Consumer Personality Analysis: Tailoring Marketing Strategies for Diverse Segments

Mingyuan Zhou[a]
BSc Computer Science, University of Birmingham, Birmingham, U.K.

Keywords: Analysis, Market Segment, Enterprise, Customer.

Abstract: The main aim of this study is to enhance the application of customer personality analysis in devising personalized marketing strategies. Firstly, cluster analysis is employed to segment customers based on their purchasing behaviour and demographic characteristics, aligning marketing efforts with the unique preferences of identified market segments. Secondly, association rules mining is utilized to identify product similarity patterns among market segments, guiding targeted promotional strategies. Thirdly, the effectiveness of these tailored strategies is evaluated by comparing and analyzing participation indices of different market segments. This study not only contributes to academic discourse by applying and evaluating contemporary data analysis methods in real-world settings but also offers practical insights for enterprises seeking to enhance marketing efficiency. The experimental results underscore the importance of detailed customer segmentation and the potential of personalized marketing to enhance customer satisfaction and loyalty. By showcasing the applicability and impact of these strategies in real-world scenarios, this study emphasizes their pivotal role in the ever-evolving marketing landscape, providing a fresh perspective for enterprises to comprehend and engage their diverse customer base.

1 INTRODUCTION

In the field of modern marketing, it is not only an advantage but also a necessity to understand complicated consumer behaviour. In this case, the concept of customer personality analysis has become an important tool (Ghorbani, 2022). By dissecting the diverse personalities of customers, businesses can transition from a one-size-fits-all marketing strategy to highly tailored approaches. This shift is not merely a tactical change but a strategic realignment towards more empathetic and effective marketing. The essence of customer personality analysis lies in its ability to uncover the unique preferences, behaviours, and expectations of different customer segments (Smith, 2020). It provides a nuanced perspective, through which companies can observe customers, so as to formulate more targeted and resonant marketing communication strategies. The importance of this analysis transcends traditional marketing boundaries and provides a blueprint for enterprises to establish deeper ties with the audience. Therefore, this field has received great attention in academic and practical fields, and it is worth a comprehensive review and exploration.

The evolution of customer personality analysis is marked by the integration of complex data analysis techniques and psychological opinions (Alves, 2023). Early efforts in this field mainly relied on demographic and transaction data to segment customers. However, with the advent of big data and advanced analysis, the focus has shifted to more subtle methods that consider a wider range of customer data. Recent research has adopted machine learning algorithms, including clustering and classification techniques, to determine customer segmentation with higher accuracy (Abdolvand, 2015)(Akhondzadeh, 2015). In addition, association rule mining has been used to reveal the patterns of customers' buying behaviour and their opinions on product preferences and cross-selling opportunities (Chan, 2011). Social media analysis also plays a key role, enabling marketers to use unstructured data for sentiment analysis and trend judgment (Khatoon, 2021)(Batrinca, 2015). These methods emphasize the growth trend of data-driven customer insight, go

[a] https://orcid.org/0009-0009-1046-3061

Zhou, M.
Consumer Personality Analysis: Tailoring Marketing Strategies for Diverse Segments.
DOI: 10.5220/0012925400004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 243-249
ISBN: 978-989-758-713-9

beyond the traditional segmentation model and turn to dynamic analysis based on behaviour. It is worth noting that the literature reflects the consensus on the transformative impact of these analysis-driven strategies in enhancing customer engagement and personalization (Bell, 2018)(Anshari, 2019). These studies combine the rigour of analysis and market acumen and lay a solid foundation for promoting the development of customer personality analysis.

