Group 5: Customer Segmentation Using Machine Learning

Thirumalareddy Abhishek Reddy
Aditya Loth
Muzafar Ahmad Wani
Nishant Kiran Valvi
21111062, 21111004, 20111271, 21111407
{tareddy21, adityal21, muzafarwan20, nkiranv21}@iitk.ac.in
Indian Institute of Technology Kanpur (IIT Kanpur)

Abstract

A customer is a critical aspect in the success of any business.Building a better relationship with customer help the organizations in increasing profit and customers satisfaction.The potential value of a customer to a company is a core ingredient in decision-making about marketing strategies. Customer Segmentation is one such strategy , which helps in identifying groups of similar customers based on their interactions with the products and then effectively implementing various marketing strategies for the suitable customers. In this work we implemented five clustering algorithms for the dataset to compare the performance. The algorithms implemented are K-Means, K-Means++, MeanShift, Agglomerative,Spectral. For our case Agglomerative showed the best performance measured by Silhouette scores. We implemented six classification algorithms namely Logistic Regression, SVC, Decision Trees, Random Forest, KNN and Ridge Classifier and three use cases.

1 Introduction

In the big data era, an important trend in managing customer relationship is the process of personalization and customization of sales and services.With the advancement of the big data era another important development is the customer transaction data being saved for analysis. Every shopping mall, e-commerce portal now collects data using IOT devices like POS machines, scanning,RFID tags, bagging,checkout systems etc. All these data collecting devices collect data with different set of attributes and then these attributes are merged to form a dataset for analysis. Customers are provided with appropriate products, convenient service and better customer care. However, big data also brings numerous novel challenges due to the volume, velocity and variety of huge data. To be more precise, volume refers to the amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing. Since, for any specific activity like a marketing campaign, targeting the whole customer base all at once is almost always impossible, communities with similar buying behaviors have to be identified effectively, so that the business activity can concentrate on a few clusters at a time. For companies, customer segmentation used to be done by business experts based on certain rules drawn from past experiences. Recently data analytics have been brought into business processes, in particular, into customer analytics with customer segmentation being an important problem. The main advantage of the data analytics approach is the ability to adapt better to the fast-changing business environment. In today's world, certain business rules that were effective just last year might become obsolete in the coming year.

Customer segmentation serves a very important role to unravel hidden patterns stored in the company's database. Customer segmentation helps to segment the customers with similar patterns into similar clusters, making easier for the business to handle the large customer base.Customer Segmentation means grouping the customers based on marketing groups which shares the similarity among customers.So as to provide different types of customers with distinctive marketing methods and improve their satisfaction as well as profits.It allows companies to visualise what actually the customers are buying which will prompt the companies to better serve their customers resulting in customer satisfaction. A customer doesn't buy products of every company out there so then why should companies should target every customer. Segmentation can influence the marketing strategy by identifying the products which are good for each segment and identify the highest selling product for each segment. This information will influence the decision making. Clustering comes under unsupervised learning, having ability to find clusters over unlabelled dataset.The objects with the possible similarities remain in a group that has less or no similarities with another group. There are a number of clustering algorithm over which like k-means, hierarchical clustering, DBSCAN clustering etc. In this project we implemented K-means clustering to segment the customer dataset into different clusters.

1.1 Goals

The purpose of this project is to explore the use of different clustering algorithm for identifying and understanding user behaviour on mall dataset to increase customer satisfaction and also increase profits for the business owner . The aim is to develop an application that can group and classify users based on their behavioural segmentation factors where in the segmentation is based on customer's behaviour pattern with a particular business or a website. E.g Spending habits, purchasing habits, browsing habits,interaction with the brand,loyalty to a brand which then could be used to improve the mall's marketing strategies.

