

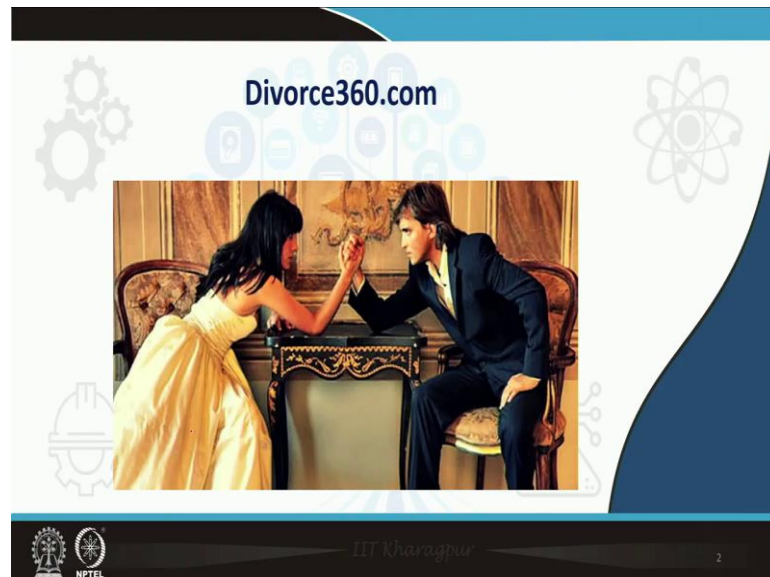
Management Information System
Prof. Saini Das
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Module - 02
Foundations of Business Analytics
Lecture – 09
Data Analytics Tools and Techniques

Hello everybody! So, currently, we are at the 4th session in the module – ‘foundations of business analytics’. This is the 4th and the final session of this particular module. In the previous lecture, we had discussed about introduction to data mining; what data mining is; how it differs from the knowledge discovery in databases process. And, we had also seen a lot of examples related to role of analytics and data mining in business.

So, today we will be talking more about the various data analytics tools and techniques that are used in analytics in business.

(Refer Slide Time: 01:09)



So, proceeding with today's topic you know before we move ahead, I wanted to share couple of very interesting role of you know analytics, in the world of business, in the world of prediction, in the world of you know forecasting.

So, this particular example though in the previous session we had discussed a lot of examples. In this particular session, will be focusing on a few more interesting and very-very; you know current recent applications of analytics in business.

So, this example divorce dot com is a particular website, which is used to predict whether a couple is going to have divorce or not in the future. So, you may consider it to be a boon, you may consider it to be otherwise, but it is there.

So, this particular application will take into consideration a lot of factors such as the income of the couple, the education level of the couple, the age of the couple, the age at which the couple got married and you know the years that have passed since the marriage.

So, all of these factors would be the predictors, which would actually go ahead and predict whether a couple is going to what is the probability that a couple is going to have a divorce in the future. So, this is a very interesting application in it is own right and uses a lot of analytics behind it.

(Refer Slide Time: 02:43)

The slide features a central image of a pregnant woman in a light blue top, with a red and white target graphic overlaid on her belly. The title 'Target predicts customer pregnancy' is positioned above the image. Below the image is a YouTube link: <https://www.youtube.com/watch?v=XH1wQEGROg4>. The slide is decorated with various icons: gears, a target, a chemical structure, and a person. In the bottom right corner, there is a small video inset of a woman with glasses, wearing a yellow and red sari, who appears to be the speaker. The bottom of the slide includes the NPTEL logo and the text 'IIT Kharagpur'.

So, this particular example is another sensational example of the role of analytics in business. So, I am sure all of you are aware of target which is a very popular retail giant. So, target it seems you know this happened a couple of years back wherein a lot of customers go to target to finish their regular shopping.

So, one fine morning a father who was using the mail id of his teenage daughter, suddenly, observed that the daughter was receiving a lot of baby coupons from Target on her mail id. So, he was furious because his daughter was a teenager and possibly she had just graduated from school. So, he called up Target and actually bashed up there the person who picked up the phone; so, also the manager because this was not right in his opinion.

So, target was also at that moment you know not sure as to what had happened. So, target apologized and the matter ended there itself, but it did not end there. So, what happened is you know a week later the father again called up target. And, the father apologized in turn saying that you know there had been certain developments at his own home that he was not aware of.