The primary objective of this study is to enhance the utilization of customer personality analysis in crafting tailored marketing strategies. Specifically, it employs cluster analysis to categorize customers based on their purchasing behaviour and demographic attributes, aligning marketing efforts with the distinct preferences of identified market segments. Additionally, association rules mining is utilized to unveil patterns of product similarity among these segments, guiding targeted promotional initiatives. Furthermore, the efficacy of these customized strategies is assessed through a comparative analysis of participation indices across different market segments. This research contributes not only to academic discourse by applying and evaluating contemporary data analysis techniques in real-world settings but also furnishes actionable insights for enterprises aiming to bolster marketing effectiveness. The empirical findings underscore the significance of meticulous customer segmentation and underscore the potential of personalized marketing to bolster customer satisfaction and loyalty. By showcasing the practical applicability and impact of these strategies, this study underscores their pivotal role in the evolving marketing landscape, offering a fresh perspective for enterprises to comprehend and engage their diverse customer base.

This essay is structured as follows: The first chapter lays the foundation by discussing the core concepts and methodologies of customer personality analysis. The second chapter delves into the details of the applied techniques and their rationale. The third chapter presents the analysis of the experimental results, followed by a discussion. The final chapter concludes the study, reflecting on the findings and their implications for both theory and practice.

2 METHODOLOGIES

2.1 Dataset Description and Preprocessing

This study utilizes a dataset from Kaggle's "Customer Personality Analysis" that provides a composite view of customers' purchasing behaviours over two years (Gaurav, 2024). It includes demographics, campaign responses and transactional data across product categories. The basic attributes include income, education, marital status and total amount spent on different products. In preprocessing, missing values (especially missing values in the "Income" field) were estimated by using median values to maintain distribution integrity. Categorical variables such as 'Education' and 'Marital Status' were encoded to facilitate algorithmic interpretation, and feature selection based on correlation ensures that variables that have a significant impact on customer segmentation are retained. The Countplot of Marital Status with Education Level is shown in Figure 1.

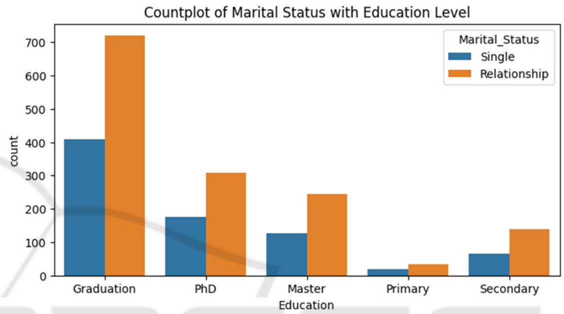

Figure 1: Countplot of Marital Status with Education Level (Photo/Picture credit: Original).

2.2 Proposed Approach

The proposed approach in this study constitutes a continuous process of extracting, analyzing, and interpreting customer data to formulate targeted marketing strategies. The initial step involves data preprocessing, entailing the cleaning and structuring of raw data. Techniques such as imputation and standard encoding are employed to standardize the dataset for analysis. Additionally, categorical columns are encoded using label encoding, and outliers are addressed using the Interquartile Range (IQR) method.

Subsequently, data clustering is performed using the K-means algorithm, an unsupervised learning technique, to categorize customers into distinct market segments based on similar traits and behaviours. This segmentation enables the identification of diverse needs and preferences among various customer groups. At the heart of the approach lies the utilization of the Apriori algorithm for model training, aimed at uncovering relationships between different customer attributes and their purchasing patterns. These patterns offer insights into customer preferences, facilitating the prediction of future

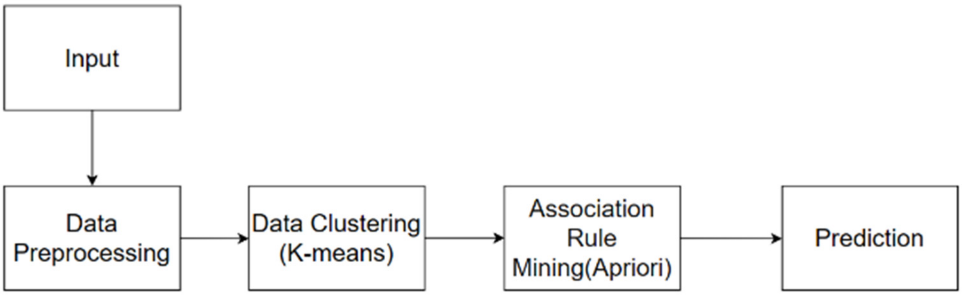

Figure 2: The Model Pipeline (Photo/Picture credit: Original).

