



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

Building Prediction Model using Market Basket Analysis

Roshan Gangurde¹, Dr. Binod Kumar², Dr. S. D. Gore³

Research Scholar, Department of Computer Science, Savitribai Phule Pune University, Pune, Maharashtra, India¹

Director, Jayawant Institute of Computer Applications, Pune, Maharashtra, India²

Rtd. Professor, Department of Statistics, Savitribai Phule Pune University, Pune, Maharashtra, India³

ABSTRACT: In the recent years, analyzing shopping baskets turned out to be very appealing to retailers. Sophisticated technology made it possible for them to collect information of their customers and what they purchase. The introduction of electronic point-of-sale expanded the utilization and application of transactional data in Market Basket Analysis (MBA). In retail business, analyzing such information is exceedingly valuable for understanding purchasing behavior. Mining purchasing patterns allows retailers to adjust promotions, store settings and serve consumers better. Predictive analysis is an advanced branch of data engineering which generally predicts some occurrence or probability based on data. Predictive analytics uses data-mining techniques in order to make predictions about future events, and make recommendations based on these predictions. The process involves an analysis of historic data and based on that analysis to predict the future occurrences or events. A model can be created to predict using Predictive Analytics modelling techniques. The form of these predictive models varies depending on the data they are using. Predictive Analytics is composed of various statistical & analytical techniques used to develop models that will predict future occurrence, events or probabilities. Predictive analytics is able to not only deal with continuous changes, but discontinuous changes as well. Classification, prediction, and to some extent, affinity analysis constitute the analytical methods employed in predictive analytics.

KEYWORDS: Association Rule Mining, Market Basket Analysis, Predictive Modelling,

I. INTRODUCTION

Predictive analytics is composed of two words predict & analysis, but it works in reverse *viz.* first analyze then predict. It is human nature to want to know and predict what the future holds. Predictive analytics deals with the prediction of future events based on previously observed historical data by applying sophisticated methods like machine learning. The historical data is collected and transformed by using various techniques like filtering, correlating the data, and so on.

Prediction process can be divided into four steps:

- (1) Collect and pre-process raw data;
- (2) Transform pre-processed data into a form that can be easily handled by the selected machine learning method;
- (3) Create the learning model (training) using the transformed data;
- (4) Report predictions to the user using the previously created learning model.

An essential goal in data mining is to create and enhance the precision of predictive models, and a basic challenge toward this end lies in the discovery of new features, inputs or predictors. This paper illustrates how rules generated from Market Basket Analysis (MBA) may be utilized to improve predictive models.

The goal of data mining process is the extraction of information from large data sets, transform such information into some understandable structure for future application [1]. It is a process of using different techniques to find useful patterns or models from data. This process is use to select, explore and model large amount of data. The extracted

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 2, February 2017

patterns from the database are then used to build data mining models, and can be used to predict performance and behavior with high accuracy.

If this data is properly integrated, analyzed, and interpreted could offer crucial insights to best serve our greatest societal priorities in different sectors such as healthcare, economics, education, and employment.

A. DATA MINING TASKS

Data mining satisfies its main goal by identifying valid, potentially useful, and easily understandable correlations and patterns present in existing data. This goal of data mining can be achieved by modelling it as either predictive or descriptive nature.

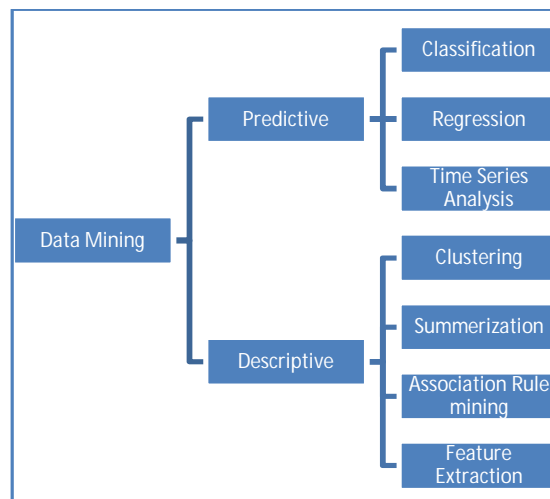


Figure 1: Data Mining Techniques

The predictive model works by making a prediction about values of data, which uses known results found from different datasets. The tasks include in the predictive data mining model includes classification, prediction, regression and analysis of time series. The predictive data-mining model predicts the future outcomes based on past records present in the database or with known answers.

