SHODH SAMAGAM

ISSN : 2581-6918 (Online), 2582-1792 (PRINT)

Customer Behavior Analysis and Prediction using Machine Learning

Sandeep Rajani, Research Scholar, Department of Computer Science
Birendra Goswami, (Ph.D.) Department of Computer Science
Sai Nath University, Ranchi, Jharkhand, INDIA

ORIGINAL ARTICLE


Authors
Sandeep Rajani
Birendra Goswami (Ph.D.)

shodhsamagam1@gmail.com

Received on : 09/06/2023
Revised on : -----
Accepted on : 16/06/2023
Plagiarism : 06% on 09/06/2023

ABSTRACT

Modern consumer behaviour models are frequently built on data mining of customer data, and each model is created to provide a specific response to a specific inquiry at a specific moment. It can be challenging and uncertain to predict client behaviour. Thus, choosing the appropriate method and strategy is essential when creating consumer behaviour models. It is challenging to alter a prediction model once it has been created for marketing reasons in order to decide precisely what marketing actions to take for each customer or group of customers. Despite this formulation's complexity, the majority of customer models are actually quite straight forward. Due to this necessity, the majority of consumer behavior models overlook so many important variables that the forecasts they produce are typically not particularly accurate. Using a typical online retail store as a data source, this article seeks to construct an association rule mining model to forecast consumer behavior and identify significant trends from the customer behavior data.

KEY WORDS

Association Rule Mining, Apriori, Digital Market, Consumer Behavior, Machine Learning.

INTRODUCTION

Digital marketing which includes mobile phones display advertising and any other digital medium (Parsons et al., 1998) and Jerry et al., 2002), is the practise of promoting goods or services through digital technologies, mostly on the Internet. Most commonly, this phrase refers to data-driven marketing ways brands and companies use technology for marketing that has evolved digital marketing.

Digital marketing applications are becoming more common and effective as marketing strategies and daily life embrace more digital platforms and as individuals utilise digital gadgets rather than go to physical stores (Yasmin et al. 2015). 2019 (ige et al).

Digital marketing strategies like campaign marketing influencer marketing, content automation, search engine optimisation (SEO), and search engine marketing (SEM), among others, have been extensively studied in literature. Through their smartphones, tablets, game consoles, and any other applications, services, and channels available on these devices, consumers are always linked to the internet. Big data analytics are utilized by retail banks to prevent fraud.

Big data analytics is the act of looking at enormous and diverse data sets (big data) to find hidden patterns, undiscovered correlations, market trends, customer preferences, and other valuable information that can assist organisations in making more knowledgeable business decisions. Big data, according to Chen et al. (2014), is the continuously growing data flood in terms of volume, variety, velocity, and complexity that is produced in the current digital eco-system. Customers' online purchasing patterns, website clicks, social media activity logs, smart linked gadgets, geo-location features, etc. are all used to create big data sets about them. Advanced big data analytics solutions offer fresh methods for tackling some of the most important marketing requirements and producing outstanding results. (2000) Sagiroglu et al. These technologies have the power to modernise conventional marketing roles and enhance the way crucial marketing operations are carried out. To provide a holistic picture of each client's activity, marketers are compiling the data generated from a range of live consumer touchpoints. Marketers may optimise customer segmentation models and use the information to create customer engagement strategies and raise customer value by analysing this massive quantity of solutions have the ability to significantly impact marketing in the customer management space Marketing analytics is the act of acquiring and analysing data about a certain market to help make decisions about how to allocate resources to maximise return on investment. Three factors are considered while evaluating the market: the type of customer, the products they are buying, and how their purchasing patterns change over time (2007 Hauser).

In order to inform or prepare data for further analysis descriptive analytics aims to present a portrayal or summary view of facts and figures in an understandable format EMC (Education Service, 2015). Data aggregation and data mining are the two basic methods used to retell historical events. It presents historical facts in a straightforward understandable manner for the advantage of a sizable business audience. Thanks to descriptive analytics, data is presented ina way that can be understood by a variety of business readers. A variety of business readers can easily understand data when it is presented and described using descriptive analytics.

As data driven businesses continue to use the results of descriptive analytics to enhance their supply chains and decision-making capabilities, data analytics will become more distinct from predictive analytics. Predictive analytics will become more prevalent or rather a mashup of hypotheses models, and optimisation will become more prevalent in data analytics consumer behaviour modelling examines consumer group behaviours in order to predict how comparable customers will act in similar situations.

