

A PREDICTIVE APPROACH IN ANALYSIS OF CONSUMER BEHAVIOR AND SATISFACTION

Dr. Punam Bhoyar

Indira Institute of Management, Pune.

Abstract

Organizations today want to be extrapolative they want to gain information and insights on every minutia of customers. Predictive analytics is one of the tools which has proved to be cornerstone in customer centricity. Though analytics is used for performing the customer analysis for decades, the manual approach of data management and analysis have constrained the functionalities. Professionals now leverage the individual information from demographics to purchase history and perform predictive analytics for optimal decision making, business marketing and thus business growth. Here we apply ensemble learning method for classification based on the monthly income and spending behaviour of customers. Secondly, we analyse customer satisfaction by applying unsupervised learning clustering method based on similarity through different performance measures.

Keywords: Predictive analytics, customer segmentation, K-means algorithm, classification algorithm.

1. Introduction

Consumer behaviour is the study of consumer attitudes and preferences that determines their buying behaviour. Predictive analytics plays a crucial role for market researchers by giving accurate predictions. The breakthrough in the big data platforms and data integration has facilitated the decision making and alleviated the operations. Research on consumer behaviour have enriched business decisions. It has always been pivotal to businesses to understand how consumers pick or eliminate a product. This understanding is paramount in business marketing and growth (Sundareswaran, Kamaraj, Elangovan, & Kruthikkha, 2022). Consumer behaviour analysis focuses on Psychological Factors, Demographic Factors and Social Factors.

During the 21st century, the Indian industries have undergone radical changes not only in the field of technology but also in efficiently managing the consumers. Analytics has introduced many innovative tools to manage consumers in a prolific way. It has helped unleash the true power of the analyzing the consumer behaviour and thus has taken businesses to the next level. Analytics aids the contemporary ways of consumer behaviour analysis and thus contributes to increasing consumer satisfaction levels by making it quantifiable. Analytics technologies can aggregate and analyze vast amounts of data on consumers, drawing relationships between demographics, buying habits and targeted marketing techniques.

Analytics services in delivering crucial information to ones responsible for strategy and operations so they can confidently engage in decision-making. Assists interpret consumer behaviour, data patterns and trends and shares the insights that lead to business decisions that support in increasing consumer satisfaction and deciding on the effective marketing strategies.

Exhibits the status of your consumers, what is happening and what needs to happen next. Analytics effectuates predictive thinking and hence makes the marketing strategies ambitious and relevant.

Predictive analytics is generating vision through insights which are superior and customized. The primary objective of predictive analytics in consumer behaviour analysis, functions in every aspect of the effective decision making and thus predict various aspects of consumer behaviour.

This research is focused on analysing and predicting consumer behaviour by using unsupervised algorithm followed by supervised methods for prediction.

2. Related Study

Consumer behaviour analysis and behaviour prediction has various aspects, which require methods to get insights into the patterns that assist the businesses in the identification of marketing strategies as per the different customer segments identified. Customer data integrated from various sources gives insights on what they view and review, which shows their interests and preferences. Predictive mining algorithms are then applied to identify behaviours in the future. Research has shown that consumer behaviour is difficult to predict, even for experts in the field however, the predictive machine learning algorithms have given a different dimension in identifying how consumers make decisions. The study of consumer behaviour is diverse and varied (Levy S, 2005). It enables CRM decisions making and reinforces progress of CRM strategies in a consumer-centric economy (Mirzaei T& Iyer L, 2014).

Surendro, K in their research proposed predictive analytics to predict customer behavior by using behavior informatics and analytics approach to generate deeper insights into customer behavior which has improved business decision making (Surendro K. 2019). Malter, Maayan S., et al. in their research proposed recommendations on how to use the study of consumer consumption to generate thought provoking consumer behaviour research questions. They made predictions on consumer segmentation twenty years down the line through polls on scholars (Malter, Maayan S., et al,2020). Predicting customer behaviour to identify and implement correct marketing strategies is an uncertain and tough task. For these predictions to be reliable Orogun & Onyekwelu used association rule mining on online retail store data. They could derive significant trends to predict customer behaviour (Orogun, A & Onyekwelu, 2019). Churn Prediction Model is a predictive model that figures the likelihood of losing the customer. Businesses can retain customers by using churn modelling. There are different types of churns used in the ecosystem Velu, A. (2021). Churn Analytics helps in maintaining low churn rate and thus helps increasing profits. Rohan B & Satyajee used the Apriori algorithm and built a recommendation system, to recommend products based on similarity in purchase patterns (Rohan Bali, Satyajee Srivastava, 2020). Choudhari, A & Potey, M. have implemented a hybrid model using hybrid Fuzzy unordered rule induction algorithm (FURIA) with Fuzzy C-Means Clustering for customer churn analysis (Choudhari, A. S., & Potey, M.,2018).

