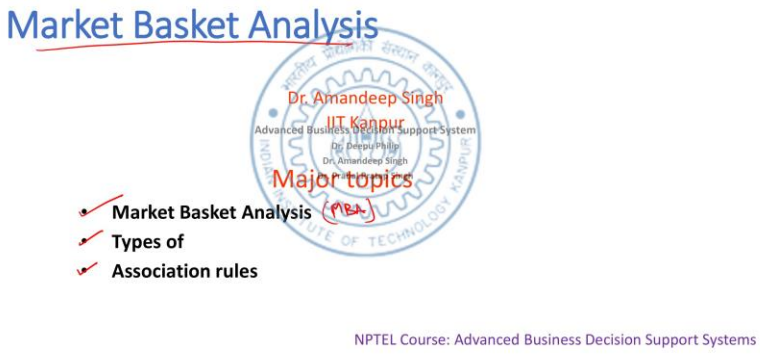
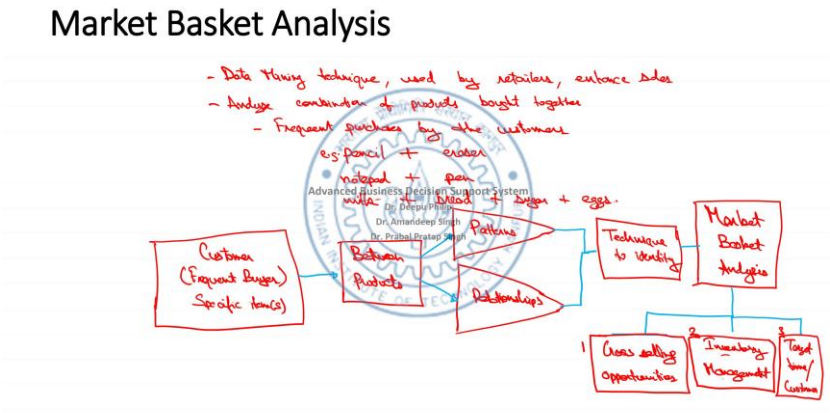
Advanced Business Decision Support Systems
Professor Deepu Philip
Department of Industrial Engineering and Management Engineering
Indian Institute of Technology, Kanpur
Professor Amandeep Singh
Imagineering Laboratory
Dr. Prabal Pratap Singh
Indian Institute of Technology, Kanpur
Lecture 41
Market Basket Analysis

Welcome to the second lecture of the last week of the course advanced business decision support systems. I have just given brief introduction to big data analytics in the first lecture in this week. As I said, I am going to talk about the Market Basket Analysis in this lecture and also, I will take some examples using the excel sheet in the coming lecture.



So, I will talk about the Market Basket Analysis, the Types of it and some Association rules that is the probability rules and relationships between the different products in the basket. Market Basket Analysis, I will call it 'MBA' is a data mining technique. So, it is used by retailers to increase sales so that they better understand the customers, they better understand that what is impact on that the customers have so that the large sets of the data which are available with them with the previous history that is there the big data that is available with them.

They put them into groupings kind of a clusters so that the likely to be purchased to gather products are put into a single basket. I will take certain examples and we will try to understand what is Market Basket Analysis.



So, as we know nowadays, a lot of Machine Learning programs are there, there we have the languages like Python. On real world data set, we can use the similar languages which helps the retail industry in many different modern ways and we can imagine forecasting the performance of sales to identify the potential buyers. There are many applications where this could be used.

So, Market Basket Analysis and its various components are important to understand what kind of specific pattern would be there in the future that the customer will take from the specific retail store.

So, it is by definition, data mining technique which is used by retailers to enhance sales or to better organize their data or their stock into specific set pattern. So, it helps them to deep understand the customer purchasing patterns. This method helps the examination of substantial data sets.

Certain historical purchase records are there in order to unravel the product groupings and to identify the item which tend to be brought together. There are certain examples that could be taken. If one would purchase pencil, it might purchase also eraser along with that. If one purchase a notepad, it might also purchase pens along with that.

