ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

Using Data Mining to Improve Consumer Retailer Connectivity

Vishakha Prashant Ratnaparkhi, Sahil Siddharth Jambhulkar

Research Scholar, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India

Research Scholar, Department of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune, India

ABSTRACT: Online shopping is still in its nascent stage in India but growing at a fast pace. To continue its growth it is significant to understand the user's preferences. Analysis of consumer's behavior with respect to online shopping consists of detailed information about consumers past purchases as well as prediction of future purchases. In recent times, customer behaviour models are typically based on data mining of customer data, andeach model is designed to answer one question at one point in time. Predicting customer behaviour is an uncertain a difficult task. Thus, developing customer behaviour models requires the right technique and approach. Once aprediction model has been built, it is difficult to manipulate it for the purposes of the marketer, so as to determineexactly what marketing actions to take for each customer or group of customers. Despite the complexity of thisformulation, most customer models are actually relatively simple. Because of this necessity, most customer behaviour models ignore so many pertinent factors that the predictions they generate are generally not very reliable. This paperaims to develop an association rule mining model to predict customer behaviour using a typical online retail store for
Data collection and extract important trends from the customer behaviour data. Many consumers prefer online shopping. Day-to-day busy schedule made many consumers to visit online e-commerce websites for shopping. This saves time and cost of the consumer. With the growth of the e-commerce websites retailers tend to fail to attract more and more consumers. Consumers no longer feel difference between e-shopping and offline shopping. Researcher proposed a system of connecting the consumer and the retailer. This system creates a bridge between consumer and retailer. Researcher had implemented an effective data mining algorithm to analyze new patterns and trends. This system will gather data from the customer behavior pattern and is supplied to the retailers, so that retailers will able to know the new patterns and trends. With these information retailer can approach targeted customer and can constantly interact with those consumers that retailer is exactly looking for. This system helps retailer to keep constant connectivity among the retailers and the consumers. In this system Researcher had used data mining algorithm that helps the retailer to discover new patterns and trends. The system updates the retailers with new trends and patterns This system helps to improve the sales and business of the retailer. System will also help the retailer to know about the updated price of the product as well as new trend in market.

KEYWORDS: E-commerce, M-commerce, Data mining, Consumer Behavior, and Personal perceived values, Website quality, Machine Learning

I. INTRODUCTION

E-commerce "basically stands for electronic commerce which relates to a website that sells products or services directly from the site with the help of a shopping cart or shopping basket system and payments can be done through cards, e-banking and cash on delivery. It helps customers to buy anything form a pen to an insurance policy from the comfort of their home or office and gift it to someone sitting miles apart just by click of the mouse. It offers various benefits to businesses for example, easy reach to fast growing online community, providing unlimited shelf place for products and services, merging global markets at low operating costs. Ease to access internet is the major factor in rapid

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

adoption to E-commerce. For popularization of E-commerce in India the main essential factors are safe and secure payment modes. Even though there are various benefits in shopping online but just like every coin has two sides, there exist various reasons for not shopping online for example lack of trust, security concerns, uncertainty about the product and service quality, delay or non-delivery of goods, and lack of touch-and-feel shopping experience." Mobile Commerce (M-commerce) is the subset of electronic-commerce, which includes all e-commerce transactions carried out using a mobile device. Basically, M-commerce is the way of doing business in a state of motion. "M-commerce depends on the availability of mobile connectivity. M-commerce offers multiple advantages like ubiquity, personalization, flexibility, and distribution, instant connectivity, immediacy. There are many ways in which businesses, government and people benefit from m-commerce like-

1 Selling a product or service which is information based (delivery directly to mobile devices) or location based

2 Improving productivity by gathering time critical information (reports, photographs) and SMS based up-to-date information.

3 The ability to access information on mobile, at affordable cost can change people's lives and livelihoods in rural areas (Latest on the weather report or health services). It can be used as the medium to educate and create awareness among the rural people.

4 Usages of Internet on mobile devices have lead to information access overcoming geographical barriers and removed the training cost of mobile technology.

