups based on common characteristics. Customer segmentation, according to Wikipedia, is a strategy for splitting a market into distinct groups. All of them are the same. The data is used in an analysis called customer segmentation. [1]segment customers using a variety of variables. The easiest and most common demographic factors include age, gender, family, education level, and income. For segmentation variables in socio-cultural, geographic, psychographic, and behavioural contexts are some of the other major variables employed in segmentation. [2],given a in-depth of the various clustering algorithms grouped under clustering ,density, grid-based, and algorithms , considering the characteristics of Big Data such as size, noise, dimensionality .[3] looked into the importance of client segmentation utilising clustering techniques as a major CRM feature. The pros and disadvantages of the commonly used Hierarchical Clustering algorithms and K-Means clustering were examined. Finally, the thought of developing a hybrid strategy is represented by combining the two strategies above, which has the potential to out perform the different approaches.

[4], Decision makers use many variables to segment customers. Demographic variables such as age, gender, family, education level and income are the easiest and common variables for segmentation. Socio-cultural, geographic, psychographic and behavioural variables
are the other major variables that are used for segmentation.

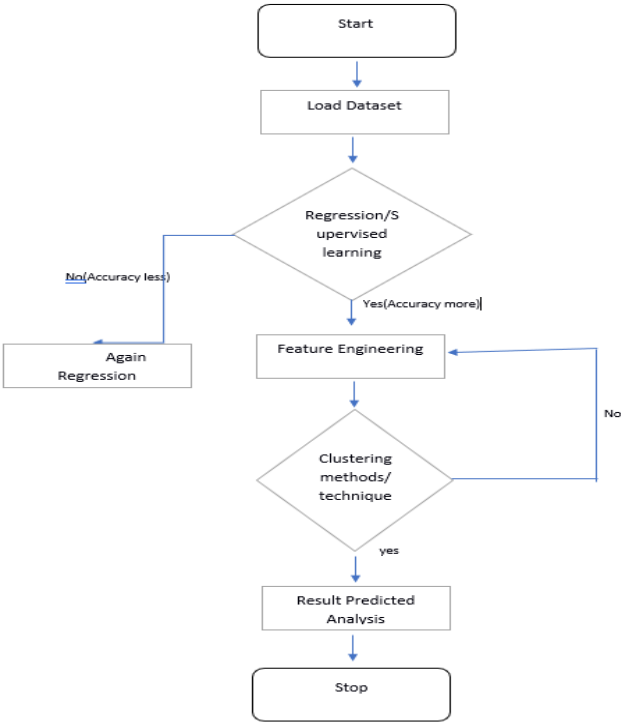
[5], presented various clustering algorithms taking into account the characteristics of Big Data such as size, noise, dimensionality, algorithm calculations, cluster shape and presented a brief overview of the various clustering algorithms grouped under partitioning, hierarchical, density,
grid-based and model-based algorithms.

[6] explored the necessity of segmentation of the customers using clustering algorithms as the core functionality of CRM. The mostly used K-Means and Hierarchical Clustering were studied and the advantages and disadvantages of these techniques were highlighted. At last, the idea of creating

a hybrid approach is addressed by integrating the above two strategies with the potential to surpass the individual designs.

[7], Merged clustering of fuzzy c-means and genetic algorithms to cluster, steel industry customers, by using the LRFM variables (length, recency, frequency, monetary value) system, customers were divided into two clusters

IV. .BLOCK DIAGRAM



A. Data Set

The data set which is used for supervised and Unsupervised learning is in the form of word document and it will be converted into excel sheet and we imported to interpreter using some libraries.

B. Regression / Supervised learning

Using the data which is imported using pandas is used for regressions. In data we use different columns for regression. If the accuracy didn't match, then we will take another columns. We repeat this process until we get best accuracy.