Akalya Devi, Ghanasiyaa, Harshini have taken consumer segmentation to the next level by analysing customer sentiments by building a product recommendation system (Sundareswaran G et.al, 2022)

3. Methodology

The unsupervised K-means clustering algorithm is an effective data classification technique used for classification of data. The k-means algorithm partitions data into k-clusters based on the values of the features used for classification. On the clustered data supervised random forest algorithm is applied to predict the clusters.

3.1 Unsupervised Algorithm for Data Classification

K-Means algorithm proposed by J.B.MacQueen is based similarity within the cluster and dissimilarity between the clusters (Li & Wu, 2012). This method forms k optimal clusters based on some criterion. It selects k points as cluster centres and the remaining datapoints will fall in either of the k clusters depending on their distance from the k - cluster centres. The distance between the cluster centres and datapoints is found by using the distance measures like the Euclidean distance.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Where,

- (x_1, y_1) are the coordinates of the cluster centre.
- (x_2, y_2) are the coordinates of the sample datapoint.
- d is the distance between the cluster centre and the sample datapoint.

The value of k that is the number of clusters is decided based on the elbow method. This method gives a plot of number of clusters to the WCSS (within cluster sum square) and choosing the elbow as the number of clusters to use.

Silhouette score is the method of interpretation and validation of consistency within clusters of data. The technique provides a graphical representation of how well each object has been classified. The silhouette value is a measure of how similar an object is to its own cluster(cohesion) compared to other clusters(separation). The value ranges from -1 to +1. A high value indicates that the object is well matched on its own cluster and poorly matched to neighbouring clusters. Negative value of silhouette indicates that the elements are placed in the wrong cluster.

3.2 Random Forest Method for Prediction

Random forest is a supervised learning algorithm which is used for both classification as well as regression. This algorithm creates a robust forest of trees. From the predictions given by each of the trees the algorithm selects the best solution by voting method. It is an ensemble method which reduces the over-fitting by averaging the result.

4. Results And Discussion

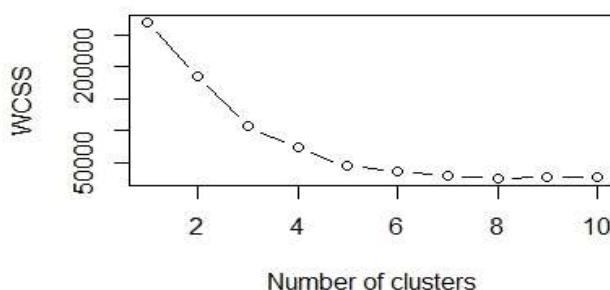
Customer dataset for prediction was taken from Kaggle datasets which has 1000 observations. Data Pre-processing is applied to remove all inconsistencies and noise from the dataset. The dataset is divided where, 70 % of the dataset is considered for training the model and 30% to test the model. The customer dataset is shown in Figure 1.

Figure 1 Customer Dataset

Emp -id	Gender	Age	Annual Income(K)	Spending Score	Average order Value	Source
1	Male	19	15	39	1.73	Online
2	Male	21	15	81	1.40	Online
3	Female	20	16	6	0.93	offline
4	Female	23	16	77	1.61	Online
5	Female	31	17	40	0.55	Online