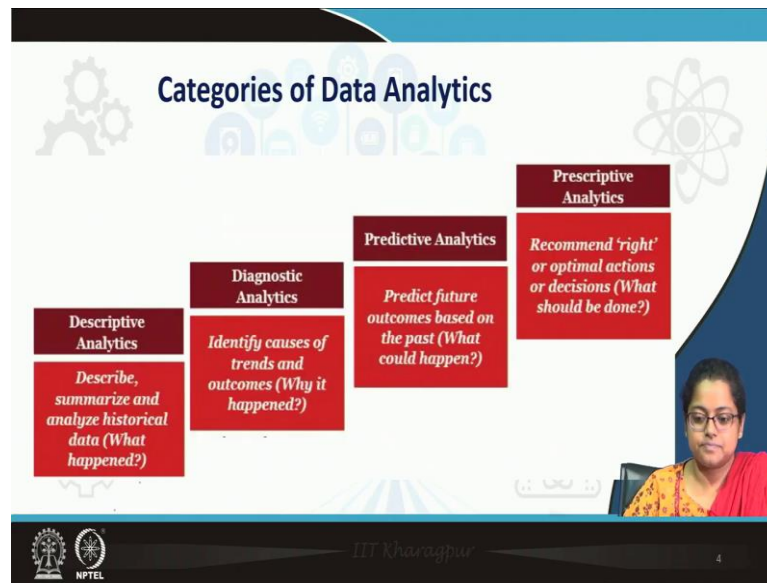
And, his daughter had not shared or revealed in news about her pregnancy ah, but it seems that target had predicted that the girl was that the teenager was pregnant, from her purchasing habits, from products that she purchased and had predicted that she was pregnant even before the family members knew of it.

So, this created ah you know huge backlash from the customers. And, they were absolutely unhappy with what target had done. Because in their opinion using customer data and trying to predict something related to a person's health is absolutely against ethics and is will not be accepted by the audience. So, they were not happy and target had to apologize.

So, here is the link I would have ideally wanted you to watch the video, but here is the link I think you can watch the video at your leisure. It talks about the entire detail of the process, what are the various parameters are related to the teenagers shopping habits, that target took into account while predicting the customer's pregnancy. So, you can have a list of what could be the input variables and what the output is right.

So, this is another example of where a particular retailer had utilized the power of analytics even before you know others knew of it and had utilized it in a not so ethical manner.

(Refer Slide Time: 06:08)



So, let us proceed. So, all of these examples some of which I had shared in the previous session and some of which I am sharing now, through light on the fact that data analytics is something that is an integral part of any organization today and organizations cannot do without it.

So, let us now focus on the various categories of data analytics. So, the first category that we will focus on is descriptive analytics. Descriptive analytics describes, summarizes and analyzes historical data, to find out what happened in certain scenarios. It just, it predominantly describes and summarizes the data and does nothing more than that.

So, that is the role of descriptive analytics we will soon see each of these categories of analytics with their examples and we will try to understand their various technologies or techniques.

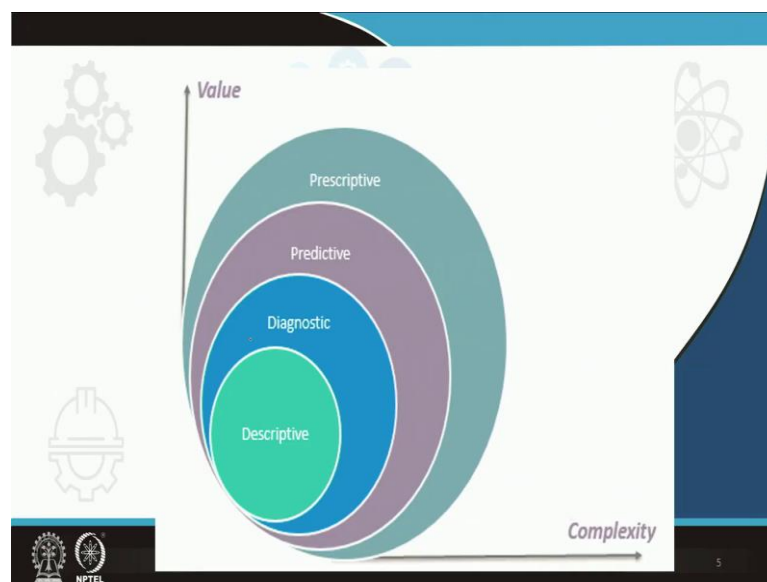
The next category of analytics is diagnostic analytics. As the name suggests diagnostic analytics tries to identify the causes, and trends, and outcomes. So, causes of trends and outcomes basically. So, why did it happen? If something happened, what was the root cause or what was the primary reason behind why it happened? So, this is the second category.

The third category and of course, the most popular of all the 4 categories of analytics is predictive analytics, wherein you try to predict the future outcomes based on the past. So,

what could happen in the future based on what had happened in the past? So, this has a very-very important role to play in business because from demand forecasting to forecasting the number of customers who will show up at your showroom, the customers will respond to a particular advertisement; all of them can be predicted using predictive analytics techniques.

The fourth one is known as prescriptive analytics techniques and as the name suggests again, it recommends the right or the optimal set of actions or decisions. So, what should be done given a particular scenario? So, this is the fourth and the final category of data analytics that is used in the world of business.

(Refer Slide Time: 08:35)



Now, this is a graph that plots the various categories of analytics on 2 parameters. The first is the complexity which is present on the x axis as you see and the second is the value that you can derive out of this kind of analytics.

So, complexity with respect to the procedures, with respect to the techniques, with respect to the algorithms, and the value with respect to the benefits that you would derive out of applying this particular category of analytics to a particular set of problems.