The descriptive model mostly identifies patterns or relationships in datasets. It serves to explore the properties of the data examined earlier and not to predict new properties. The descriptive model encompasses task to perform as Clustering, Association Rules, Summarizations, and Sequence Analysis. The descriptive data-mining model discovers patterns in the data and understands the relationships between the attributes represented by the data.

The data-mining task is further divided into the following two approaches:

B. SUPERVISED OR DIRECTED DATA MINING

The goal of supervised or directed data mining is to use the available data as like predictive data-mining to build a model that describes one particular variable of interest in terms of the rest of the available data. The user selects the target field and directs the computer to determine estimate, classify or predict its value.

In supervised learning, data is modeled from training data to find patterns within the data which can then be used to predict a label or value, given some set of parameters. Supervised learning is the process of creating predictive models using a set of historical data that contains the results we are trying to predict. The type of data determines whether this is done using a Regression or a Classification algorithm.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

C. UNSUPERVISED OR UNDIRECTED DATA MINING

In unsupervised or undirected data mining, however variable is singled out as the target as like the descriptive mining technique. The objective is fairly to build up some relationship among all the variables in the data. The user requests the computer to recognize patterns in the data that might be noteworthy. Undirected modelling is used to clarify those patterns and relationships found in the data.

Unsupervised learning refers to the problem of trying to find hidden structure in unlabeled data. Unlike supervised algorithms, unsupervised algorithms do not learn from historical data with known labels, hence, they perform without any supervision. Standard unsupervised techniques include clustering, characterization, association rule mining, and change and deviation detecting techniques.

II. PREDICTIVE DATA MINING MODEL

The major objective of data mining is to build a model that can be used to predict the occurrence of an event. The model builders will extract knowledge from historic data and represent it in such a form that the resulting model can be applied to new situations. The process of analyzing data sets extracts useful information on which to apply one or more data mining techniques in order to discover previously unknown patterns within the data, or find trends in the data which can then be used to predict future trends or behaviours.

A predictive analytical model is built by data mining tools and techniques. The first step consists of extracting data by accessing massive databases. The data thus obtained is processed with the help of advanced algorithms to find hidden patterns and predictive information. Though there is an obvious connection between statistics and data mining, methodologies used in data mining have originated in fields other than statistics.

The core element of predictive analytics is the 'predictor', a variable that can be measured for an individual or entity to predict its future behavior. For example, a credit card company may consider age, income, and credit history as predictors to determine the risk factor in issuing a credit card to an applicant. Multiple predictors can be combined into a predictive model, which is then used to forecast future probabilities with an acceptable level of reliability. In predictive modeling, data is collected, a statistical model is formulated, predictions are made, and the model is validated (or revised) as additional data become available [2].

Predictive analytics combines business knowledge and statistical analytical techniques which, when applied to business data, produce insights. These insights help organizations understand how people behave as customers, buyers, sellers, distributors, and so on.

Multiple related predictive models produce insights for making strategic company decisions such as exploring new markets, acquisitions, and retentions; finding up-selling and cross-selling opportunities; and discovering areas that can improve security and fraud detection. Predictive analytics indicates not only what to do, but also how and when to do it, and to explain 'what-if' scenarios [3].

The significance of predictive analytics is self-evident. The more you comprehend customer behavior and motivation, the more powerful your marketing will be. The more you understand why a few customers are reliable and how to pull in and hold diverse customer segments, the more you can create significant, convincing messages and offers.

Predicting customer buying and product preferences and habits requires an analytical framework that empowers you to find meaningful patterns and relationships within customer information keeping in mind the end goal to accomplish better message focusing on and drive customer esteem and reliability.

Marketing and sales experts are starting to capture and analyze a wide range of sorts of client information attitudinal, behavioral, and transactional related to purchasing and product preferences to make predictions about future buying behavior.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Today's challenging environment is driving more businesses to investigate predictive analytics. Commonly used by market researchers when analyzing survey data, predictive analytics can also be applied in real-time scenarios, such as customizing offers to consumers or improving an online customer experience.

A. PREDICTIVE ANALYSIS

The future of data mining lies in predictive analytics. Predictive Analysis is an analytic model that consolidates historical data, statistical algorithms, and information technology to propose potential conclusions. The principle target of this analytic model is to 'predict' what will occur in a given situation in light of significant investigation of the past. It analyzes the "How?" and "Why?" factors behind each occurrence in regard to a subject in order to decide the next probable move. It therefore empowers frontline decision makers with significant understanding so that they can build more revenue, maximize efficiency, avoid pit-falls and so forth... The potential outcomes to enhance your business are endless.

