IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 22-27

Machine Learning for Market Basket Analysis through Association Rules

Devashri Raich[1], Bireshwar Ganguly[2], Madhavi Tota[3]
[1]*Assistant Professor, Dept. of IT, RCERT,Chandrapur*
[2]*Assistant Professor, Dept. of IT, RCERT,Chandrapur*
[3]*Assistant Professor Dept. of IT, RCERT,Chandrapur*

Abstract: *In the internet world, we all are surrounded with tons of data around us. "Where there is data, there are data mining applications"- which are application driven. Market Basket Analysis is connected with the most popular and successful application of Business Intelligence. (BI). When we deal with BI, the profit or loss depends on finding the exact correlation or association between the items sold. Market Basket Analysis is defined as discovery of frequent patterns and uncovers the association between items. Machine learning and advanced analytics combined with modern BI platform can efficiently discover new patterns and further help in analytics. Machine learning helps improve the performance of computers based on the data. It helps automatically recognize complex patterns and make intelligent decisions based on the data.*

Keywords: *Data Mining, Association, Correlation, Machine Learning, Business Intelligence.*

I. Introduction

Market Basket Analysis is one of the key methods utilized by substantial retailers to reveal relationship between things. It works by searching for blends of things that happen together every now and again in exchanges. To put it another way, it enables retailers to recognize connections between the things that individuals purchase.

Affiliation rules are comprehensively used to analyze retail holder data or trade data and are required to recognize the solid principles found in return data using interest-based, thought based measures. solid tenets. When we understand that customers who get one thing are most likely going to buy another thing, it is serviceable for the association to grandstand the things together or to make the buyers of one thing the potential objective for another. ..By focusing on clients definitely known to be potential purchasers, showcasing effectiveness increments significantly, in the case of advertising appears as store screens, index plan or direct offers to clients. This is the motivation behind shopping basket examination - to improve the viability of showcasing and deals strategies by utilizing client information officially accessible to the organization.

Market basket analysis is the most widely recognized method in data mining. Data mining is the way toward removing beforehand obscure data from vast arrangements of data. Today, data mining is utilized by numerous ventures, including fund and keeping money. The Bank is a marketing administration that can utilize data mining to break down client datasets and create factual profiles of inclination for individual clients for items and administrations.

In bank direct promoting space, there are a few information mining procedures can be utilized for ordering advertising administration, for example, choice tree, naive Bayes classifier, support vector machine, classification and association rule mining and six-sigma strategy.

The further part of paper describes how machine learning can be applied in practice, by relatively simple methods to find purchasing patterns among customers and how these can be converted to actionable insight for online or other retailers regardless of their size.

Market Basket Analysis through Association Rules:

Association Rule Mining represents finding association rules that are simple If/Then statements that help discover relationships between independent relational databases or other data repositories.

Most AI calculations work with numeric datasets and subsequently will in general be scientific. Be that as it may, affiliation rule digging is reasonable for non- numeric, all out information and requires only somewhat more than straightforward checking.

Affiliation rule mining is a procedure which intends to observe once in a while happening precedents, connections, or relationship from datasets found in different sorts of databases, for example, social databases, value-based databases, and different types of archives.An association rule has two parts:

- an antecedent (if) and
- a consequent (then).

An antecedent is something that's found in data, and a consequent is an item that is found in combination with the antecedent. Have a look at this rule for instance:
"If a customer buys bread, he's 70% likely of buying milk."

Association rules are made by altogether breaking down information and searching for successive on the off chance that/at that point designs. At that point, contingent upon the accompanying two parameters, the critical connections are watched:

1. **Support**: Support indicates how frequently the if/then relationship appears in the database.
2. **Confidence**: Confidence tells about the number of times these relationships have been found to be true.