1.2 Disposition

The rest of the report is organized as

1. Section 2: This section has the literature survey we did for the project.
2. Section 3: Presents the proposed idea.
3. Section 4: Presents the methodology followed in implementing the project.
4. Section 5: Presents the results from our implementation of the project.
5. Section 6: Presents analysis of the results you have obtained and directions on future work.
6. Section 7: Presents the conclusions of our project.
8. Section 8: Presents the contributions of each member of the group in the project.

2 Related Work

For centuries customer relation management has relied on certain characteristics of customers such as length(how long a customer has been with the business), recency, frequency, monetary value and profit to analyze customer behaviors and perform customer segmentation. With the emergence of online shopping and E-commerce,only these prove inadequate for market segmentation. The last two decades have seen a lot of new techniques been used to improve the increasingly more complex customer base. First data mining was used to find the hidden patterns in data. Then Machine Learning algorithms were used to cluster the data into different clusters based on certain features. Kansal et al [1] implemented three clustering algorithms K-Means, Agglomerative, and Meanshift on one dataset to segment the customers and compared the results of clusters obtained from the algorithms.They used two features, one is mean of amount spent by a customers and second one is average of the customer's visit into the shop annually The results showed that there is not much significant difference in K-means and Agglomerative clustering performance as these two algorithms were able to cluster the data well than Mean shift algorithm. The features selected in this case were the reason for not getting good performance.

Abidar et al. [2] proposed a new model based on RFM model Recency, Frequency, and Monetary and K-mean algorithm. RFM is a method used to give every customer a significant value and then applying k-mean clustering for every parameter and sorting them allow us to calculate an overall score,a formula is used to calculate it based on clustering results.Based on these scores customers were segmented to three segment (High, Mid and Low segment). The results showed better performance as the features selected RFM give us a better picture of the customer data. Hung et al [3] presented the implementation of agglomerative clustering algorithm in the R programming language to perform customer segmentation on credit card data sets to determine the appropriate. Using the results, the analysts did promote appropriate marketing strategies that are more profitable. The drawback of this method is quite slow and hardware dependent.

Ngyuen [4] utilized a highly successful deep neural network called autoencoder to extract information from customer database and construct input features for the segmentation process. The segmentation is then accomplished by a unsupervised probabilistic clustering algorithm that iterates (until a stopping criterion is met) to create meaningful customer clusters. No previous work in customer segmentation has ever brought together these techniques. The results of the implementation showed very good results for the dataset used. This model has great potentials in extending to more complex business data. Its mechanism accommodates an incorporation of new data sources such as customer history of transactions/interactions with certain services or products.

Kishana et al. [5] presented is a report of k-means clustering technique and SPSS(Statistical Package for the Social Sciences) Tool to develop a real time and online system for a particular super market to predict sales in various annual seasonal cycles. The model received inputs directly from sales data records and automatically updated segmentation statistics at the end of day's business. Singh et al. [6] proposed a model for segmenting a group of customers in online market by using the predictive neural network approach and statistical analysis. The model is based on the parameters like product reviews, products, buying pattern, viewing pattern and time based segments and clustering techniques in the sector of electronic commerce. The results showed the distinct brands per customer and evaluation of unigrams, bigrams and trigrams have been done as a part in the feature extraction. Also, neural network is used for finding out the most preferred brands which the customer prefers and thereafter the statistical analysis is done from the predicted data.

Ezenkwu et al. [7] presented a MATLAB implementation of the k-Means clustering algorithm for customer segmentation based on data collected from a mega business retail outfit. Training of the algorithm was done on a two-feature dataset. Based on the clustering done business decision were taken in terms of risky situations such as credit relationship with its customers; identification of products associated with each segments and how to manage the forces of demand and supply; unravelling some latent dependencies and associations amongst customers, amongst products, or between customers and products which the business may not be aware of; ability to predict customer defection and which customers are most likely to defect; and raising further market research questions as well as providing directions to finding the solutions. Kamthania et al. [8] proposed a model for formulating business strategies based on the user's interest and location. Principal Component Analysis (PCA) followed by k-means clustering. PCA maps the number of correlated variables into a smaller set of uncorrelated variables called the principal components. The model works for small businesses very well but is not that efficient for larger businesses.