Maybe the sales of the notepads could be may be 5000 PCs per month or so, along with them may be 2000 pens are also sold which are sold together to the same customer. Sometimes, may be both together notepad and pens being sold.

So, when we put them together, the customers who have picked notepad, who have picked pens, the customers who have picked pencils and along with them, it may have picked erasers. May be along with 4 pencils, 1 eraser is being purchased or may be for the kids, along with 1 pencil, 1 eraser is being purchased. So, these patterns could be identified. So, this becomes a data mining technique.

So, we analyze the combination of products bought together. So, this technique gives a careful study of purchases done by customers in a supermarket. This concept identifies the pattern of frequent purchases by the customers. I would say, pencil plus eraser as an example or may be notepad plus pen and so on.

So, data mining concepts are used for sales and marketing. Here, to provide the customer service, to improve cost, to improve cross selling opportunities, to increase direct mail response rates, customer attention is there. So, we use certain rules here for the Market Basket Analysis before that let us try to see how are the patterns changing.

So, we have customer or I would say frequent buyer, I am talking about a specific item or set of items. So, it picks between products then we need to see whether there are certain patterns in what they purchase or certain relationships between them.

Then, there has to be technique to identify these patterns and relations identify pattern relations which are mentioned here. This technique is our Market Basket Analysis. Let me try to connect them. We have customers that purchase between specific products and between the products, there could be certain patterns. The way they purchase and there could be certain relationships between what they purchase.

These are to be connected or identified using a technique which is Market Basket Analysis and it is of certain kinds. So, it is cross-selling opportunities. Cross-selling means, if you know in the set of Nataraj Pencils, 10 pencils if you take it in a packet, they give you a sharpener along with it. So, this is Market Basket Analysis for the 10 pencils, one sharpener is given along with that some of the companies might also have been giving the eraser.

Along some of the companies might be giving if you purchase 2 notepads, a pen along with it. So, this is cross-selling opportunities that is, along with selling one set of the items, another similar item which is purchased accordingly is given. Then, because they are all connected to each other, there is a relationship between them or the pattern that is, the collective purchase is there.

So, we can have better inventory management, what to store, where to store, how to store them together or so. Then, specific customers might be there may be close to the start of the batches of students or start of the school. May be in the month of April, when the classes start for the kids from class may be nursery to senior secondary or so. At that point, the sale of notepads would be very high; at that point, the sale of pens and pencils would be high. So, to have a specific target, I would say, target time or customer, this is also set while having a proper or legitimate or statistically proven Market Basket Analysis. Let me talk about the three types which I have put here 1, 2 and 3 in a little detail.

First is Cross-Selling Opportunities. This is the type when we try to understand the association rule for data mining that is, what are the patterns which are associated or what are the things which are associated together. Let me talk about Grocery store along with maybe if I putting examples here, I would say, along with milk, people might also be purchasing bread or they might also be purchasing sugar with it or they might also be purchasing exit, etcetera and so on. So, Cross-Selling Opportunities are identified using certain association rules which I will talk about in the next slide. There are certain algorithms available for it.

So, in cross selling opportunities, one we identify the associations or relationship between items in a transactional data set. That means, we try to extract the frequently bought together items. And, we try to capture the dependencies between them so that, we have a good set or optimized set of the cross-selling opportunities.