AI and Machine Learning:

In the previous five years, numerous innovations have moved from ideas to the real world. They have made energizing new open doors for organizations simultaneously. Technologies such as artificial intelligence (AI), machine learning (ML) and data analytics capabilities are now being embedded into practical applications to increase business intelligence (BI), improve choice help, give more prominent and quicker preparing productivity, find new vulnerabilities all the more quickly, and make critical cost investment funds and gainfulness. ML speaks to a key development in the fields of software engineering, information examination, programming designing, and AI. ML is established on the premise that machines ought to almost certainly learn and adjust through involvement (Rouse, 2017). Developing volumes of different information, progresses in computational preparing, and moderate

information stockpiling have all added to a re- flooding enthusiasm for ML in the course of recent years. Ventures, substantial and little, around the globe are beginning to utilize ML to change basic procedures, especially in regions such as natural language processing (NLP), text classification and mining, emotional/behavioural analysis, and image recognition.

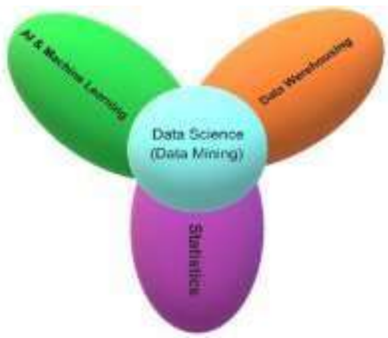


Figure1: Data Science

II. Literature Review

As per [1], in late 2016, MIT Technology Review led an online overview of 375 qualified respondents from an assortment of enterprises including technology, business administrations, money related administrations, and so forth., with respect to the utilization of ML. 60% of respondents (which included CEOs, presidents, and chairmen) in this worldwide overview showed they had effectively actualized ML procedure and had committed their organization to a continuous interest in ML.

More than 33% viewed them as at a develop phase of actualizing ML activities. Some 45% of respondents announced achievement in gaining a superior comprehension of client and imminent practices, needs, and wants. Most of these respondents originated from the USA and Canada (40%), Europe (13%), India (6%), and the UK (5%). As per the study, ML gave significant commercial center points of interest, empowering adopters to gain a focused edge (MIT Technology Review, 2016).

The paper composed by Radha K.Mahapatra [2] thusly, gives a review of AI methods and talks about their qualities and shortcomings with regards to mining business information. A review of information mining applications in business is given to explore the utilization of learning methods. Guideline enlistment (RI) was observed to be most well known, trailed by neural systems (NNs) and case-based thinking (CBR). Most applications were found in monetary regions, where forecast of things to come was a prevailing undertaking classification.

This paper [3] checks the execution of various arrangement procedures to choose the one with the most precise outcomes for characterization of bank direct advertising dataset. The paper picks four regularly utilized systems from various arrangement methods, two strategies from choice tree and the rest from AI. Choice tree comprise J48-unite calculation and LAD tree calculation while AI comprise spiral premise work system and bolster vector machine. In paper[4] the creators show that arrange based methodology can succinctly separate impact among items, mitigating the need to seek through huge arrangements of affiliation rules. This paper builds up an intriguing quality measure for networks of items and demonstrates that it detaches valuable, noteworthy networks.

III. Proposed Methodology

In this paper an algorithm is designed using machine learning concepts and association rule mining. This will extract relevant patterns and analyze the data which is frequently used. Different modules used are explained in the next section of this paper.

IV. Modules and Description

1. Preparing a Machine Learning-Ready Dataset for Market Basket Analysis

BigML Associations can help recognize which pairs (or gatherings) of things happen together more as often as possible than anticipated. A common use case for affiliation rule disclosure is advertise crate investigation, where the objective is to discover the items that are typically acquired together by clients.

BigML's Associations can yield such fascinating relationship from your dataset as standards, which are communicated as a blend of fields and their qualities. The field esteems for which you need to discover the affiliations can be either be masterminded in various fields (when the quantity of components per example is fixed), or they can be put away all together in a one of a kind field. This last field group is alluded to as a things field in BigML and it takes into account numerous qualities per case isolated by an assigned separator.