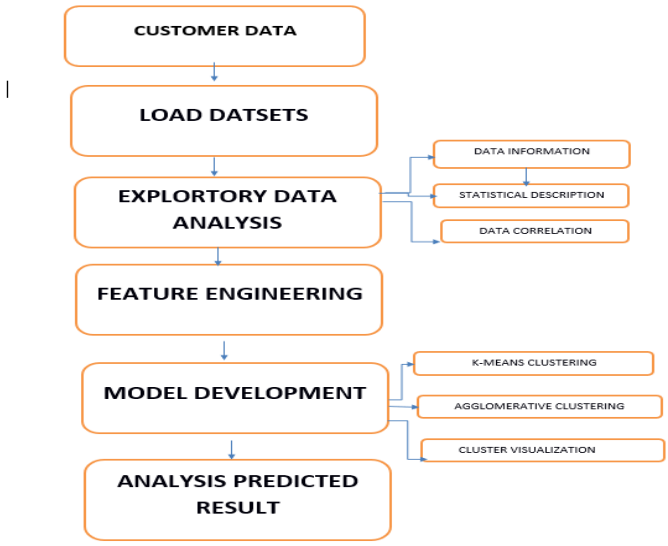
C. Feature Engineering

In this step we remove some unwanted empty columns and empty rows and also we try to insert some predicted answers empty answers. We predict some answers and filled it in the empty answers which requires to clustering techniques.

D. Clustering / Unsupervised Learning.

In clustering, we cluster the data into some different groups based on the similarity of the customers, If clustering is not done in proper way then we wont get our analysis in right way. So ,we repeat this process until the clustering get done in right way.

V. METHODOLOGY AND ALGORITHM



A. System Architecture

The goal of this project is to increase the profits and branding name of the marts and malls, correctly estimate client segmentation. Machine learning models are used in conjunction with conventional methods to improve the accuracy of client segmentation estimates in malls. Various machine learning models are employed**.**

B. Workflow of the system

Step1: Customer data

First step is to collect the customers data through different sources.

Step2: Load Datasets

The data which is collected through different sources are added in excel file and we import to the complier using pandas.

Step3: Exploratory Data Analysis
In this step we use some supervised algorithm to check the accuracy and data corelation. we will do data information and statistical description, then data corelation. If the accuracy is not matched or less, then we repeat the process for different columns.

Step4: Feature Engineering
In this step , we do the process of choosing, changing and transforming the raw data into columns that can be used for supervised learning and unsupervised learning.

Step5: Model Development
In this step , we undergo some clustering techniques like K-means clustering , agglomerative clustering and cluster visualization where we cluster the customer into different groups based on some similarities.

Step6: Analysis Predicted Result
In this step we get to know the prediction analysis of the all steps where it gives the result of the clustering of the clustering of the customers which helps for the marketing team

Clustering:
Clustering is the technique of dividing customers into different groups based on some similarity of the customers data.

C.Algorithms and Models:-
Linear Regression and Logistic regression:-
If the dataset is small or minimal then we can use the linear regression and it is used to find the best fit line that accurately predict the output.
If the dataset is larger then we can use logistic regression for calculating the accuracy. These are the two supervised algorithms are used in this project.
Coming to clustering techniques we used **K-means clustering and agglomerative clustering.**
K-means clustering algorithm determines the best value for k center points and also assigns each data point to its nearest k-center.
Agglomerative clustering algorithm makes most similar clusters together and merge them.

VI. Experimental Results

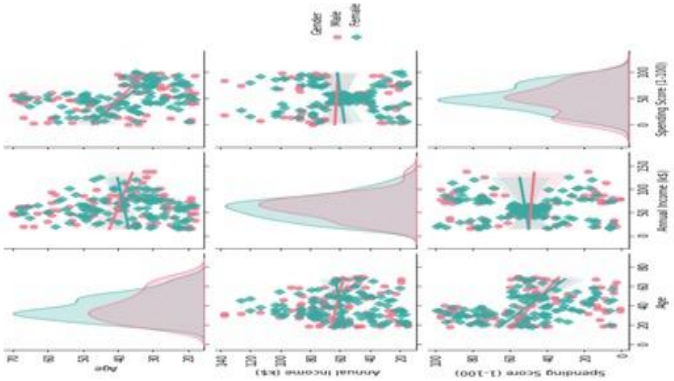
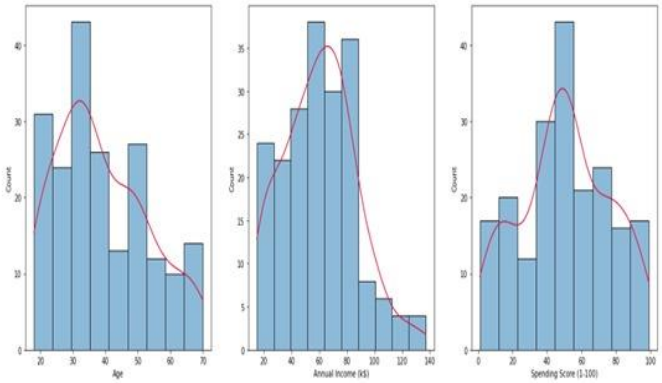


Fig1:- pair plot

From the above pairplot we can see that green colour has highest ratio than pink colour as the there are



more female customers than male .