So, here we see descriptive analytics as we move from descriptive to prescriptive analytics in the particular order that is given in the graph, the complexity increases and at the same time the value also increases. So, descriptive analytics is the simplest of all the

4 categories of analytics and it also gives you the least value. Though that value is also very important, in business because that can also be used in multiple, multiple to develop multiple number of strategies, as we will see with certain examples.

But of course, as we move from descriptive to diagnostic to predictive and eventually to prescriptive the complexity of the algorithms increase. And, at the same time the value that you derive out of each of these analytics techniques in a particular set of problem also increases right.

(Refer Slide Time: 10:09)

Descriptive Analytics Applications

- › Most shoppers turn towards right when they enter the a retail store.
- › Conversion rate of women shoppers is higher than male shoppers among electronic gadgets purchasers (**Radio Shack**).
- › Strawberry pop-tarts sell 7 times more during hurricane compared to regular period (**Wal Mart**).
- › Women car buyers prefer women sales person.

IIT Kharagpur
NPTEL

So, let us move ahead certain descriptive analytics applications. So, some of the most popular applications, you know these are some examples of what descriptive analytics could do to your business. Most shoppers turn towards the right when they enter a retail store.

So, this could be an outcome, or this could be a descriptive analytics outcome that would help shoppers to design the layout of their store, to place products in a particular fashion that they want, and take multiple measures related to store layout. Also you know they can take measures related to customer representative, or you know a helper, a customer helper, or a customer service agent, being present at a particular point in the store. So, that more shoppers are able to access him or her easily.

The second application here, conversion rate of women shoppers is higher than male shoppers, among electronic gadget purchases. So, this could again give you know the owners of electronic gadget stores, multiple strategies to deal with.

Third one strawberry pop tarts sell 7 times more during hurricanes compared to regular period. So, of course, if there is a forecast of hurricane you know what to do as a store selling strawberry pop tarts.

Finally, a women car buyers prefer women sales person. So, this is kind of a very interesting insight. And, all of these insights were derived from sets huge bulks of data that were collected from various stores. So, departmental stores, from electronic gadget purchases, from different environments or different scenarios data was collected, and descriptive analytics was performed to come up with these insights.

(Refer Slide Time: 12:09)

Diagnostic Analytics Applications

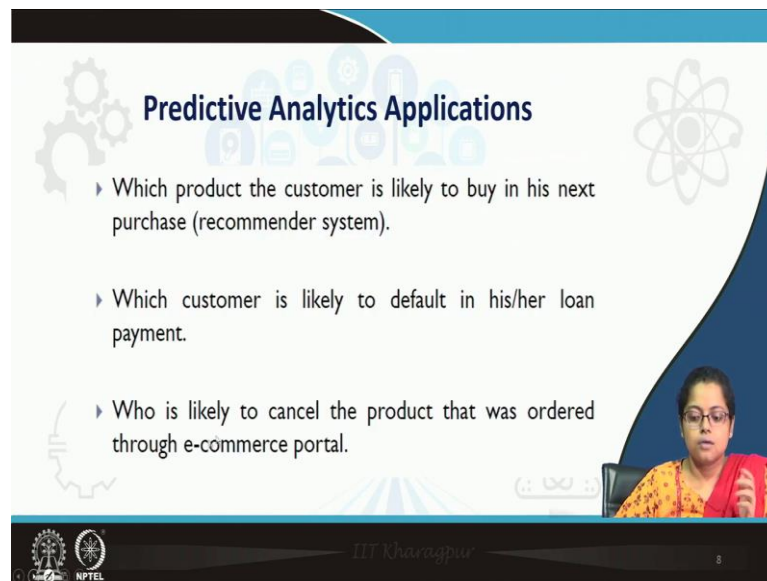
- Why customers liked your social media campaign or why they didn't?
- Why certain products were popular at a certain time, at a certain place?

The slide features a background with a stylized tree of nodes and various icons including gears, a hard hat, a circuit board, and a molecular structure. At the bottom left are the logos for IIT Kharagpur and NPTEL. The text 'IIT Kharagpur' is centered at the bottom, and the number '7' is in the bottom right corner.

Now, moving on to diagnostic analytics applications, some examples given here are; why customers liked your social media campaign or why they did not like it? So, if you have come up with a particular social media campaign, maybe to popularize your, you know spring sale of electronic garments, or spring sale of electronic goods or spring sale of garments. Why did the customers actually happen to like your social media campaign? So, this would be taking a step back and analyzing, what actually happened right. And, the reasons as to why it actually happened.

Similarly, why certain products were popular at a certain time and at a certain place? You might have observed that, some products all of a sudden became become a sensation, maybe in a particular geography. So, what is the reason? Why did it become a sensation at that particular time in that particular geography? So, these are all things that should be analyzed in depth to find out the root causes or the reasons.

(Refer Slide Time: 13:20)



Predictive Analytics Applications

- ▶ Which product the customer is likely to buy in his next purchase (recommender system).
- ▶ Which customer is likely to default in his/her loan payment.
- ▶ Who is likely to cancel the product that was ordered through e-commerce portal.

The slide features a blue and white color scheme with icons of gears and a person. A video feed of a woman in a yellow and red sari is visible in the bottom right corner. The footer includes the IIT Kharagpur and NPTEL logos.