purchasing behaviours. The final step involves prediction, wherein a Logistic Regression model serves as a predictive tool to allocate new customers to established clusters. Figure 2 illustrates the streamlined process pipeline, guiding the flow from data preprocessing to prediction.

2.2.1 Cluster Analysis (KMeans Clustering)

The attempt at segmentation began with KMeans clustering, which is an unsupervised learning model renowned for its efficacy in dividing datasets into distinct, non-overlapping subsets or clusters. The basic idea of K-means clustering is that given the data points in the data set, the algorithm tries to find K centres (or centroids), which can best represent the data. The algorithm works iteratively: initialize the centres, assign each data point to the nearest centre, and then recalculate the centres according to the data points assigned to them. This process is repeated until the centre point no longer changes significantly or reaches the preset number of iterations. The Elbow Method is used to control the implementation, which is a heuristic method for determining the optimal number of clusters by locating the point where the decreasing rate in the within-cluster Sum of Squared Errors (SSE) sharply changes, similar to the elbow. This phase culminated with the silhouette analysis to confirm the model's clustering consistency and separation acuity.

2.2.2 Association Rule Mining (Apriori Algorithm)

Apriori algorithm is the cornerstone of data mining field, which is used to extract frequent itemsets from the data pool and export association rules. It operates on a foundational principle of the 'Apriori property', which assumes that all non-empty subsets of a frequent itemset must also be frequent. This characteristic significantly optimizes the search by reducing the number of itemsets considered in this study. The algorithm uses the breadth-first search method and constructs an itemset layer with length k from frequent itemsets with length k-1 by employing the downward closure lemma. This iterative process continues until further expansion is impossible.

The frequent itemsets criteria include Support, Confidence and Lift. Support is the fraction of transactions that contain an itemset.

$$Support(X,Y) = P(XY) = \frac{number(XY)}{number(AllSamples)} \quad (1)$$

Confidence reflects the probability that X appear in transactions that contain Y.

$$Confidence(X \leftarrow Y) = P(X \mid Y) = \frac{P(XY)}{P(Y)} \quad (2)$$

Lift represents the ratio of the probability of containing X at the same time under the condition of containing Y to the probability of the occurrence of X population.

$$Lift(X \leftarrow Y) = \frac{P(X \mid Y)}{P(X)} = \frac{Confidence(X \leftarrow Y)}{P(X)} \quad (3)$$

The utility of Apriori in understanding customer behaviour lies in its ability to reveal hidden patterns in purchasing behaviour and provide opportunities for cross-selling and up-selling by exploring the possibility of purchasing items together.

2.2.3 Predictive Modelling (Logistic Regression)

This study uses logistic regression, a generalized linear model on predictive modelling that is a part of supervised learning in machine learning. The natural log of the outcome's probability of occurring is converted into a logit variable by fitting a logistic function to a collection of data. This is how the model operates. It functions on the odds ratio, transforming linear combinations of predictors through the logistic function, thus generating threshold probabilities to assign class labels. The process encompassed the standard practices of feature scaling, cross-validation, and hyperparameter optimization, to enhance model performance and mitigate overfitting. The bound of the Logistic function is between 0 and 1. Therefore, it is useful to solve binary classification tasks like predicting customer engagement or purchase intentions.