2. Load the packages:

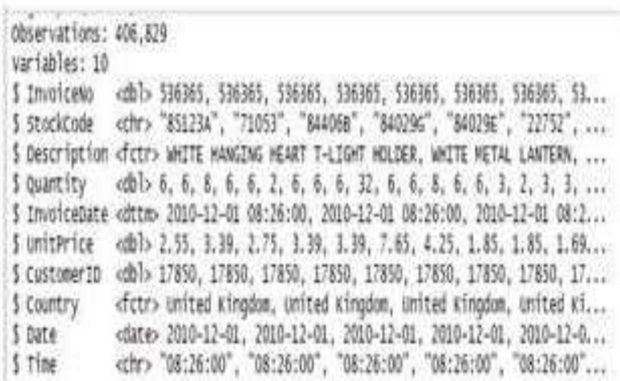
This module imports the following packages

library(tidyverse) library(readxl) library(knitr) library(ggplot2)
library(lubridate) library(arules) library(arulesViz) library(plyr)

3. Data pre-processing and exploring:

retail <- read_excel('Online_retail.xlsx') retail <- retail[complete.cases(retail),] retail <- retail %>%
mutate(Description = as.factor(Description))
retail <- retail %>% mutate(Country = as.factor(Country))
retail$Date <- as.Date(retail$InvoiceDate) retail$Time <- format(retail$InvoiceDate,"%H:%M:%S")
retail$InvoiceNo <- as.numeric(as.character(retail$InvoiceNo)) glimpse(retail)

After preprocessing, the dataset includes 406,829 records and 10 fields: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country, Date, Time.



```
Observations: 406,829
Variables: 10
$ InvoiceNo   <dbl> 536365, 536365, 536365, 536365, 536365, 536365, 536365, 53...
$ StockCode   <chr> "85123A", "71053", "84406B", "84029G", "84029E", "22752", ...
$ Description <fctr> WHITE HANGING HEART T-LIGHT HOLDER, WHITE METAL LANTERN, ...
$ Quantity    <dbl> 6, 6, 8, 6, 6, 2, 6, 6, 6, 32, 6, 6, 8, 6, 6, 3, 2, 3, 3, ...
$ InvoiceDate <dttm> 2010-12-01 08:26:00, 2010-12-01 08:26:00, 2010-12-01 08:2...
$ UnitPrice   <dbl> 2.55, 3.39, 2.75, 3.39, 3.39, 7.65, 4.25, 1.85, 1.85, 1.69...
$ CustomerID  <dbl> 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17850, 17...
$ Country     <fctr> United Kingdom, United Kingdom, United Kingdom, United Ki...
$ Date        <date> 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-01, 2010-12-0...
$ Time        <chr> "08:26:00", "08:26:00", "08:26:00", "08:26:00", "08:26:00"...
```

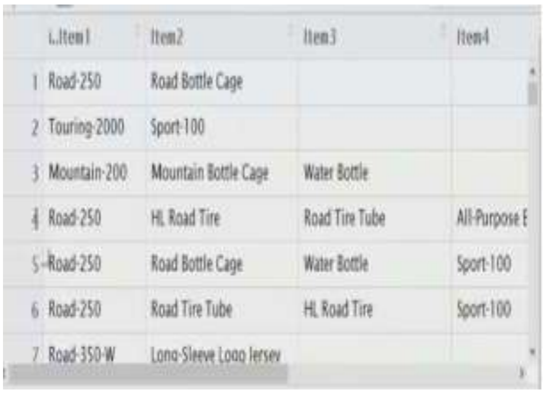
Figure 2. Preprocessing Data

4. Association rules designing:

More often than not in market basket analysis there are numerous things and every client purchase few of them. That prompts an inadequate grid and most characterization techniques have issues managing them.A regular strategy to examine market basket information is utilizing Association rules which are executed by first finding incessant thing sets. we create an Association to discover the items that are normally obtained together. Affiliations can be designed in multiple ways. By default, the search strategy is Leverage (the difference between the probability of the rule and the expected probability were the items statistically independent).
You can use association rules on order to learn on market basket in the following way:

- Add item for every class
- Add the owner class to its basket (e.g, transform {milk, beer} basket of a student in to the basket {milk, beer, student}
- Use a frequent item set algorithm to find them
- Remove all frequent item sets that don't have a class member in them. Such sets can't help deduce the class.
- From the frequent items set find rules that imply the classes (with good enough support and confidence
- Use the these rules that match a basket in order to predict class