The goal of this study is to use modelling to overcome the difficulty of predicting customer behaviour across digital platforms The practise of developing a mathematical construct to reflect the typical behaviours exhibited in particular customer groups in order to forecast how comparable customers would behave in similar circumstances is known as customer behaviour modelling. Customer behaviour models frequently start with customer data mining, and each model is developed to address a specific issue at a certain point in time. (2009) Tadajewski; (1985) Sheth.

2013 (Fullerton) The massive volume of data must be handled and examined in order to extract knowledge that supports cost reduction and decision-making.

In order to uncover hidden knowledge or patterns in processed data, data mining offers a range of tools and procedures. Personnel can use this secret information to help them make delicate judgments. The main objective of data mining is to take information from a data set and organise it so that it may be used later (Fayyad et al., 1996). For a phrase that is stated as:

$$y = f(x) \quad \text{-----------------------} \quad (eqn\ 1)$$

Where y (observed) is the dependent variable and x is a set of independent variables Data mining offers details about the type of customer qualities or behaviour can be thought of as the parameter x, while trends between attributes are represented by the parameter y. Data mining reveals how x and y are related.

Important data mining fields include healthcare, customer relationship management, and fraud or anomaly detection. (2005) Koh et al.

Association rule mining (ARM), a fundamental technique that extensively explores through big datasets for hidden patterns, is one of the typical data mining tasks. A pair (X, Y) of sets of characteristics is referred to as an association rule (AR), where X is the antecedent and Y is the consequent of the rule X Y. According to the rule if X occurs, Y will follow (Hanetal, 2007). Association rule mining is a technique used in operating databases to find motivational linkages, recurring patterns, and connections between groups of items. They provide risk factor interactions and are interpretable. In order to handle the problem of forecasting consumer behaviour, this study investigates association rule mining utilising the Apriori algorithm. Based on the client behaviour data, it will be done to assess the way well these algorithm operate.

Literature Survey

A paper titled Task Completion Approach for Modelling Purchase Behaviour at an ECommerce Website was published by Sismeiro et al in 2004 The goal of this study is to develop and estimate a model of online purchase using click stream information from a website that sells cars. The model links user activities and exposure to material while they are on the webpage to their likelihood of making an online purchase by using a Bayesian methodology.

Sequence mining for predicting customer behaviour in telecommunications was introduced in Eichinger et al. (2006). The difficult task of forecasting client behaviour, which is crucial for service-oriented firms is what inspired the article The suggested sequence mining method which permits taking into consideration both temporal trends and historical data Sequence mining is used with decision tree analysis to create a composite classifier In the field of sequence mining a hashing techniqueextended tree data structure and a new version of a traditional algorithm are both introduced The authors of this study develop a tree data structure and method for sequence mining. Hair (2007) presented the topic of knowledge production in marketing and the function of predictive analytics. The study included an overview of predictive analytics summarised its effects on marketing knowledge production and made recommendations for the field s future advances for both businesses and scholars. The ability to transform information into knowledge is essential for survival in an economy built on knowledge.

Advanced Analytics

Potential and challenges was studied by Bose (2009). Data analyses based on advanced analytics give businesses a detailed picture of their operations and clients In the context of predictive or advanced analytic, the article examined the usage of these thre mining technologies (data, text, an we) as well as the chal lenges associated with their efficient implementation and maintenance. To examine their current condition, problems, and challenges learned through their practises, a variety of recently published research literature on business intelligence (BI), predictive analytics, and data, tax, and web mining is studied.

Data mining techniques were used to analyse customer behaviour in Gupta et al. (2012). Detailed information on clients, past purchases, and projections of future purchases are now necessary in order to achieve customer satisfaction, which can no longer be attained with a basic listing of marketing contacts Using data mining techniques, this article gives a business and technological overview of the field and explains how to maximise client profitability. Data mining may strengthen and redefine client relationships when used in conjunction with sound business procedures and auxiliary technology.

Data mining methods were applied in Nejad et al. (2012) to boost the effectiveness of the customer relationship management (CRM) procedure. This study demonstrates that by combining CRM and data mining approaches, it is possible to increase CRM effectiveness and provide a quick and efficient response to client requests. Thus, the writers of this study examine key CRM and data mining ideas in order to do this. The authors think that data mining methods can boost CRM's effectiveness. Organisations can identify the customer's data patterns by using data mining. Consequently, it helps business owners to more effectively provide their services and goods.