2.3 Implementation Details

The execution of the model depends on Kaggle's Python notebook, utilizing Pandas for data preprocessing and Scikit-learn for modelling. Data visualization is done using the Seaborn and Matplotlib libraries. In the preprocessing process, the classified data is encoded into digital format, 0and the continuous variables are normalized by standard scale. Hyperparameter adjustment (especially for KMeans clustering) involves iterative experiments to finally determine the number of clusters, based on the elbow method of silhouette score and optimal granularity. The hyperparameter of the logistic regression model keeps the default value, which provides reliable baseline performance. However, association rules mining by Apriori algorithm does not need complex parameter adjustment because it depends on the support and confidence thresholds.

3 RESULTS AND DISCUSSION

3.1 Clustering Results

In this study, the application of KMeans clustering in the dataset illustrates different customer groups based on purchasing behaviours and demographics. Figure 3 shows the silhouette scores for cluster numbers from 2 to 9, and this is a method to determine the optimal number of clusters by comparing the mean intra-cluster distance with the mean nearest-cluster distance. The results demonstrated that the two-cluster solution has the highest silhouette score, indicating that the structure is firm and the cluster density is appropriate. This is confirmed by the elbow method shown in Figure 4, in which the sum of squared distances within clusters is stable, and additional clusters do not significantly improve the model fit.

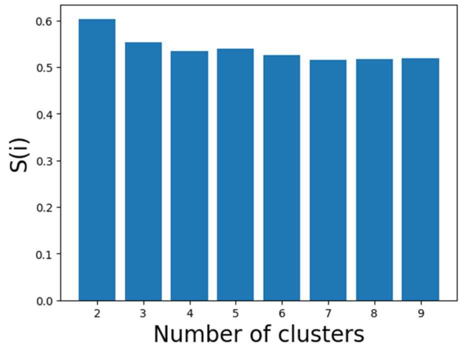

Figure 3: Silhouette Scores for Determining Optimal Cluster Number (Photo/Picture credit: Original).

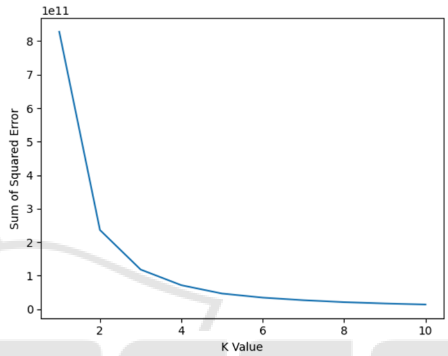

Figure 4: Elbow Method Visualization for KMeans Clustering (Photo/Picture credit: Original).

As shown in Figure 5, 6 and 7, the final distribution of the cluster shows that the customer base is almost evenly distributed between the two primary segments. Cluster 1, identified as the 'Highly Active Customers,' includes customers who spend more in various product categories, indicating that this market segment has greater engagement and brand loyalty. Cluster 2 is called 'Least Active Customers' and consists of individuals with lower overall expenses, indicating that this group is either less involved, price-sensitive or may shop infrequently.

The differences between these clusters are of great significance to targeted marketing strategies. For example, customers in Cluster 1 may be more receptive to quality products and loyalty programs, while customers in Cluster 2 may be more responsive to discount promotions and value-oriented products. Enterprises can tailor marketing campaigns according to the unique characteristics and needs of each cluster, so as to allocate resources more effectively.

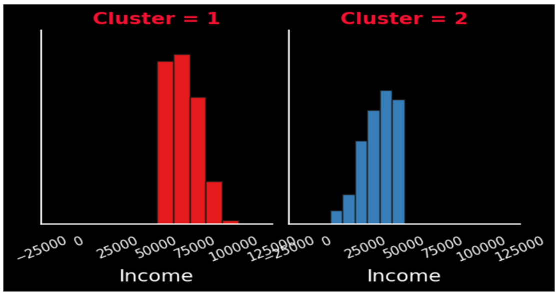

Figure 5: Distribution of Customer Clusters in Income (Photo/Picture credit: Original).

Figure 6: Distribution of Customer Clusters in Expenses (Photo/Picture credit: Original).