Predictive analytics is used to determine the probable future outcome of an event or the likelihood of a situation occurring. It is the branch of data mining concerned with the prediction of future probabilities and trends. Predictive analytics is used to automatically analyze large amounts of data with different variables; it includes clustering, decision trees, market basket analysis, regression modelling, neural nets, genetic algorithms, text mining, hypothesis testing, decision analytics, and more [2].

The power of information technology has transformed traditional statistical prediction methods, providing users with intuitive, attractive, and interactive ways to use predictive models. Predictive Models use identified consecutive results to determine the future values of the target variables.

Recognizing purchasing principles is vital for each effective business. Transactional data is utilized for mining helpful information on co-buys and conforming promotion and advertising as needs be. The notable arrangement of beer and diapers is only a case of an association rule found by information researchers. Mining association rules from transactional data will furnish us with important data about co-occurrences and co-buys of products. A few customers may buy a single product during a shopping trip, out of curiosity or boredom, while others purchase more than one item for productivity reasons.

III. MARKET BASKET ANALYSIS: REVIEW

Supermarkets utilize data to better comprehend shopper needs and, at last, boost their overall spend. One of the key techniques used by the big retailers is called Market Basket Analysis (MBA), which reveals relationship between products by searching for blends of products that frequently co-occur in transactions. At the end of the day, it permits the stores to recognize relationships between the products that individuals purchase. For instance, customers that purchase a pencil and paper are probably going to purchase an eraser or ruler.

"Market Basket Analysis allows retailers to identify relationships between the products that people buy."

A. WORK ENVIRONMENT OF MBA

To carry out an MBA, we first need a data set of transactions. Each transaction represents a group of items or products that have been bought together and often referred to as an "itemset". For example, one itemset might be: {pencil, paper, staples, rubber} in which all of these items have been bought in a single transaction.

In an MBA, the transactions are analysed to identify rules of association. For example, one rule could be: {pencil, paper} => {rubber}. This means that if a customer has a transaction that contains a pencil and paper, then they are likely to be interested in also buying a rubber.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

Before acting on a rule, a retailer needs to know whether there is sufficient evidence to suggest that it will result in a beneficial outcome. We therefore measure the strength of a rule by calculating the following three metrics (Other metrics are also available, but these are the three most commonly used):

Support: The percentage of transactions that contain all of the items in an itemset (e.g., pencil, paper and rubber). The higher the support the more frequently the itemset occurs. Rules with a high support are preferred since they are likely to be applicable to a large number of future transactions.

Confidence: The probability that a transaction that contains the items on the left hand side of the rule (in our example, pencil and paper) also contains the item on the right hand side (a rubber). The higher the confidence, the greater the likelihood that the item on the right hand side will be purchased or, in other words, the greater the return rates you can expect for a given rule.

Lift: the probability of all of the items in a rule occurring together (otherwise known as the support) divided by the product of the probabilities of the items on the left and right hand side occurring as if there was no association between them. For example, if pencil, paper and rubber occurred together in 2.5% of all transactions, pencil and paper in 10% of transactions and rubber in 8% of transactions, then the lift would be: $0.025/(0.1*0.08) = 3.125$. A lift of more than 1 suggests that the presence of pencil and paper increases the probability that a rubber will also occur in the transaction. Overall, lift summarises the strength of association between the products on the left and right hand side of the rule; the larger the lift the greater the link between the two products.

Support of a product or product bundle indicates the popularity of the product or product bundle in the transaction set. Higher the support, more popular is the product or product bundle. This measure can help in identifying driver of traffic to the store. Hence, if Barbie dolls have a higher support then they can be attractively priced to attract traffic to a store. Confidence can be used for product placement strategy and increasing profitability. Place high-margin items with associated high selling (driver) items. If Market Basket Analysis indicates that customers who bought high selling Barbie dolls also bought high-margin candies, then candies should be placed near Barbie dolls. Lift indicates the strength of an association rule over the random co-occurrence of Item A and Item B, given their individual support. Lift provides information about the change in probability of Item A in presence of Item B. Lift values greater than 1.0 indicate that transactions containing Item B tend to contain Item A more often than transactions that do not contain Item B.

To perform a Market Basket Analysis and identify potential rules, a data mining algorithm called the 'Apriori algorithm' is commonly used, which works in two steps:

1. Systematically identify itemsets that occur frequently in the data set with a support greater than a pre-specified threshold.
2. Calculate the confidence of all possible rules given the frequent itemsets and keep only those with a confidence greater than a pre-specified threshold.