3 Proposed Idea

Our proposed idea is shown in the Figure 1. We selected a data set of a mall which have details of a purchased product. After doing pre-processing and feature extraction, we will select features for clustering. The features will be used to group customers using 5 clustering algorithms. After making inferences from the cluster, we will implement 6 classification models to classify new customers. Based on the cluster of a customer and most purchased and correlated product, we will perform various use cases and send various offers via email.

4 Methodology

4.1 Data Collection

The dataset we are using for this project is from a Shopping Mall in London. It was collected by using different kinds of IOT devices used in Malls like POS machines, scanners, checkout systems. The data from all these devices is merged together to form one dataset for analysis. Figure 1 shows a view of the dataset we are using. The dataset has total 8 attributes namely InvoiceNo, StockCode, Description, Quantity,InvoiceDate, UnitPrice, CustomerId, Country.

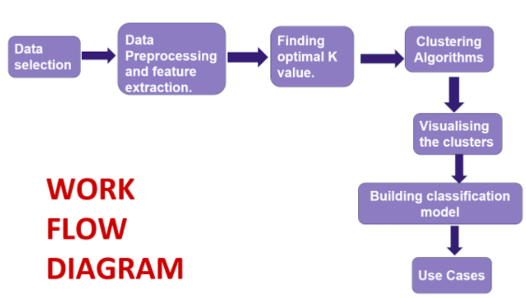

Figure 1: Work Flow Diagram

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Figure 2: Dataset

4.2 Data Pre-processing

Data pre-processing is the process of converting the raw data into useful and meaningful data to be used for further analytics and algorithms. It involves dealing with missing values, normalization of data, selection of relevant features etc. We have started with removing the null values and then selected the features best suited.

4.2.1 Data Cleaning

This step involves dealing with missing and irrelevant parts in our dataset. For missing values we can do two things: either remove the tuple from the data completely or fill those missing places with the most probable value. Most probable value can be mean or mode. In our dataset the attribute ”Description” had 1454 and attribute ”CustomerID” had 135080 missing values. We chose the former approach and ignored the rows containing missing values.

4.2.2 Feature Extraction

Feature extraction is a process of deriving new features from the already existing featuers and then ultimately using those new features discarding original ones. The dataset has 8 inbuilt features: ”InvoiceNo”,”StockCode”,”Description”,”Quantity”,”InvoiceDate”,”UnitPrice”,”CustomerID”,”Country”. We used ”Quantity” and ”UnitPrice” to create a new attribute namely ”Amount”. This ”Amount” was later used to find average spendings for each customer and named it attribute ”AvgSpend”. We used the ”InvoiceDate” and ”CustomerID” to derive three attributes ”frequency”,”recency” and ”meanDays”. Then we found the top 5 most bought products using ”StockCode” and ”Quantity” in order to use them as attributes. Now for each customer ID we had 9 attributes: ”AvgSpend”,”frequency”,”recency”,”meanDays”,”85123A”,”2 and ”20725”.

We derived new features from the inbuilt features. The features are known as RFM(Recency,Frequency, Monetary).

1. Recency - how recently the customer did his purchase?
2. Frequency – How often do they purchase?
3. Monetary Value – How much revenue they generate?.

4.Mean Days -is the average number of days customer has waited for his next purchase.

5. Also include top 5 products of each cluster as a features.

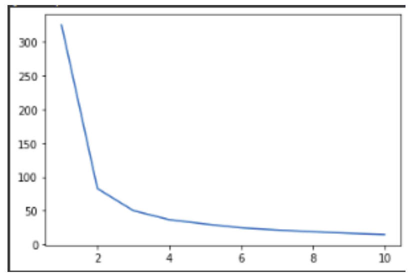

Figure 3: Cost vs K values for RFM

4.2.3 Data Normalisation

The range of these attribute values vary too much on scale. We used data normalisation (min-max scaling) to re-scale and shift those values between 0 and 1. It involves subtrating the minimum value of attribute from each entry and then dividing by difference of maximum and minimum values for each attribute. After this all the values in data ranges between zero and one.