1.1 Background

Digital marketing is the marketing of products or services using digital technologies, mainly on the Internet, whichincludes mobile phones, display advertising, and any other digital medium(Parsons et al., 1998), (Jerry et al.2002). Thisterm is mostly referred to data-driven marketing. Currently, digital marketing has changed the way brands andbusinesses use technology for marketing. As digital platforms are increasingly incorporated into marketing plans andeveryday life, and as people use digital devices instead of visiting physical shops, digital marketing applications arebecoming more prevalent and efficient (Yasmin et al. 2015), (ige et. Al 2019). Digital marketing techniques such assearch engine optimization (SEO), search enginemarketing (SEM), influencer marketing, content automation, campaign marketing, etc have been greatly researched on in literature. Consumers are constantly connected digitally allthe time, through their smart devices, tablets, gaming consoles and every application, service and channel accessiblethrough these devices. Retail banks use big data analytics for fraud prevention. Big data analytics is the process ofexamining large and varied data sets (big data) to uncover hidden patterns, unknown correlations, market trends,customer preferences and other useful information that can help organizations make more-informed business decisions.Chen et al. (2014) refer to big data as the ever-increasing data deluge in terms of volume, variety, velocity andcomplexity that is being generated in today's digital eco-system.

Big data sets are generated around customers based ontheir online purchase behaviour, website clicks, social media activities log, smart connected devices, geo-locationattributes, etc. Sophisticated analytics solutions for big data provide new approaches to addressing some of the keymarketing imperatives and delivering impressive results. (Sagiroglu et al. 1998).These solutions can transform traditional marketing roles and improve how to execute essential marketing functions.Marketers are collecting the data produced from a variety of live customer touch-points to paint a complete picture ofeach customer's behaviour. Analyzing this large amount of data in motion enables marketers to fine-tune customersegmentation models and apply the insights to develop customer engagement strategies and improve the value ofcustomer. Multiple big data applications are showing tremendous potential for driving marketing impact in thecustomer management domain. Marketing analytics involves information gathering and processing of a particularmarket in order to aid decisions on where to spend the budget to gain more value. Three factors come into mind inmarket analytics namely, who is the customer, what are they buying and how the buying changes with time

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

(Hauser2007). Descriptive Analytics seeks to provide a depiction or summary view‖ of facts and figures in an understandable format, to either inform or prepare data for further analysis (EMC Education Service, 2015).
It uses two primarytechniques, namely data aggregation and data mining to report past events. It presents past data in an easily digestibleformat for the benefit of a wide business audience. Descriptive Analytics helps to describe and present data in a formatwhich can be easily understood by a wide variety of business readers. As data-driven businesses continue to use theresults from Descriptive Analytics to optimize their supply chains and enhance their decision-making powers, DataAnalytics will move further away from Predictive Analytics towards Prescriptive Analytics or rather towards a mashupof predictions, simulations, and optimization. Customer behaviour modelling identifies behaviour among groups ofcustomers to predict how similar customers will behave under similar circumstances. This paper is geared towardssolving this modelling problem of predicting customer behaviour across digital platform. Customer BehaviourModelling is the formulation of a mathematical construct to represent the common behaviour observed amongparticular groups of customers in order to predict how similar customers will behave under similar circumstances.

1.2 Motivation

There is an increasing focus on data mining, which has been defined as the application of data analysis and discovery algorithms to large databases with the goal of discovering (predictive) models (Gregory Piatetsky-Shapiro,et al, 1996). Business Week (Berry 1994) estimated that over half of all retailers are using or planning to use database marketing, and those who do use it have good results.Retail stores are massively collecting large volumes of data in their daily business operations and the retail shops in themselves stoke hundreds of thousands of items. One leading retail shop in Uyo had more than 250,000 unique items and posted an average of 7,000 transactions per day. This volume of data was a sea of sitting knowledge that can yield strategic businessintelligence when extracted. Transactions at these stores were routinely captured at the point of sale. Furthermore major retail shops and many more were using Customer loyalty cards largely to retain their clients. These cards equally provided a means of understanding the customers' bio information and buying characteristics. Indeed they provided a means of understanding the value of the customer to the shop and how to reach them.With data mining analysis, retailers can drive more profitable advertisement and promotions (database driven), attract more customers into the stores, increase the size and value of basket purchases, improve loyalty card promotions with longitudinal analysis, test and learn by using a market place as a laboratory, empower planners and vendors to make smarter decisions, march inventory to needs by customizing layouts, assortments and pricing to the local demographics.