The thresholds at which to set the support and confidence are user-specified and are likely to vary between transaction data sets.

Market Basket Analysis (MBA), otherwise called affinity analysis, is an approach to discover items likely to be bought together. The introduction of electronic point of sale frameworks has led to accumulation of expansive amount of data. Basic, yet capable - MBA is a reasonable method to recognize cross-sell opportunities. An exemplary illustration is toothpaste and fish. It appears that individuals who eat fish are more inclined to brush their teeth directly subsequent to completing their dinner. So, why it is imperative for retailers to get a good grasp of the product affinities? This information is critical to appropriately plan for promotions because reducing the price on some items may cause a spike on related high-affinity items without the need to further promote these related items.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 2, February 2017

B. APPLICATIONS OF MBA

The analysis can be applied in various ways:

- Develop combo offers based on products sold together
- Organize and place associated products/categories nearby inside a store
- Determine the layout of the catalog of an ecommerce site
- Control inventory based on product demands and what products sell together

In order to gain better insights, differentiate Market Basket Analysis based on

- Weekend vs weekday sales
- Month beginning vs month-end sales
- Different seasons of the year
- Different stores
- Different customer profiles

Based on the content and value of the basket, it is useful to classify the trip. Variables such as total basket value, number of items, number of category X vs. category Y items, help in developing rules to map each of the baskets to a previously defined classification. Understanding what kind of shopping trips a customer performs at a particular store at a particular time is critical for planning purposes. This data provides a unique window into what is happening at the store and enables advanced applications such as labor scheduling, product readiness and even temporary layout changes.

IV. PREDICTION MODEL USING MBA

A prediction model is made up of a number of predictors, which are variable factors that are likely to influence future behaviour or results. Any classifier that can be used for determining a particular class for a test object is part of predictive modelling. Therefore market basket analysis can be used for prediction modelling.

If we use the results of the MBA to predict future behavior of customers, then it could be called a prediction model. Ex. if you are using market basket analysis to come up with product bundles then you are basing past purchase behaviour of customers to predict future purchase behaviour, which is a predictive model

V. RELATED WORK

SachinKamley[4] has developed a model for finding interesting patterns for stock market dataset using association rule mining. The model is helpful for stock brokers, investors to earn maximum profits for each trading. He developed a method to learn investment plan for stock brokers and investors and understanding the market conditions.

Raorane [5] has observed that Market Basket Analysis (MBA) can be used in managing the product placement on the shelves in the supermarket. This method can prove to fetch more profit to the seller. Thus the Data Mining tool can be used to improve the strategy in placement of the product on the shelf by using the Data mining tools. Using MBA, the frequent transactions made by the customers have been analyzed using the support and confidence of the customers in buying associated items.

Grishma [6] analyses the patterns of consumer buying behaviour of products of a lifestyle store. The developed software proves to be useful for retailers to understand the purchasing behaviour of their customers and gives valuable insights relating to the formation of the basket. It helps in product assortments, refilling of the stocks for the likely items sold, make promotions based on likely items sold with a particular category, bundling of the products, give discounts to prompt the customers to buy the products.

In [7] research, Frequent Pattern Growth algorithm is used, for efficient mining of frequent pattern. FP-Growth is applied on dataset and generated the association rules for the sports store which care for profit maximization at wholesale and retailer level. The results generated will help the owner of the sports store how to increase the sales and earn profits. The findings of this paper are applied to optimize marketing and sale of sports equipment.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Debaditya [8] attempted to develop a method using association rule mining to find out the most preferable and popular genres which can be represented as movie business's trend. The study can predict the possible movie trend based on the genres. These trends for future months can be treated as guidelines for the production houses to drive movie business towards more profitability.

OmprakashChandrakar [9] presented an application of data mining in higher education. Association rule mining is applied to analyze the performance of students in their examinations and predicts the outcome of the forthcoming examination. This prediction allows student and teacher to identify the subjects which need more attention even before the commencement of semester. The rules discovered through association rule mining are used to predict the outcome of the forthcoming examination. This prediction may be used to guide the student at the very beginning of their semester by identifying the subjects in which they need to focus more.

KarandeepKaur [10] developed a novel algorithm based on association rule learning to precisely predict the result of a coronary illness examination and the act of this novel calculation could turn out to be useful for the medicinal specialists and experts for precisely anticipating the coronary illness. Associative rule mining, works best for the cases where no example or test outcome are available on the grounds that associative rule mining is an unsupervised learning calculation.