4.2.4 Anova Analysis

Out of these 9 features we intended to select the most relevant ones. For this task we used Analysis of Variance(ANOVA). It is statistical tool which consist of various models and their procedures. In particular we used F-test of ANOVA to determine the top 5 relevant attributes which helps in forming clusters in our data. It was done using f_classif and SelectKBest from sklearn.featureselection. This analysis gave the result that attributes "AvgSpend","frequency","meanDays" and "recency" are more relevant than others. The work was continued using these attributes only.

4.3 Clustering

It is a unsupervised ML algorithm which divides the dataset into different clusters. There are various clustering algorithms available but we started with most popular one i.e. K-Means Clustering. For that we need to find optimal K value first which was done using Elbow method. We implemented another version of K-means we call it K-Means ++

4.3.1 Elbow Method

For clustering we need to find the optimal K value first. One of the way to decide K value is to use Elbow method. We can do that by computing the cost of K-Means model with different values of K. It can be observed that the cost decreases as the value of K increases. But at a certain value of K the cost stops decreasing drastically. That value can be considered as the optimal K value. We have used elbow method for two datasets, one with "Avgspend","frequency", and "recency" values only and another with added "meanDays" values. Plotted graphs for first and second datasets are Figure 3 and Figure 4 respectively. The graph of K v/s cost forms a shape of an elbow as in the figures. From both the graphs you can infer that optimal K value should 3. But there is a significant difference in cost value for both the datasets. Figure 5 shows the cost comparisons for RFM features and RFM + MeanDays features. Cost of having 3 clusteres using RFM is 49.9 and cost of 3 clusters with RFM+meandays is 135.63. That is why we discared MeanDays feature. Including meanDays also affects our accuracy.

4.3.2 K-means

K-Means algorithm is a centroid based clustering technique. This technique cluster the dataset to k different cluster having an almost equal number of points. Each cluster is k-means clustering algorithm is represented by a centroid point. The initial k-centroids were picked randomly from the data points.

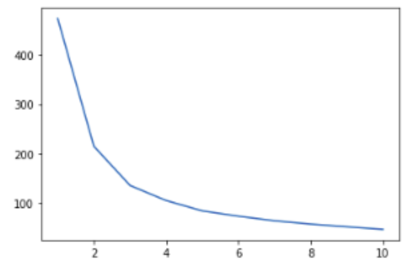

Figure 4: Cost vs K values for RFM+Mean Days

```
#in both using mean days and without using mean days optimal k is 3
print(cost1[2],cost2[2])

49.93351925902856 135.63964506190402
```

Figure 5: Cost

4.3.3 K-means ++

K-Means++ is a smart centroid initialization technique and the rest of the algorithm is the same as that of K-Means.

4.3.4 Mean-Shift

Mean-Shift assigns the data points to the clusters iteratively by shifting points towards the mode (mode is the highest density of data points in the region, in the context of the Meanshift). As such, it is also known as the Mode-seeking algorithm.

4.3.5 Agglomerative

Also known as bottom-up approach or hierarchical agglomerative clustering (HAC). The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram.

4.3.6 Spectral

Spectral clustering is a graph partitioning problem. Data points are treated as nodes of a graph. The nodes are then mapped to a low-dimensional space that can be easily segregated to form clusters. No assumption is made about the shape/form of the clusters. The goal of spectral clustering is to cluster data that is connected but not necessarily compact or clustered within convex boundaries.

4.4 Data After Clustering

Figure 6 shows how the clustered data looks like. Classification models will be built on the clustered data. The clustered data has one extra feature ”clno” which is the cluster number the customer id belongs to.