A study on social media and digital marketing was conducted in 2014 by Tiago et al. In light of digital marketing, a sizeable amount of the related research is currently more customer-focused than firm-focused. In order to better comprehend digital marketing and social media usage, as well as their advantages and disadvantages, this study adopts the perspective of the firm. This study adopts the perspective of the firm to make it easier to grasp social media usage, digital marketing, and their uses, as well as their advantages and disadvantages, in order to address the flaw of the firm's over-reliance on customer behaviour. The difficulties and options for marketing in the digital age were discussed in Leeflang et al. The enormous benefits that digital marketing offers have received a lot of attention, but the actual difficulties that businesses are having while turning digital have received little attention. Based on the findings of a survey among a convenience sample of 777 marketing executives worldwide, the authors of this paper highlight these difficulties.

Analytics and data mining for consumer behaviour were presented by Haastrup et al. in 2014. Accurate profiles are being created by identifying requirements and interests through the monitoring of client behaviour, enabling businesses to provide customers with what they want, when they want it, improving customer happiness and keeping them coming back for more. Data mining provides the connection between transactional and analytical systems, which have been developing independently in largescale information technology. The apriori method, rule induction techniques, and association rule mining are all compared in this market base analysis research.

Methodology

For this article, network data that describes consumer activity when making purchases from online retailers was gathered. Invoice number, stock code, item description, quantity, invoice date, unit price, customer ID, and country of purchase are a few examples, but there are many more. They were examined after being obtained from the UCI repository. The online retail store dataset has roughly 500,000 rows and eight (8) attributes. The input variables for the suggested model are these characteristics of client behaviour. The method involves mining client behaviour and purchasing rules using association analysis. The Apriori algorithm is used to put the technique into practise. The architecture frame work for association rule based consumer behaviour prediction is shown in Figure 1.

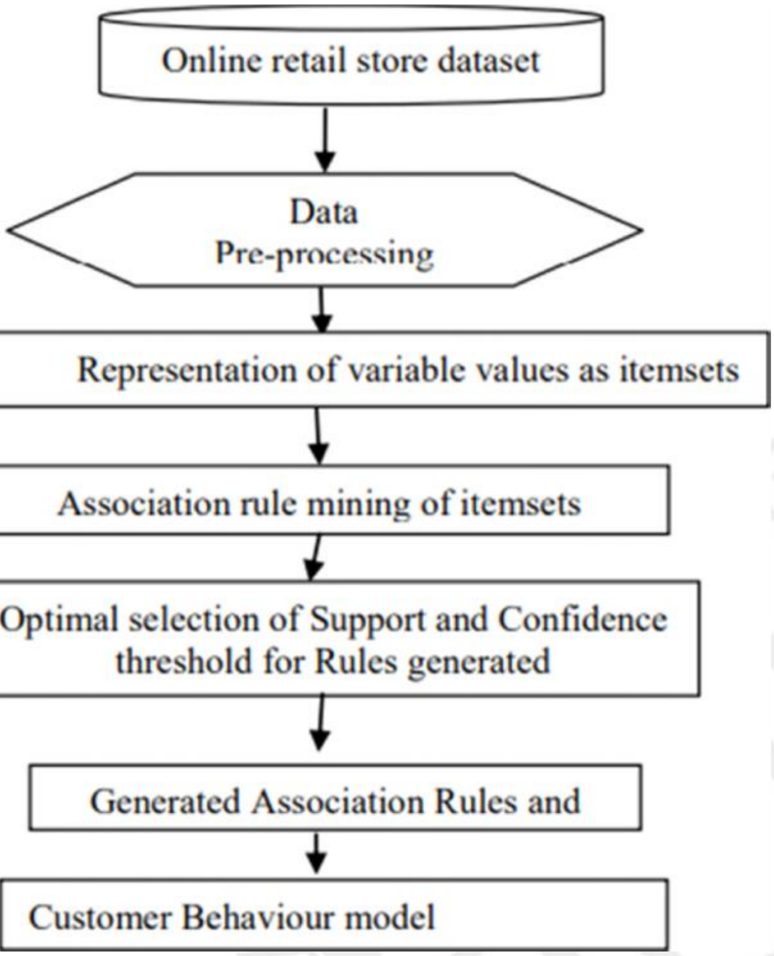

Figure 1: Customer behavior prediction model using rule mining approach architecture

Dataset Description

All online customer transactions for a globally based and officially registered nonstore online retailer that took place between December 1, 2010, and December 9, 2011, are included in the dataset, which is a structured transnational data set. The company primarily offers one-of-a-kind gifts for every occasion. The company has a large number of wholesalers as clients. The customer behaviour data from online retail purchases includes eight variables with both continual and symbolic features. The invoice number, the first attribute, has a nominal value and is a 6 digit integral number that is specifically assigned to each transaction. This code denotes a cancellation if it begins with the letter “c.” The stock code, which defines the product (item) code, is the second attribute. It holds a nominal value, a 5-digit integral number uniquely assigned to each distinct product. The third gives the description of the product (item) name, while the fourth is the quantity of each item purchased per transaction. The fifth attribute is the invoice date and time of each transaction. Unit price, which is the product price per unit, is the sixth attribute. The seventh attribute is the customer ID, or customer number, a 5-digit integral number uniquely assigned to each customer. The last attribute holds the name of the country where the customer resides. Table 1 shows the different behavioral features.