Before using any rule mining algorithm, we need to transform the data from the data frame format, into transactions such that we have all the items bought together in one row.



	i..Item1	Item2	Item3	Item4
1	Road-250	Road Bottle Cage		
2	Touring-2000	Sport-100		
3	Mountain-200	Mountain Bottle Cage	Water Bottle	
4	Road-250	HL Road Tire	Road Tire Tube	All-Purpose E
5	Road-250	Road Bottle Cage	Water Bottle	Sport-100
6	Road-250	Road Tire Tube	HL Road Tire	Sport-100
7	Road-350-W	Long-Sleeve Logo Jersey		

Figure 3. Online transaction

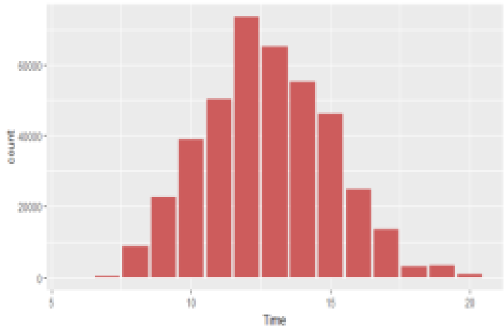


Figure 4: Shopping time distribution

Above figure shows that there is a clear bias between the hour of day and order volume. Most orders happened between 10:00–15:00.

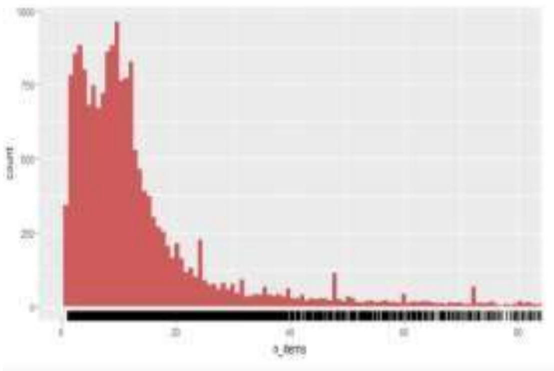


Figure 5: Number of items per invoice distribution

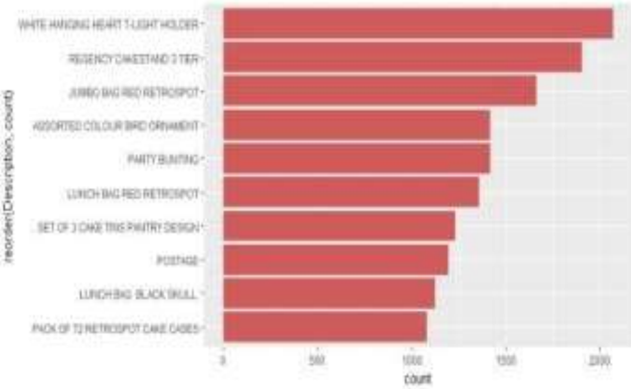


Figure 6: data showing best sellers

V. Conclusion

There are a few components fuelling the advancement of ML. Advances in information accumulation, investigation arrangements, equipment and programming, and computational power have changed the field. ML is compromising numerous

Occupations that require long periods of training, for example, design acknowledgment or picture analysis done by pathologists and radiologists. The key point is that conveying a ML technique in a business association requires an exceedingly qualified devoted group of information researchers to assume on this liability, which may not be a venture an association is happy to take on. ML is bound to be the domain of huge national and universal organizations. Be that as it may, diminishing expense of framework gave in the distributed computing, expanding availability and advancement of logical calculations, joined with new markets for buying information, will open ML to little scales tasks. In this paper the outcome demonstrates the continuous examples dug for a given informational index. This paper likewise depicted how to build a ML-Ready dataset for BigML Associations from various source documents. Diverse approaches to assemble Associations between items were additionally found. There are additionally different sorts of pursuit systems when constructing an Association Discovery. On the off chance that we need to discover connections between abnormal items, utilizing Lift, and driving a base help of 0.1%, the subsequent standards in the new affiliation are very extraordinary.