Moving on predictive analytics applications and as I mentioned predictive analytics has a whole range of applications in business. So, in recommender systems, which is today a part of almost every online platform or every e-commerce platform, you would observe that predictive analytics has a major role to play.

As to based on the customers' prior purchases, or based on the history of the demographic background, or the history of purchases of customers who are very similar to a particular customer, which product is the customer likely to purchase in his next visit? Right.


So, based on that you can actually recommend products to him or you can actually stock up your products in your warehouse, when the customer comes online you can actually tell him that you must have seen in Amazon. See ah particular customer comes to buy a product. And, Amazon prompts that customer similar to him or her has brought a set of other products which Amazon actually recommends to the first customer.

So, similarly which customer is likely to default in his or her loan payment? This could be very insightful to an insurance company or you know to a banking organization specifically. If, in order which gives out loans to if the if the banking organization can actually predict in advance, that a particular customer is likely to default in his or her loan payment, the banking organization might be very cautious as to whether the customer should be actually given the loan or not.

Finally, our third example is related to cancellations of products that are ordered on e-commerce portals. So, research suggests that on e-commerce portals, cancellations or returns, amount to almost 30 percent of the products that are actually purchased. So, returns create a huge loss for e-commerce portals.

And, if they are able to predict in advance, which customers are likely to cancel products. They can be very cautious, when selling products to such customers, or they can you know sometimes take measures or strategies multiple strategies like blacklisting certain customers, or taking other measures, to prevent or minimize the cancellations or the returns, because at the end of the day their goal is to minimize the losses.

(Refer Slide Time: 15:55)



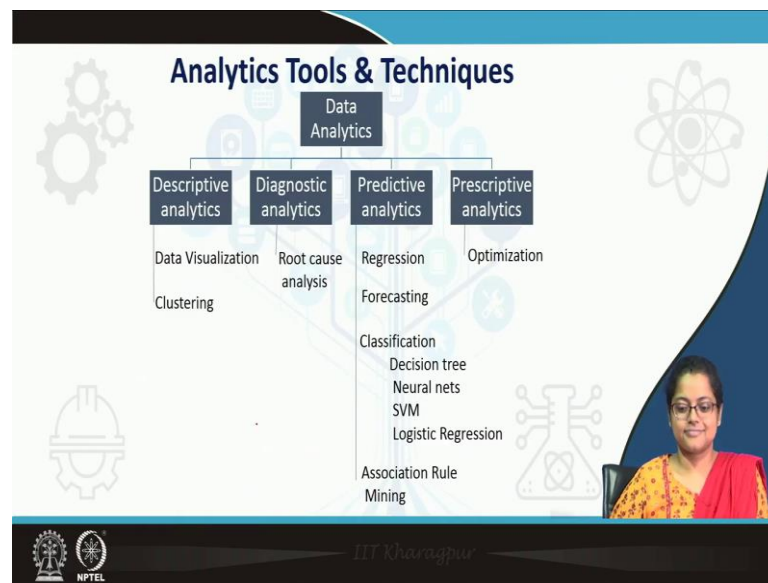
The slide is titled "Prescriptive Analytics Applications" and features a list of five bullet points. The background is white with a blue header and footer. There are decorative icons: a gear, a network of nodes, and a molecular structure. A small video inset in the bottom right corner shows a woman with glasses and a red shawl. The footer includes the IIT Khargpur logo and the NPTEL logo.

- ▶ What is the optimal product mix?
- ▶ What is the optimal route for a delivery truck.
- ▶ Best markdown pricing for fashion products.
- ▶ Optimal assignment of aircraft to flight.
- ▶ How to manage the fleet of vehicles owned by a company for employee drop and pick up?

Moving ahead prescriptive analytics application. So, these are some examples. What is the optimal product mix, if you are trying to sell a bucket or an array of products? What is the optimal route for a delivery truck? So, prescriptive analytics helps you to optimize your; you know your goal or your objective based on certain constraints or parameters as

you can see. Best markdown pricing for fashion products, optimal assignment of aircraft to flights. And, how to manage fleet of vehicles owned by a company for employee drop and pick up? So, all of these applications have a very important role played by prescriptive analytics.

(Refer Slide Time: 16:44)