1.3ProblemStatement

That most retail shops are sitting on enormous amount of information on their databases is evident with the sprawling of retail shops and the massive queues of customers transacting within these shops. The introduction of customer loyalty cards, the acceptance of usage of visa and credit cards are additional tools that capture customer transaction behavior and demographics. Most managers plan their product placement and replenishment, promote their product and organize advertisement based largely on their experience rather than on the basis of database driven intelligence.

1.4 Scope of the Paper

Retailers always assume risk every time they make decisions around buying, replenishment, advertising, promotions and assortment planning. These decisions need not be based on experience or instincts. A 1 % lift in sales or 0.01% improvement in margin can tip the balance between success, survival or failure. Every retailer's top-line sales and success require constant fine tuning of controls available to the retailer. Most retailers still suffer from the old age retail problem of stocking too much of the wrong item and not enough of the right one. The right product moves and the wrong one sits until it is marked down. Customer's life is further complicated when he or she cannot get enough quantities of a popular product. Data mining therefore will leverage retailers on smarter decision making process.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

1.5 Objective of the Paper

The main goal is to exploit data mining techniques to perform basket analysis with a view to using the knowledge mined to improve on sales and assortment planning.

1. To examine algorithms for association rule mining
2. To extract interesting and useful patterns from a retail shop
3. To develop models based on generated patterns.

II. LITERATURE REVIEW

Sismeiro et al. (2004) presented a paper on Modelling purchase behaviour at an e-commerce web site: A task completionapproach. The aim of this paper is to develop and estimate a model of online buying using click stream datafrom a Web site that sells cars. The model predicts online buying by linking the purchase decision to what visitors doand to what they are exposed while at the site using Bayesian technique.

In Eichinger et al.(2006), Sequence mining forcustomer behaviour predictions in telecommunications was presented. The paper is motivated by the challengingproblem of predicting the behaviour of customers, which is important for service oriented businesses. The proposedsequence mining approach, which allows taking historic data and temporal developments into account as well. In orderto form a combined classifier, sequence mining is combined with decision tree analysis. In the area of sequence mining,a tree data structure is extended with hashing techniques and a variation of a classic algorithm is presented. In this paperthe authors extended a tree data structure and approach for sequence mining. Knowledge creation in marketing: the roleof predictive analytics was presented in Hair (2007).

The paper provided an overview of predictive analytics, summarizehow it is impacting knowledge creation in marketing, and suggest future developments in marketing and predictiveanalytics for both organizations and researchers. Survival in a knowledge-based economy is derived from the ability toconvert information to knowledge.

Bose (2009) did a study on advanced analytics: opportunities and challenges.Advanced analytics-driven data analyses allow enterprises to have a complete or view of their operations and customers.The paper investigated these three (data, text and web) mining technologies in terms of how they are used and the issuesthat are related to their effective implementation and management within the broader context of predictive or advanced analytics. A range of recently published research literature on business intelligence (BI); predictive analytics; and data,text and web mining is reviewed to explore their current state, issues and challenges learned from their practice.

InGupta et al.(2012), analysis of customer behaviour using data mining techniques was carried out. Achieving customersatisfaction is no longer satisfied with a simple listing of marketing contacts, but wants detailed information aboutcustomers, past purchase as well as prediction of future purchases. This paper discusses a business and technologicaloverview of data mining and outline how to optimize Customer profitability through data mining application, along withsound business processes and complement technologies, data mining can reinforce and redefine Customer relationship.