The model developed by [11] is a model of crime mapping using association rule mining for criminal based on geographical and demographic factors. It examined the occurrence of crime at a specific location. The predictive crime mapping is one of the solutions that can be used to analyse the relatively high future crime location that can improve the crime prevention implementation.

Samuel Musungwini [12] investigated the relationship between Product, Place, Promotion and Price (4Ps) in Market Basket Analysis (MBA) and established how the 4Ps can be applied as a tool for competitive advantage in the small scale retail sector.

Association Rule Mining is used for prediction of stock market [13]. Prediction depends on technical trading indicators and closing prices of the stock. Rules are defined according to signal generated by each technical trading indicator and mapped across the current date query to generate the signals like buy, sell or holds the shares. It also increased the accuracy of the prediction system by accepting the accurate closing prices of the stock.

VI. CONCLUSION AND FUTURE WORK

Predictive modeling offers the potential for firms to be proactive instead of receptive. Predictive modeling using transactional data create particular challenges which need to be carefully addressed to develop valuable models.

With MBA, leading retailers can drive more profitable advertising and promotions, attract more customers, increase the value of market basket and much more. Consumers, planners, merchandisers and store administrators have started to recognize how this new era of easy-to-use market basket analysis tools helps to work more intelligently and compete more successfully. Our future work would be to design and develop intelligent prediction models to generate the association rules that can be adopted on recommendation system to make the functionally more operational. Better and effective rule mining techniques can be used for better performance of the recommendation system.

RECOMMENDATION

The knowledge generated from Market Basket Analysis is useful for the organizations engaged in retailing business for their decision making process.



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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

REFERENCES

1. J. Han and M. Kamber. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA, 2001
2. <http://www.articlesbase.com/strategic-planning-articles/predictiveanalytics-1704860.html>
3. M Zaman, Predictive analytics; the future of businessintelligence www.mahmoudyoussef.com
4. SachinKamley, ShaileshJaloree, R.S. Thakur, "An Association Rule Mining Model for Finding the Interesting Patterns in Stock Market Dataset", International Journal of Computer Applications (0975 – 8887), Volume 93 – No 9, May 2014
5. Raorane A.A., Kulkarni R.V, Jitkar B.D., "Association Rule – Extracting Knowledge Using Market Basket Analysis", Research Journal of Recent Sciences, ISSN 2277-2502, Vol. 1(2), 19-27, Feb.2012
6. GrishmaKapadia, KavitaKalyandurgmath, "Market Basket Analysis of Consumer Buying Behaviour of a Lifestyle Store", International Conference on Technology and Business Management March 23-25, 2015
7. HarpreetKaur, Kawaljeet Singh, "Market Basket Analysis of Sports Store using AssociationRules", International Journal of Recent Trends in Electrical & Electronics Engg., ISSN: 22316612, Volume 3, Issue 1, pg: 81-85, Dec. 2013.
8. Debaditya Barman, NirmalyaChowdhury, "Movie Business Trend Prediction using Market Basket Analysis", International Journal of Computer Applications (0975 – 8887), Volume 74– No.9, July 2013
9. OmprakashChandrakar, Jatinderkumar R. Saini, "Predicting Examination Results using Association Rule Mining", International Journal of Computer Applications (0975 – 8887), Volume 116 – No. 1, April 2015
10. KarandeepKaur, PoonamdeepKaur, LovepreetKaur, "International Journal of Computer Techniques", ISSN: 2394-2231, Volume 2 Issue 5, Page 72, Sep – Oct 2015
11. SitiAzirahAsmai, NurIzzatulAbidahRoslin, RosmizaWahida Abdullah, Sabrina Ahmad, "Predictive Crime Mapping Model using Association Rule Mining for Crime Analysis", International Symposium on Research in Innovation and Sustainability 2014 (ISoRIS '14), Malacca, Malaysia, ISSN 1013-5316, 15-16 October 2014
12. Samuel Musungwini, Tinashe Gwendolyn Zhou, Raviro Gumbo, TinomudaMzikamwi, "The Relationship Between (4ps) & Market Basket Analysis. A Case Study of Grocery Retail Shops in Gweru Zimbabwe", International Journal of Scientific & Technology Research, Volume 3, Issue 10, ISSN 2277-8616, October 2014
13. Shubhangi S. Umbarkar, S. S. Nandgaonkar, "Using Association Rule Mining: Stock Market Events Prediction from Financial News", International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Volume 4 Issue 6, June 2015