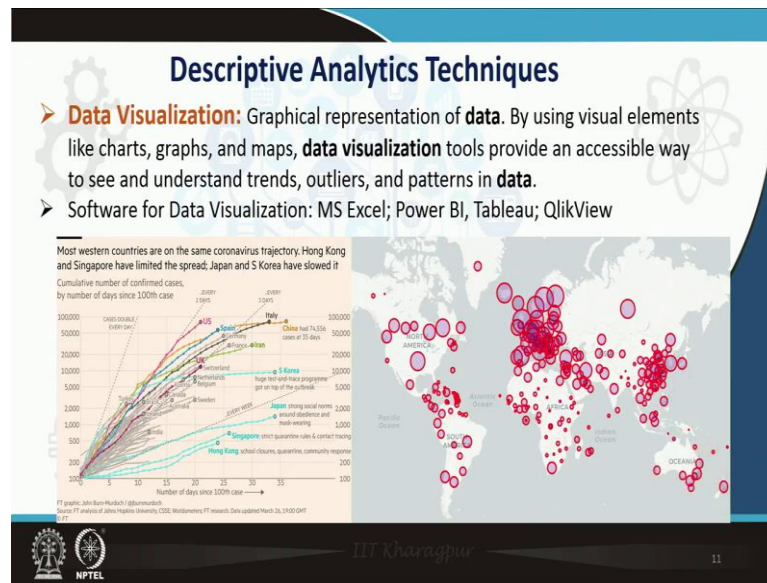


Now, moving on each of these 4 categories of analytics do have certain tools and techniques that are used such as, you know we will see the tools and techniques that are used for each of them. Descriptive analytics, the tools that are primarily used here are data visualization and clustering.

So, we will talk about each of these in detail in the subsequent slides. A diagnostic analysis root cause analysis is performed; it is the primary technique here. Predictive analytics has again a host of techniques – regression; forecasting; classification. Classification is a very popular technique, which again has multiple you know techniques that are used for it. Such as decision trees, neural networks, support vector machines, logistic regression, Bayesian classifiers to name a few.

And, then finally, we will talk about association rule mining, which is another very important and interesting technique that is used in predictive analytics. Finally, for prescriptive analytics, the primary tool or technique that is used is optimization.

(Refer Slide Time: 18:03)



So, moving on descriptive analytics, we have already mentioned that descriptive analytics has you know two primary techniques; one is data visualization and the second is clustering. So, data visualization you know it is a graphical representation of data. And as we had mentioned the; you know in one of the previous sessions; you know graphs, or pictures; they speak a 1000 words. So, and in terms of data, you know, they speak a million of data.

So, if you have huge amounts of data, simply looking at the data might not give you enough insights. But, if you are able to analyze the data, or represent that particular set of data using graphs or charts, that would give the in reader or the or the or the a person who is looking at your graph and chart much better insights just by having a glance at the data, in terms of the graphs or charts.

So, by visual by using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way, to see and understand trends, outliers, and patterns in the data. So, data visualization tools, such as charts, graphs, maps, would help you understand a lot of trends in the data, outliers or exceptions in the data, and certain patterns that are emerging out of the data.

So, here we see certain examples. Today, the examples that you know data visualization has applications in every area of business, but today we will focus more on Covid-19 pandemic data right.

So, you see here that, the first graph here shows a cumulative number of confirmed cases by number of days since the 100 case. So, various countries and this graph or this particular graph, you must have seen in multiple places.

So, it is using data of the number of cases, cumulative number of cases. Since, the 100 case in multiple countries across the world and coming up with insight such as which country has tried to has flattened the curve, which country is going you know the cases are doubling every day, in which particular country cases are doubling every 3 days and so on.

So, this can provide very important insights to the governments of particular nations to find out you know what, what could be the problems at their end due to which may be the number of Covid-19 cases are multiplying at a huge rate, compared to the other nations. Similarly, the second graph that you see here.

So, the second graph that you see here. In this particular graph this is this is you know representation of spatial data. So, what it does is that on a on the map of the world, it is trying to show the distribution of Covid-19 positive cases across various countries and continents. And, the bubbles actually represent the number of cases, the larger the size of the bubble, the larger is the sorry the number of cases in that particular geography.

So, so, similarly data visualization can be applied in these here we just simply saw on the map a map and a graph, you can also see multiple charts that are available, and you know several other that there are bubble graphs, there are you know scatter plots, there are say bar charts, pie charts. So, all of these charts and graphs would help a user or a reader get much better insights from data, than simply looking at data residing in a database.

And, there are a lot of software available for data visualization as you see here, Microsoft Excel is I think the most popular one though it is a it is it primarily helps you, obtain a very static representation of data. The others here that I have mentioned, Power BI, Tableau, Qlik View, these are more recent and more powerful tools for data visualization, because they help you, obtain a very dynamic visualization of the data.

Having said that MS Excel is by far the most you know, widely used and it has a lot of applications with respect to data visualization; ok.

(Refer Slide Time: 22:59)

Descriptive Analytics Techniques (contd..)

- **Clustering:** Given a set of data points, each having a set of attributes, find clusters such that:
 - data points in one cluster are more similar to one another
 - data points in separate clusters are less similar to one another.
- Application in Market Segmentation: Subdivide a market into distinct subsets of customers based on their geographical and lifestyle related information where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix.

Intercluster distances are maximized

Intracuster distances are minimized

IIT Kharagpur

NPTEL

12

Moving ahead, the second technique under descriptive analytics we will talk about is clustering.

So, clustering basically says that, given a set of data points each having a set of attributes. So, there could be two attributes, or there could be three attributes or more number of attributes. So, in case there are two attributes you would be able to see the clusters in a two dimensional space.

So, given a set of data points each having a set of attributes, find clusters such that, data points in one cluster are more similar to one another, and data points in a separate clusters are less similar to one another.

So, if you look at this graphic here on the right, you would see a three dimensional space. So, data here is organized you know it has three attributes and all the three attributes are used to form clusters of data such that, inter cluster distances are maximized, which again we you know talks about this particular point, data points in separate clusters are less similar to one another.