4.5 Visualizing The Clusters and Making Inferences

We have used seaborn.pairplot to visualize the different scatter plots between the attributes ”freq”,”recency” and ”avgspend” shown in the Figure 7. It can be observed that most of the customers have frequency close to 0. The few portion of customers who are frequent have low average spend value and high recency. The recency value clearly divides the customers into clusters. High recency value tells you are in cluster 0 whereas low recency value says you are in cluster 1. The ones in cluster 2 have average

	id	freq	rec	avgspend	clno
0	12347.0	0.024291	0.994638	0.099186	0
1	12348.0	0.012146	0.798928	0.072380	0
2	12349.0	0.000000	0.951743	0.283126	0
3	12350.0	0.000000	0.168901	0.053869	1
4	12352.0	0.040486	0.903485	0.022632	0

Figure 6: Clustered Data

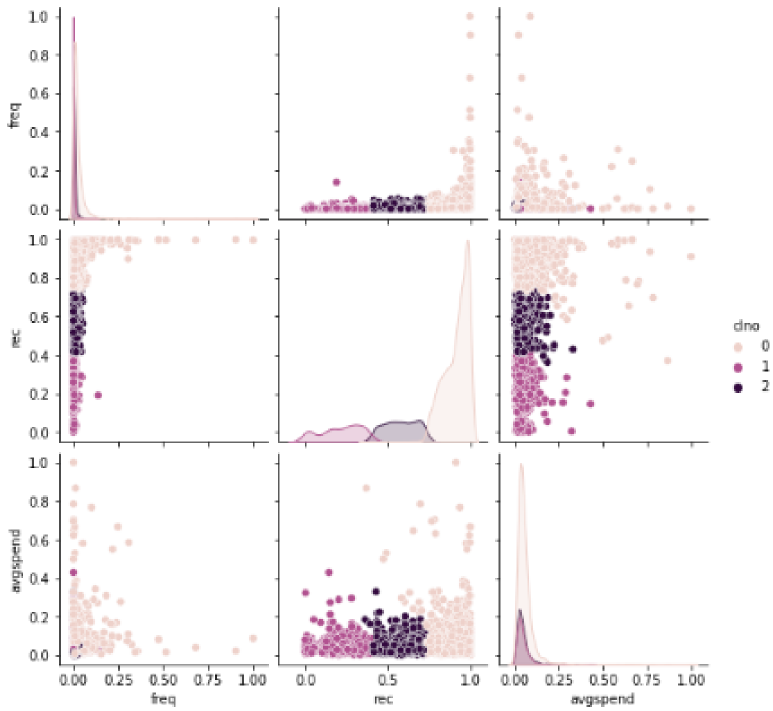

Figure 7: Pairplot

recency value. If we consider the average spend value you can see that most customers belong to low to medium category. The only ones whose average spend is high are less frequent and belong to cluster 0. We can infer that they have made a single purchase but huge one. Figure 8 shows the average RFM values for each cluster.

4.6 Classification Models

For the use cases we need to create a classification model. We have used different classification techniques namely Logistic Regression, SVC, Decision Trees, Random Forest, KNN and Ridge Classifier. We compared accuracy scores of all these algorithms and found apart from Ridge classifier all were giving good results.

4.6.1 Model Training and Validation(Hyperparameter Tuning)

For classification task we need to define our independent variables and dependent variables(output). The independent variables were the attributes "avgspend","rec" and "freq" and the output label was

	clno	rec	freq	avgspend
0	0.0	0.912584	0.022052	0.055709
1	1.0	0.215143	0.002299	0.042875
2	2.0	0.575987	0.006320	0.043740

Figure 8: Average RFM Values of each Cluster

the cluster number namely column "clno". Then we ramdomly split the data into training set(70%) and testing set(30%). Then the classification was done on our dataset using each algorithm one by one. For better results we also used K-Fold cross validation with number of folds as 6. The accuracy was calculated on each fold and average was returned. Each algorithm has its own set of hyperparameters, for example we have K in KNN and max_depth in Decision trees. We checked the accuracy levels for different values of hyperparameters and chose the best one.