In Nejad, et al.(2012), data mining techniques was used to increase efficiency of customer relationship management(CRM) process. This study shows that it is possible to improve CRM efficiency, to have an effective and rapid responseto customer needs, by integrating CRM and data mining techniques. Thus in achieving this, the authors investigate majorconcepts of CRM and data mining in this study. The authors believe that using data mining techniques in CRM canimprove CRM's efficiency. Using data mining organizations can identify the customer's data patterns. So, it enables thebusiness owners to better offer of their services and products.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

Tiagoet al.(2014) did a study on digital marketing andsocial media. Currently, a significant portion of the associated research is focused more on the customer than on the firmin light of digital marketing. This study adopts the perspective of the firm to facilitate an understanding of digitalmarketing and social media usage as well as its benefits and inhibitors. To redress the shortcoming of over reliance oncustomer behaviour than the firm, this study adopts the perspective of the firm to facilitate an understanding of digitalmarketing and social media usage as well as its benefits and inhibitors.

In Leeflang et al.(2014), Challenges andsolutions for marketing in a digital era was presented. A great deal of attention has been focused on the tremendousopportunities digital marketing presents, with little attention on the real challenges companies are facing going digital. Inthis study, the authors present these challenges based on results of a survey among a convenience sample of 777marketing executives around the globe.

Haastrup et al. (2014) presented Customer Behaviour Analytics and Data Mining.Through analysis of customers' behavior, accurate profiles are being generated by specifying needs and interest andallowing business to give customers what they want it, when they want, leading to a better customer satisfaction therebykeeping them to come back for more. While large-scale information technology has been evolving separate transactionand analytical systems, data mining provides the link between the two. The paper makes a comparative study ofAssociation Rule Mining, Rule Induction Technique and Apriori Algorithm in market base analysis.

III. PROPOSED METHODOLOGY

Network data that describes the behaviour of customers on an online retail store purchases are sourced for this paper.They include but not limited to; invoice number, stock code, item description, quantity, invoice date, unit price, customerID and country of purchase and sourced from UCI repository and studied. The online retail store dataset contains eight (8)attributes and about 500,000 rows. These customer behaviour attributes are the input variables to the proposed model.The approach used involves the use of association analysis in mining customer behaviour purchase rules. The techniqueis implemented using Apriori algorithm. Figure 1 shows the architecture framework for customer behaviour predictionusing association rule based approach.

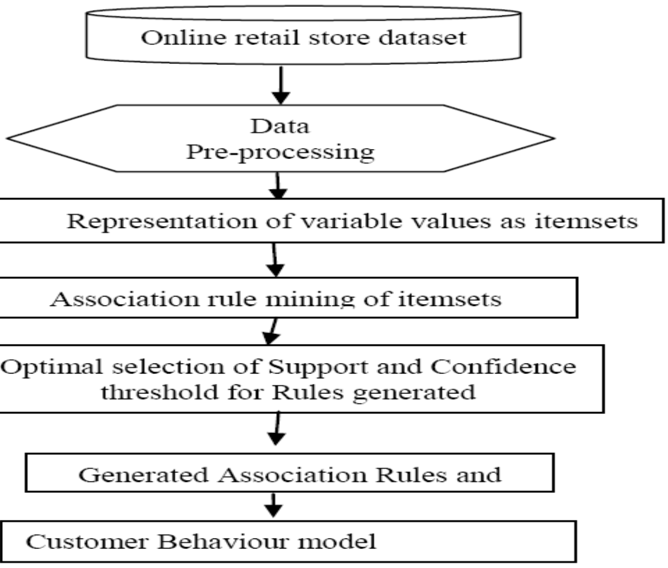

Figure 1: Customer behaviour prediction model using rule mining approach architecture.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

3.1 Advantages of the proposed project:-

1. System helps retailer to keep constant connectivity among the retailers and the consumers.
2. With the help of data mining algorithm, system will display new trends and patterns.
3. System will help to discover new trends and patterns in market.
4. This system helps to improve the sales and business of the retailer.
5. System will also help the retailer to know about the updated price of the product as well as new trend in market.

Disadvantages:

- Users who don't have internet connection can't access the system.
- We have to create e –commerce website where retailer can access the system.

3.2 Application:

This application can be used by many users who prefer shopping.