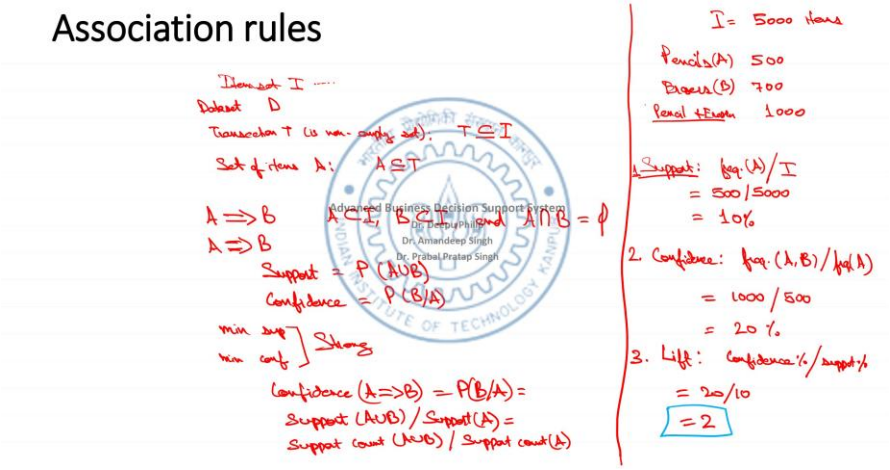
Then comes the second type that is the Inventory Management. When I say Inventory Management, we also need to understand the Sequence Analysis, the way the customers are purchasing the things. So, what we try to do? We try to analyze the sequence of items purchased. As I said, in the month of the October and November, when festival season is there, some specific set of the items may be the gift items or so are to be put into the basket close to the reach of the customer or may be in the month of the April, more notepads more pens are to be put into the basket. So, analyze the sequence of the items to be purchased, this is important and this is taken from the purchase histories. You will appreciate it more when I will try some examples in the MS Excel and we will try to see, how do we connect different kind of items together and try to make a basket out of it. So, these purchase histories are very important to understand the sequence.

So, that we uncover certain patterns in the temporal sequence of the purchases and we identify the frequently occurring sequences. That is, we try to predict the next item of the

customer and we try to set the inventory system accordingly. I would say, for the customer.

Then comes the third type where we set the target time in customer that is, we try to optimize the market basket. This goes little beyond in just identify the associations of the products and it helps to have product recommendations. When I say recommendations, we are trying to optimize these selling patterns. You call it selling patterns from the retailer viewpoint or you call it the purchase patterns from the customer's viewpoint.

So, we use mathematical optimizations or statistical techniques or certain models or algorithms to determine the most effective product bundles, pricing schemes, maybe the promotional offers to maximize profitability.



Now, let us try to see certain association rules that is the Probability rules when we are trying to talk about the MBA. Association rules, if I say, let me have an item set 'I', so as it has certain items in it and I have a data set 'D' which is the database of transactions, where each transaction 'T' is a non-empty set such that, T is contained in I, then each transaction associated with an identifier is called as TID.

So let A be the set of the items A, where T is the transaction that is set to contain A that is, A ⊆ T (that is, T ⊆ A). Then Association rule if I say, A is associated with B, where A ⊆ I, B ⊆ I and A ∩ B = ∅. Then, the rule of association of A and B could be said in three ways.

Number one is Support. When I say Support, which means support given to B by A. $Support = P(A \cup B)$ and also there is a conditional probability when there is a specific percentage of transaction in D that is containing A and also it contains B so it becomes confidence as well. So, $Confidence = P\left(\frac{B}{A}\right)$

So, these rules that satisfy both minimum support threshold are called min sup rules. I will try to explain them when I will try to work on the excel sheet. It is a minimum confidence threshold is also there which is called as min conf. The rule that satisfies both min sup and min conf are known as Strong that is, we have

$$Confidence = (A \Rightarrow \mathrm{B}) = \mathrm{P}\left(\frac{\mathrm{B}}{\mathrm{A}}\right) = Support\,\frac{A \cup \mathrm{B}}{\mathrm{Support(A)}}$$
$$= \mathrm{Support\ Count}\frac{(\mathrm{A} \cup \mathrm{B})}{\mathrm{Support}(A)}$$

So, put it in simple terms let us try to take an example here. Simply, let me say the examples which I took there are pencils, erasers and few of them taken together. Let me say out of 5000 items = I. There are 500 pens which are sold, there are 700 erasers which are sold and both pencil + eraser together is also purchased 1000. So, let us try to see what do we have here in now support and confidence.