So, inter cluster distances are maximized whereas, intra cluster distances are minimized which means that within a cluster distances between the data points are as minimal as possible, but across clusters or between clusters the algorithm tries to maintain as much of distance as possible.

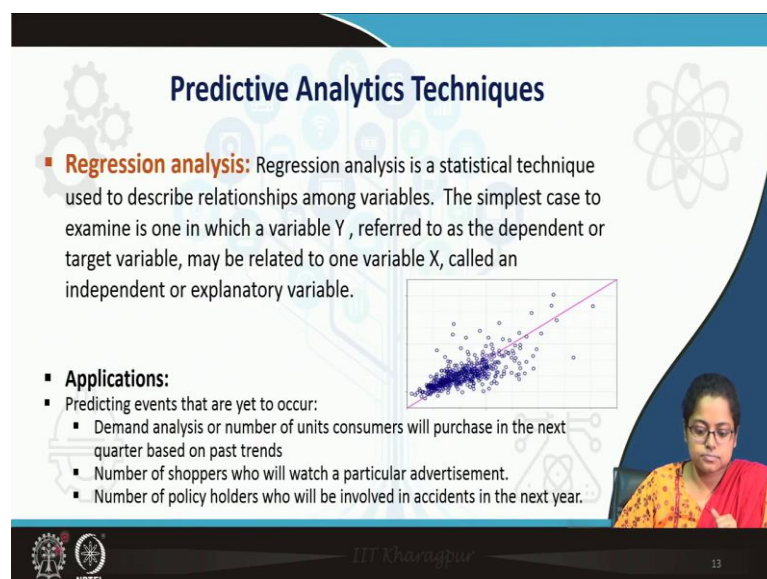
So, clustering has got a lot of you know techniques such as we have a hierarchical clustering k means clustering, but we will not discuss the details of each of these techniques or algorithms in this course, because that is beyond the scope of this course. So, for each of these techniques that I am discussing today, you can take up several other courses related to them.

But, in this particular course, this course will give you the flavor of the various techniques that are available around you with respect to data analytics, and which you can study in depth much later on. And, currently what you have to do is understand the techniques in brief and their applications to business. This particular course talks about that.

So, applications of clustering in market segmentation. Subdivide a market into distinct subsets of customers based on their geographical and lifestyle related information, where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix or a campaign.

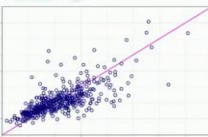
So, what is mentioned here is that, if you have data pertaining to your customers related to same three parameters. I am just taking an example say, income level, education status, and say age of customers. You can actually; ah you know form clusters such that each of those clusters can be targeted by the market years with separate marketing campaigns. So, that is the predominant application of clustering in business.

(Refer Slide Time: 26:14)



Predictive Analytics Techniques

- **Regression analysis:** Regression analysis is a statistical technique used to describe relationships among variables. The simplest case to examine is one in which a variable Y , referred to as the dependent or target variable, may be related to one variable X , called an independent or explanatory variable.



- **Applications:**
 - Predicting events that are yet to occur:
 - Demand analysis or number of units consumers will purchase in the next quarter based on past trends
 - Number of shoppers who will watch a particular advertisement.
 - Number of policy holders who will be involved in accidents in the next year.

13

Coming to the next technique which is we are moving on from descriptive to predictive analytics. So, the first technique here that we will talk about is regression analysis. A regression analysis has a huge application in business. So, it is used in almost every business, every organization that you see around you, use regression analysis to solve some problem on the other.

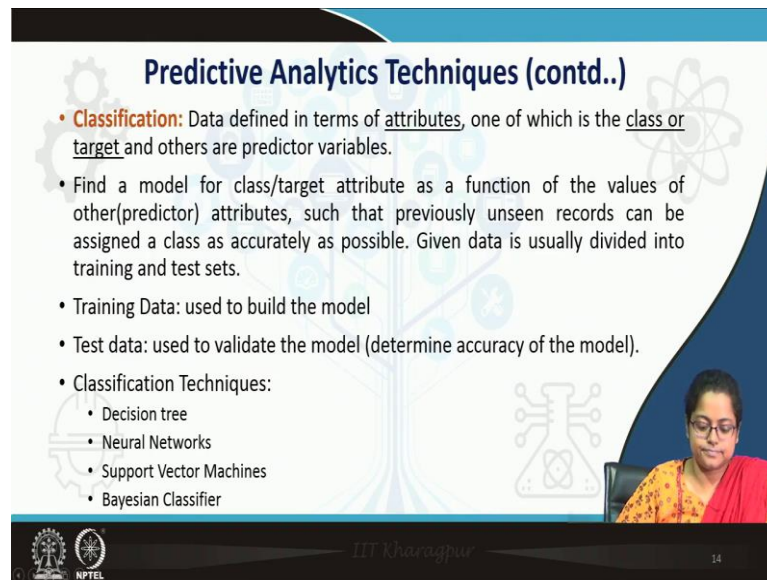
So, regression analysis is a statistical technique, used to describe relationships among variables. The simplest case to examine is one in which a variable X , referred to as the dependent or the target variable may be related to one variable. So, a variable Y , which is the dependent variable or the target variable, may be related to one variable X , which is called the independent or the explanatory variable.