Fig2:-

From above figure we can see that the distribution of numerical features, discrete features , continuous features and categorical features .

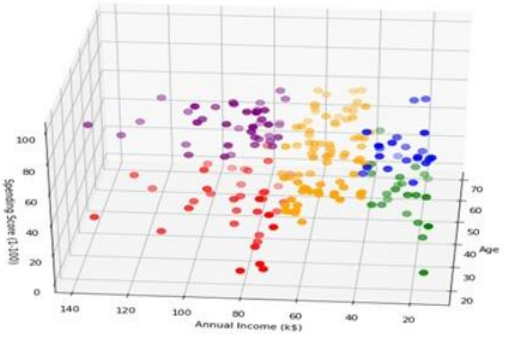


Fig3:-

From the above figure we can see that the customers are divided into five different groups:-

Cluster with red colour indicates that consumers or customers are of low annual income and low spending scores. In this case the marketing team is of low interested in this kind of people.

Cluster with purple colour indicates that customers or consumers are of low annual income and high spending scores . In this case the marketing team are not concentrating more but still they wont lose them.

Cluster with pink colour indicates that customers or consumers have average annual income and spend average spending scores. In this case the marketing team uses some techniques to increase their spending scores.

Cluster with blue colour indicates that customers or consumers having high income and high spending scores. The marketing team uses some techniques to increases the spending scores.

Cluster with green colour indicates that customers with high income but low spending scores. The marketing team try to add new facilities so they can attract these kind of people and meets their needs.

VII. Compliance with the society

Our project is indeed useful to society (Business, People) as our analysis predict the result which can help both the customers as well as business. This analysis can help the market to increase their sales and gain profits. People also get benefit from our analysis in which most buying customers can get product recommends, rewards and so on.

VIII. Conclusion

Merchandisers have constant obstacles in the ever-changing market around. Retailers must recognise the compelling value proposition of a "buyers' market." Consumers have a extended range of purchasing experiences to choose from, but no one can possibly capture them all. As a result, it is up to supervision to define their market prediction and pointing their efforts on solving the problems of that market. Technology, demographics, customer position, and the emergence of a global country are all conspiring to change the game. How careful merchants are about the new values, expectations, and demands of the shopper will determine how successful they are in the twenty-first century. To boost the mall's productivity, this analysis will generate a prediction based on customer segmentation.

IX. FEATURE ENHANCEMENT

This project can be further can be make use for product recommender system which can benefit the customer and also the products by recommending the product to the consumers. By this the revenue of the marts or malls can be increase and also customers to that malls or marts are increased in an continuous way . Customers are benifited with many facilities like getting more discounts, offers…etc.

X. References

[1]. T. Sajana, C. M. Sheela Rani and K. V. Narayana "A Survey on Clustering Techniques for Big Data Mining", Indian Journal of Science and Technology, Volume 9, Issue 3, Jan 2016.

[2]. Azarnoush Ansari, Arashss Riasi, "Customer Clustering Using a Combination of Fuzzy C-Means and Genetic Algorithms", International Journal of Business and Management, Volume 11, Issue 7, 2016.

[3]. Pedro Quelhas Brito, Carlos Soares, Sérgio Almeida , Ana Monte,Michel Byvoet, "Customer segmentation in a large database of an online customized fashion business", Robotics and Computer Integrated Manufacturing ,2015.

[4] I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.

[5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 881-892, 2002.

[6] MacKay and David, "An Example Inference Task: Clustering," Information Theory, Inference and Learning Algorithms, Cambridge University Press, pp. 284-292, 2003.

[7] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining Concepts and Techniques", Third Edition.

[8] S. Dasgupta and Y. Freund, "Random Trees for Vector Quantization," IEEE Trans. on Information Theory, vol. 55, pp. 3229-3242, 2009.

[9]Puwanenthiren Premkanth, ―Market Segmentation and Its Impact on Customer Satisfaction with Especial Reference to Commercial Bank of Ceylon PLC.‖ Global Journal of Management and Business Research Publisher: Global Journals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1