3.3 Module Working

- **Admin Login:** Admin can access the authorized modules by login to the system using his credentials.
  - Add Products: Admin can add products by entering product details like product image, cost, and description.
  - View User: Admin can view registered user details.
  - View Retailer Details: Admin can retailer details who had registered to the system.
  - View Products: Admin can view products which are added into the database.
  - View Feedback: Admin can view feedback of the user.
- **User Login:** User must register with his details and system will provide him with id and password. He must use this user id and password to login to the system.
  - View Products: User can view the products and their cost.
  - Products Details: User must select the product of his choice and view further details of that product.
  - Add To Cart: User can add products into cart, if he wants to purchase the product.
  - Make payment: System will display total cost. User can make payment by selecting the mode of payment.
  - Add Feedback: Customer can add feedback about any product.
- **Retailer Login:** Retailer can access the system using his credentials.
  - Frequent Items: System will check frequent items purchased by the consumer. System will display those products to retailers.
  - Sales Prediction: System will predicate the sales of the retailer.
  - User Details: Retailer can view user details for contacting the user.

3.4 Dataset Description

The dataset is a structured transnational data set which contains all the online customer transactions occurring between**01/01/2020 and 18/10/2020** for an international based and registered non-store online retail. The company mainly sellsunique all-occasion gifts. Many customers of the company are wholesalers. The online retail purchase data has customerbehaviour data with 8 attributes that have both continuous and symbolic attributes. The first attribute invoice numberholds nominal value, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', itindicates a cancellation. The second attribute is stock code describing the product (item) code. It holds nominal value, a5-digit integral number uniquely assigned to each distinct product. The third gives the description of product (item) namewhile the fourth is the quantity of purchase of each item per transaction. The fifth attribute is the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

invoice date and time ofeach transaction. Unit price: which is the product price per unit price is the sixth attribute. The seventh attribute is theCustomer ID or Customer number, a 5-digit integral number uniquely assigned to each customer. The last attribute holdsthe name of Country where customer resides. Table 1 shows the different behavioral features.

3.5 Data pre-processing

The dataset has 3 numeric and 8 nominal data type. All data fields are converted to uniform format which is numeric forefficient analysis. Feature selection is an important step in the pre-processing stage. The feature selection method used isthe Filter method.After the uniform representation of the variables as numeric values, feature selection process will becarried out on the input variables to investigate significance of the input variables to the output. The feature selectiontechnique to be used is the correlation coefficient analysis as shown in equation below.

$$r = \frac{n\left(\sum V_i B\right) - \left(\sum V_i\right) * \left(\sum B\right)}{\sqrt{\left[n\sum V_i^2 - \left(\sum V_i\right)^2\right] * \left[n\sum B^2 - \left(\sum B\right)^2\right]}} \qquad \ldots\ldots(1)$$

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

IV. ANALYSIS OF ALGORITHMS

Finding interesting and rare patterns in the dataset is based on association analysis. Association analysis is a set of toolsused to find valuable relationships in a large set of data. This analysis is based on Apriori principle which states that if an item set is frequent, then all of its subsets are frequent‖. Association rules suggest that a strong relationship existsbetween two items. An illustration is shown with an example in table extracted from the raw data. The association ruleis a pair (X, Y) of sets of symptoms defined as; The rule states that if the set of customer behaviour feature(s) on the antecedent part (X) occurs, then the behaviourstates on the consequent part (Y) does happen. In general, a set of behaviour feature items, such as X or Y, which aredisjoint, is the behavior feature item set. In table 2,a hypothetical six rows table of 7 features is shown.

$$X \rightarrow Y\text{, where } X, Y \subseteq I \text{ and } X \cap Y = \phi \qquad \ldots\ldots(2)$$

4.1 Apriori algorithm

The Apriori algorithm performs frequent item set mining and association rule learning over the database records. It isused to implement the association rule mining. Its task is to find frequent sets of attribute characteristics that commonlyoccur together. The algorithm scans the network dataset for frequent item sets. The basic approach to finding frequentitem sets using the algorithm is shown in the listing. The transaction data set will then be scanned to see which sets meetthe minimum support level. Sets that don't meet the minimum support level will get tossed out. The remaining sets willthen be combined to make item sets with two elements. Again, the transaction dataset will be scanned and item setsnot meeting the minimum support level will be removed. This procedure will be repeated until all sets are pruned out.Generation of the patients' item sets is carried out by setting a function to create an initial set, and scan the datasetlooking for items that are subsets of transactions. Steps for scanning the dataset are described as;