So, in regression analysis we see how various you know one or more independent variables represented as X_1, X_2, X_3, X_4 in turn try to explain a dependent variable Y . So, again regression has a lot of applications so, some of the applications that we have discussed here are predicting events that are yet to occur.

So, demand analysis or number of unit's consumers will purchase in the next quarter based on the past trends. So, regression tries to predict the future based on past data.

The second example that we are talking about here, pertains to number of shoppers who will watch a particular advertisement. So, again based on that you can take certain measures or strategies. And, finally for an insurance company number of policy holders, who will be involved in accidents in the next year. This is again a prediction based on the past.

(Refer Slide Time: 28:13)



Predictive Analytics Techniques (contd..)

- **Classification:** Data defined in terms of attributes, one of which is the class or target and others are predictor variables.
- Find a model for class/target attribute as a function of the values of other(predictor) attributes, such that previously unseen records can be assigned a class as accurately as possible. Given data is usually divided into training and test sets.
- Training Data: used to build the model
- Test data: used to validate the model (determine accuracy of the model).
- Classification Techniques:
 - Decision tree
 - Neural Networks
 - Support Vector Machines
 - Bayesian Classifier

IT Kharagpur 14

Now, moving on classification; classification is another technique which has a very important role to play in predictive analytics. It is a very important predictive analytics technique. Data is defined in terms of attributes one of which is the class or the target and others are predictor variables.

So, if you have a set of data, then one attribute is called the class or the target variable, target attribute, which you are trying to predict with respect to certain predictor variables or independent variables that you have.

So, what classification does is find a model for class or target attribute, as a function of the values of other predictor attributes such that, previously unseen records can be assigned a class as accurately as possible, given data is usually divided into training and test sets. Training data is used to build the model, and test data is used to validate the model, or determine the accuracy of the model.

So, what happens is in brief. If, you have a set of data with certain predictor variables and a target or a class variable, you will use, you will divide the entire data set into two categories; training and test data. The training data will be used to so, you know build the model and test data will be used to validate or test the model.

Usually the training data and the test data, you know the entire data set is divided into both of these in a certain ratio. So, it could be 70, 30 or 80, 20 and it depends on the nature of your problem.

So, again if you have to understand this particular technique in detail, you have to take up a course on classification. Classification techniques so, there are certain techniques that are you multiple techniques that are used for classification, decision tree, neural networks, support vector machines, Bayesian classifiers to name a few.

(Refer Slide Time: 30:25)

Predictive Analytics Techniques (contd..)

- Application 1: Given old data about fraudulent customers and their background, predict whether new customer will commit fraud or not.

Predictor attributes → Target

| Tid | Predictor attributes | | | | Target | | | | |
|-----|----------------------|----------------|----------------|-------|--------|----------------|----------------|-------|--------|
| | Refund | Marital Status | Taxable Income | Cheat | Refund | Marital Status | Taxable Income | Cheat | Actual |
| 1 | Yes | Single | 125K | No | No | Single | 75K | ? | No |
| 2 | No | Married | 100K | No | Yes | Married | 50K | ? | Yes |
| 3 | No | Single | 70K | No | No | Married | 150K | ? | No |
| 4 | Yes | Married | 120K | No | Yes | Divorced | 90K | ? | No |
| 5 | No | Divorced | 95K | Yes | No | Single | 40K | ? | Yes |
| 6 | No | Married | 60K | No | No | Married | 80K | ? | No |
| 7 | Yes | Divorced | 220K | No | | | | | |
| 8 | No | Single | 85K | Yes | | | | | |
| 9 | No | Married | 75K | No | | | | | |
| 10 | No | Single | 90K | Yes | | | | | |

Training Set → Learn Classifier → Model

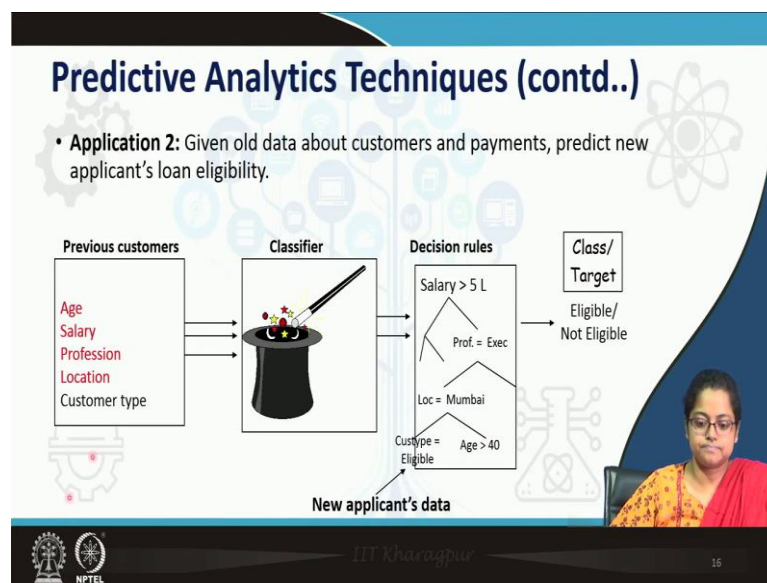
Test Set → Predicted by algorithm

So, we will quickly see two examples or applications of pred of classification technique in business. So, this is first application here, talks about given old data above fraudulent customers and their background, predict whether new customers will commit fraud or not. So, this particular technique has certain predictor attributes we see, the predictor attributes are refund marital status and taxable income of the.. customer. And, the class or the target attribute is whether the person will cheat or not or commit fraud or not and this is the training data set.