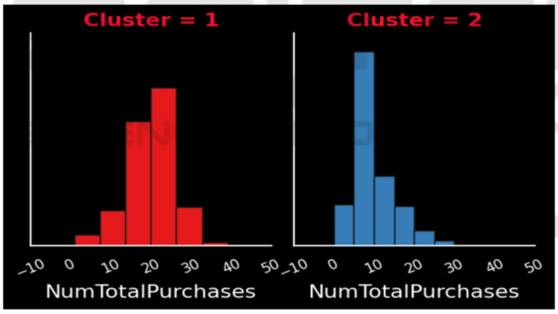

Figure 7: Distribution of Customer Clusters in Total Purchases (Photo/Picture credit: Original).

3.2 Association Rule Mining with the Apriori Algorithms

Apriori algorithm plays an important role in mining association rules in customer segmentation. By creating various customer segments according to age, income and participation duration, the deep relationship between different customer behaviours and product preferences is revealed. As seen in Table 1, the segments range from 'Senior' customers with a 'Medium to high income' who are 'Old customers' to 'Adults' with a 'Low income' classified as 'New customers'. This difference is very important for understanding the basic patterns that drive purchasing decisions.

This study uses the processed dataset with encoded categorical variables and the algorithm revealed key relationships. For example, there is a strong link between 'Highly Active Customers' in wine purchases and those in the 'Senior' age group with 'Medium to high income'. This correlation not only verifies the segmentation strategy but also provides information for targeted marketing initiatives. For instance, a 'Senior' segment that shows high activity in wine purchases can be approached by high-quality wine products.

Furthermore, the duality of the 'Cluster' column indicating the activity level allows a clear comparison between the most active and the least active customers. This insight gained from mining association rules guides accurate and effective marketing activities, optimizes resource allocation and improves customer satisfaction. These findings emphasize the utility of Apriori algorithm in identifying complex consumer patterns in a multi-dimensional dataset.

Table 1: Information from the Preprocessed Dataset.

Education	Marital_Status	Income	Kids	Expenses	TotalAcceptedCmp	NumTotalPurchases	Age	day_engaged	Cluster	Age_group	Income_group	dayengaged_group
0	1	58138	0	1617	1	25	67	4136	1	Senior	Medium to high income	Old customers
0	1	46344	2	27	0	6	70	3586	2	Senior	Low to medium income	New customers
0	0	71613	0	776	0	21	59	3785	1	Mature	High income	Discovering customers
0	0	26646	1	53	0	8	40	3612	2	Adult	Low income	New customers
2	0	58293	1	422	0	19	43	3634	1	Adult	Medium to high income	New customers

3.3 Predictive Modeling and Evaluation

In this study, Logistic Regression is applied to predict whether customers would fall into Cluster 1 or Cluster 2, and they are classified according to their buying behaviour. As shown in Figure 8, the confusion matrix shows the performance of the model, with the X-axis representing the predicted cluster and the Y-axis representing the actual cluster.

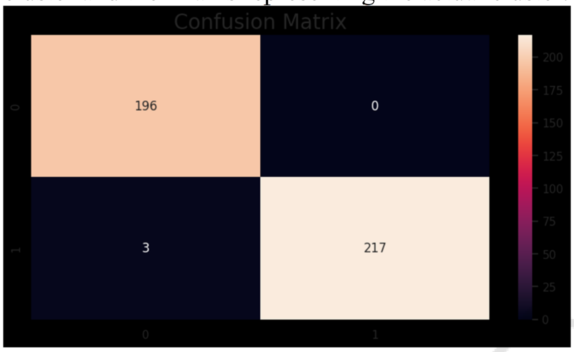

Figure 8: Confusion Matrix of Logistic Regression Predictions (Photo/Picture credit: Original).

The results obtained from the model evaluation show excellent accuracy of 98% to 100% and similar high recall. The overall accuracy of the model is about 99.28%, and the F1 scores that balance accuracy and recall are above 99%. These metrics show a highly reliable model, which can accurately segment customers according to the data.