4.2 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 itemsets of length k
Step 3: Scan the dataset to see if each customer attribute itemset is frequent
Step 4: Keep frequent attribute itemsets to create itemsets of length k+1
The step 3 (Scan dataset) in Algorithm 1 which is an iterative process is further describes as;
For each transaction in tran the dataset:
For each customer attribute itemset, cus:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

Check to see if pat is a subset of tran
If true, increment the count of cus
For each customer attribute itemset:
If the support meets the minimum, keep this item
Return list of frequent attribute itemsets

Listing 2

```
fromnumpy import *
defloadDataSet():
return [[11, 13, 14], [12, 13, 15], [11, 12, 13, 15], [12, 15]]
def createP1(dataSet):
P1 = []
for transaction in dataSet:
for item in transaction:
if not [item] in P1:
P1.append([item])
P1.sort()
return map(frozenset, P1)
defscanD(D, Pk, minSupport):
ssCnt = {}
fortid in D:
for pat in Pk:
ifpat.issubset(tid):
if not ssCnt.has_key(can): ssCnt[pat]=1
else: ssCnt[pat] += 1
numItems = float(len(D))
retList = []
supportData = {}
for key in ssCnt:
support = ssCnt[key]/numItems
if support >= minSupport:
retList.insert(0,key)
supportData[key] = support
returnretList, supportData
```

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

Listing 2 contains three functions: loadDataSet(), which creates a simple dataset for testing, createP1(), and scanD().

The function createP1() creates P1 which is a customer behaviour feature itemset of size one. In the algorithm, P1 iscreated and a scan is carried out in the dataset to see if these one itemsets meet the specified minimum supportrequirements. The itemsets that meet the minimum requirements become L1. Then L1 gets combined to become P2 andP2 will get filtered to become L2.Filtering out the sets is implemented in the following code snippet as shown in listing 3.

Listing 3:

```
defaprioriGen(Lk, k): #creates Pk
retList = []
lenLk = len(Lk)
fori in range(lenLk):
for j in range(i+1, lenLk):
L1 = list(Lk[i])[:k-2]; L2 = list(Lk[j])[:k-2]
L1.sort(); L2.sort()
if L1==L2: #if first k-2 elements are equal
retList.append(Lk[i] | Lk[j]) #set union
returnretList
defapriori(dataSet, minSupport = 0.5):
P1 = createP1(dataSet)
D = map(set, dataSet)
L1, supportData = scanD(D, P1, minSupport)
L = [L1]
k = 2
while (len(L[k-2]) > 0):
Pk = aprioriGen(L[k-2], k)
Lk, supK = scanD(D, Pk, minSupport) #scan DB to get Lk
supportData.update(supK)
L.append(Lk)
k += 1
return L, supportData
```

The code snippet in listing 3 contains two functions:apriori() and apriori(). The main function is apriori() which callsaprioriGen() to create network feature itemsets: Pk. The function aprioriGen() takes a list of frequent itemsets, Lk, andthe size of the itemsets, k, to produce Pk. For example, it will take the itemsets {0}, {1}, {2} and produce {0,1} {0,2},and {1,2}.The apriori() function takes up a dataset and a support number to generate a list of customer behaviour

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

feature itemsets.This works by first creating P1 and then taking the dataset and turning that into D, which is a list of sets. The mapfunction is used to map set() to every item in the dataSet list. Next, the scanD() from listing 4.1 is called to create L1and placed inside a list, L. List L will contain L1, L2, L3,..., Ln. Subsequent lists, L2, L3,..., Ln is fetched using thewhile‖ loop which creates larger lists of larger itemsets until the next-largest itemset is empty.Setting support of 70% (0.7) generates four frozensets after the preceding command is entered in the Python shell.

V. RESULTS AND DISCUSSION

The implementation of the model is done in Python development environment. Python is a great language for datamining applications because it has clear syntax and easy to implement big data analysis requiring large datasets. Anotherrationale for choice of this tool is that Python also has an interactive shell, which allows for viewing and inspection ofelements during implementation. The model evaluation is carried out based on execution time and generated associationrules.