Now, this is used to train the model, which will be used to test the test data. So, these are new records that are coming up for which you do not know the outcome variable, that is the target variable, but you will try to predict it using the classification algorithm this. So, this is the; what is there in your test data and this is what is predicted by your algorithm.

So, how well your; you know predicted algorithm predicts the test data set; the target variable of the test data set determines the accuracy of your model. So, depending on this you will actually be able to understand whether a customer is going to cheat or not in the future. And, based on that you will actually find out you know, whether it is whether it is what sort of a customer he is, whether you should take some precautionary measures before handling that particular customer and several other strategies.

(Refer Slide Time: 32:11)



Another example here, talks about given old data about customers and payments, predict new applicant's loan eligibility. So, similar to the previous example, here we have four predictor variables, age salary profession and location of the customer, and customer type in black here is the target variable.

So, the classifier or the model is built, which tries to predict the loan eligibility in terms of eligible or not, which is the class or the target variable of the new applicant.

So, it comes up with rules if it is using a decision tree algorithm, such as if the customer salary is greater than 5 lakhs and profession is an executive, location is Mumbai, then the customer is eligible for the loan.

So, this is how using certain algorithms or classifiers classification techniques a company is able to predict in advance an applicant's loan eligibility. And, if the company

determines through the algorithm, that the customer is not eligible for a loan he may take certain precautionary measures.

(Refer Slide Time: 33:21)

Predictive Analytics Techniques (contd..)

- **Association Rule Mining:** Finding frequent patterns:
 - Analysing item sets in customers basket or transactions and identifying frequently occurring item sets, which can be basis for recommendations.
 - Known as **Market Basket Analysis** in Retail.

Previous Baskets

- Egg, beer, sugar, bread, diaper
- Egg, beer, cereal, bread, diaper
- milk, beer, bread
- cereal, diaper, bread

Current Basket

- Beer, X1

Applications:

- Store layout: Place the frequently occurring items either in close proximity or as far apart as possible; give combo offers.
- Warehousing and inventory management.

NPTEL

17

Finally, moving on to the last application here that I will discuss under predictive analytics technique is association rule mining. So, this has a very important application in marketing. It is used to find frequent patterns, by analyzing item sets in customer's baskets or transactions, and identifying frequently occurring item sets, which can be basis for recommendations.

So, you may remember that in one of the previous sessions, we had discussed about you know so, a lot of young fathers buy diaper and beer together on Fridays right. The algorithm that is used for predicting it is association rule mining, which it is also used for market basket analysis. So, this particular technique is also known as market basket analysis, which tries to tell or predict based on a customer's past purchases, what are the products that the customer is going to purchase in the future in bulk?

So, here you see if these are the previous baskets where customers have purchased products. So, if in the in the current basket, if customer is purchasing beer, what else will he purchase along with it?

So, based on this if you identify the 2 products or 3 products are purchased together, you can take multiple measures such as it has a very important rule in store layout. So, what

can be done is if you know that beer and diaper are going to be purchased together. Either place the frequently occurring items in close proximity. So, that people who are walking by will pick up both the products together or place them as far apart as possible in the store.

Say suppose, in the entrance you place beer and you place diaper at the exit. So, that you know the customer who is buying beer you know that, through the algorithm you know that definitely the customer will also purchase diaper. So, in effect what will happen is customer will walk through the entire store and maybe pick up a lot of other products on impulse.

And, you can also give combo offers that will of course, encourage them to pick up more, you know buy more of those products, which are which they buy in bulk. And, of course, this has a role in warehousing and inventory management, because you can stock these products accordingly. If, you know that on Friday evenings a lot of customers will buy diaper and beer together you can store them accordingly.

So, association rule has a very important role in market basket analysis in marketing. So, so, as I mentioned each of these techniques that I discussed here, I have been discussed in brief and I have tried to give you some applications in business of these techniques. But, if you have to again I would reiterate, that if you have to understand the algorithms or the techniques behind each of these, you need to go in depth and read them and understand them in detail in other courses.

So, with that I would like to end this session on, you know data mining and data analytics; and of course, databases, data warehouses. I hope you have gained some sort of understanding about how data is stored in organizations; and, then they are stored in databases, data warehouses; how you derive business intelligence out of data in data warehouses; and, eventually how data analytics and data mining has a very important role in organizations and in businesses.

So, with that I would like to thank you for being with me through this entire session. See you in the next lecture.

Thank you!