Table 1: Online retail store transaction dataset description

S/N	Name of features	Description
1	InvoiceNo	a 6-digit integral number uniquely assigned to each transaction
2	StockCode	a 5-digit integral number uniquely assigned to each distinct product
3	Description	Product (item) name
4	Quantity	The quantities of each product (item) per transaction
5	InvoiceDate	The day and time when each transaction was generated
6	UnitPrice	Product price per unit currency
7	CustomerID	a 5-digit integral number uniquely assigned to each customer
8	Country	the name of the country where each customer performs a transaction

Eight nominal and three numeric data types are present in the collection. For effective analysis, all data fields are transformed to a standard format, which is numerical. In the preprocessing stage, feature selection is a crucial step. The filter approach is employed for feature selection. After the variables have been uniformly represented as numerical values; a feature selection method will be applied to the input variables to determine how important they are to the output. The feature selection technique to be used is the correlation coefficient analysis as shown in equation below.

$$r = \frac{n\left(\sum V_i B\right) - \left(\sum V_i\right) * \left(\sum B\right)}{\sqrt{\left[n\sum V_i^2 - \left(\sum V_i\right)^2\right] * \left[n\sum B^2 - \left(\sum B\right)^2\right]}}$$

Where r = Pearson correlation coefficient, V_i is variable (attribute) state value, B is target behavior of the customer, n is total number of transactions (data points) in the data.

Apriori Algorithm

Over the database records, the Apriori algorithm performs frequent item set mining and association rule learning. To execute association rule mining, use this. Its job is to identify frequently occurring groupings of attribute traits that frequently occur toget. The programme looks for common item sets in the network dataset. The listing illustrates the fundamental methodology for utilising the algorithm to locate common item groupings. The sets that fulfil the minimum support level will subsequently be identified by scanning the transaction data set. Sets that fall short of the required degree of support will be discarded. The final step is to merge the remaining sets to create item sets with two elements. Steps for scanning the dataset are described as:

Listing 1: Apriori algorithm

Step 1: While the number of customer attribute items in the set is greater than 0:

Step 2: Create a list of customer attribute item sets of length k

Step 3: Scan the dataset to see if each customer attribute item set is frequent

Step 4: Keep frequent attribute item sets to create item sets of length k+1

Following is a detailed explanation of algorithm 1's iterative step 3 (scan dataset): Each transaction in the dataset is as follows: Determine whether Pat is a subset of Tran for each set of customer attribute item sets. If so, raise the count for every set of customer attribute item. Keep this item if the support is sufficient. Give me a list of commonly occurring attribute item sets.

Results and Discussion

In a Python programming environment, the model is implemented. Python has a straight forward syntax and is simple to implement for big data analysis, which requires massive datasets, making it a useful language for data mining applications. Python also offers an interactive shell that enables the viewing and inspection of elements during implementation, which is another justification for adopting this technology. Based on execution time and automatically produced association rules, the model is evaluated.

Performance Analysis

Due to their extensive applicability in the majority of related literature, three evaluation criteria were chosen for this study. They consist of the number of regularly occurring item sets, the number of rules generated, and the execution time (in seconds). The graph of the frequency of items against the minimal support is shown in Figure 2.