The author finds its high accuracy that shows the model is robust and effective and has great potential to assist strategic marketing initiatives. The small number of misclassifications (3 out of 416) further emphasizes the predictive ability of the model and ensures that marketing resources are effectively allocated to target customers. This modelling can provide operational insights and realize personalized marketing strategies, thus significantly improving customer engagement and conversion rates.

4 CONCLUSIONS

This study introduces a comprehensive method for consumer personality analysis, leveraging a Kaggle dataset to unveil customer purchasing behaviours. A methodology is proposed to deeply investigate rich consumer data, utilizing the synergy of KMeans clustering, the Apriori algorithm, and logistic regression. KMeans clustering is employed to partition the dataset into distinct segments, unveiling inherent groupings within the consumer base. The Apriori algorithm adeptly mines the intricate associations between consumer segments and their purchasing tendencies. Subsequently, logistic regression is utilized to predict cluster membership, enabling the measurement of potential purchasing decisions among consumers. The integration of these methods yields a robust model capable of analyzing and interpreting the complex dynamics of customer interactions.

Extensive experiments are conducted to evaluate the proposed method. The experimental results demonstrate the model's effectiveness, with high clustering accuracy and forecasting modelling affirming the efficacy of segmented customer profiles. The utilization of association rules further solidifies the advantages of analysis and furnishes concrete, data-driven insights for strategic marketing endeavours. Looking ahead, future research will consider individual consumers and their influence on purchasing patterns as the primary objective. The focus will be on enhancing the accuracy of market segmentation and customizing marketing strategies with greater precision. The vision of this study is to explore the dynamic relationship between evolving consumer characteristics and market trends, guiding enterprises to successfully engage with customers in a more nuanced manner.

REFERENCES

Abdolvand., Neda., Amir, A., and Mohammad, A., (2015). Performance management using a value-based customer-centered model. International Journal of Production Research. vol. 53(18). pp: 5472-5483.

Akhondzadeh, N., Elham., and Amir, A., (2015). Mining the dominant patterns of customer shifts between segments by using top-k and distinguishing sequential rules. Management decision. vol. 53(9). pp: 1976-2003.

Alves, G., Miguel., and Tobias, M., (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. Information Systems and e-Business Management. vol. 21(3). pp: 527-570.

Anshari., Muhammad., et al. (2019). Customer relationship management and big data enabled: Personalization & customization of services. Applied Computing and Informatics. vol. 15(2) pp: 94-101.

Batrinca., Bogdan., and Philip, C., Treleaven., (2015). Social media analytics: a survey of techniques, tools and platforms. Ai & Society. vol. 30. pp: 89-116.

Bell., David., and Chidozie, M., (2018). Data-driven agent-based exploration of customer behavior. vol. 94(3) pp: 195-212.

Chan, C-C., Henry., Chi-Bin, Cheng., and Wen, C., Hsien., (2011). Pricing and promotion strategies of an online shop based on customer segmentation and multiple objective decision making. Expert Systems with Applications. vol. 38(12). pp: 14585-14591.

Gaurav, B., (2024). Kaggle dataset. https://www.kaggle.com/code/gbiamgaurav/consumer-behaviour-analysis/

Ghorbani., Mijka., Maria, K., and Andrea T., (2022). Consumers' brand personality perceptions in a digital world: A systematic literature review and research agenda. International Journal of Consumer Studies. vol. 46(5), pp: 1960-1991.

Khatoon., Sajira., and Varisha, R., (2021). Negative emotions in consumer brand relationship: A review and future research agenda. International Journal of Consumer Studies. vol. 45(4). pp: 719-749.

Smith., Trevor, A., (2020). The role of customer personality in satisfaction, attitude-to-brand and loyalty in mobile services. Spanish Journal of Marketing-ESIC. vol. 24(2) pp: 155-175.