References

[1]. Mitchell, T. (1997) Machine Learning, McGraw Hill, ISBN: 0- 07-042807-7

[2]. Business data mining — a machine learning perspective IndranilBose[1a]Radha K.Mahapatra[b], Information & Management Volume 39, Issue 3, 20 December 2001, Pages 211-225

[3]. A Comparison of Different Classification Techniques for Bank Direct Marketing K. Wisaeng, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013

[4]. Chappell, D. (2015) Introducing Azure Machine Learning: a Guide for Technical Professionals, Chappell & Associates [online] http://www.davidchappell.com/writing/ white_papers/Introducing- Azure-ML-v1.0--Chappell.pdf (accessed 11 December 2017). Costantino, M. and Coletti, P. (2008) Information Extraction in Finance, p.13, WIT Press, Southampton, Boston. D'Onfro, J. (2015)

[5]. Business Insider [online] http://www.businessinsider.com/ pinterest-acquisition-2015-1 (accessed 10 October 2017). Das, K. and Beher, R.N. (2017)

[6]. 'A survey on machine learning: concept, algorithms and applications', International Journal of Innovative Research in Computer and Communication Engineering, February, Vol. 5, No. 2, pp.1301–1309. Data Warehouse (2017) Tech-Faq [online] http://www.tech-faq.com/data-warehouse.html# (accessed 21 November 2017).

[7]. Dedić, N. and Stanier, C. (2016) 'An evaluation of the challenges of multilingualism in data warehouse development', in 18th International Conference on Enterprise Information Systems – ICEIS, p.196. Deloitte (2017)

[8]. Machine Learning: Things are Getting Intense [online] https://www2.deloitte.com/ content/dam/Deloitte/global/Images/infographics/technologymedia telecommunications/gxdeloitte-tmt-2018-intense-machine-learning-report.pdf (accessed 11 November 2017).

[9]. Devi, I., Karpagam, G.R. and Kumar, B.V. (2017) 'A survey of machine learning techniques', Int. J. of Computational Systems Engineering, Vol. 3, No. 4, pp.203–212.

[10]. Khanum, M., Mahboob, T., Ghafoor, H.A. and Sehar, R. (2017) 'A survey on unsupervised machine learning algorithms for automation, classification, and maintenance', International Journal of Computer Applications, Vol. 119, No. 13, pp.34–39. Khaytin, A. (2017)

[11]. 'The four stages of machine learning implementation', eeNews Europe, January, p.9 [online] http://www.eenewseurope.com (accessed 10 February 2018). Krishna, D., Albinson, N. and Chu, Y. (2017)

[12]. Managing Algorithmic Risks: Safeguarding the use of Complex Algorithms and Machine Learning, Deleoitte [online] https://www2.deloitte.com/ content/dam/Deloitte/us/Documents/risk/us-risk-algorithmic- machine-learning-riskmanagement.pdf (accessed 10 February). Kumar, A. and Sarkar, B.K. (2017)

[13]. 'A case study on machine learning and classification', Int. J. of Information and Decision Sciences, Vol. 9, No. 2, pp.179–208. Lebied, M. (2017) Top 10 Business Intelligence and Analytics Trends for 2018, Business Intelligence [online]

https://www.datapine.com/blog/business-intelligence-trends/ (accessed 10 February 2018).

[14]. MIT Technology Review (2016) Machine Learning: the New Proving Ground for Competitive Advantage [online] https://s3.amazonaws.com/files.technologyreview.com/ whitepapers/MITTR_GoogleforWork_Survey.pdf (accessed 10 February 2018).

[15]. Cortes C, Vapnik V. 1995. Support-vector networks. Machine Learning, 20:273-297.

[16]. Izabella V. Lokshina, Barbara J. Durkin, Cees J. M. Lanting, "Data Analysis Services Related to the IoT and Big Data:Potential Business Opportunities for Third Parties", Proceedings of the 50th Hawaii International Conference on System Sciences | 2017

[17]. Dagmar Klopper, "The Possibilities and Challenges of the Application and Integration of the Internet of Things for Future Marketing Practice", 7th IBA Bachelor Thesis Conference, July 1st, 2016, Enschede, The Netherlands.