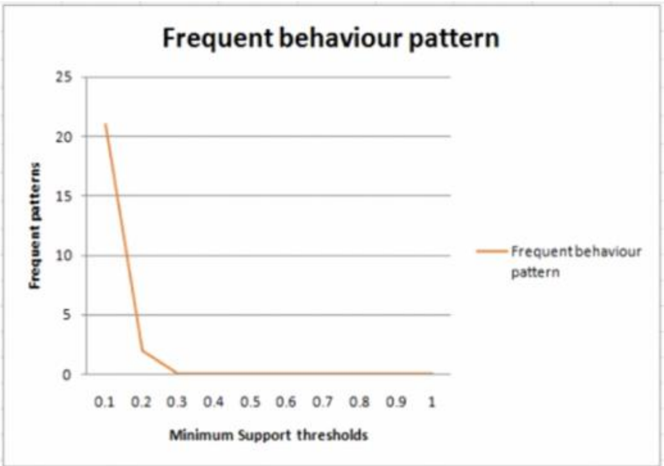

Fig. 2: Plot of number of frequent patterns generated based on minimum support levels

REFERENCES

1. Bose, R. (2009). Advanced analytics: Opportunities and challenges. *Industrial Management & Data Systems* 109(2):155-172.DOI: 10.1108/02635570910930073

2. Chen, M., Mao, S. & Liu, Y. (2014). Big Data: A review. Mobile NetwAppl, Mobile Networks and Applications, *Springer*-Verlag New York, Inc. Secaucus, NJ, USA. Volume 19 Issue 2, Pages 171-209 DOI 10.1007/s11036-013-0489-0.

3. Eichinger, F., Nauck, D.D., &Klawonn, F. (2006). Sequence Mining for Customer Behaviour Predictions in Telecommunications.for marketing in a digital era. *European Management Journal*, 32(1),pp 112. https://doi.org/10.1016/j.emj.2013.12.001

4. EMC Education Service (2015). *Data Science Big Data Analytics: Discovering, Analyzing,* Visualizing and Presenting Data Published by John Wiley Sons, Inc. ISBN: 978-1-118-87613-8,January 2015, 432 Pages

5. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) *From Data Mining to Knowledge Discovery: An Overview*. In: Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, Cambridge.Pp 37-5

6. Fullerton, R.A. (2013). The Birth of Consumer Behavior: Motivation Research in the 1950s, In *Journal of Historical Research in Marketing*, Vol. 5, No. 2, 2013, pp.212-222

7. Gupta, G. & Aggarwal, H. (2012). Improving Customer Relationship Management Using Data Mining.

International Journal of Machine Learning and Computing, Vol. 2, No. 6, December 2012. Pp. 874-878

8. Haastrup, A. V., Oladosu, O. A., Okikiola, F. M., Oladiboye, O. E., &Ishola, P. E. (2014). Customer behaviour analytics and data mining. *American Journal of Computation*, Communication and Control 2014; 1(4): 66-74
9. Hair Jr, J. F. (2007). Knowledge Creation in Marketing: The Role of Predictive Analytics. *European Business Review* 19(4):303-315. DOI: 10.1108/09555340710760134
10. Han, J., Cheng, H., Xin, D. & Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining Knowledge Disc* (2007) 15:55–86 DOI10.1007/s10618-006-0059-1.
11. Hauser, W. J. (2007) "Marketing analytics: the evolution of marketing research in the twenty first century", *Direct Marketing: An International Journal*, Vol. 1 Issue: 1, pp.3854, https://doi.org/10.1108/17505930710734125 .
12. Ige A.O, Akingbesote A.O, Orogun A.O. (2019). Trust appraisal in a conventional e-market environment: A review. *Canadian Open Information Science and Internet Technology Journal*. Vol. 1, No. 1, April 2019, pp. 1 -9
13. Jagdish N. Sheth and Chin Tiong Tan, Singapore: Association for Consumer Research, Pages: 5-7.
14. Jerry, W. &Vijan, M.(2002). Digital Marketing, *Symphonya*, Emerging Issues in Management, n. 1,2002,,http://symphonya.unimib.it/article/view/2002.1.04wind.mahajan. pp. 43-54.
15. Kiani, G. R. (1998),"Marketing opportunities in the digital world", *Internet Research*, Vol. 8 Issue 2 pp. 185 – 194
16. Koh, H. C. & Tan, G. (2005). Data Mining Applications in Healthcare. *Journal of Healthcare Information Management,* Vol. 19, No. 2. Pp 64-72.
17. Leeflang, P. S., Verhoef, P. C., Dahlström, P., &Freundt, T. (2014). Challenges and solutions for marketing in a digital era, *European Management Journal* 32 (2014) 1–12
18. Nejad, M. B., Nejad, E. B. &Karami, A. (2012). Using Data Mining Techniques to Increase Efficiency of Customer Relationship Management Process. Research, In *Journal of Applied Sciences*, Engineering and Technology 4(23): 5010-5015,2012 ISSN: 2040-7467
19. Parsons, A., Zeisser, M. &Waitman, R. (1998). Organizing today for the digital market tomorrow. *Journal of interactive marketing*, volume 12 / number 1 / winter 1998

****