Figure 2 Optimal Number of Clusters

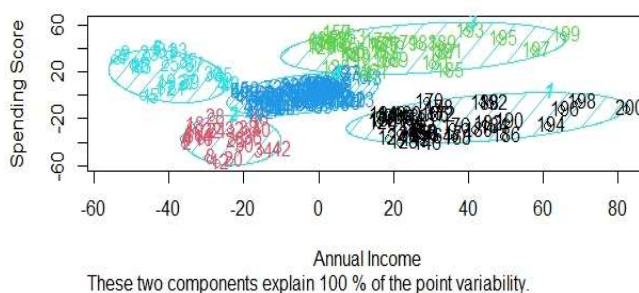
The Elbow Method



The elbow curve in Figure 1 provides important knowledge on the optimum number of clusters. This graph shows that optimum number of clusters would be 4 or 5. Thus, with the optimum number of clusters shown by elbow method, the clusters will be distinctive i.e. there will be maximum similarity within the cluster and dissimilarity between the clusters.

Figure 1:K- Clusters

K-Clusters



The 5 clusters shown in Figure 2 communicates the spending behaviour of customers with different income levels. Cluster 1 has customers with low income and low spending, cluster 5 shows customers with high and high spending behaviour etc.

Silhouette scores for different number of clusters are shown in Figure 4. Thus 5 clusters are shows maximum Silhouette score and is thus optimum.

Figure 4: Silhouette Scores for clusters

Number of Clusters	Silhouette scores
4	0.4931963
5	0.5539320
6	0.5397610
7	0.5281944
8	0.4585134

Taking all variables as predictors the cluster number as the target predictive analytics applied on the dataset. Random forest algorithm was used to predict the cluster where the customer falls into. The classes were almost balanced and hence accuracy was used to find the performance of the prediction model. Accuracy in prediction was found to be 84%.

5. Conclusion

With the use of Unsupervised k-means clustering algorithm the customers were segmented based on their income and spending behaviour. Identifying patterns in the data through the unsupervised algorithm will help businesses make valuable and interesting decisions. The impact of Customer segmentation on business depends on its accuracy. The Silhouette score for 5 clusters was found to be considerably high. Based on the important information generated from customer interactions predications are made using the random forest algorithm. The prediction accuracy was found to be 84 % which is significantly high. The prediction results can help device strategies for deeper understanding of purchase decisions of customers.

6. Future Enhancement

For predictive approach for consumer behaviour analysis various other approaches apart like churn bagging and boosting can be used. Accuracy scores of different methods can be compared to choose the one with best accuracy scores.

References

1. Levy, S. J. (2005). The evolution of qualitative research in consumer behavior. *Journal of Business Research*, 58(3), 341-347.
2. Mirzaei, T., & Iyer, L. (2014). Application of predictive analytics in customer relationship management: A literature review and classification.

3. Sundareswaran, G., Kamaraj, H., Sanjay, S., Devi, A., Elangovan, P., & Kruthikkha, P. (2022). Consumer Behavior Analysis. *International Journal of Research and Applied Technology (INJURATECH)*, 2(1), 82-90.
4. Choudhari, A. S., & Potey, M. (2018, August). Predictive to prescriptive analysis for customer churn in telecom industry using hybrid data mining techniques. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE.
5. Surendro, K. (2019, March). Predictive analytics for predicting customer behavior. In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)* (pp. 230-233). IEEE.
6. Malter, M. S., Holbrook, M. B., Kahn, B. E., Parker, J. R., & Lehmann, D. R. (2020). The past, present, and future of consumer research. *Marketing Letters*, 31(2), 137-149.
7. Velu, A. (2021). Customer Churn Management Using Predictive Modeling–A Machine Learning Approach. *Journal of Emerging Technologies and Innovative Research*, 8(4).
8. Sundareswaran, G., Kamaraj, H., Sanjay, S., Devi, A., Elangovan, P., & Kruthikkha, P. (2022). Consumer Behavior Analysis. *International Journal of Research and Applied Technology (INJURATECH)*, 2(1), 82-90.
9. Rohan Bali, Satyajee Srivastava(2020). Understanding Customer Behaviour with Machine Learning. *International Journal of Recent Technology and Engineering (IJRTE)*, ISSN: 2277-3878 (Online), Volume-8 Issue-6
10. Orogun, A., & Onyekwelu, B. (2019). Predicting consumer behaviour in digital market: a machine learning approach. *International Journal of Innovative Research in Science, Engineering and Technology*, Vol. 8, Issue 8.