We started with Random Forest classifier which has hyperparameters "max_depth" and "number_of_estimators". We checked 100 different combinations of both these values and for each combination cross validation was performed. The combination which gave highest accuracy was "max_depth"=9 and "number_of_estimators"=70. Then again we repeated the same process with Decision Tree classifier with "max_depth" as hyperparameter. We checked for 10 different values of max_depth and fount the optimal value as "max_depth"=4. For KNN the hyperparameter is K(number of neighbours to be considered). After checking for 10 different values we found the optimal value as 5. In case of SVC there are three hyperparameters we took into account: C (Regularization parameter), gamma (kernal parameter) and kernel. We had to try out 160 different combinations to find the best one and it turned out to be "C"=600, "gamma"=1 and "kernel"=rbf. For Logistic Regression we have C (Regularization parameter) as hyperparameter and tried 8 values. Optimal "C" was 300. At last for Ridge classifier hyperparameter was "alpha" whose 5 different values were tested and optimal value was 1.

4.6.2 Model Evaluation

For evaluating the different models we calculated individual accuracy scores on our test data points. Classification report was also generated for each model which includes Precision, Recall, F1-score and Support. The hyperparameters of models were set to their best value. On comparing the accuracy score of all these algorithms we found that apart from Ridge classifier all were giving good results. In particular we can say that K-Nearest Neighbour(KNN) works exceptionally well among the others but the problem with it is it is a lazy algorithm. It predicts most accurately but takes time to do so . Second best algorithm in terms of accuracy is Random Forest. Its accuracy level is almost equal to KNN and works faster. So for classification we decided the Random Forest as our final classification model.

4.7 Use Cases

We have implemented three use cases for this project

4.7.1 Product Suggestions

In the first use-case, for a particular customer, we find the cluster to which it belongs and the most purchased product in that cluster. If the customer has not purchased any product from the top product in that cluster then we suggest that product to him.

4.7.2 Market Basket Analysis

Market Basket Analysis is used to find an association between the products. It finds the product which is most frequently brought together by the customers. We have used aprior algorithm to find these associations. We apply market basket to the clusters first and then we suggest the offers.

4.7.3 Combo Offers

In this use case, we find the most purchased product by the customer and the least sold product in the store. And we have provided a combined offer on both products.

5 Results

Silhouette Coefficient It is a method of interpretation and validation of consistency within clusters of data.It is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from -1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters.

Clustering Results We have implemented five different clustering algorithms, K-Means,K-Means++, Mean shift, Spectral clustering, and Agglomerative clustering on RFM features of customers. We have calculated the silhouette coefficient for these algorithms. Table 1 shows the results. Based on the silhouette score, we have selected the Agglomerative Clustering algorithm.

Classification Results We have implemented six algorithms: Random Forest, Decision Tree, KNN, support vector classifier, Logistic Regression, and Ridge Classifier for classifications. Table 2 shows the results. The result from the clustering algorithm is used as the training set. Based on the accuracy of these algorithms, best classification model is K Nearest Neighbours but we do not take it since it is a lazy algorithm and after that the best is random forest.

Algorithm	Silhouette Score
K-Means	0.609185
K-Means++	0.610002
Mean-Shift	0.606873
Spectral	0.503549
Agglormerative	0.610704

Table 1: Silhouette Score Comparison.

Classification Method	Train Accuracy	Test Accuracy
Random Forest	1.000000	0.997708
Decision Tree	1.000000	0.996528
K Nearest Neighbours	0.999210	0.998450
Support Vector	0.996411	0.996528
Logistic Regression	0.989286	0.995045
Ridge Classifier	0.666508	0.666667

Table 2: Classification Method Comparison .