So, when I say support, it is the frequency of specific item, whatever we were talking about let me say, I will be talking about the support for pencils.

$$Support: \frac{frequency(A)}{I}$$

$$\frac{500}{5000} = 10\%$$

Then comes the next association point. Confidence is that how confident are we that the eraser would be purchased also with the pencil that is the number of erasers or pencil which are purchased together are 1000 here.

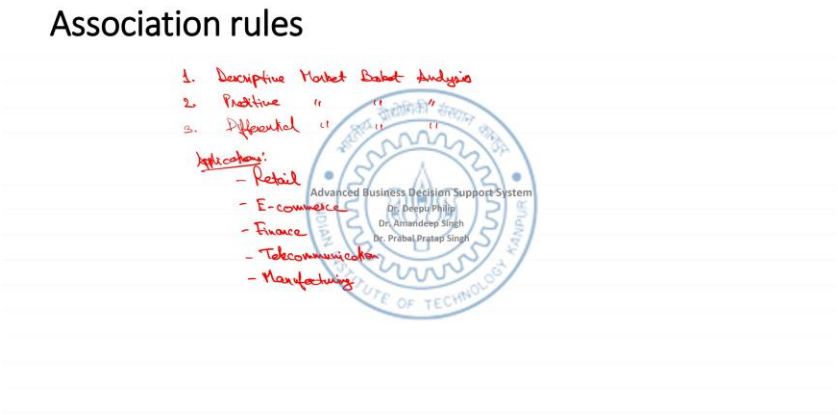
$$Confidence: \frac{frequency(A,B)}{frequency(A)}$$

$$\frac{1000}{500} = 20\%$$

Also, we need to now design the basket. For that, we need to understand one more point here that is known as Lift. What is Lift? Lift is the ratio of the sales, the lift value if it is below 1, it means the combination is not so frequently bought by the consumers, the lift number has to be more than one. So, it is the talking about the confidence only.

$$Lift: \frac{Confidence\%}{support\%} = \frac{20}{10} = 2$$

So, in this case, the probability of buying both pencil and eraser together is high when compared to a transaction of 9 digital items which are sold. So, here we can see the overall view of the data mining using the Market Basket Analysis.



So, these rules could be used for various different kinds of the applications. It could be used for the descriptive Market Basket Analysis only. When I say Descriptive, only describing the data is important. It is a sort of analysis which tells us about the patterns, the connections in the data whether it exists or not. So, this kind of study is used to understand the consumer behavior in general. So, when we are able to describe that we are also able to predict sometime. Predictive Market Basket Analysis that is, we try to predict the future purchases based upon the past purchasing patterns. So, this is one of the predictive Market Basket Analysis where large volumes of data are analyzed.

So, using Machine Learning algorithms, it is sorted and we create predictions on how the products would be more likely or most likely bought together by different customers. Also, we have along with it Differential Market Basket Analysis where differential market is taken that is, we try to identify variations between different markets and we try to compare the behavior of different client segments and different time horizons during the year and try to have a comparative statement to understand the better basket in the market.

The certain benefits of Market Basket Analysis which I have discussed we can enhance customer understanding; we can have improved inventory management and we can have better pricing strategies sales growth. Certain applications could be retail is definitely one of the applications that we have discussed along with retail. We can also have the online purchase systems like e-commerce, I would say applications here, retail, e-commerce that is customer buying habits in the data-driven decisions.

Then we have its applications in finance and we try to forecast the types of the investment decisions or the items the investor would like to buy in the recent future. Then, in telecommunication, it means how to evaluate the customer behavior based upon the data-driven goods and services.

So, in the telecommunication business, we can identify the happiness patterns of the customer in manufacturing as well it has certain applications that is, which products are to be produced accordingly, which material is to be taken, which vendors are to be connected, all those different applications could be there for the Market Basket Analysis. I will take some examples in the Excel sheet in the next lecture and we will try to conclude the course there. Thank you.