5.1 Performance Analysis

In this research, three evaluation criteria are used due to their widespread relevance in most related literature. Theyinclude number of frequent itemsets, rules generated, and execution time (in seconds). Figure 2 shows the chart ofnumber of frequent itemsets against minimum support.

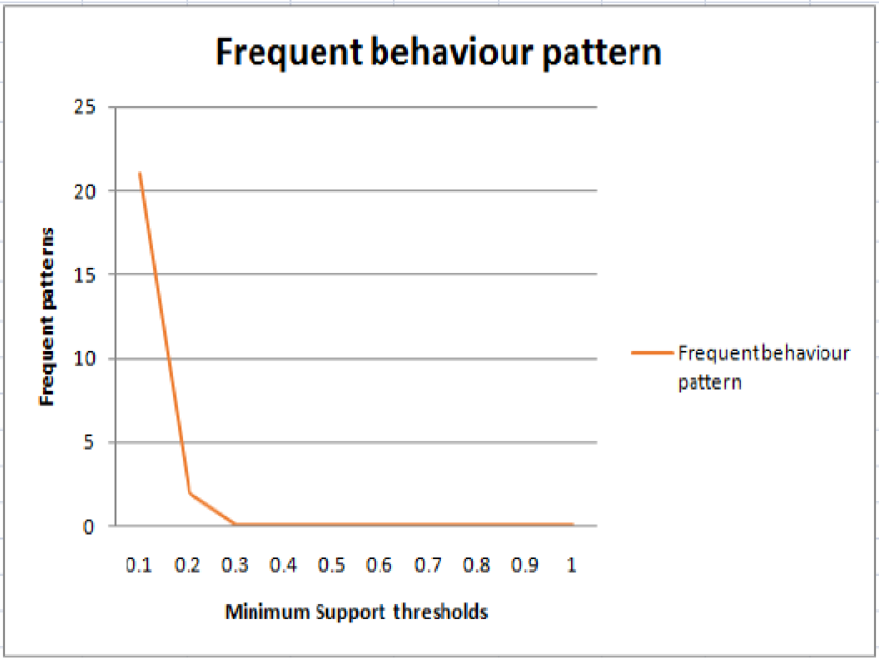

Fig. 2: Plot of number of frequent patterns generated based on minimum support levels

The chart in Figure 2 shows a decrease in total number of frequent itemsets mined as minimum support threshold isincreased. A sharp decline was recorded from a support of 0.1 to 0.2 and gradually declined up to 0.3 threshold. Thecurve almost ran parallel to the horizontal axis from a support of 0.4 to 1. Figure 3 shows a chart of number ofassociation rules mined based on minimum confidence thresholds (from 10% to 100%) respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