6 Discussion and Future Work

The results in our project showed that Agglomerative clustering performs better than K-Means, K-Means++, Mean-Shift, Spectral clustering for our dataset. All the previous work to best of our knowledge showed that K-Means performed better in almost all cases. Here the results are different than the previous work and one reason is the dataset. The datasets used for customer segmentations vary. One dataset has one set of attributes and another dataset might have another set of attributes. Based on the attributes the features for clustering are selected. The features are another reason for the performance of the algorithms. In our case we finally decided to go with RFM features but we also tried other features as well like mean_days but the results were not so good by using this feature.

The result of this project is directly linked to the model and the data at hand. Having more data and features does not always improve a model, but better data and better features certainly do. Adding more features, or at least further improving the current features could increase the performance of the model. Therefore it is of interest to investigate if more suitable features, describing user behaviour can be aggregated from the data source. Future work for this project will include

(1) Dynamic Model: The model will make the suggestions to customers while they are shopping in the Mall.

(2) Datasets: Implementing all the clustering algorithms for different datasets which help to generalize marketing strategies.

(3) Clustering Algorithms: There are many other option of clustering algorithms and classification models, and because the results are data dependent they might give better results.

7 Conclusion

The purpose of this project was to explore the different clustering algorithms as a tool for identifying user behaviour on a customer dataset. In order to achieve the purpose of the project, six classification models have been trained to categorize users into different user personas. The performance of the classification models have been analyzed using a number of evaluation metrics. The results of this study showed that Agglormerative clustering algorithm performed better than rest of the clustering algorithms and for classification models Random Forest performed better. We choose many features for clustering but Recency, Frequency, Monetary (RFM) gave best results for our dataset. For the clustered customers we implemented three use-cases.

It is worth noting that the result of the project, like most other classification problems, is tightly connected to the data available. Should a couple of different decisions have been made during the project, the result would likely have been different, better or worse.But in the end, the result of the project satisfied the our goals from the project and the purpose has been fulfilled.

8 Individual Contributions

	Abishek	Aditya	Muzafar	Nishant
Literature Survery	25%	25%	25%	25%
Data Pre-Processing	25%	25%	25%	25%
Model Framework	25%	25%	25%	25%
Feature Extraction	25%	25%	25%	25%
Clustering Implementation	10%	40%	40%	10%
Classification Implementation	40%	10%	10%	40%
Use-cases	25%	25%	25%	25%

References

[1] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer segmentation using k-means clustering," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pp. 135–139, 2018.

[2] L. Abidar, D. Zaidouni, and A. Ennouaary, “Customer segmentation with machine learning: New strategy for targeted actions,” in *Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications*, SITA’20, (New York, NY, USA), Association for Computing Machinery, 2020.

[3] P. D. Hung, N. T. T. Lien, and N. D. Ngoc, “Customer segmentation using hierarchical agglomerative clustering,” in *Proceedings of the 2019 2nd International Conference on Information Science and Systems*, ICISS 2019, (New York, NY, USA), p. 33–37, Association for Computing Machinery, 2019.

[4] S. Nguyen, “Deep customer segmentation with applications to a vietnamese supermarkets’ data,” *Soft Computing*, vol. 25, 06 2021.

[5] K. Kashwan and C. Velu, “Customer segmentation using clustering and data mining techniques,” *International Journal of Computer Theory and Engineering*, vol. 5, pp. 856–861, 01 2013.

[6] H. Singh, “Improving customer segmentation in e-commerce using predictive neural network,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 2326–2331, 04 2020.

[7] C. Ezenkwu, S. Ozuomba, and C. Kalu, “Application of k-means algorithm for efficient customer segmentation: A strategy for targeted customer services,” *International Journal of Advanced Research in Artificial Intelligence(IJARAI)*, vol. 4, 10 2015.

[8] D. Kamthania, A. Pahwa, and S. Madhavan, “Market segmentation analysis and visualization using k-mode clustering algorithm for e-commerce business,” *Journal of Computing and Information Technology*, vol. 26, pp. 57–68, 01 2018.