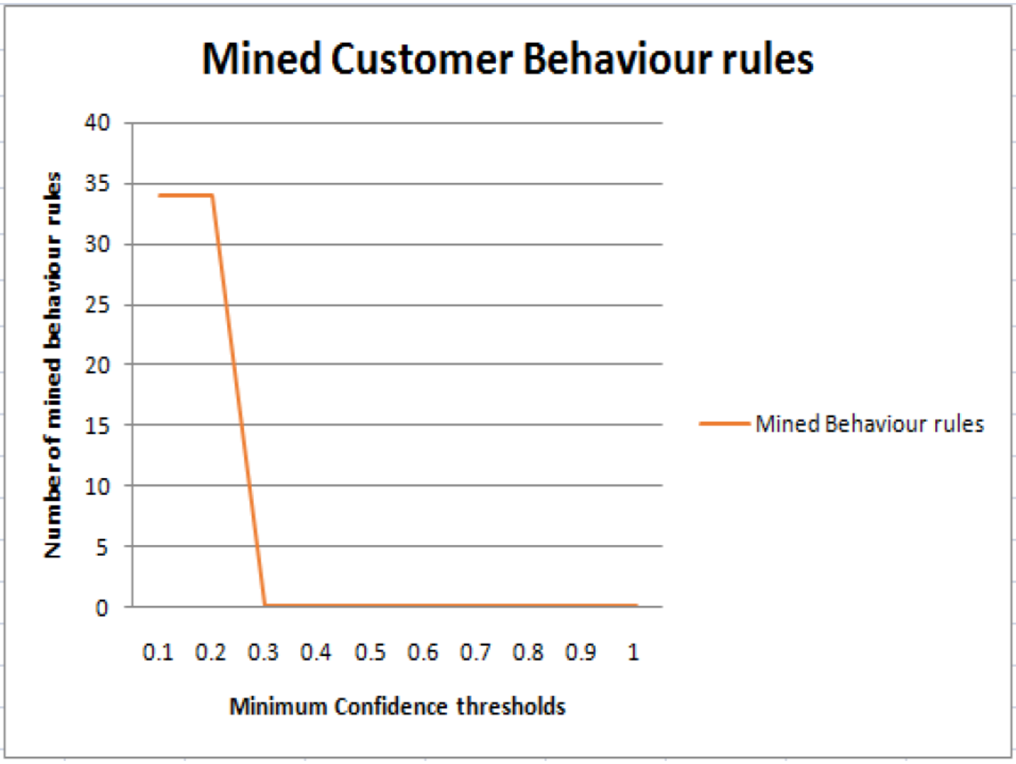

Fig. 3: Chart showing number of customer behavioral rules mined based on minimum confidence thresholds

It can be observed that the mined rules trend is a negatively sharp declining slope curve. At confidence threshold of 30%to 100%, no mined rule was recorded. Highest number of rules mined is generated at minimum confidence threshold of10% to 20%. This shows that as the confidence level is increased, optimal rule mining occurs thereby producing strongrules with efficient confidence. The strongest rules have rule confidence of approximately 100%.

VI. CONCLUSION

Data mining is primarily used today by companies with a strong customer focus - retail, financial, communication and marketing organizations. Data mining has a lot of importance because of its huge applicability. It is being used increasingly in business applications for understanding and then predicting valuable data, like customer buying actions and buying tendency, profiles of customers, industry analysis, etc. Data Mining is used in several applications like market research, customer behavior and directmarketing .This paper was able to identify frequent itemsets customer behaviour features patterns and mining association rulesbetween frequent purchase behaviour on an online store.The results from the frequent pattern mining shows thatoptimum rule generation occurred at minimum support and confidence thresholds of 0.1 and 0.2.This paper was able to design and implement a of association rule mining model for customer behaviour prediction. Itdiscovered interesting frequent customer behaviour purchasing patterns that occurred in the online retail store datasetand mined strong association rules. The performance of this model is greatly affected by the quality and dimensionalityof the dataset used and nature of feature set. Overall optimum performance of the model is peaked at minimum supportand confidence thresholds of 0.1 and 0.2 respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 12, December 2019

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 marketingin 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 JohnWiley 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 inKnowledge 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 inMarketing, Vol. 5, No. 2, 2013, pp.212-222
7. Gupta, G. & Aggarwal, H. (2012). Improving Customer Relationship Management Using Data Mining. International Journal of MachineLearning 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: AnInternational Journal, Vol. 1 Issue: 1, pp.3854, https://doi.org/10.1108/1750593071073412.
12. Ige A.O, Akingbesote A.O, Orogun A.O. (2019). Trust appraisal in a conventional e-market environment: A review. Canadian OpenInformation 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. Lim H. and Dubinsky, A., (2004). Consumers‟ perceptions of e-shopping characteristics: an expectancy-value approach, Journal of Services Marketing, 18 (7), 500-513.
17. O‟Cass, A. and Fenech, T., (2003). Web retailing adoption: Exploring the future of Internet users‟ web retailing behavior, Journal of Retailing and Consumer Services, 10 (2), 81–94.
18. COOMBS, W.T. 2010. Parameters for Crisis Communication. In: W.T. COOMBS and S.J. HOLLADAY (eds.) (The Handbook of Crisis Communication. United Kingdom: Wiley-Blackwell, pp. 17-53.)
19. Galanxhi-Janaqi, H. and Fui-Hoon Nah, F. (2004). U-commerce: emerging trends and research issues. Industrial Management & Data Systems, 104 (9), 744–755.